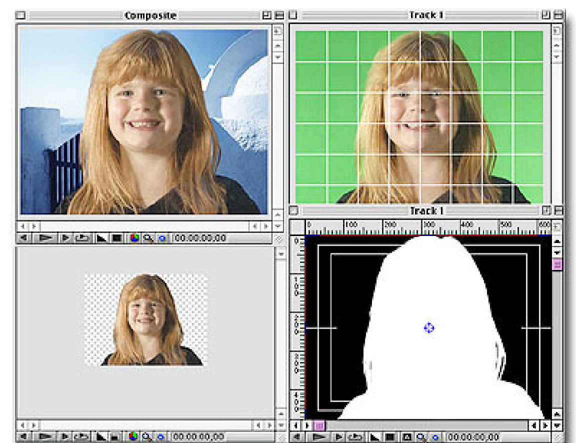
ADVANCED MULTIMEDIA CODING

Fernando Pereira

Instituto Superior Técnico



The Old Analogue Times: the TV Paradigm

- Video data modeled as a sequence of pictures with a certain number of lines
- One audio channel is added to the video signal
- Video and audio have an analogue representation
- User chooses among the available broadcast programmes

Evolving Multimedia Context ...

- More information is in **digital form**, ...
- More information is **on-line**, ...
- More information is **multimedia**, …
- Multimedia information now covers **all bitrates and all networks**
- Applications & services become **'multimedia'** …
- Applications & services become **'interactive'** …
- **Internet** is growing …

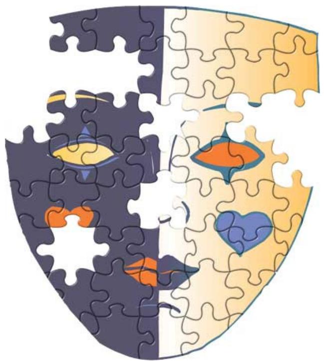
New Technologies, New Needs …

- Having multimedia information available wherever you are, covering a wide range of access conditions
- More freedom to interact with what is *within* the content
- Reusing the multimedia content, combining elements of content in new ways
- Hyperlinking from elements of the content
- Finding and selecting the information you need
- Identifying, managing and protecting rights on content
- Common technology for many types of services, notably broadcasting, communications, retrieval

Demands come from users, producers and providers !

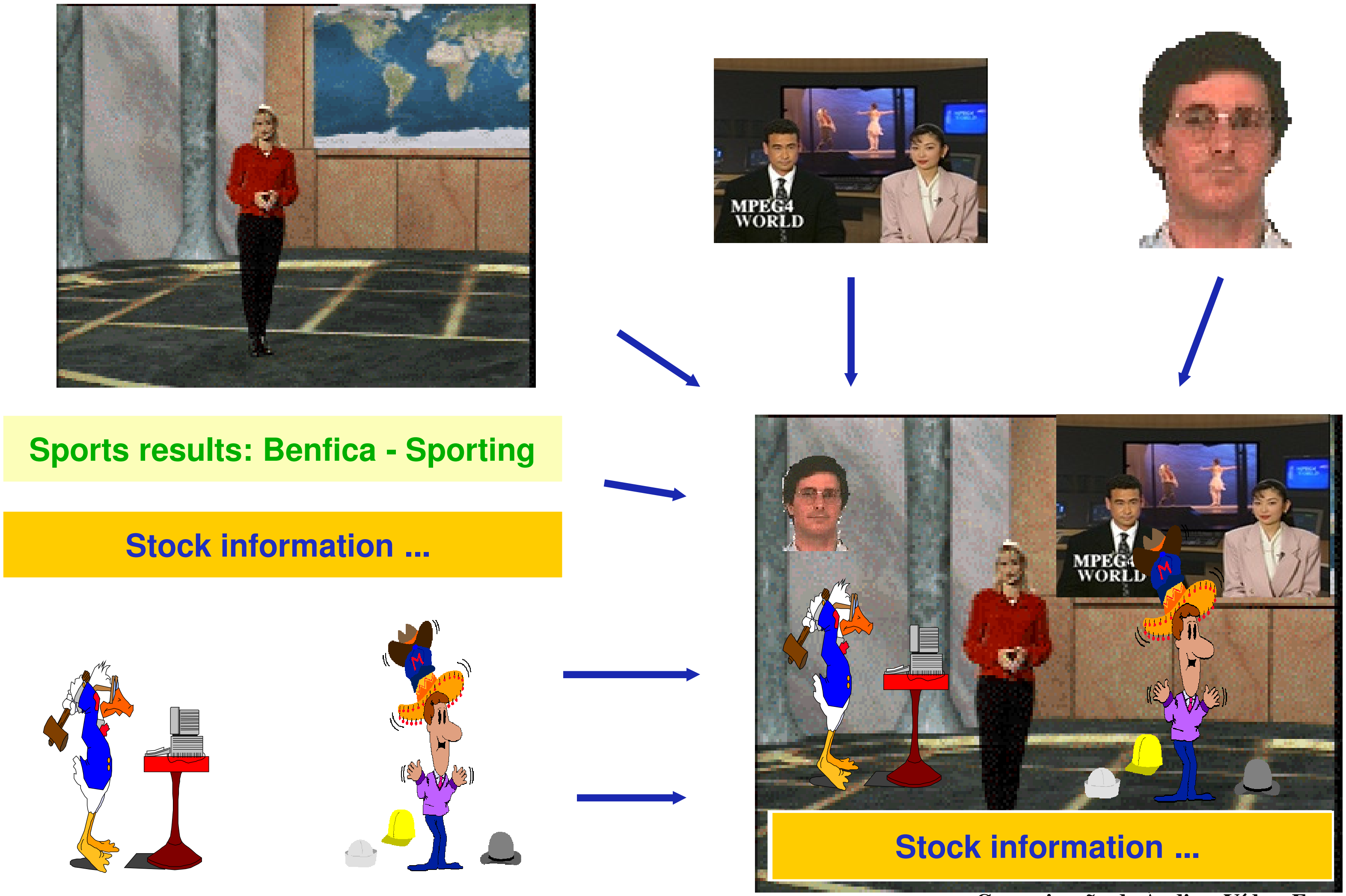
We and the World around us …

Towards the Real World: The Object-based Representation Model

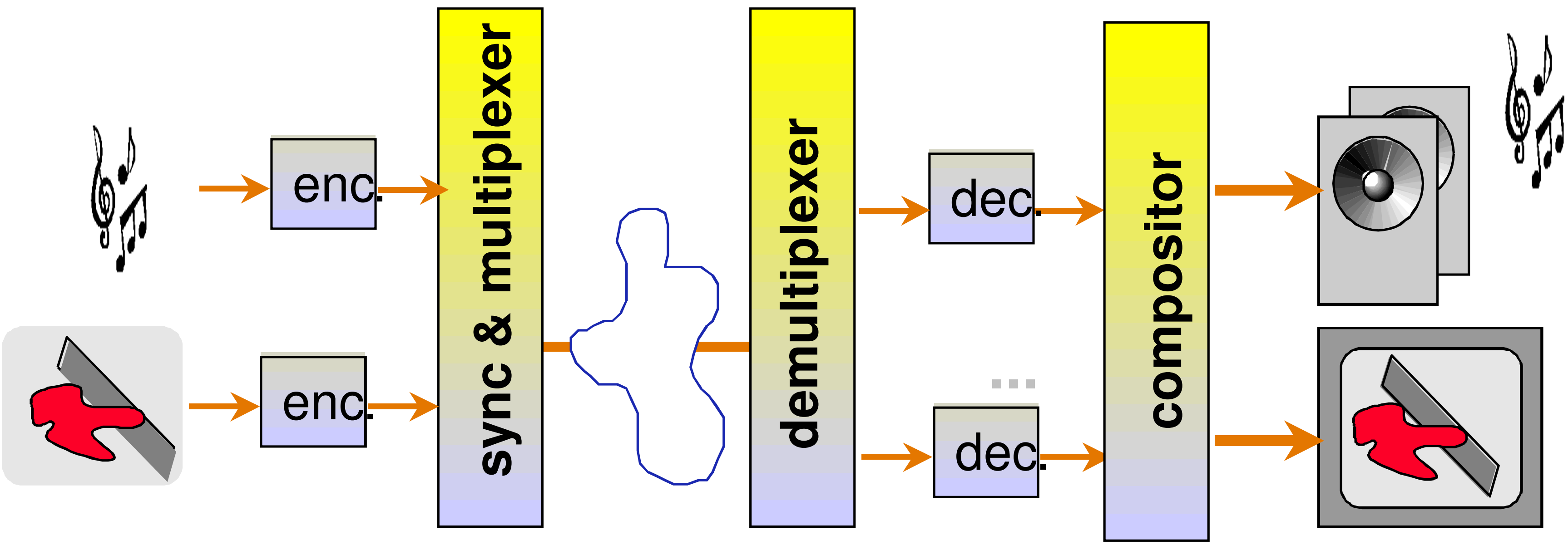


- **Audiovisual scene represented as a composition of objects**
- **Integration of objects from different nature: A&V, natural and synthetic, text & graphics, animated faces, arbitrary and rectangular video shapes, generic 3D, speech and music, ...**
- **Object-based hyperlinking, processing, coding and description**
- **Interaction with objects and their descriptions is possible**
- **Object-based content may be reused in different contexts**
- **Object composition principle is independent of bitrate: from low bitrates to (virtually) lossless quality …**

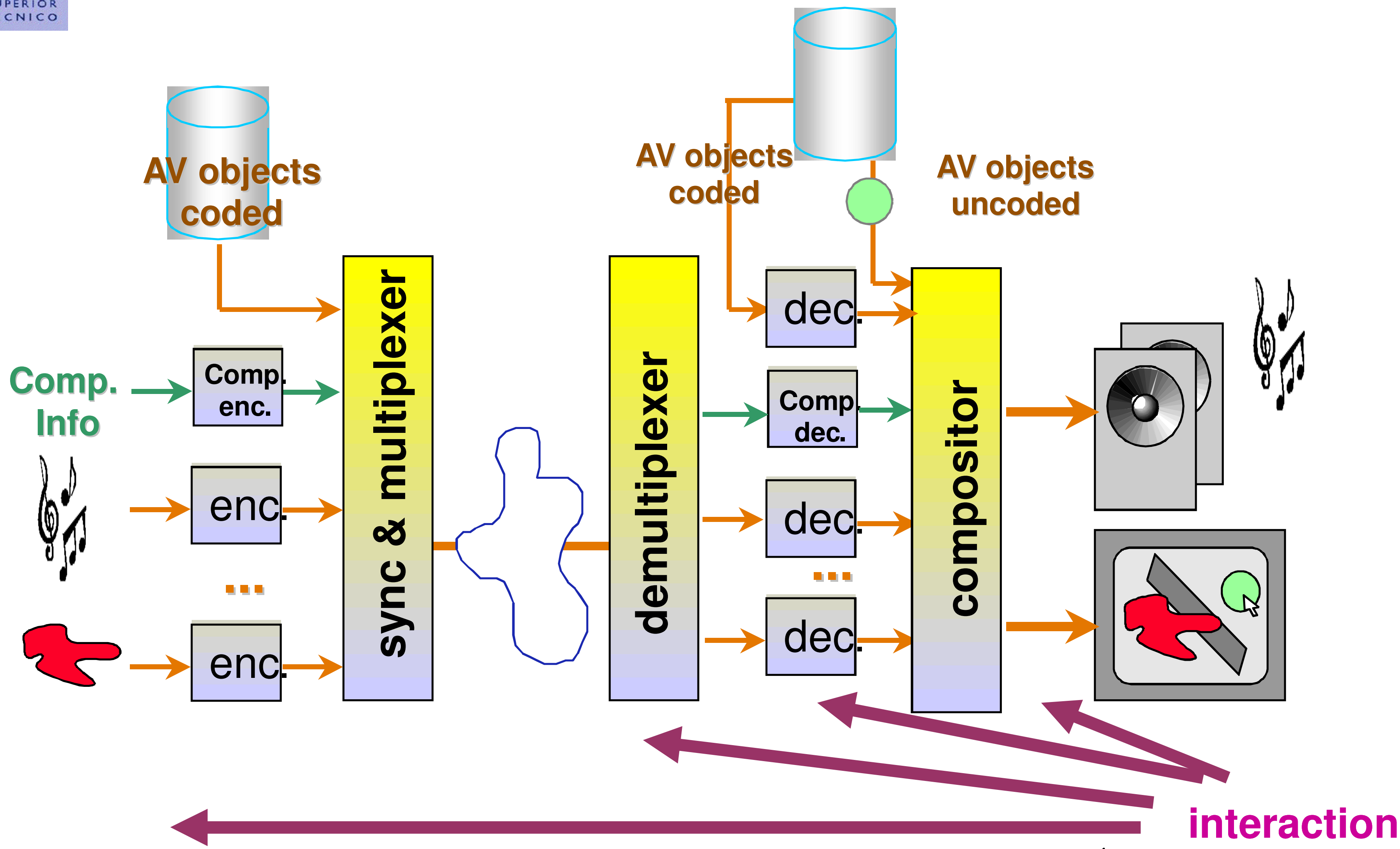
Object-based Content …



Conventional Audiovisual System

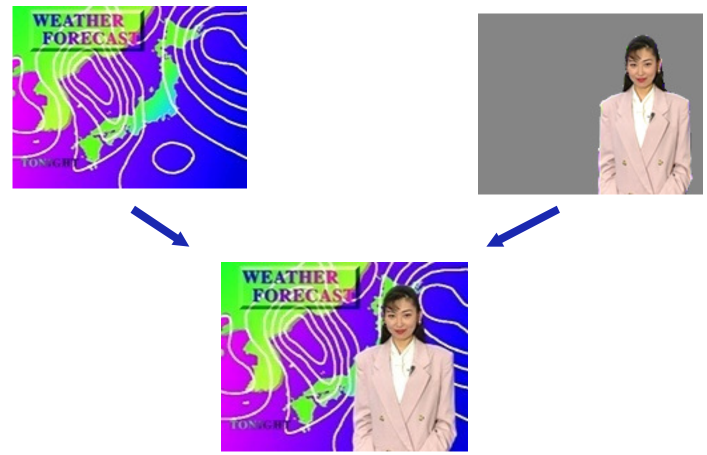


Object-based Audiovisual System

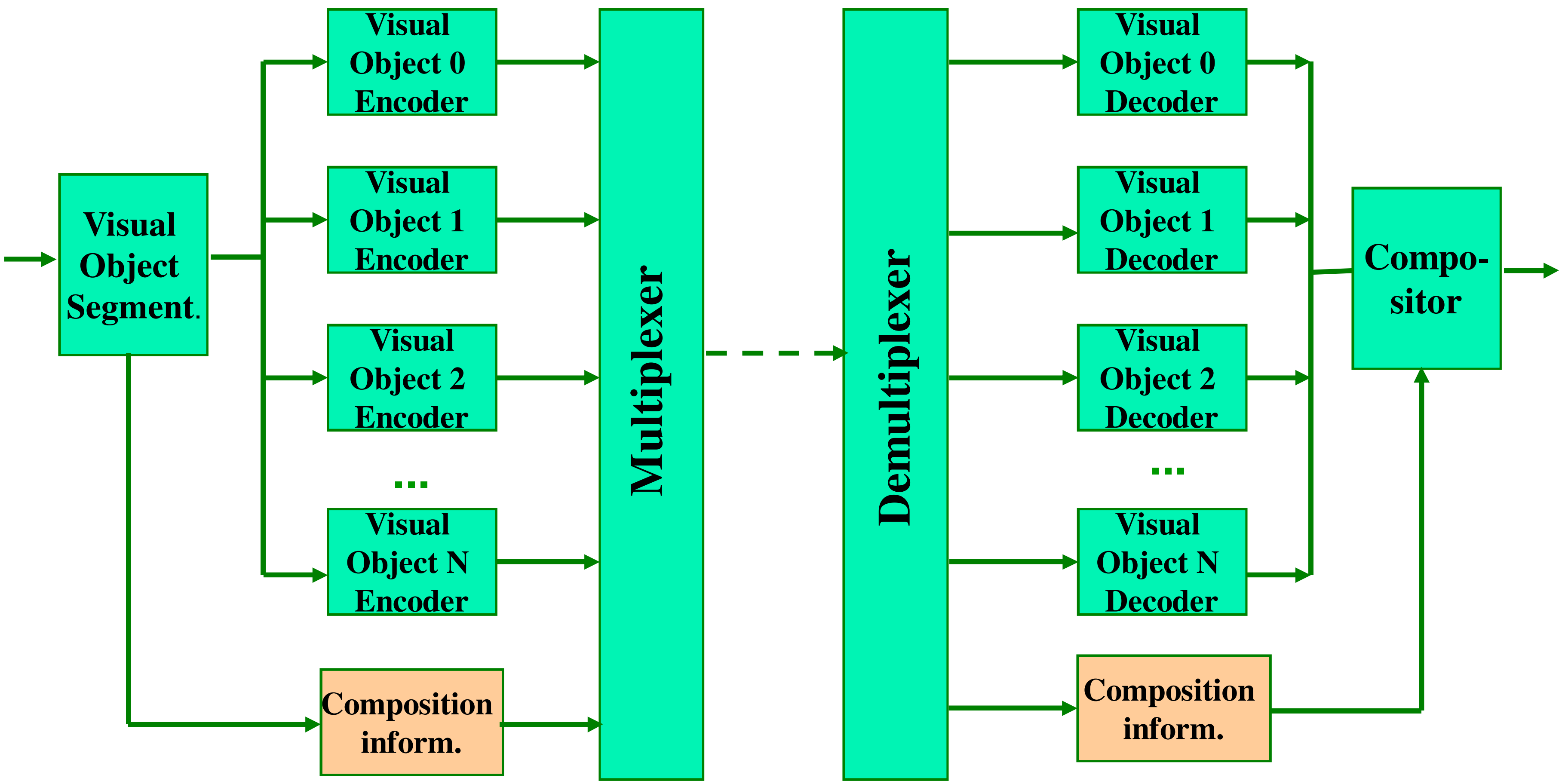


MPEG-4: Object-Based Coding Standard

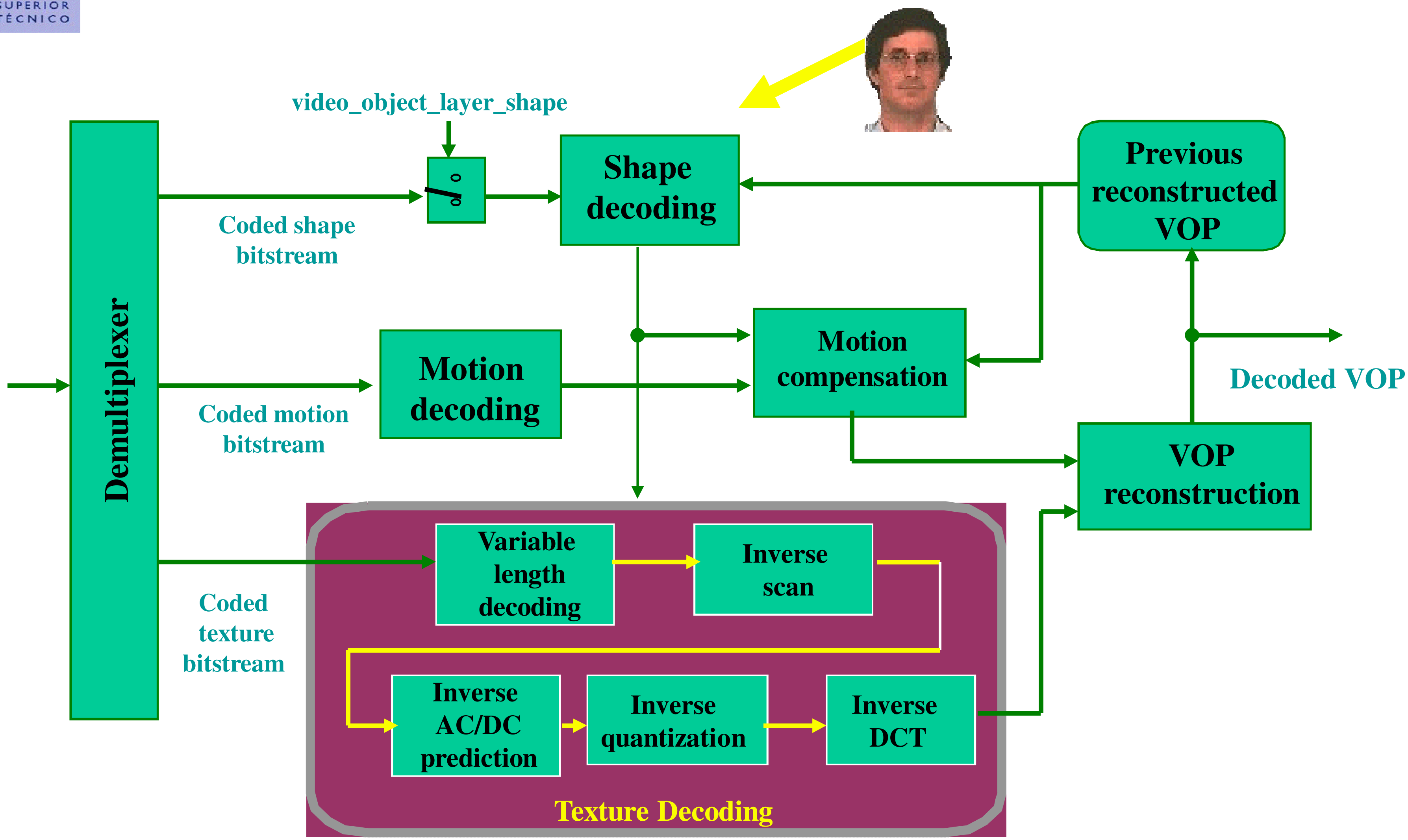
- Adopts the object-based model giving a semantic value to the data structure
- Integration of natural and synthetic content, both aural and visual
- Object-based functionalities, e.g., re-using and manipulation capabilities
- Powerful data model for interaction and personalisation
- Exploitation of synergies, e.g., between Video Coding, Computer Vision and Computer Graphics



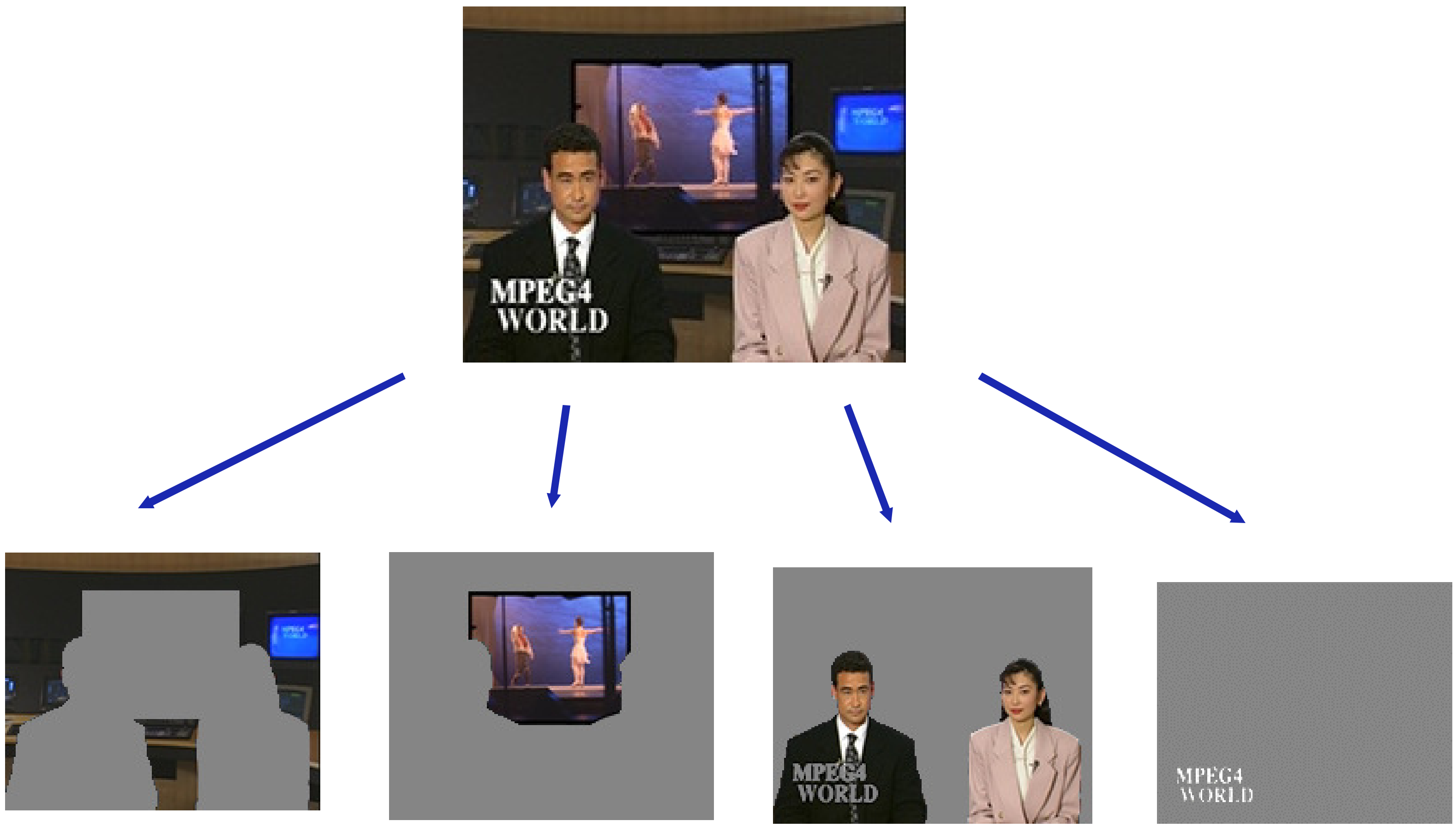
MPEG-4: Visual Coding Architecture



Basic MPEG-4 Video Decoding

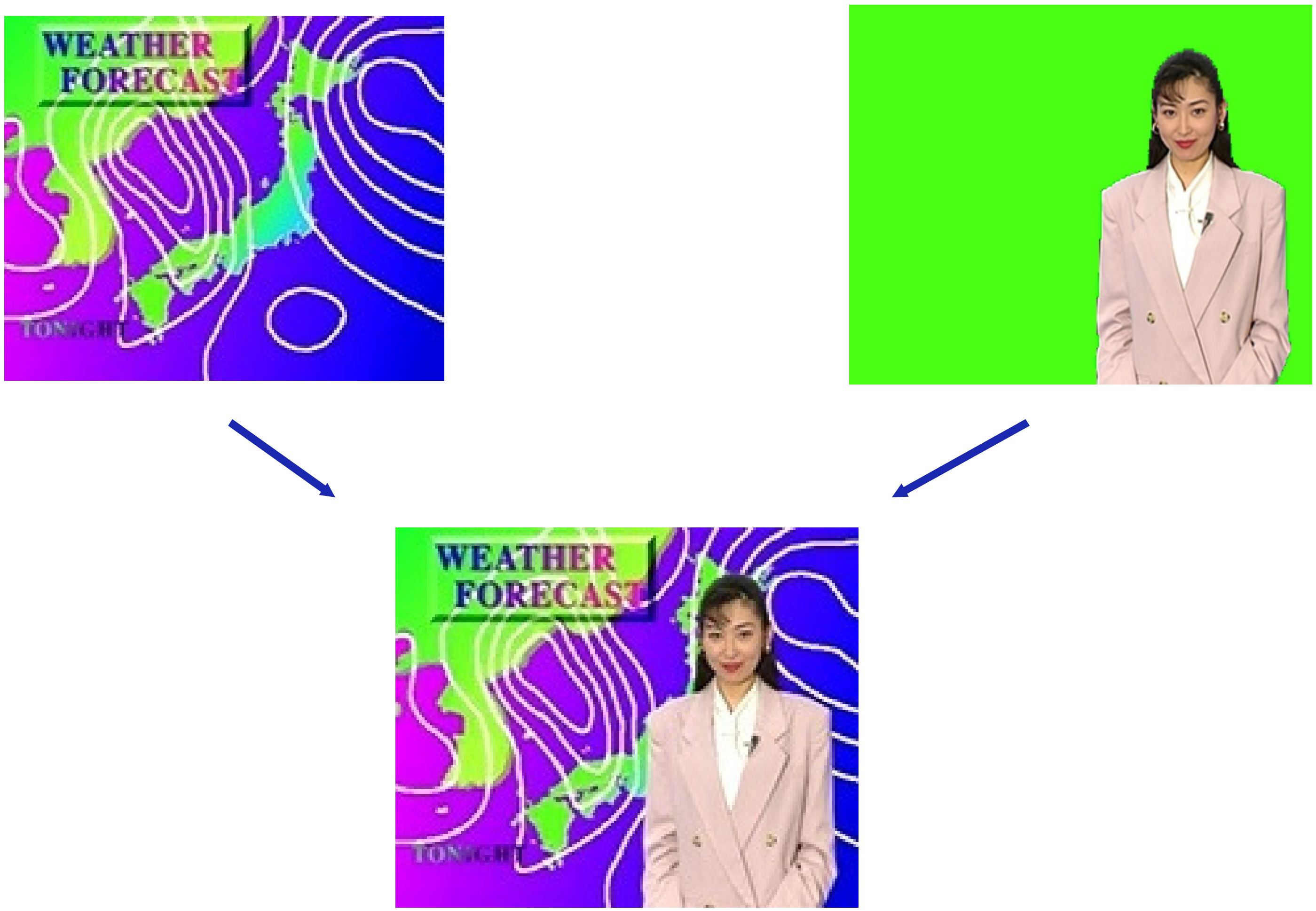


Segmentation: a Limitation or not so Much ?

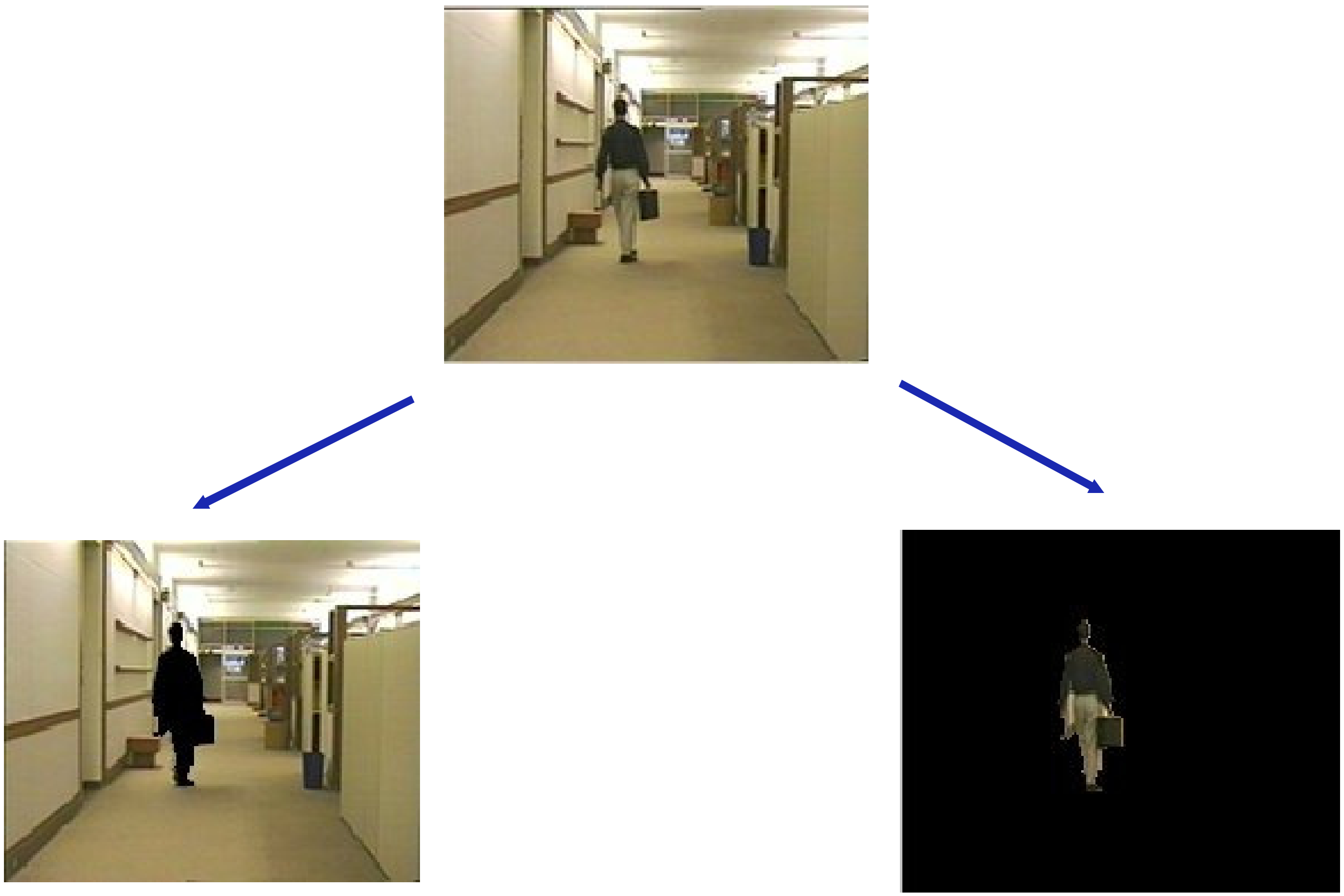


The ‘Weather’ Girl ...

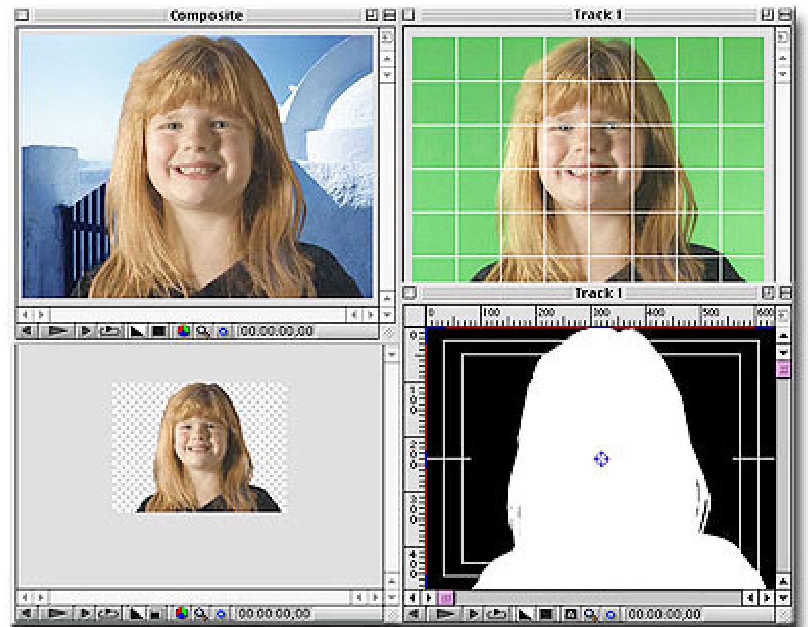
Segmentation: the Problem that Sometimes does not Exist ...



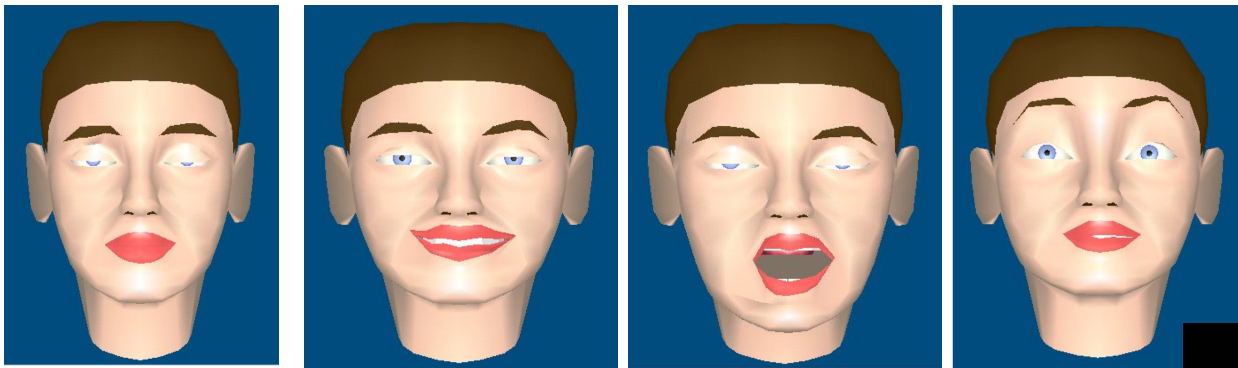
Segmentation: Automatic and Real-Time ?

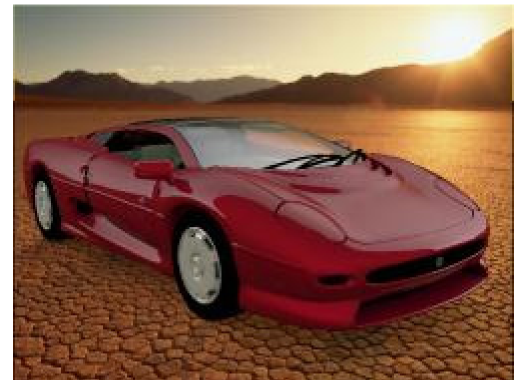


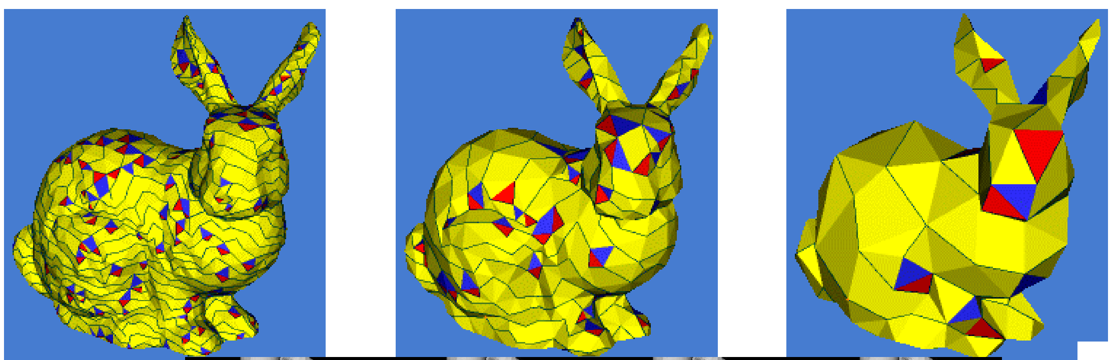
Segmentation by Chroma-Keying …

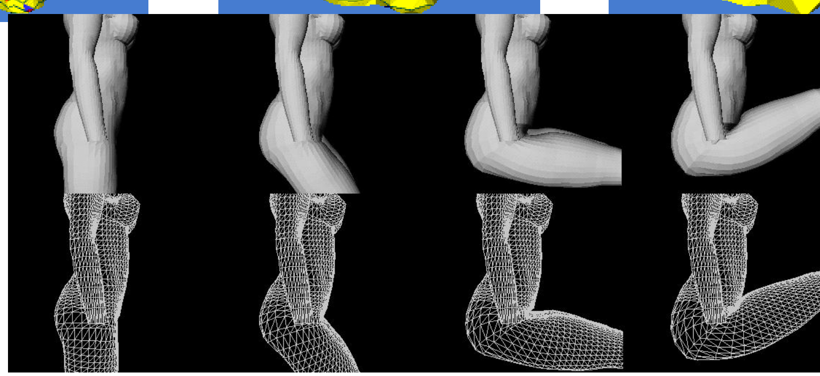


Synthetic Content: Facial Animation and More …









ıdio 'a

The MPEG-4 Tools (1): The Codecs

- Efficiently encode **video data** from very low bitrates, notably in view of low bitrate channels such as the telephone line or mobile environments, to very high quality conditions;
- Efficiently encode **music and speech data** for a very wide bitrate range, notably from transparent music to very low bitrate speech;
- Efficiently encode **text and graphics**;
- Efficiently encode **time-changing 3D generic objects** as well as some more specific 3D objects such as human faces and bodies;
- Efficiently encode **synthetically generated speech and music** as well as 3D audio spaces;
- Provide error resilience in the encoding layer for the various data types involved, notably in view of critical channel conditions;

The MPEG-4 Tools (2): Systems Tools

- Independently represent the various objects in the scene, notably visual objects, allowing to independently access, manipulate and re-use these objects;
- Compose aural and visual, natural and synthetic, objects in one audiovisual scene;
- Describe objects and events in the scene;
- Provide hyperlinking and interaction capabilities;
- Provide some means to protect audiovisual content so that only authorised users can consume it.

MPEG-4 Application Examples

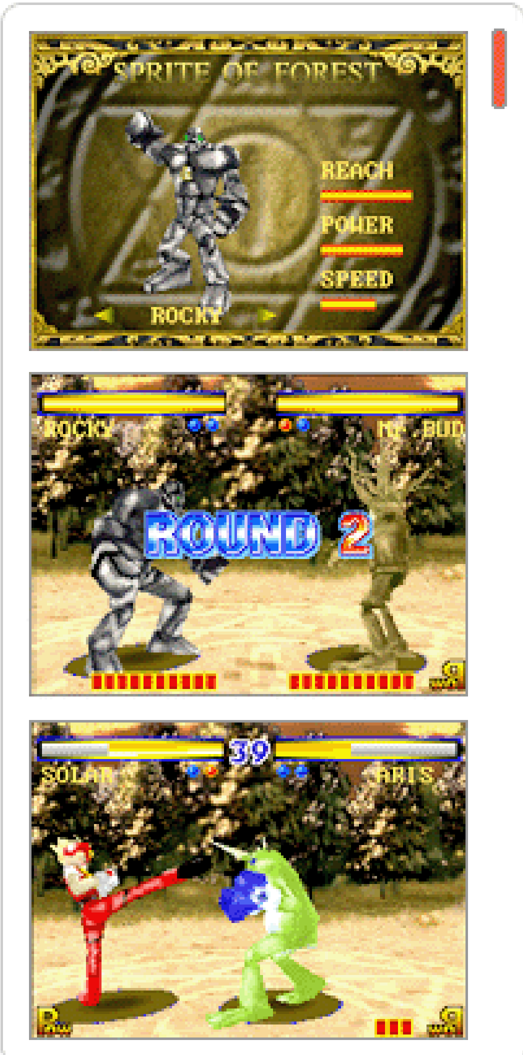
- Video streaming in the Internet/Intranet
- Advanced real-time (mobile) communications
- Multimedia broadcasting
- Video cameras
- Content-based storage and retrieval
- Studio and television post-production
- Interactive DVD
- Remote surveillance, monitoring
- Virtual meetings
- ...

The Bloomberg Case … Today !

- **Coding efficiency**
- **Automatic/manual customization of content**
- **Automatic/manual customization of screen layout based on:**
  - global content and objects, content-based AV events, language, complex user defined criteria, …

Using Objects …

3D Games for 3G … in Korea …



MPEG-4 Objects: Old is Also New ...

Video Coding in MPEG-4

There are two Parts in the MPEG-4 standard dealing with video coding:

- **Part 2: Visual (1998) – Specifies several coding tools targeting the efficient and error resilient of video, including arbitrarily shaped video; it also includes coding of 3D faces and bodies.**

- **Part 10: Advanced Video Coding (AVC) (2003) – Specifies more efficient (about 50%) and more resilient frame based video coding tools; this Part has been jointly developed by ISO/IEC MPEG and ITU-T through the Joint Video Team (JVT) and it is often known as H.264/AVC.**

Each of these 2 Parts specifies several profiles with different video coding functionalities and compression efficiency versus complexity trade-offs. Part 10 only addresses rectangular frames !

MPEG-4 Visual (Part 2) Profiles in the Market

Simple and *Advanced Simple* are the most used MPEG-4 Visual profiles !

- **The *Simple* profile is rather similar to Rec. ITU-T H.263 with the addition of some error resilience tools. There are many products in the market using this profile, notably video cameras.**

- **The *Advanced Simple* profile, more efficient, uses also global and ¼ pel motion compensation and allows to code interlaced video.**

MPEG-4 Advanced Video Coding (AVC), also ITU-T H.264

H.264/AVC (2003): The Objective

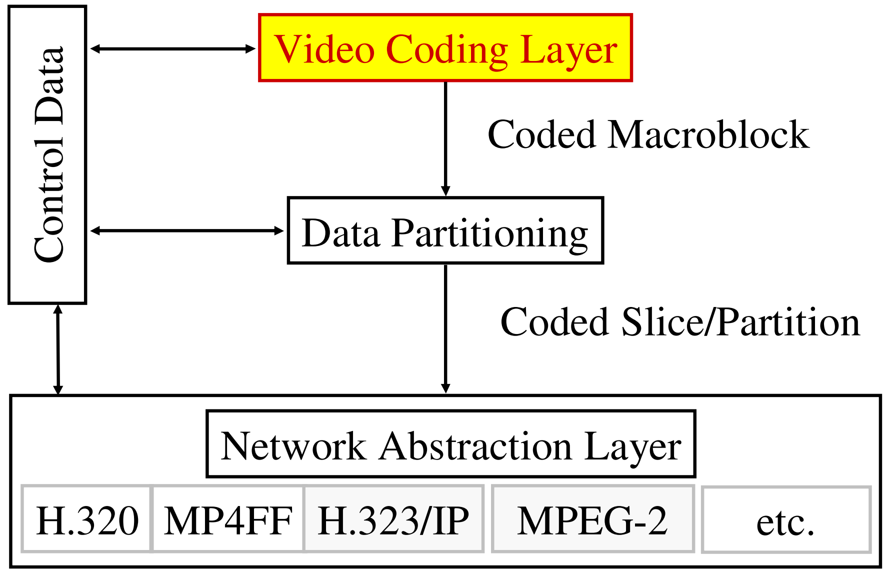
Coding of rectangular video with increased efficiency: about 50% less rate for the same quality regarding existing standards such as H.263, MPEG-2 Video and MPEG-4 Visual.

This standard (joint between ISO/IEC MPEG and ITU-T VCEG) offers also good flexibility in terms of efficiency-complexity trade-offs as well as good performance in terms of error resilience for mobile environments and fixed and wireless Internet (both progressive and interlaced formats).

Applications

- **Entertainment Video (1-8+ Mbps, higher latency)**
  - Broadcast / Satellite / Cable / DVD / VoD / FS-VDSL / …
  - DVB/ATSC/SCTE, DVD Forum, DSL Forum
- **Conversational Services (usually <1 Mbps, low latency)**
  - H.320 Conversational
  - 3GPP Conversational H.324/M
  - H.323 Conversational Internet/best effort IP/RTP
  - 3GPP Conversational IP/RTP/SIP
- **Streaming Services (usually lower bitrate, higher latency)**
  - 3GPP Streaming IP/RTP/RTSP
  - Streaming IP/RTP/RTSP (without TCP fallback)
- **Other Services**
  - 3GPP Multimedia Messaging Services

H.264/AVC Layer Structure



To address this need for flexibility and customizability, the H.264/AVC design covers:

- **A Video Coding Layer (VCL), which is designed to efficiently represent the video content**
- **A Network Abstraction Layer (NAL), which formats the VCL representation of the video and provides header information in a manner appropriate for conveyance by a variety of transport layers or storage media**

H.264/AVC Compression Gains: Why ?

The H.264/AVC standard is based on the same hybrid coding architecture used for previous video coding standards with some important differences:

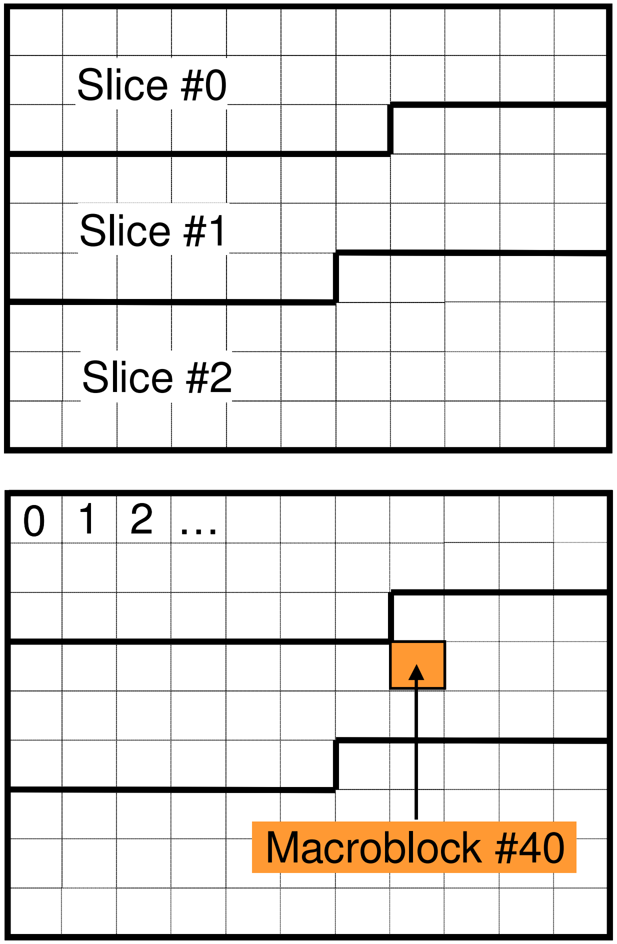
- **Variable (and smaller) block size motion compensation**
- **Multiple reference frames**
- **Hierarchical transform with smaller block sizes**
- **Deblocking filter in the prediction loop**
- **Improved, adaptive entropy coding**

which all together allow achieving substantial gains regarding the bitrate needed to reach a certain quality level.

The H.264/AVC standard addresses a vast set of applications, from personal communications to storage and broadcasting, at various qualities and resolutions.

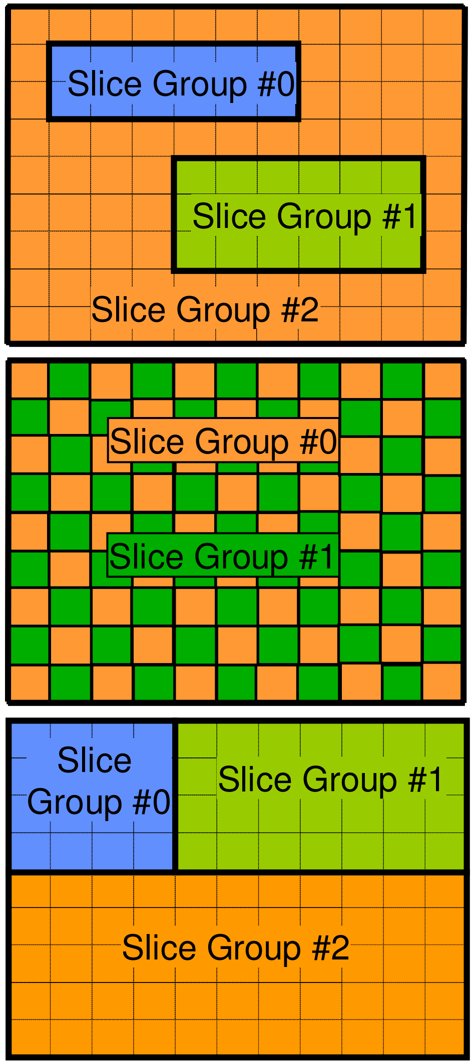
Partitioning of the Picture

- **Picture (Y,Cr,Cb; 4:2:0 and later more; 8 bit/sample):**
  - A picture (frame or field) is split into 1 or several slices
- **Slice:**
  - Slices are self-contained
  - Slices are a sequence of macroblocks
- **Macroblock:**
  - Basic syntax & processing unit
  - Contains 16×16 luminance samples and 2 × 8×8 chrominance samples (4:2:0 content)
  - Macroblocks within a slice depend on each other
  - Macroblocks can be further partitioned

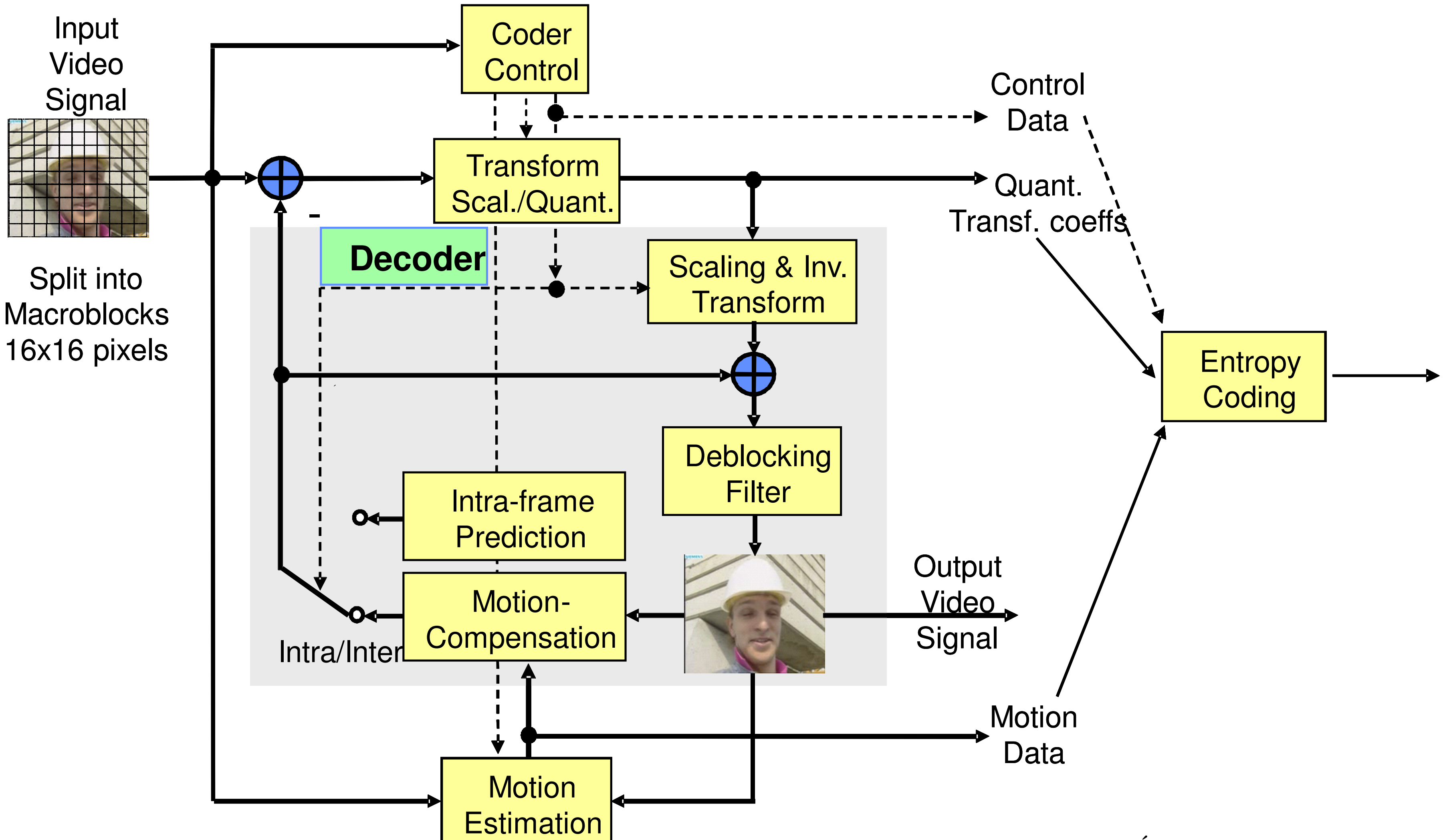


Slices and Slice Groups

- **Slice Group:**
  - Pattern of macroblocks defined by a Macroblock Allocation Map
  - A slice group may contain 1 to several slices
- **Macroblock Allocation Map Types:**
  - Interleaved slices
  - Dispersed macroblock allocation
  - Explicitly assign a slice group to each macroblock location in raster scan order
  - One or more “foreground” slice groups and a “leftover” slice group
- **Coding of Slices:**
  - I Slices: all MBs use only Intra prediction
  - P Slices: MBs may also use backward motion compensation
  - B Slices: MBs may also use bidirectional motion compensation



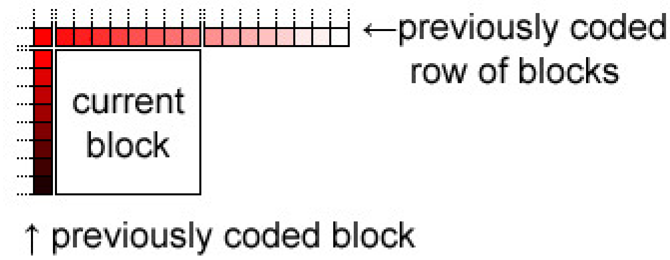
H.264/AVC Encoding Architecture



Common Elements with other Standards

- Original data: Luminance and two chrominances
- Macroblocks: 16 × 16 luminance + 2 × 8×8 chrominance samples
- Input: Association of luminance and chrominance with conventional sub-sampling of chrominance (4:2:0, 4:2:2, 4:4:4)
- Block motion displacement
- Motion vectors over picture boundaries
- Variable block-size motion
- Block transforms
- Scalar quantization
- I, P, and B coding types

Intra Prediction



- **To increase Intra coding compression efficiency, it is possible to exploit for each MB the correlation with adjacent blocks or MBs in the same picture.**
- **If a block or MB is Intra coded, a prediction block or MB is built based on the previously coded and decoded blocks or MBs in the same picture.**
- **The prediction block or MB is subtracted from the block or MB currently being coded.**
- **To guarantee slice independency, only samples from the same slice can be used to form the Intra prediction.**

This type of Intra coding may imply error propagation if the prediction uses adjacent MBs which have been Inter coded; this may be solved by using the so-called *Constrained Intra Coding Mode* where only adjacent Intra coded MBs are used to form the prediction.

Intra Prediction Types

Intra predictions may be performed in several ways:

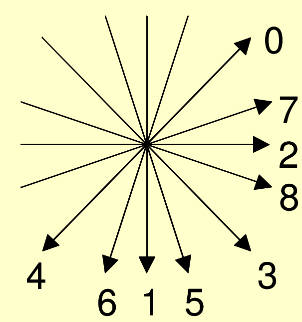
1. **Single prediction for the whole MB (Intra16×16): four modes are possible (vertical, horizontal, DC e planar) -> uniform areas !**

2. **Different predictions for the 16 samples of the several 4×4 blocks in a MB (Intra4×4): nine modes (DC and 8 direccionalmodes -> areas with detail !**

3. **Single prediction for the chrominance: four modes (vertical, horizontal, DC and planar)**

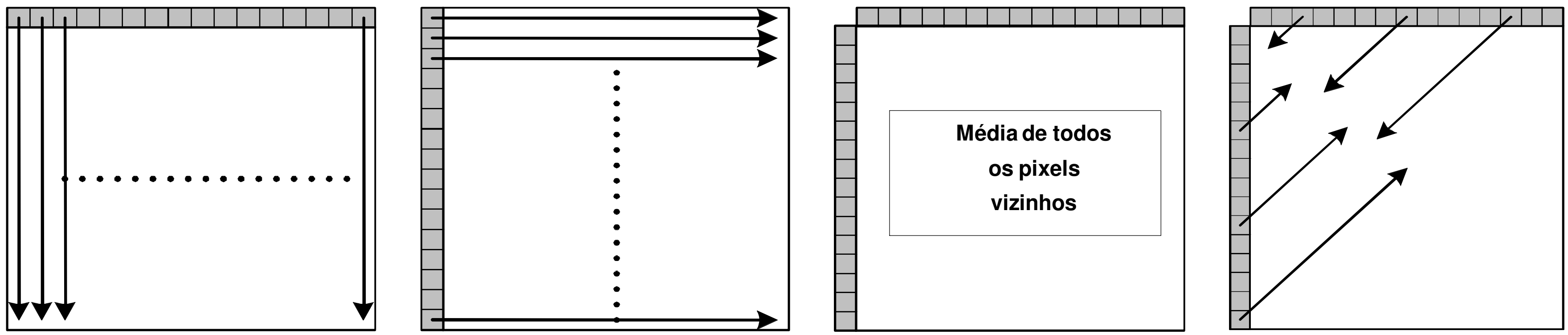
- Directional spatial prediction (9 types for luma, 1 chroma)

Q	A	B	C	D	E	F	G	H
I	a	b	c	d				
J	e	f	g	h				
K	i	j	k	l				
L	m	n	o	p				



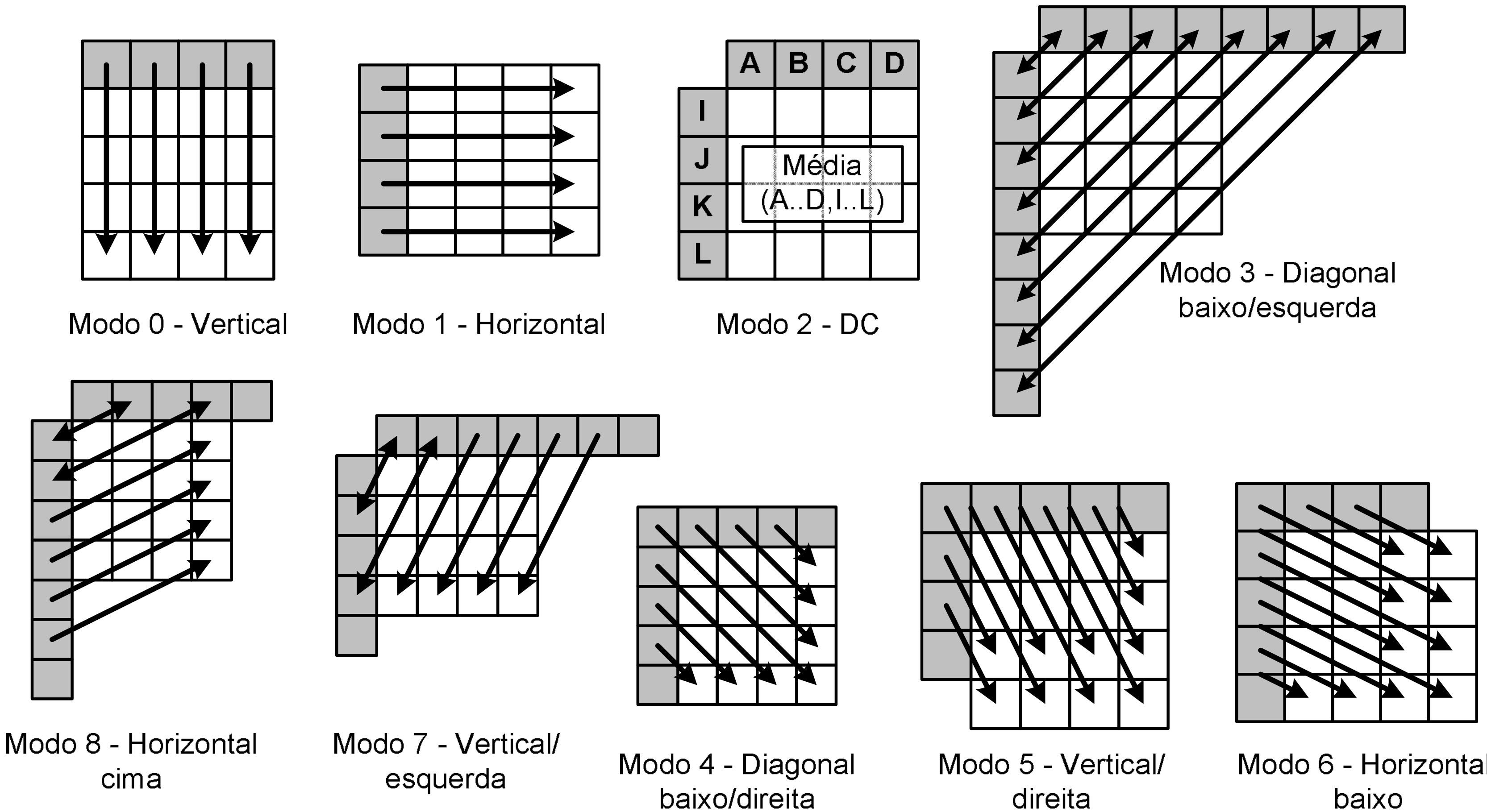
- e.g., Mode 3:
  diagonal down/right prediction
  a, f, k, p are predicted by
  (A + 2Q + I + 2) >> 2

16×16 Blocks Intra Prediction Modes

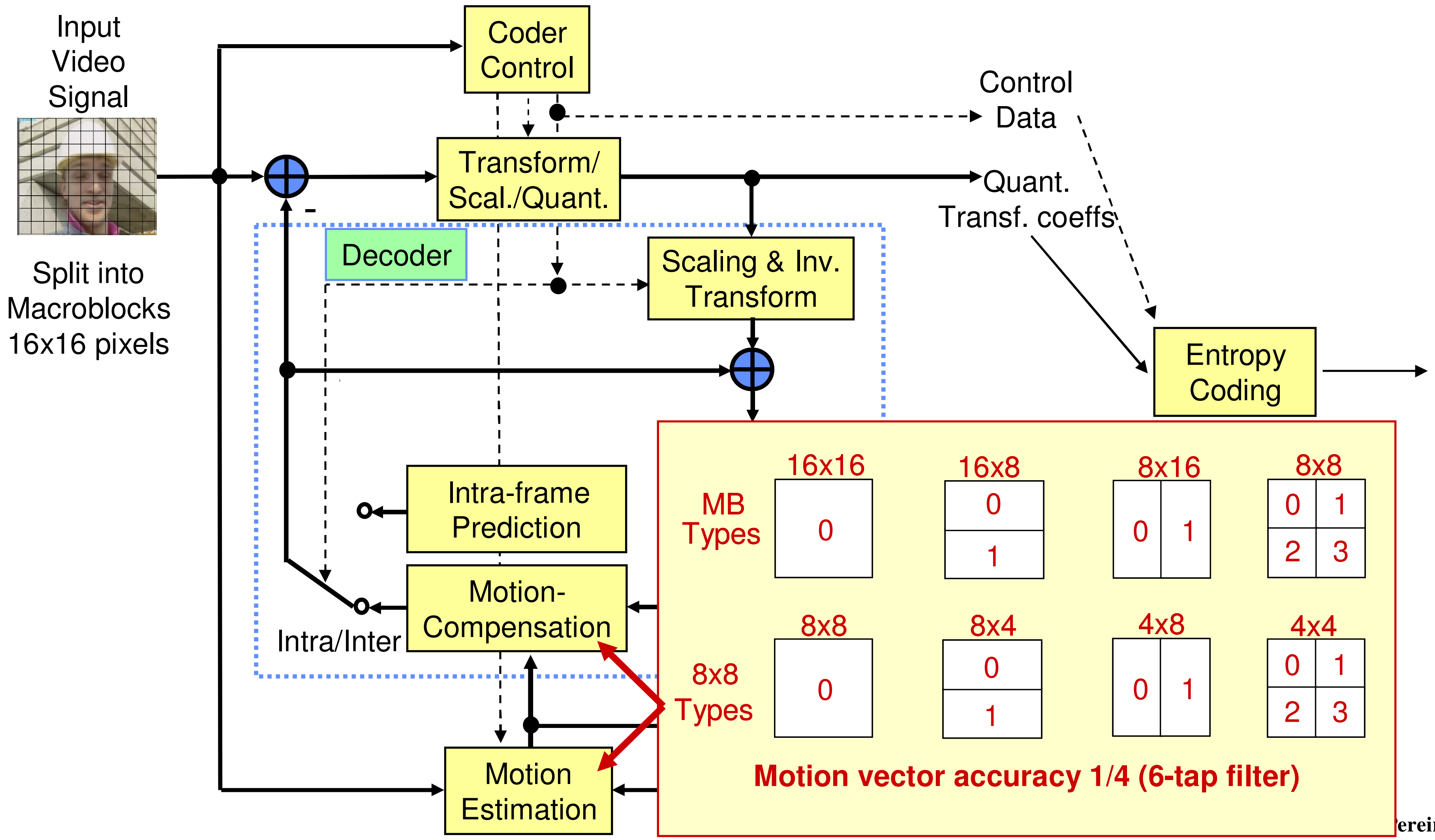


- **The luminance is predicted in the same way for all samples of a 16×16 MB (Intra16×16 modes).**
- **This coding mode is adequate for the image areas which have a smooth variation.**

4×4 Intra Prediction Directions



Variable Block- Size Motion Compensation



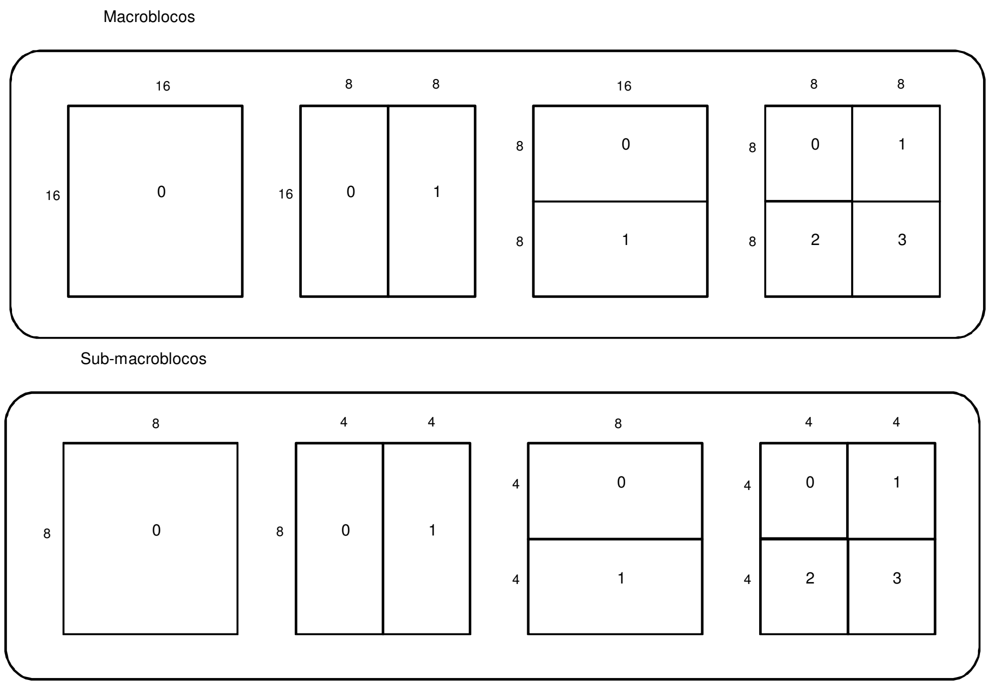
ereira

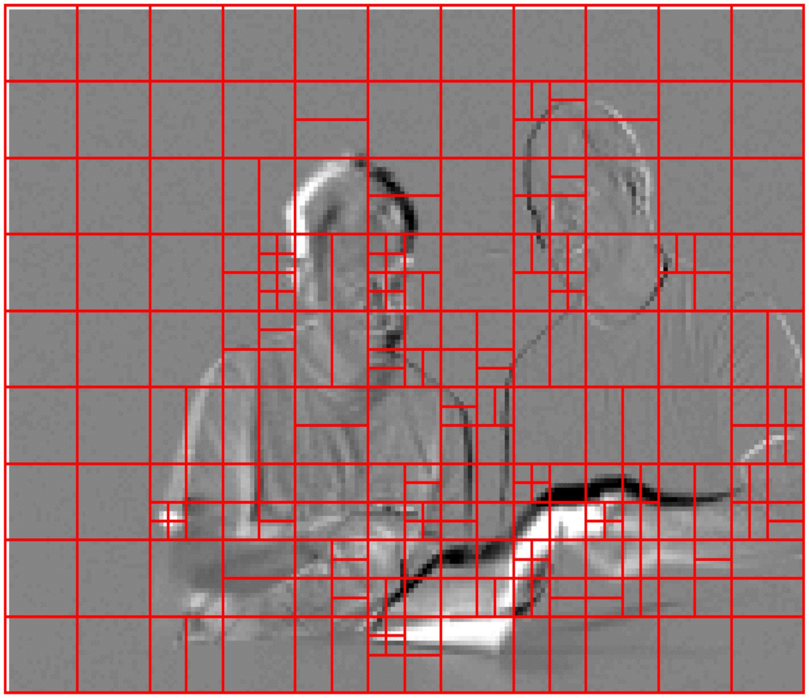
Flexible Motion Compensation

- Each MB may be divided into several fixed size partitions used to describe the motion with ¼ pel accuracy.
- There are several partition types, from 4×4 to 16×16 luminance samples, with many options between the two limits.
- The luminance samples in a MB (16×16) may be divided in four ways - Inter16×16, Inter16×8, Inter8×16 and Inter8×8 – corresponding to the four prediction modes at MB level.
- For P-slices, if the Inter8×8 mode is selected, each sub-MB (with 8×8 samples) may be divided again (or not), obtaining 8×8, 8×4, 4×8 and 4×4 partitions which correspond to the four predictions modes at sub-MB level.

For example, a maximum of 16 motion vectors may be used for a P coded MB.

MBs and sub-MBs Partitioning for Motion Compensation





Motion vectors are differentially coded but not across slices.

Intra MB

SKIP

Inter 16x16

Inter 8x8

Inter 8x16

Inter 16x8

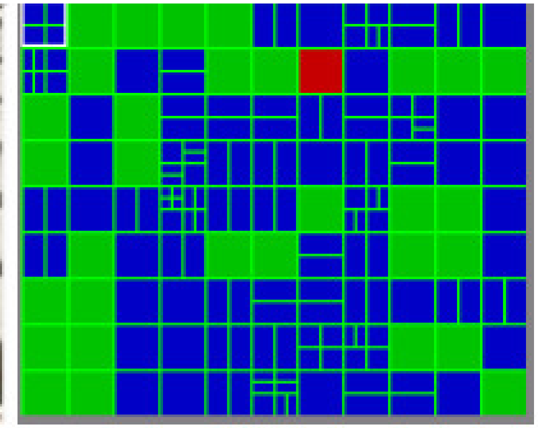
Inter 8x8 sub-MB

Inter 4x4 sub-MB

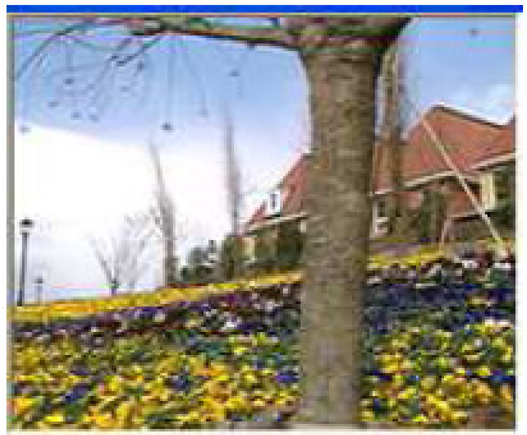
Inter 4x8 sub-MB

Inter 8x4 sub-MB

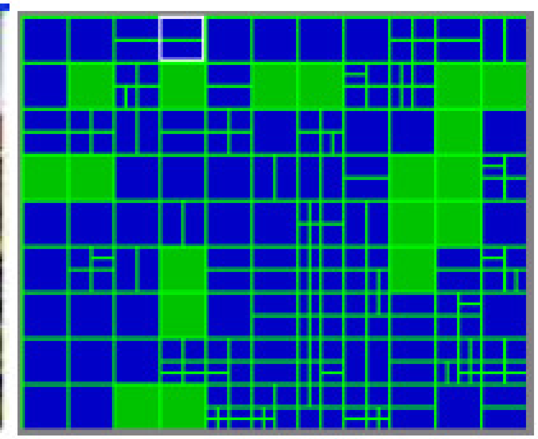
(a) Foreman second frame



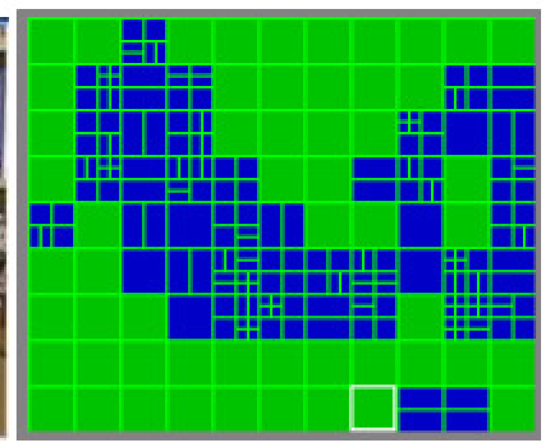
(b) Foreman second frame mode decision



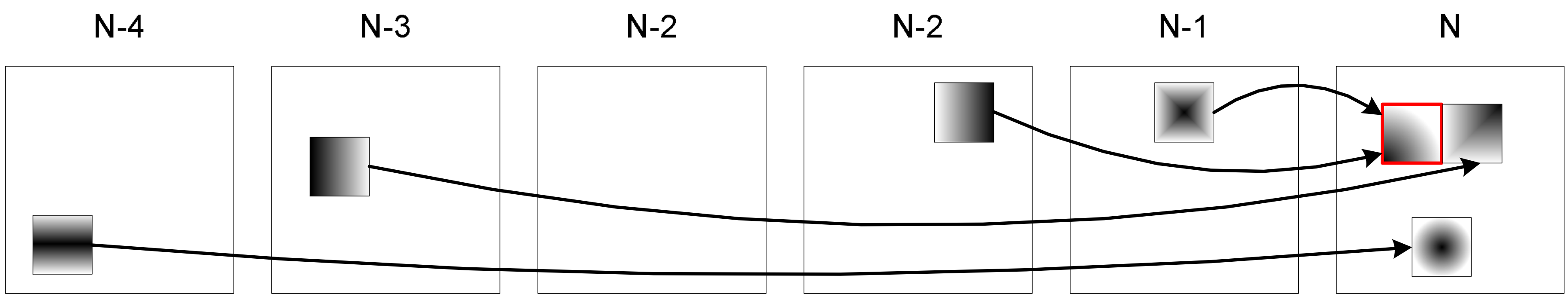
(c) Flower second frame



(d) Flower second frame mode decision



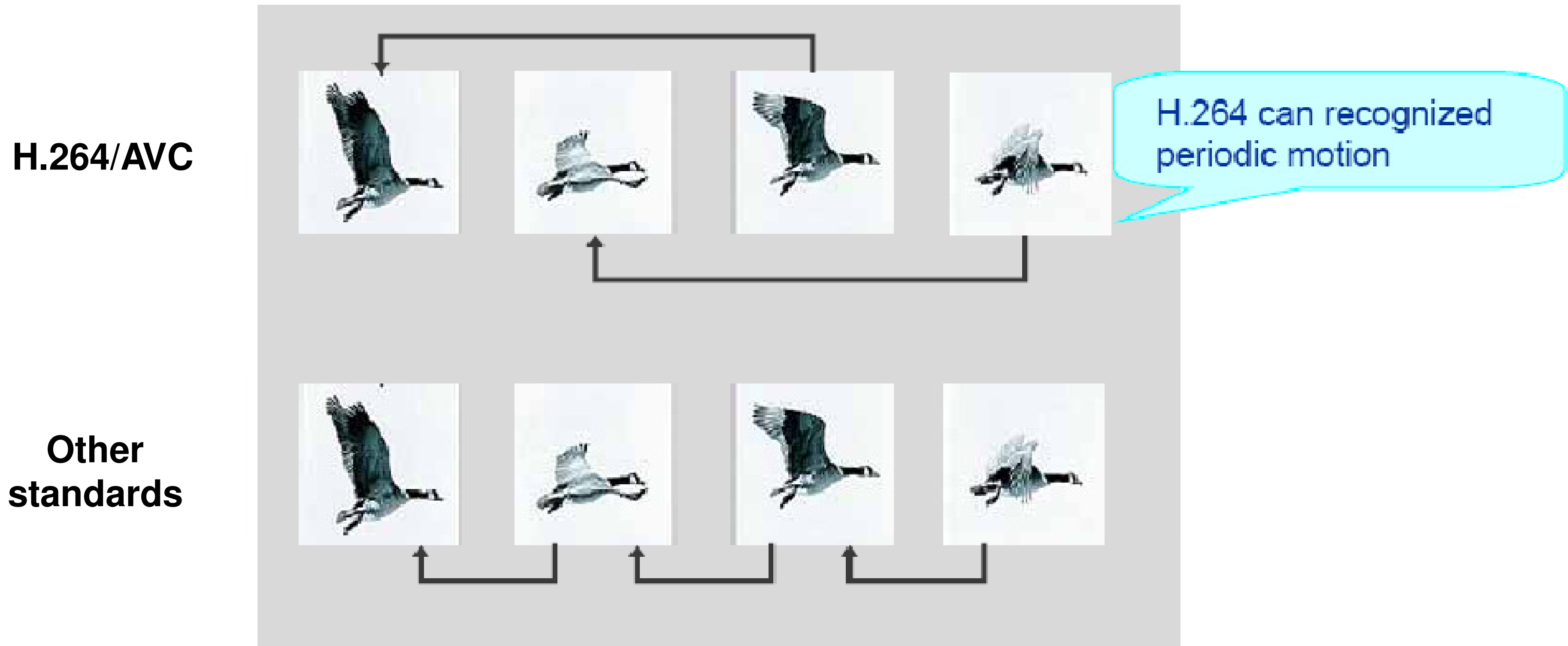
Multiple Reference Frames



The H.264/AVC standard supports motion compensation with multiple reference frames this means that more than one previously coded picture may be simultaneously used as prediction reference for the motion compensation of the MBs in a picture (at the cost of memory and computation).

- **Both the encoder and the decoder store the reference frames in a memory with multiple frames; up to 16 reference frames are allowed.**
- **The decoder stores in the memory the same frames as the encoder; this is guaranteed by means of memory control commands which are included in the coded bitstream.**

The Benefits of Multiple Reference Frames



Generalized B Frames

The B frame concept is generalized in the H.264/AVC standard since now any frame may use as prediction reference for motion compensation also the B frames; this means the selection of the prediction frames only depends on the memory management performed by the encoder.

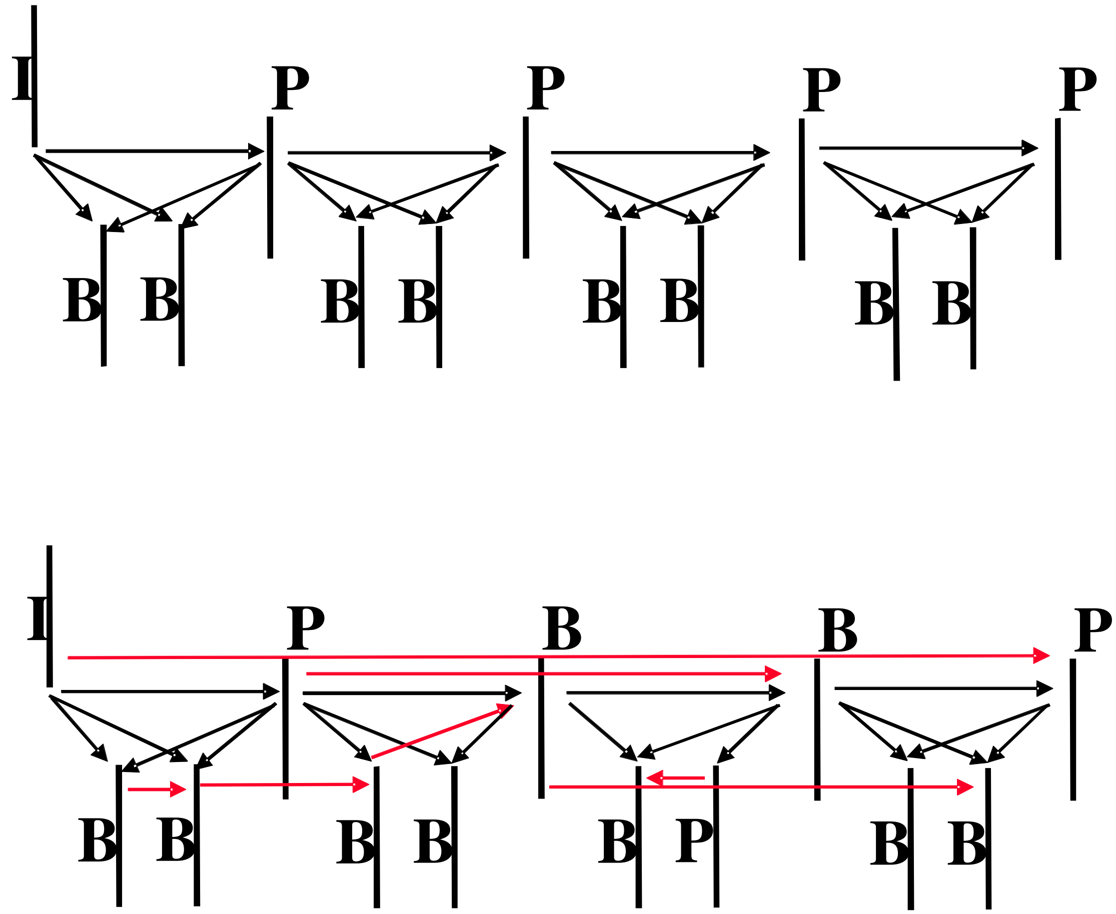
- **For B slices, some blocks or MBs are coded using a weighted prediction of two blocks or MBs in two reference frames, both in the past, both in the future, or one in the past and another in the future.**
- **B type frames use two reference frames, referred as the first and second reference frames.**
- **The selection of the two reference frames to use depends on the encoder.**
- **The weighted prediction allows to reach a more efficient Inter coding this means with a lower prediction error.**

New Types of Temporal Referencing

Known dependencies, e.g. MPEG-1 Video, MPEG-2 Video, etc.

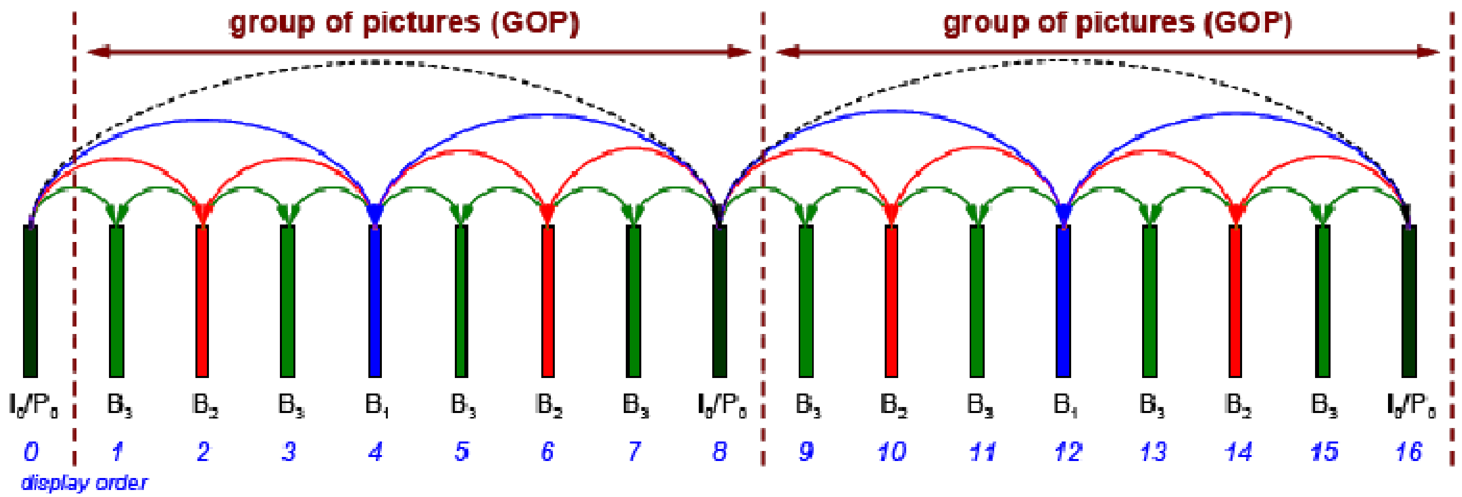
New types of dependencies:

- Referencing order and display order are decoupled, e.g. a P frame may not use for prediction the previous P frames
- Referencing ability and picture type are decoupled, e.g. it is possible to use a B frame as reference

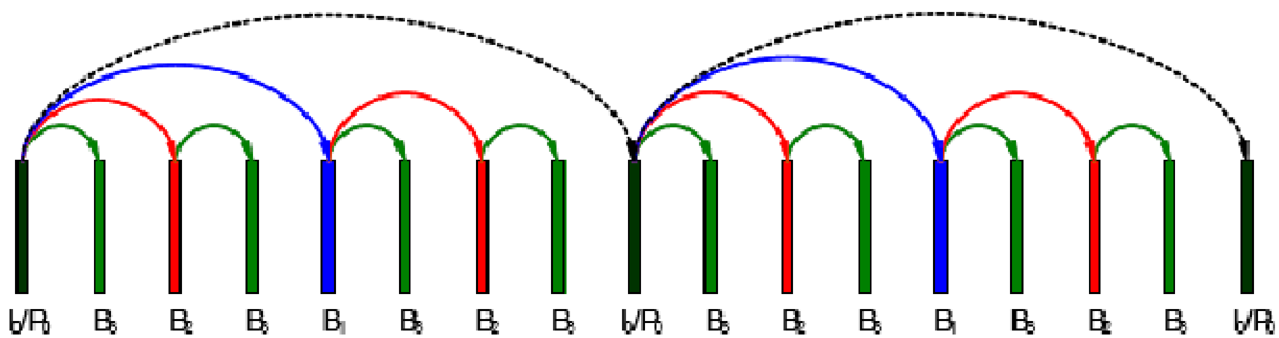


Hierarchical Prediction Structures

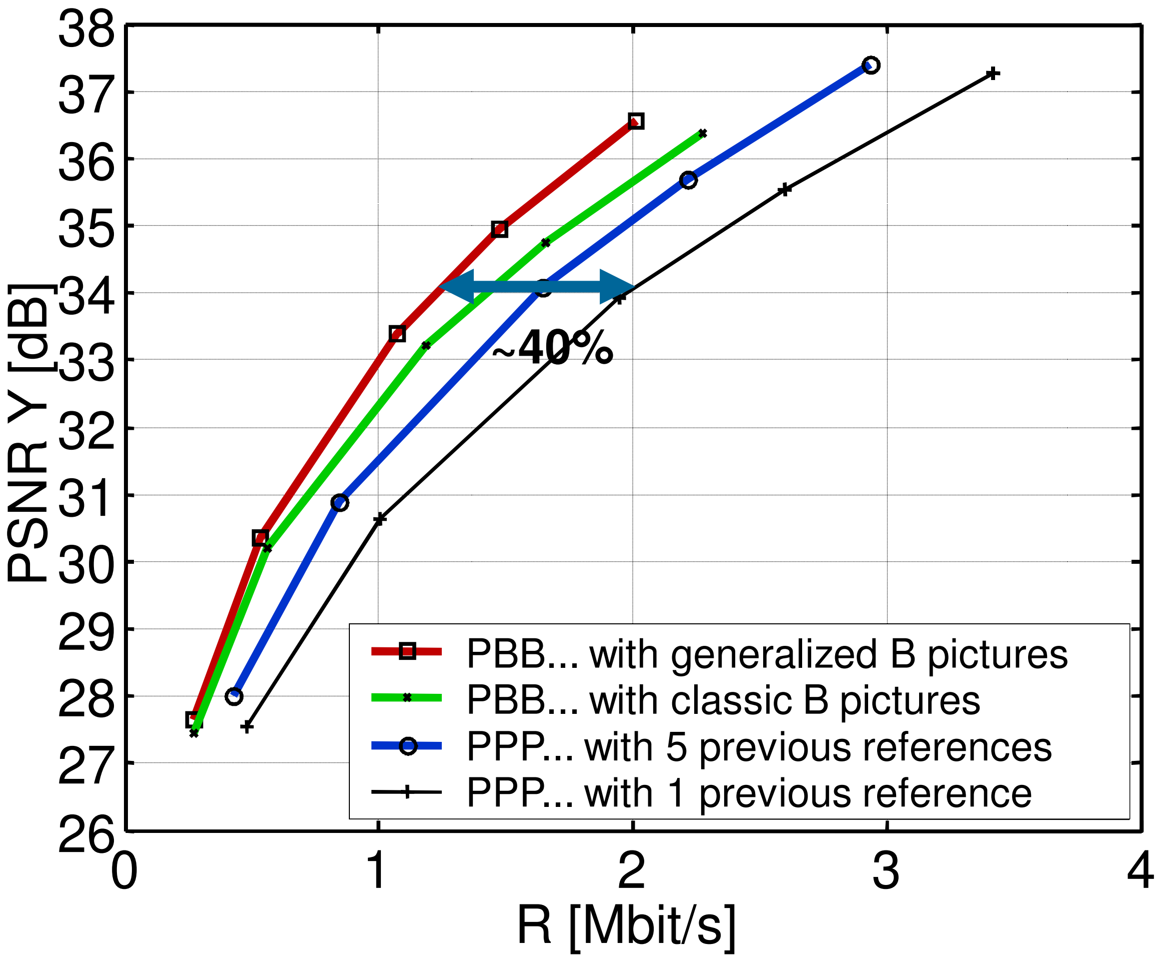
- Dyadic temporal scalability



- Low-delay prediction structure (structural delay is 0)



Comparative Performance: Mobile & Calendar, CIF, 30 Hz

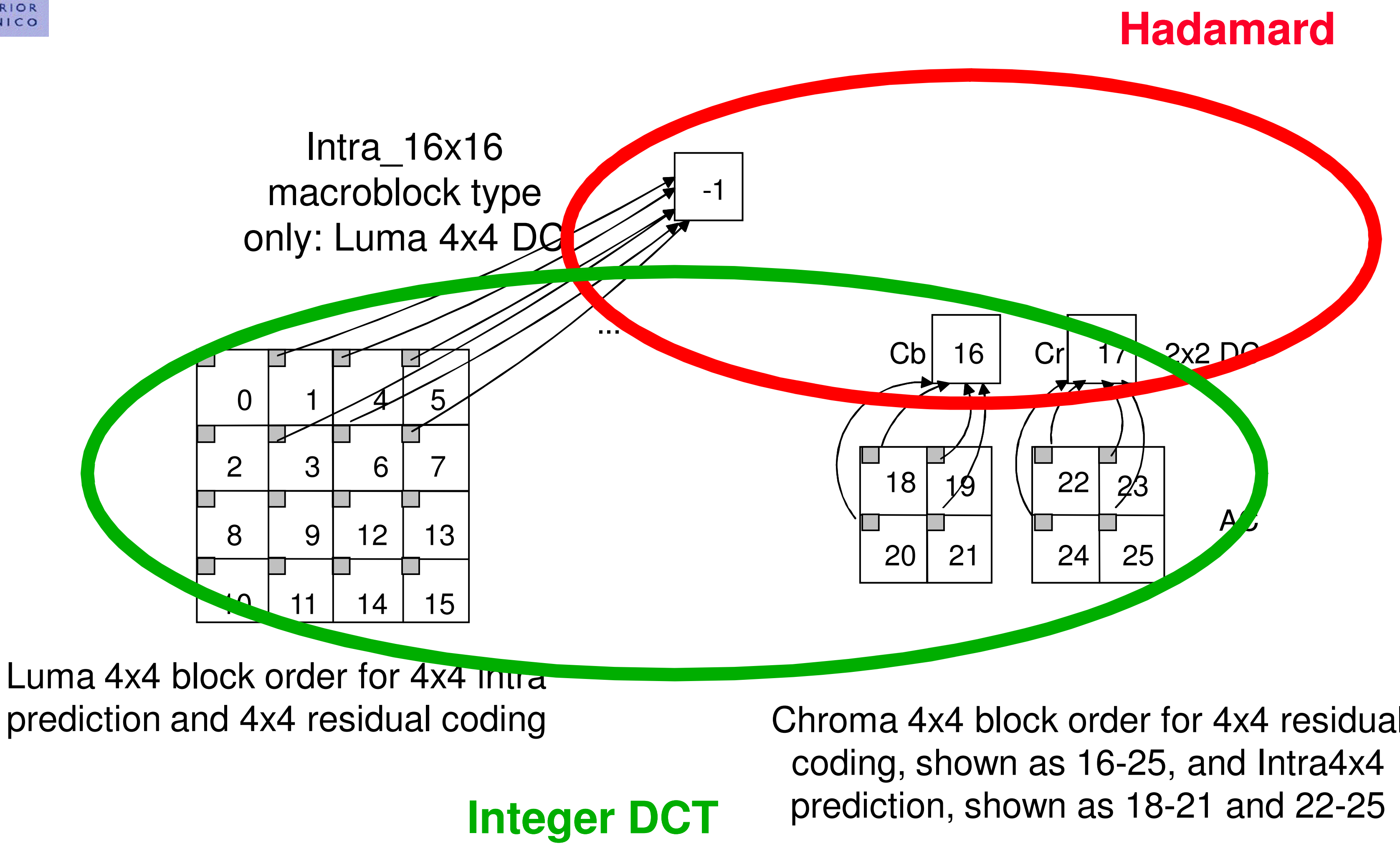


Multiple Transforms

The H.264/AVC standard uses three transforms depending on the type of prediction residue to code:

1. 4×4 Hadamard Transform for the luminance DC coefficients in MBs coded with the Intra 16×16 mode
2. 2×2 Hadamard Transform for the chrominance DC coefficients in any MB
3. 4×4 Integer Transform based on DCT for all the other blocks

Transforming, What ?



Integer DCT Transform

The H.264/AVC standard uses transform coding to code the prediction residue.

- **The transform is applied to 4×4 blocks using a separable transform with properties similar to a 4×4 DCT**

$$C_{4x4} = T_v \cdot B_{4x4} \cdot T_h^T$$

- **T_v, T_h: vertical and horizontal transform matrixes**

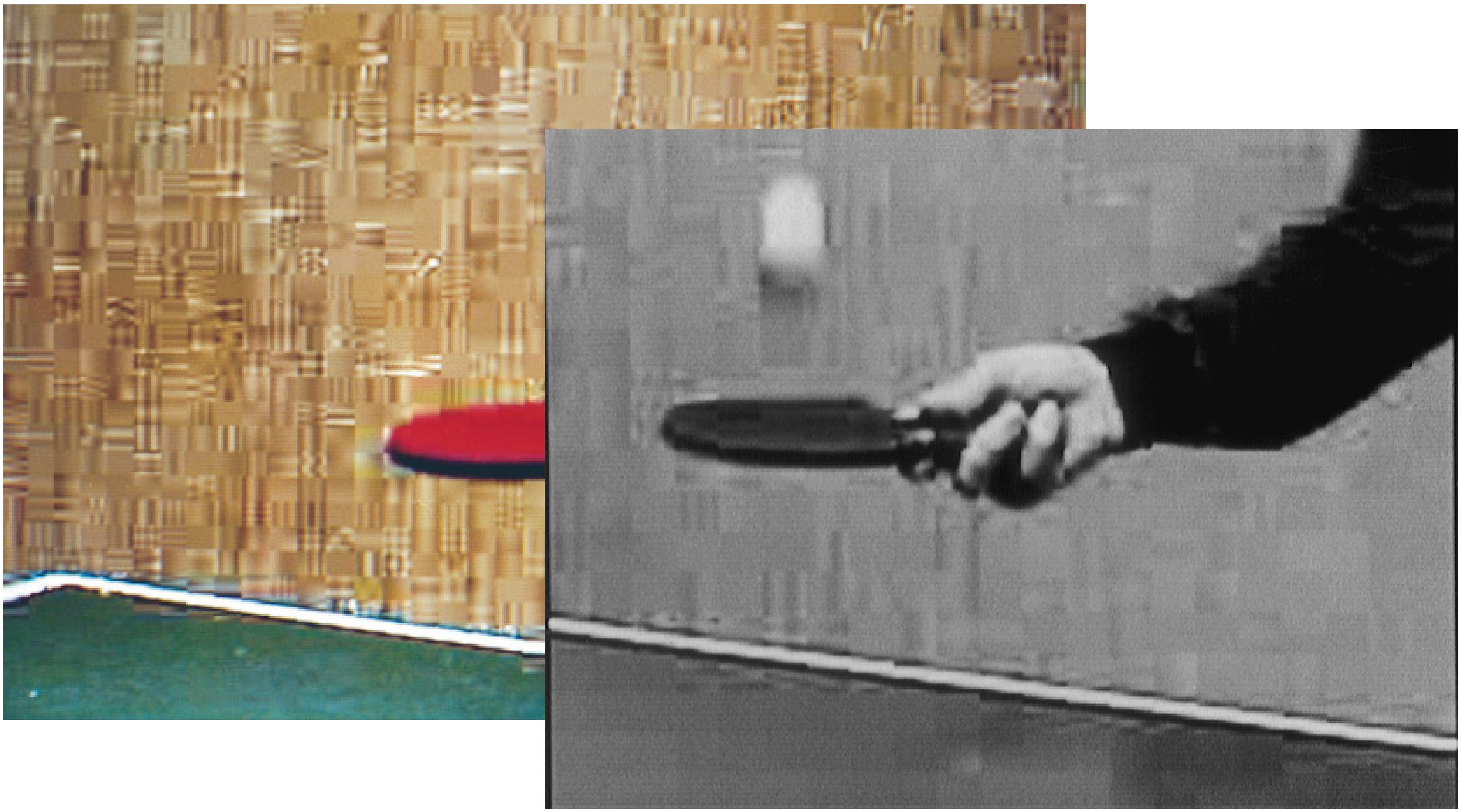
$$T_v = T_h = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

- **4×4 Integer DCT Transform**
  - **Easier to implement (only sums and shifts)**
  - **No mismatch in the inverse transform**

Quantization

- Quantization removes irrelevant information from the pictures to obtain a rather substantial bitrate reduction.
- Quantization corresponds to the division of each coefficient by a quantization factor while inverse quantization (reconstruction) corresponds to the multiplication of each coefficient by the same factor (there is a quantization error involved ...).
- In H.264/AVC, scalar quantization is performed with the same quantization factor for all the transform coefficients in the MB; some changes in this respect were made later.
- One out of 52 possible values for the quantization factor (Q_{step}) is selected for each MB indexed through the quantization step (Q_p) using a table which defines the relation between Q_p and Q_{step}.
- The table above has been defined in order to have a reduction of approximately 12.5% in the bitrate for an increment of 1 in the quantization step value, Q_{step}.

The Block Effect …



Deblocking Filter in the Loop (1)

The H.264/AVC standard specifies the use of an adaptive deblocking filter which operates at the block edges with the target to increase the final subjective and objective qualities.

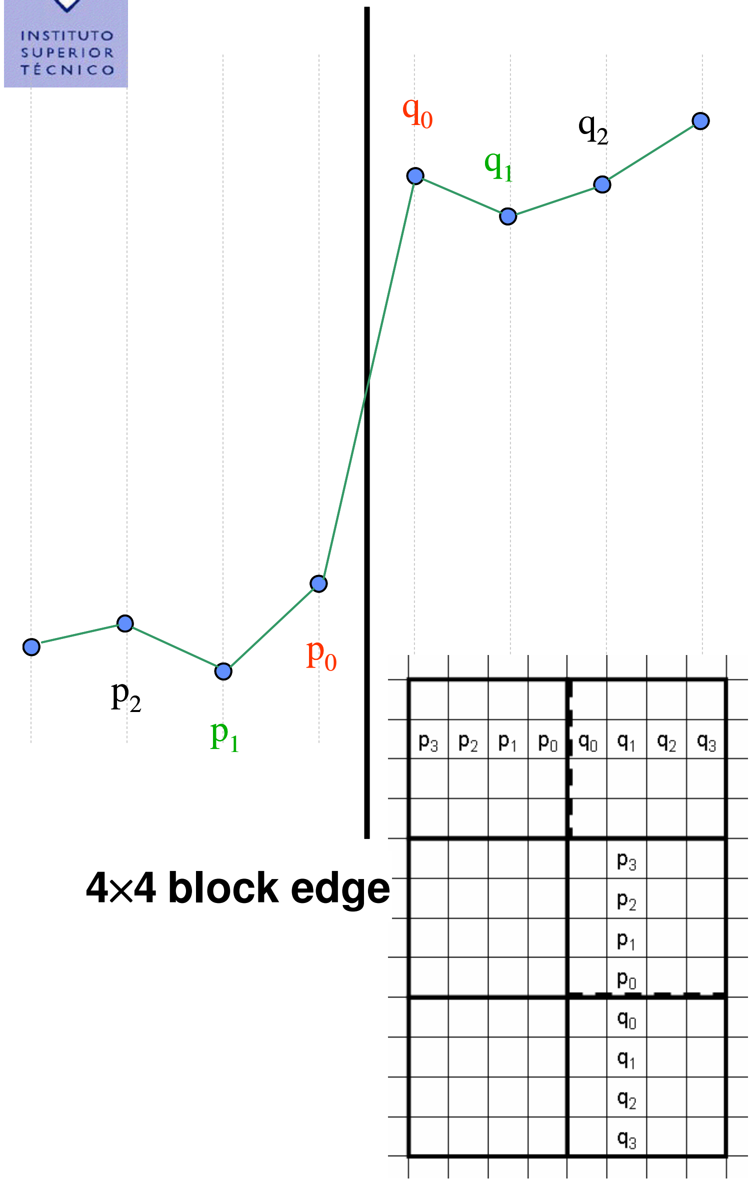
- **This filter needs to be present at the encoder and decoder (normative at decoder) since the filtered blocks are after used for motion estimation (filter in the loop). This filter has a superior performance to a post-processing filter (not in the loop and thus not normative).**
- **This filter has the following advantages:**
  - **Blocks edges are smoothed without making the image blurred, improving the subjective quality.**
  - **The filtered blocks are used for motion compensation resulting in smaller residues after prediction, this means reducing the bitrate for the same target quality.**
  - **The filter is applied to the vertical and horizontal edges of all 4×4 blocks in a MB.**

Deblocking Filter in the Loop (2)

- **The basic idea of the deblocking filter is that a big difference between samples at the edges of 2 blocks should only be filtered if it can be attributed to quantization; otherwise, that difference must come from the image itself and, thus, should not be filtered.**

- **The filter is adaptive to the content, essentially removing the block effect without unnecessarily smoothing the image:**
  - **At slice level, the filter strength may be adjusted to the characteristics of the video sequence.**
  - **At the edge block level, the filter strength is adjusted depending on the type of coding (Intra or Inter), the motion and the coded residues.**
  - **At the sample level, the filter may be switched off depending on the type of quantization.**
  - **The adaptive filter is controlled through a parameter B_s which defines the filter strenght; for $Bs = 0$, no sample is filtered while for $B_s = 4$ the filter reduces the most the block effect.**

Principle of Deblocking Filter



One dimensional visualization of an edge position

Filtering of p_0 *and* q_0 only takes place if:

1. $|p_0 - q_0| < \alpha(QP)$
2. $|p_1 - p_0| < \beta(QP)$
3. $|q_1 - q_0| < \beta(QP)$

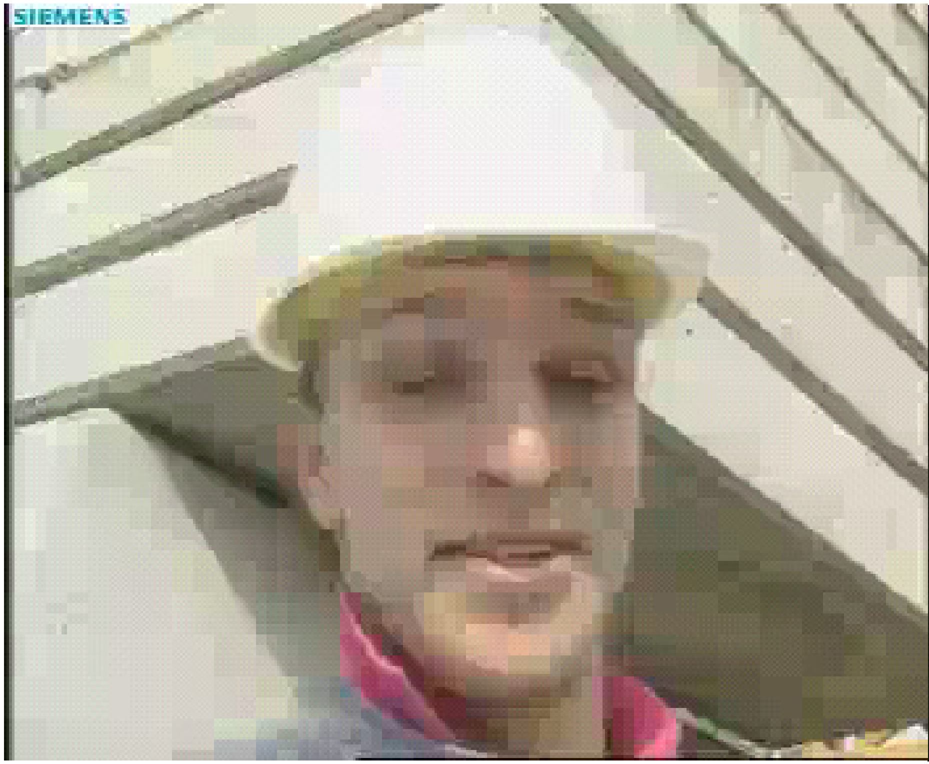
Where $\beta(QP)$ is considerably smaller than $\alpha(QP)$

Filtering of p_1 *or* q_1 takes place if additionally :

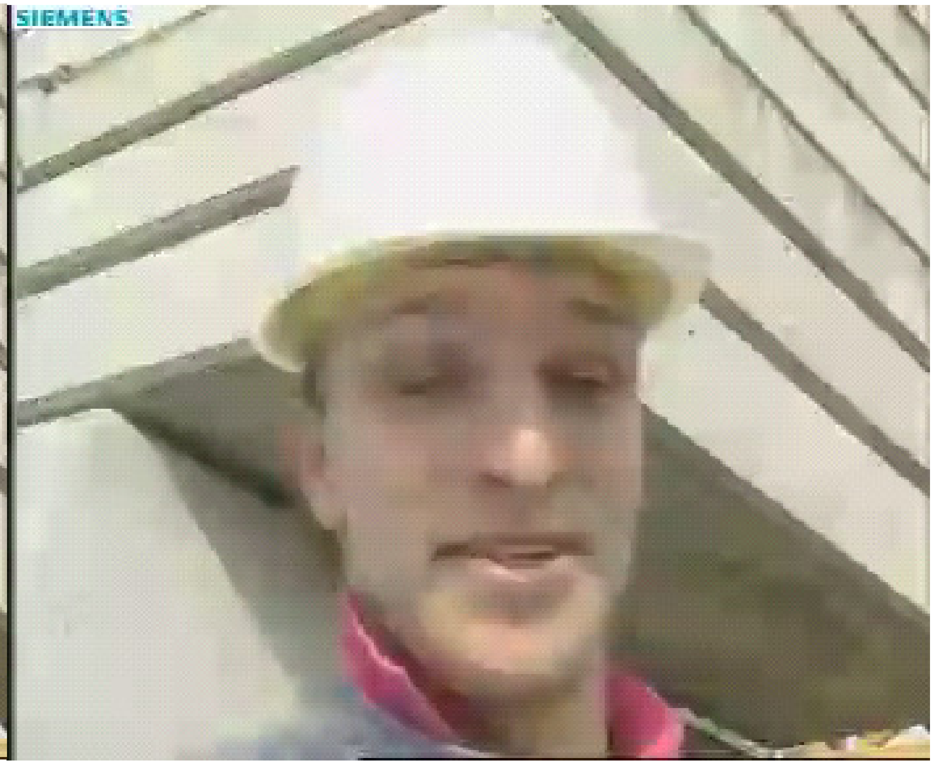
1. $|p_2 - p_0| < \beta(QP)$ *or* $|q_2 - q_0| < \beta(QP)$

(QP = quantization parameter)

Deblocking: Subjective Result for Intra Coding at 0.28 bit/sample

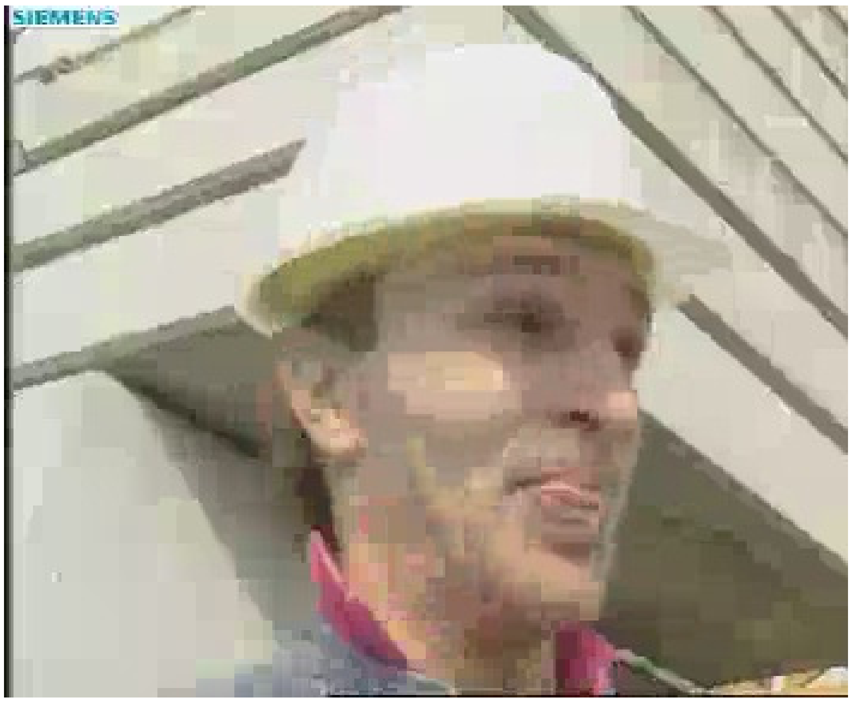


1) Without filter



2) With H.264/AVC deblocking

Deblocking: Subjective Result for Strong Inter Coding



1) Without Filter

2) With H.264/AVC deblocking

Entropy Coding

1 0 1 0 0 1 1 0 0 1 0 ...

SOLUTION 1

- **Exp-Golomb Codes are used for all symbols with the exception of the transform coefficients**
- ***Context Adaptive VLCs* (CAVLC) are used to code the transform coefficients**
  - **No *end-of-block is used*; the number of coefficients is decoded**
  - **Coefficients are scanned from the end to the beginning**
  - **Contexts depend on the coefficients themselves**

SOLUTION 2 (5-15% less bitrate)

- ***Context-based Adaptive Binary Arithmetic Codes* (CABAC)**
  - **Adaptive probability models are used for the majority of the symbols**
  - **The correlation between symbols is exploited through the creation of contexts**

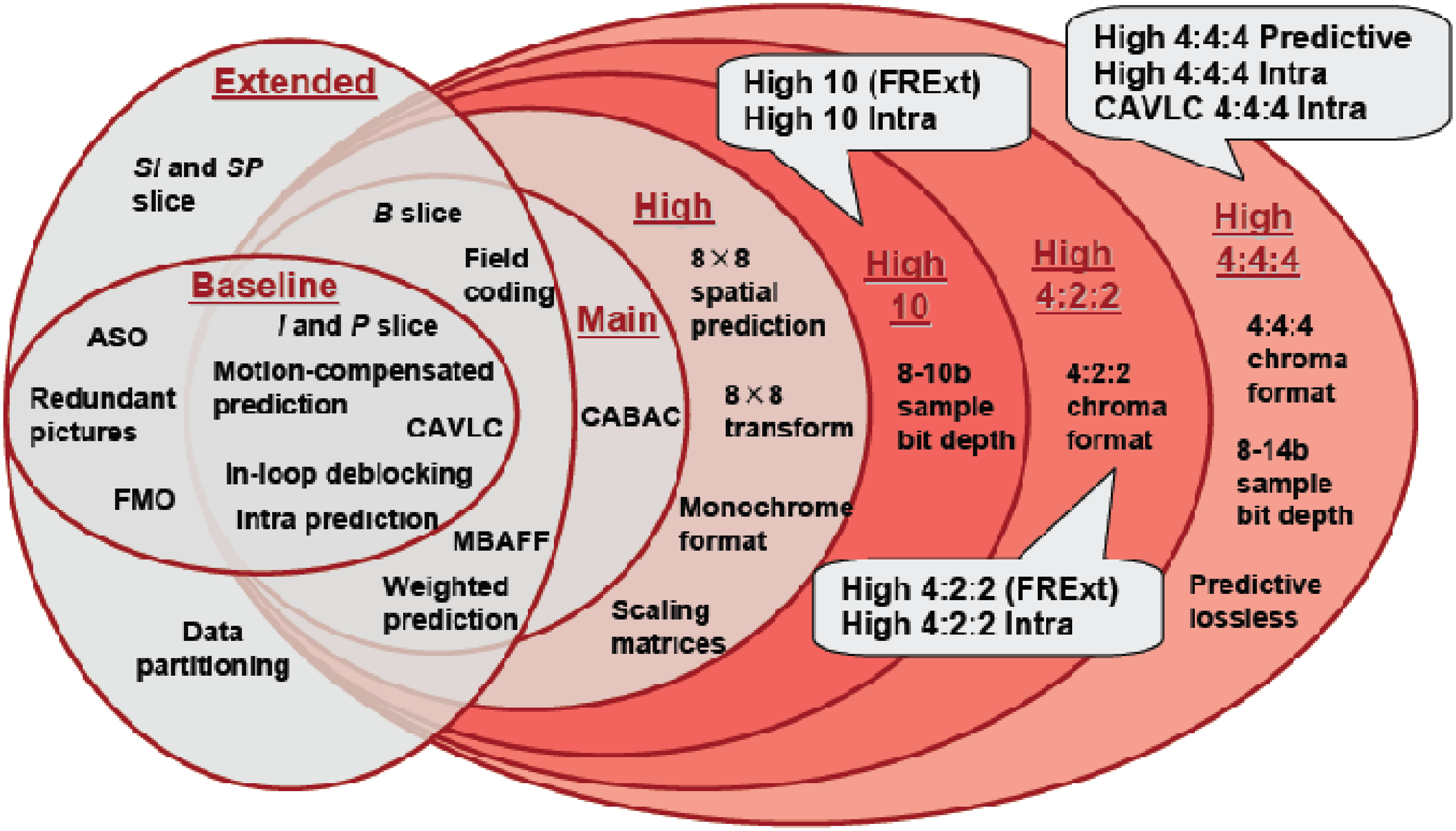
Adding Complexity to Buy Quality

Complexity (memory and computation) typically increases 4× at the encoder and 3× at the decoder regarding MPEG-2 Video, Main profile.

Problematic aspects:

- **Motion compensation with smaller block sizes (memory access)**
- **More complex (longer) filters for the ¼ pel motion compensation (memory access)**
- **Multiframe motion compensation (memory and computation)**
- **Many MB partitioning modes available (encoder computation)**
- **Intra prediction modes (computation)**
- **More complex entropy coding (computation)**

H.264/AVC Profiles …



H.264/AVC: a Success Story …

- 3GPP (recommended in rel 6)
- 3GPP2 (optional for streaming service)
- ARIB (Japan mobile segment broadcast)
- ATSC (preliminary adoption for robust-mode back-up channel)
- Blu-ray Disc Association (mandatory for Video BD-ROM players)
- DLNA (optional in first version)
- DMB (Korea - mandatory)
- DVB (specified in TS 102 005 and one of two in TS 101 154)
- DVD Forum (mandatory for HD DVD players)
- IETF AVT (RTP payload spec approved as RFC 3984)
- ISMA (mandatory specified in near-final rel 2.0)
- SCTE (under consideration)
- US DoD MISB (US government preferred codec up to 1080p)
- (and, of course, MPEG and the ITU-T)

H.264/AVC Patent Licensing

- As with MPEG-2 Parts and MPEG-4 Part 2 among others, the vendors of H.264/AVC products and services are expected to pay patent licensing royalties for the patented technology that their products use.

- The primary source of licenses for patents applying to this standard is a private organization known as MPEG LA (which is not affiliated in any way with the MPEG standardization organization); MPEG LA also administers patent pools for MPEG-2 Part 1 Systems, MPEG-2 Part 2 Video, MPEG-4 Part 2 Video, and other technologies.

MPEG LA®

Decoder-Encoder Royalties

- **Royalties to be paid by end product manufacturers for an encoder, a decoder or both ("unit") begin at US $0.20 per unit after the first 100,000 units each year. There are no royalties on the first 100,000 units each year. Above 5 million units per year, the royalty is US $0.10 per unit.**
- **The maximum royalty for these rights payable by an Enterprise (company and greater than 50% owned subsidiaries) is $3.5 million per year in 2005-2006, $4.25 million per year in 2007-08 and $5 million per year in 2009-10.**
- **In addition, in recognition of existing distribution channels, under certain circumstances an Enterprise selling decoders or encoders both (i) as end products under its own brand name to end users for use in personal computers and (ii) for incorporation under its brand name into personal computers sold to end users by other licensees, also may pay royalties on behalf of the other licensees for the decoder and encoder products incorporated in (ii) limited to $10.5 million per year in 2005-2006, $11 million per year in 2007-2008 and $11.5 million per year in 2009-2010.**
- **The initial term of the license is through December 31, 2010. To encourage early market adoption and start-up, the License will provide a grace period in which no royalties will be payable on decoders and encoders sold before January 1, 2005.**

Participation Fees (1)

- TITLE-BY-TITLE – For AVC video (either on physical media or ordered and paid for on title-by-title basis, e.g., PPV, VOD, or digital download, where viewer determines titles to be viewed or number of viewable titles are otherwise limited), there are no royalties up to 12 minutes in length. For AVC video greater than 12 minutes in length, royalties are the lower of (a) 2% of the price paid to the licensee from licensee's first arms length sale or (b) $0.02 per title. Categories of licensees include (i) replicators of physical media, and (ii) service/content providers (e.g., cable, satellite, video DSL, internet and mobile) of VOD, PPV and electronic downloads to end users.
- SUBSCRIPTION – For AVC video provided on a subscription basis (not ordered title-by-title), no royalties are payable by a system (satellite, internet, local mobile or local cable franchise) consisting of 100,000 or fewer subscribers in a year. For systems with greater than 100,000 AVC video subscribers, the annual participation fee is $25,000 per year up to 250,000 subscribers, $50,000 per year for greater than 250,000 AVC video subscribers up to 500,000 subscribers, $75,000 per year for greater than 500,000 AVC video subscribers up to 1,000,000 subscribers, and $100,000 per year for greater than 1,000,000 AVC video subscribers.

Participation Fees (2)

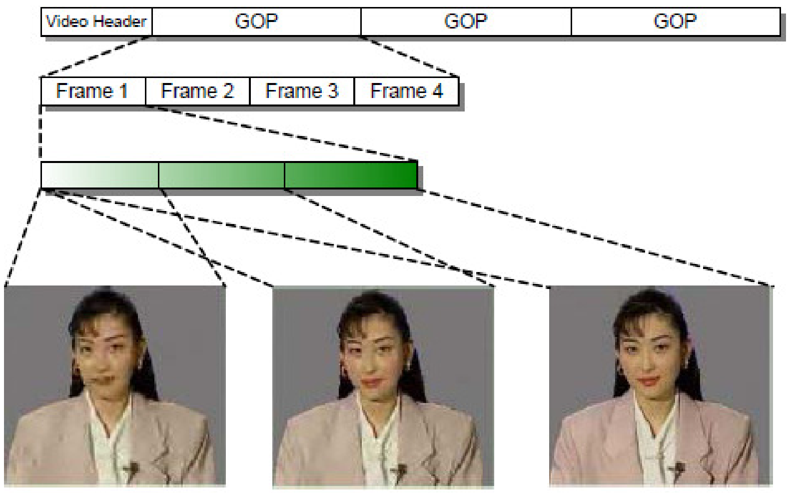
- **Over-the-air free broadcast – There are no royalties for over-the-air free broadcast AVC video to markets of 100,000 or fewer households. For over-the-air free broadcast AVC video to markets of greater than 100,000 households, royalties are $10,000 per year per local market service (by a transmitter or transmitter simultaneously with repeaters, e.g., multiple transmitters serving one station).**
- **Internet broadcast (non-subscription, not title-by-title) – Since this market is still developing, no royalties will be payable for internet broadcast services (non-subscription, not title-by-title) during the initial term of the license (which runs through December 31, 2010) and then shall not exceed the over-the-air free broadcast TV encoding fee during the renewal term.**
- **The maximum royalty for Participation rights payable by an Enterprise (company and greater than 50% owned subsidiaries) is $3.5 million per year in 2006-2007, $4.25 million in 2008-09 and $5 million in 2010.**
- **As noted above, the initial term of the license is through December 31, 2010. To encourage early marketplace adoption and start-up, the License will provide for a grace period in which no Participation Fees will be payable for products or services sold before January 1, 2006.**

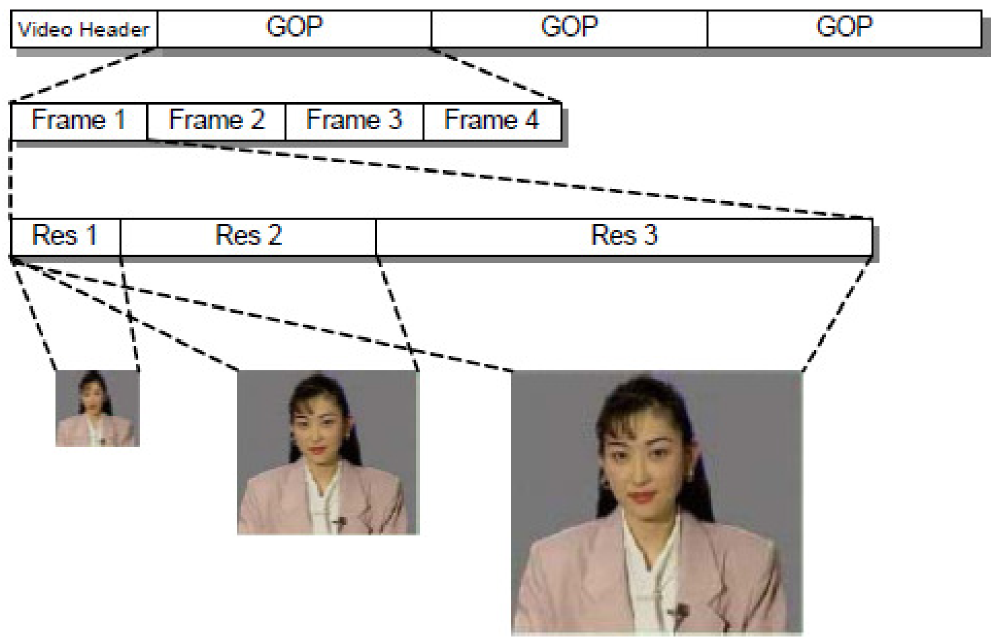
Scalable Video Coding (SVC)

An H.264/AVC Extension

An Heterogeneous World …

Quality and Spatial Resolution Scalability …

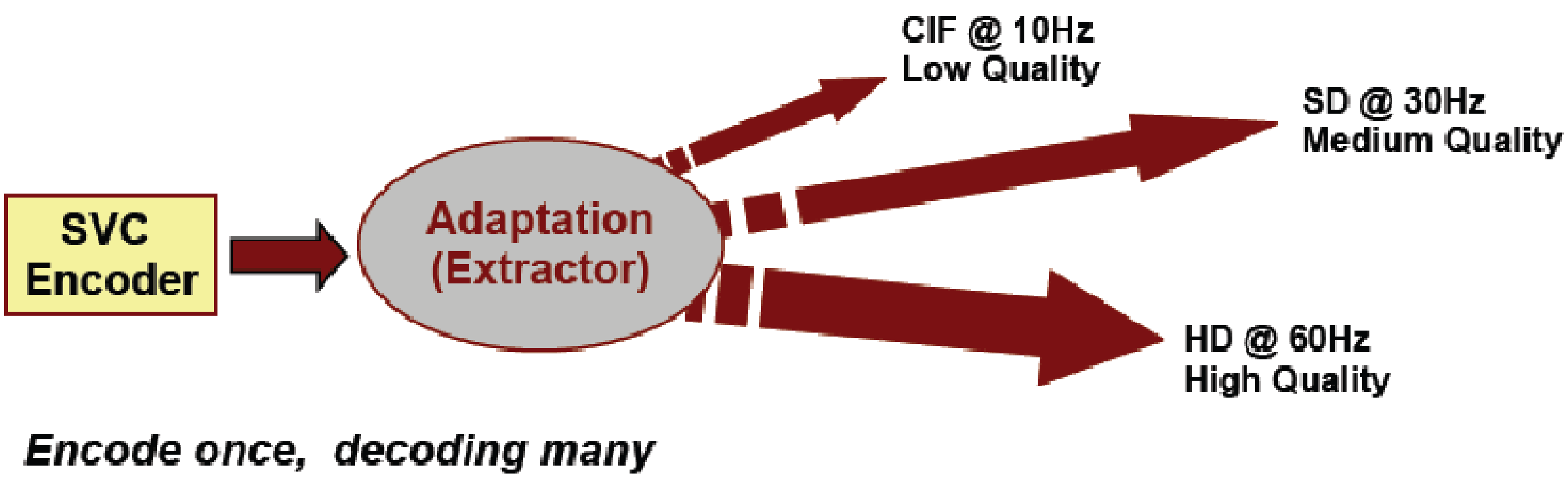




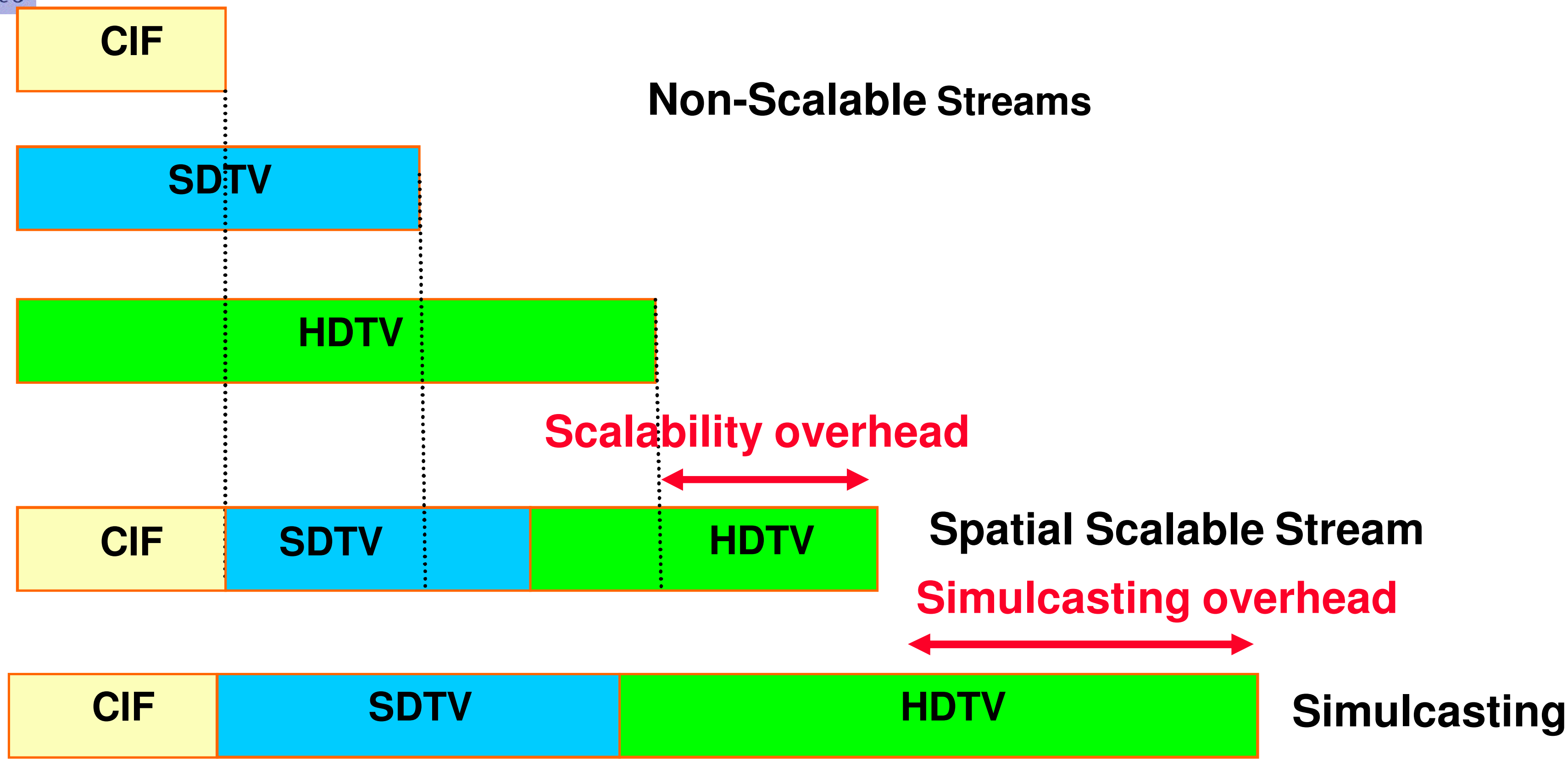
Scalable Video Coding: Objectives

Scalability is a functionality regarding the decoding of parts of the coded bitstream, ideally

1. while achieving an RD performance at any supported spatial, temporal, or SNR resolution that is comparable to single-layer coding at that particular resolution, and
2. without significantly increasing the decoding complexity.



Scalability: Rate Strengths and Weaknesses



For each spatial resolution (except the lowest), the scalable stream asks for a bitrate overhead regarding the corresponding alternative non-scalable stream, although the total bitrate is lower than the total simulcasting bitrate.

Scalable Video Coding (SVC) Challenge

The SVC standard objective was to enable the encoding of a high-quality video bit stream that contains one or more subset bit streams that can themselves be decoded with a complexity and reconstruction quality similar to that achieved using the existing H.264/AVC design with the same quantity of data as in the subset bit stream.

- **SVC should provide functionalities such as graceful degradation in lossy transmission environments as well as bitrate, format, and power adaptation; this should provide enhancements to transmission and storage applications.**

- **Previous video coding standards, e.g. MPEG-2 Video and MPEG-4 Visual, already defined codecs that were not successful due the characteristics of traditional video transmission systems, the significant loss in coding efficiency as well as the large increase in decoder complexity in comparison with non-scalable solutions.**

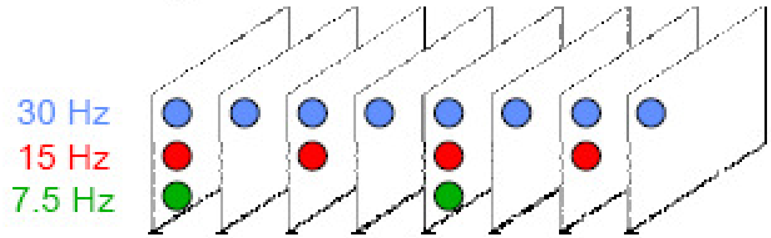
- **Alternatives to scalability may be simulcasting, and transcoding.**

Main SVC Requirements

- Similar coding efficiency compared to single-layer coding for each subset of the scalable bit stream.
- Little increase in decoding complexity compared to single-layer decoding that scales with the decoded spatio-temporal resolution and bitrate.
- Support of temporal, spatial, and quality scalability.
- Support of a backward compatible base layer (H.264/AVC in this case).
- Support of simple bitstream adaptations after encoding.

SVC Scalability Types

- **Temporal**: change of frame rate



- **Spatial**: change of frame size

- **Fidelity**: change of quality (a.k.a. SNR)

ra

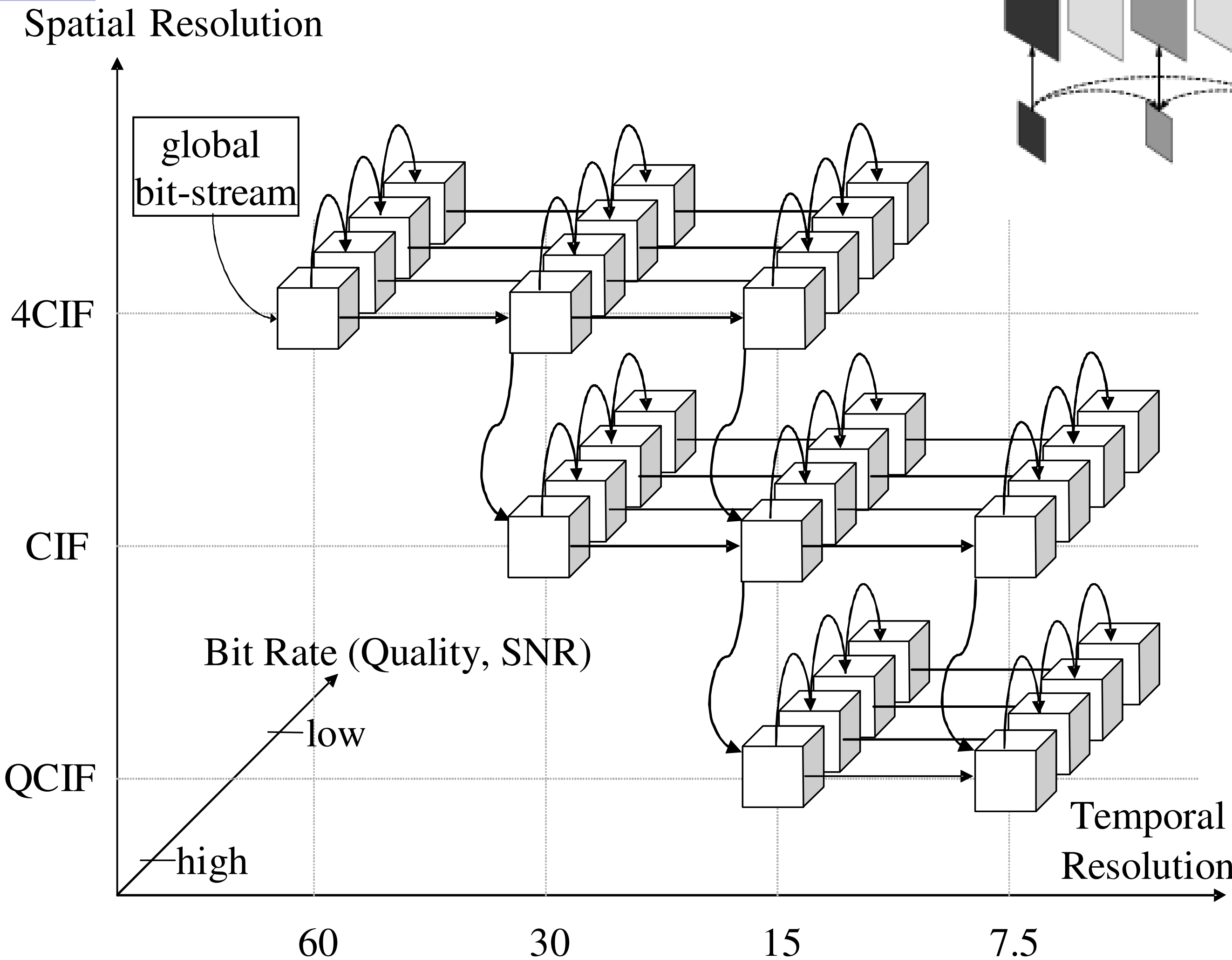
SVC Applications

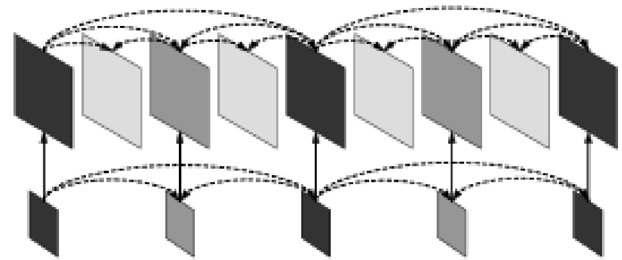
- **Robust Video Delivery**
  - Adaptive delivery over error-prone networks and to devices with varying capability
  - Combine with unequal error protection
  - Guarantee base layer delivery
  - Internet/mobile transmission

- **Scalable Storage**
  - Scalable export of video content
  - Graceful expiration or deletion
  - Surveillance DVR's and Home PVR's

- **Enhancement Services**
  - Upgrade delivery from 1080i/720p to 1080p
  - DTV broadcasting, optical storage devices

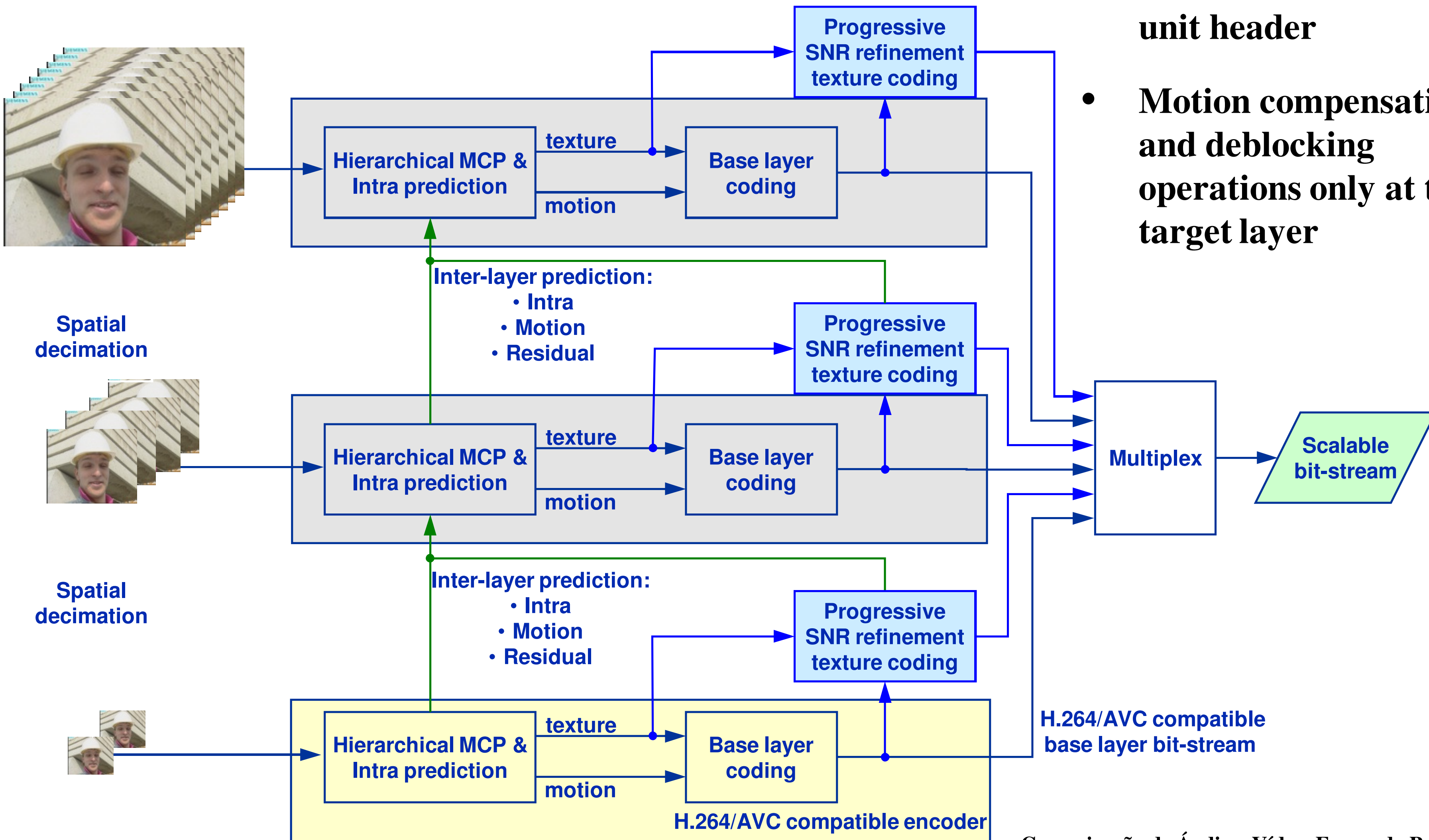
Spatio-Temporal-Quality Cube



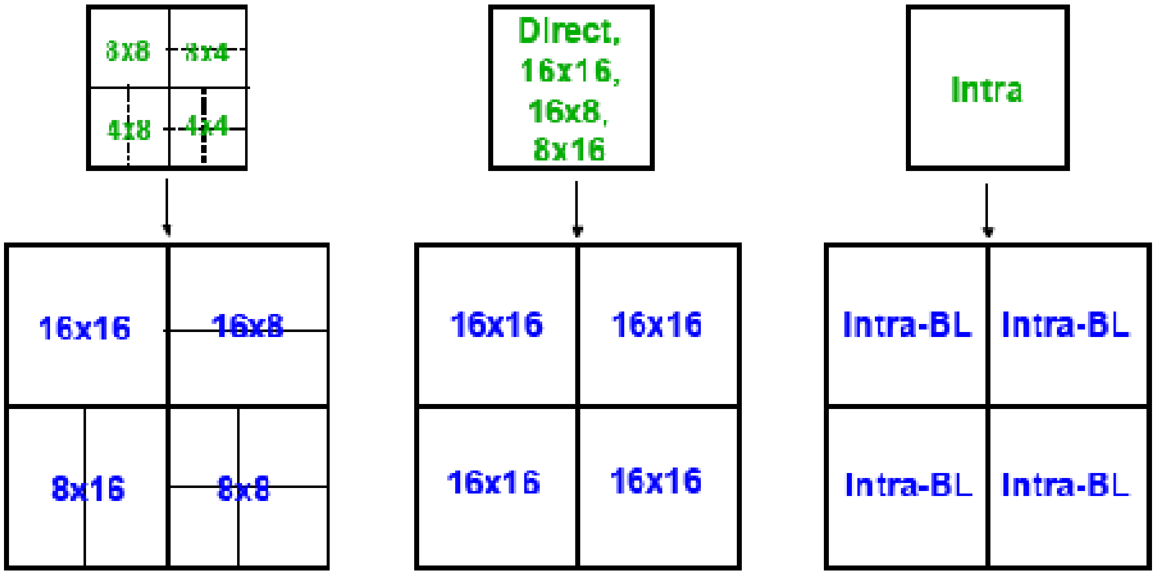


SVC Coding Architecture

- Layer indication by identifiers in the NAL unit header
- Motion compensation and deblocking operations only at the target layer



SVC Inter-Layer Prediction



The main goal of inter layer prediction is to enable the usage of as much lower layer information as possible for improving the RD performance of the enhancement layers:

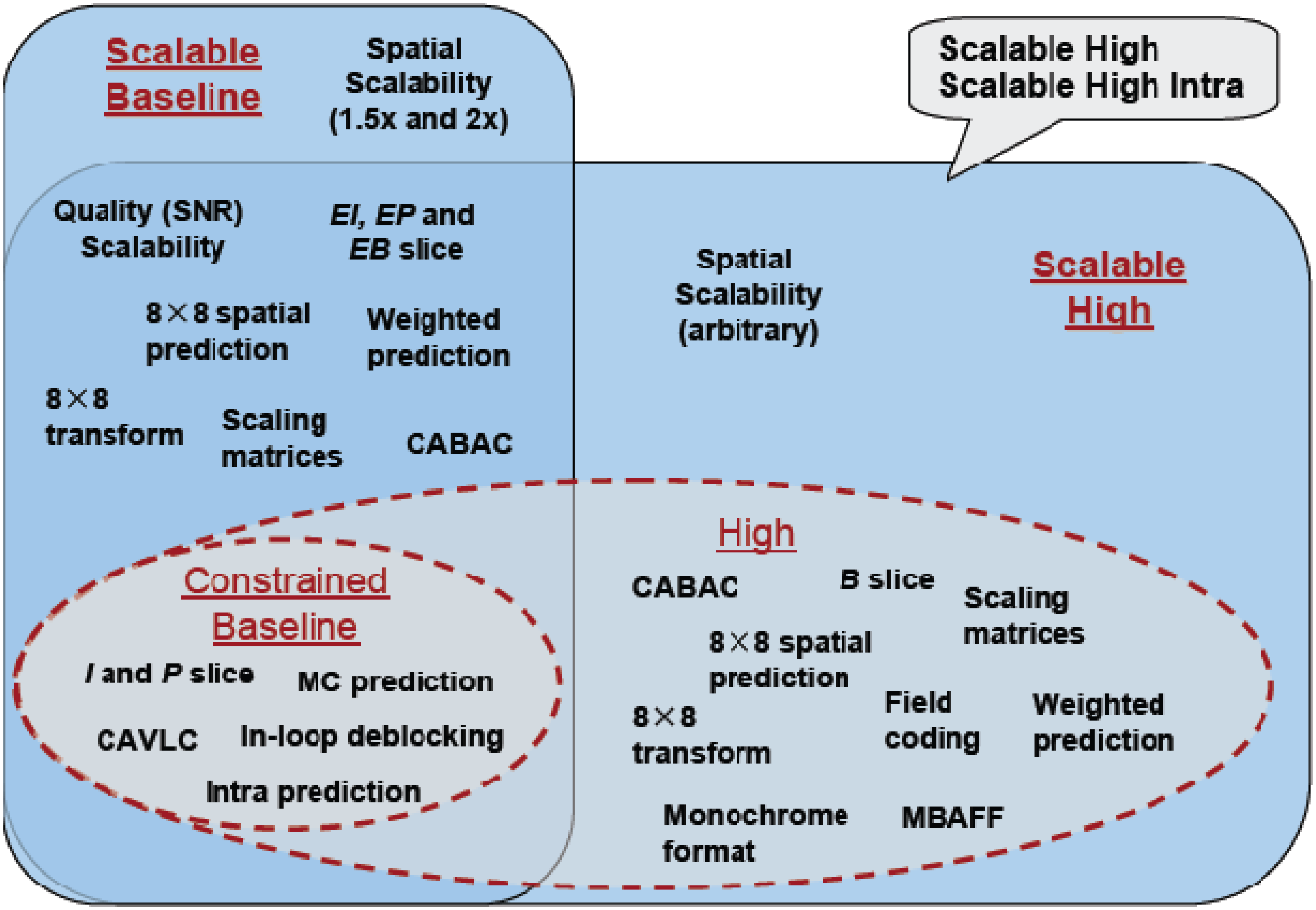
- Motion: (Upsampled) partitioning and motion vectors for prediction
- Residual: (Upsampled) residual (bi-linear, blockwise)
- Intra: (Upsampled) intra MB (direct filtering)

SVC Scalability Types: What Cost ?

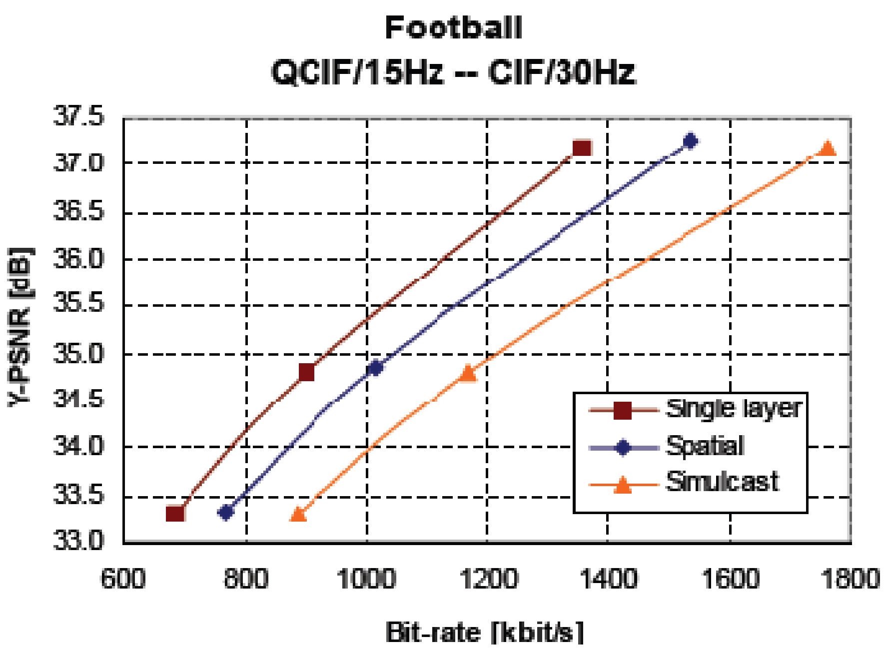
- **Temporal scalability - Can be typically achieved without losses in rate-distortion performance.**
- **Spatial scalability - When applying an optimized SVC encoder control, the bitrate increase relative to non-scalable H.264/AVC coding, at the same fidelity, can be as low as 10% for dyadic spatial scalability. The results typically become worse as spatial resolution of both layers decreases and results improve as spatial resolution increases.**
- **SNR scalability - When applying an optimized encoder control, the bitrate increase relative to non-scalable H.264/AVC coding, at the same fidelity, can be as low as 10% for all supported rate points when spanning a bitrate range with a factor of 2-3 between the lowest and highest supported rate point.**

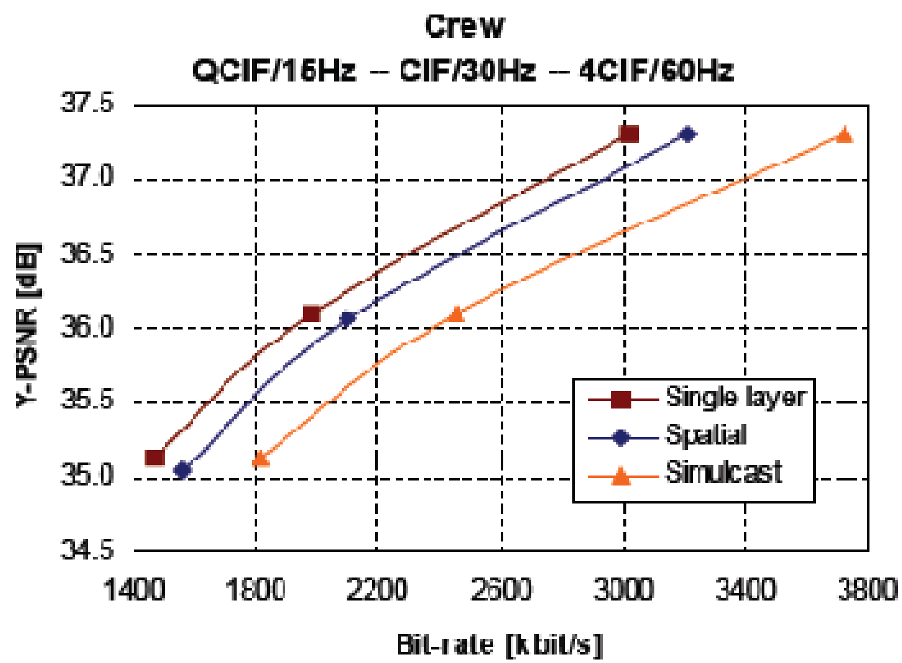
From IEEE Transactions on Circuits and Systems for Video Technology, September 2007.

SVC Profiles



SVC Performance: Spatial Scalability





- **10~15% gains over simulcast**
- **Performs within 10% of single layer coding**

[Segall& Sullivan, T-CSVT, Sept'07]

SVC: What Future ?

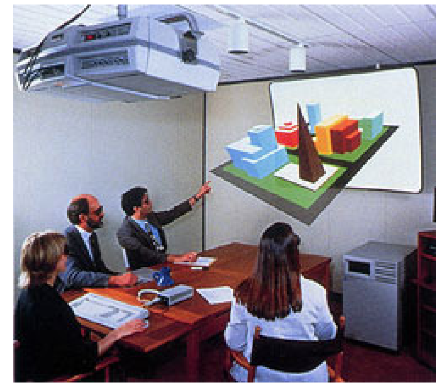
- **Technically, the standard is a great success already with some adoption**
  - ***Google*** **Gmail service**
  - ***Vidyo*** **video conferencing for the Internet**
  - **Industry appears to be open towards embracing SVC for DTV broadcast services**
  - **Specifically, enhancement of 720p to 1080p**

- **Others might be less certain, but still possible …**
  - **SVC for surveillance recorders**
  - **Lots of discussion on Scalable Baseline in ATSC-M/H**

Multiview Video Coding (MVC)

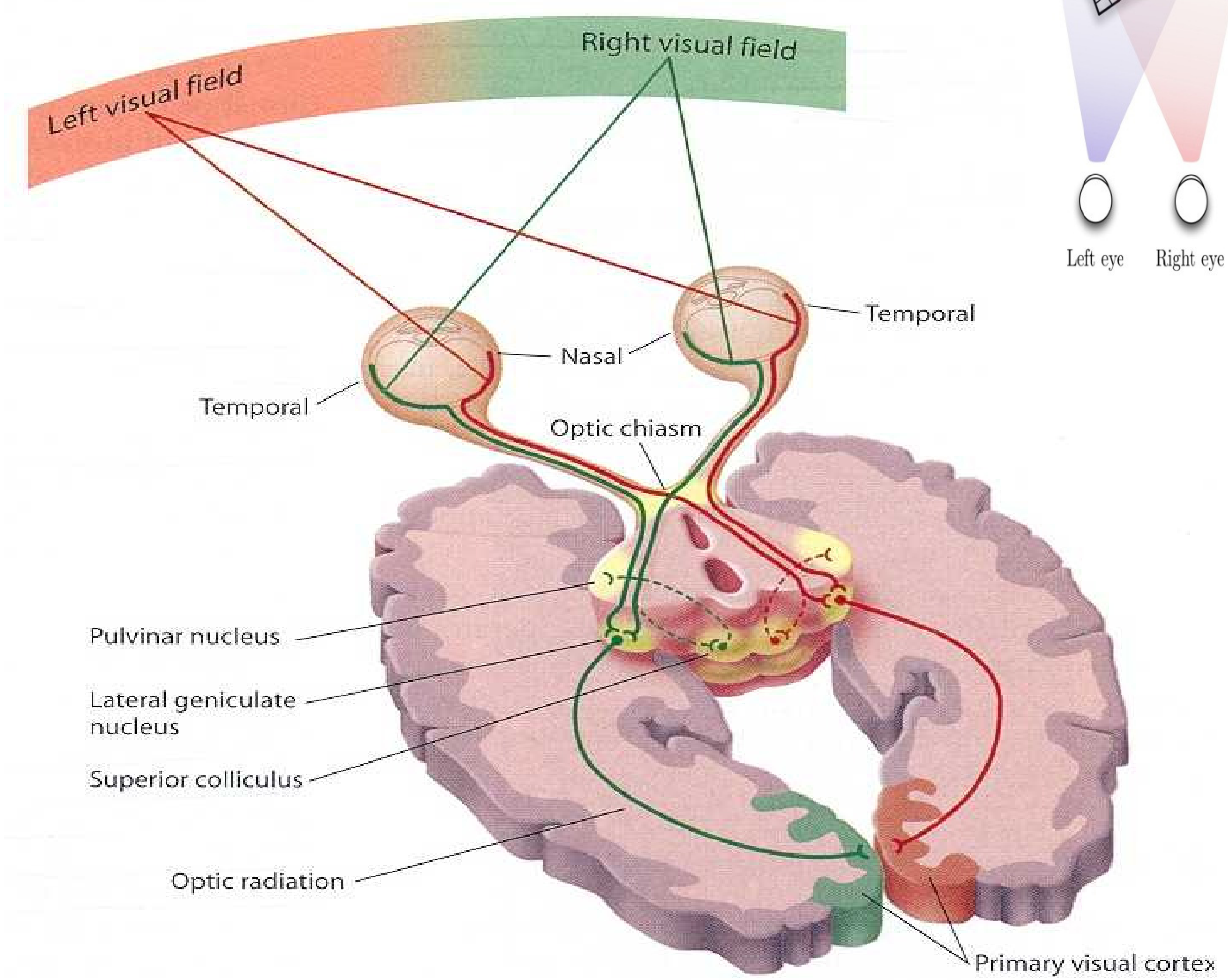
An H.264/AVC Extension

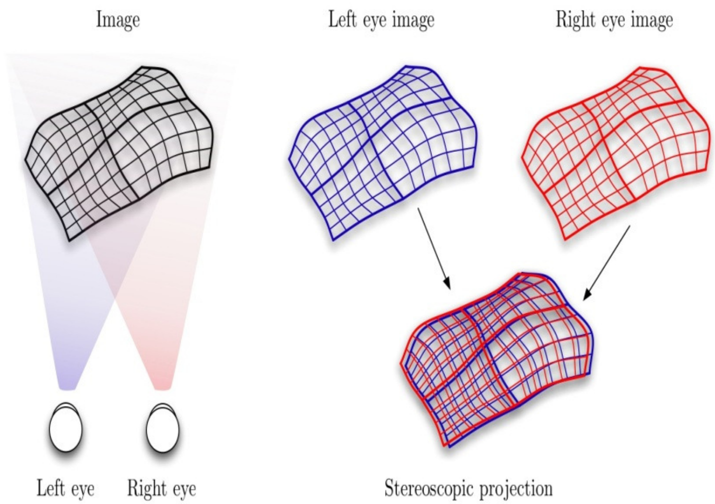
3D Worlds

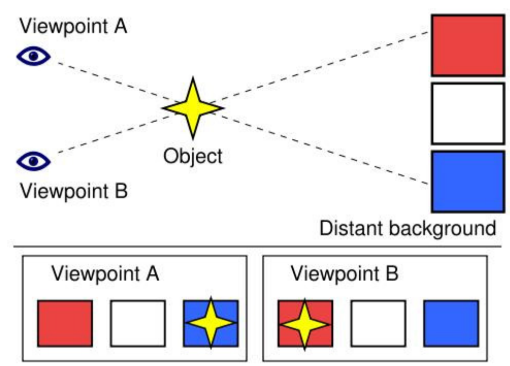


- **3D experiences may be provided through multi-view video, notably**
  - **3D video (also called stereo) which brings a depth impression of a scene**
  - **Free viewpoint video (FVV) which allows an interactive selection of the viewpoint and direction within certain ranges.**
- **May require special 3D display technology: many new products announced recently and being exhibited**
- **New 3D display technology is driving this area: no glasses, multi-persons displays, higher display resolutions, avoid uneasy feelings (headaches, nausea, eye strain, etc.)**
- **Relevant for broadcast TV, teleconference, surveillance, interactive video, cinema, gaming or other immersive video applications**

Human Visual System

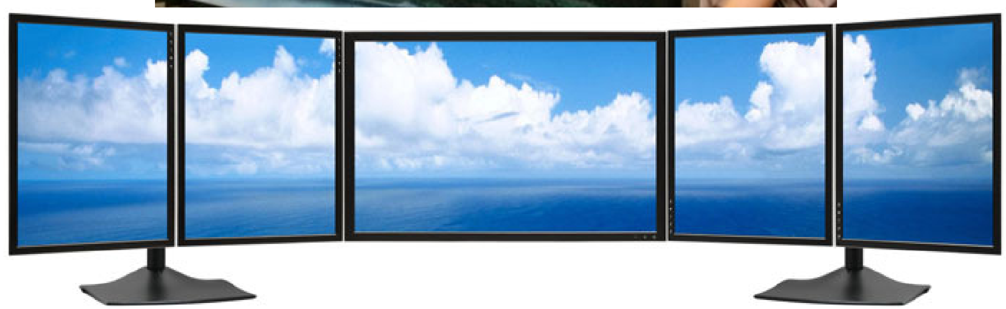






INSTITUTO SUPERIOR TÉCNICO

3D Displays: a Major Driving Force …

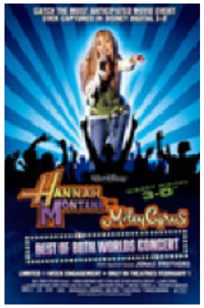


- 3D displays are maturing rapidly …
- High quality stereoscopic displays can now be offered with no added cost
- As display bandwidth increases, 3D is more attractive as a consumer choice
- Results in a wider customer base with 3D-ready HD displays

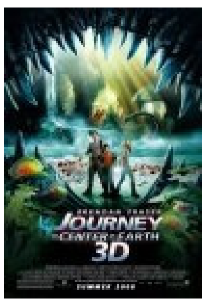
INSTITUTO SUPERIOR TÉCNICO

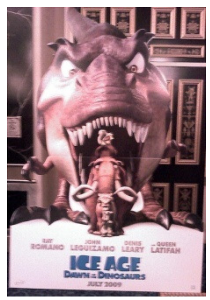
Coming 3D Content …

- Nine 3D title releases to date since 2005
  - Recent: Beowulf, Hannah Montana, U23D



- More on the way
  - Another 10 releases planned for 2009 alone

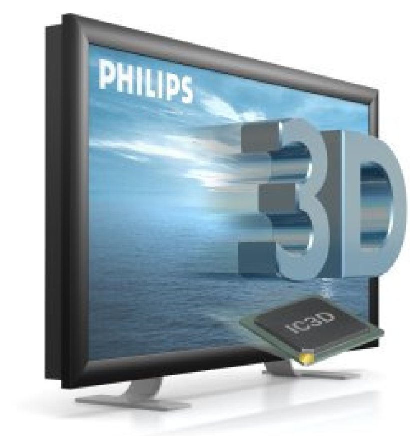




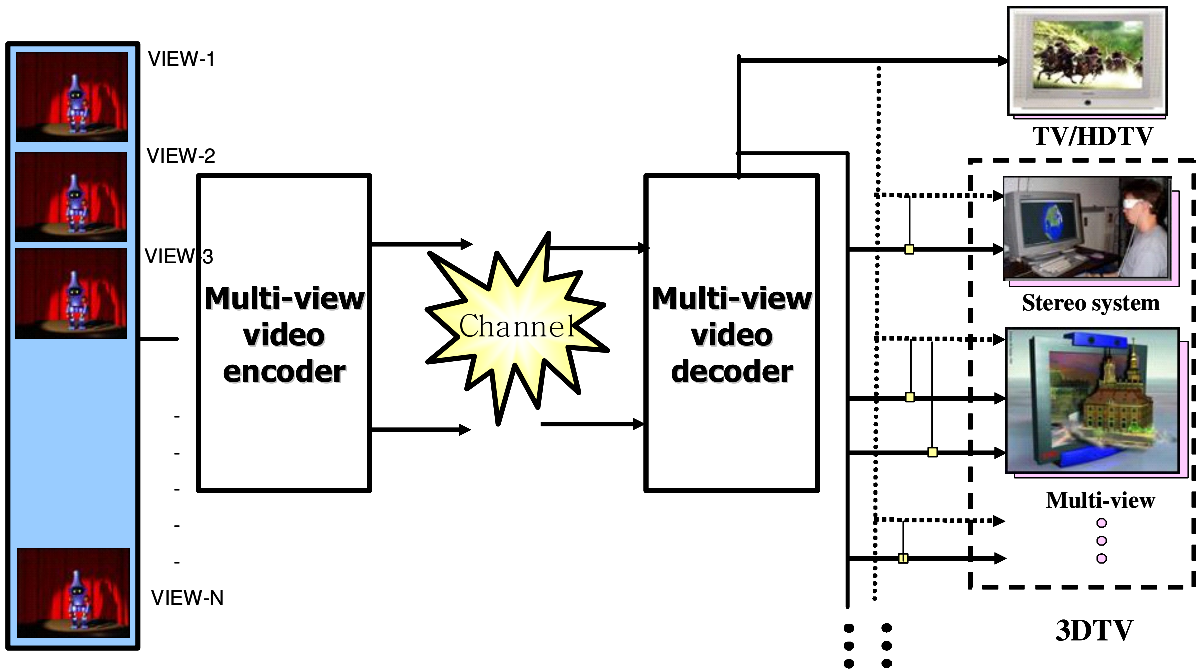
- Hollywood is now able to offer unique, high-quality immersive 3D experience in theaters
- Revenue per 3D screen is typically three times higher than traditional 2D screens
- Results in increased momentum in 3D production and growing consumer appetite for 3D content

3D Formats/Standards …

- There is much confusion in the area of 3D video formats and standards. Most formats are closely coupled to 3D display types and application scenarios.
- A universal, flexible, generic, scalable, backward compatible 3D video format/standard would be highly desirable to support any 3D video application in an efficient way, while decoupling content creation from display and application.
- Experts expect 3D television to follow much the same trajectory as HDTV did earlier this decade: a slow start, then a rapid ascent in sales once enough content exists to attract mainstream buyers.



Multi-View Video System



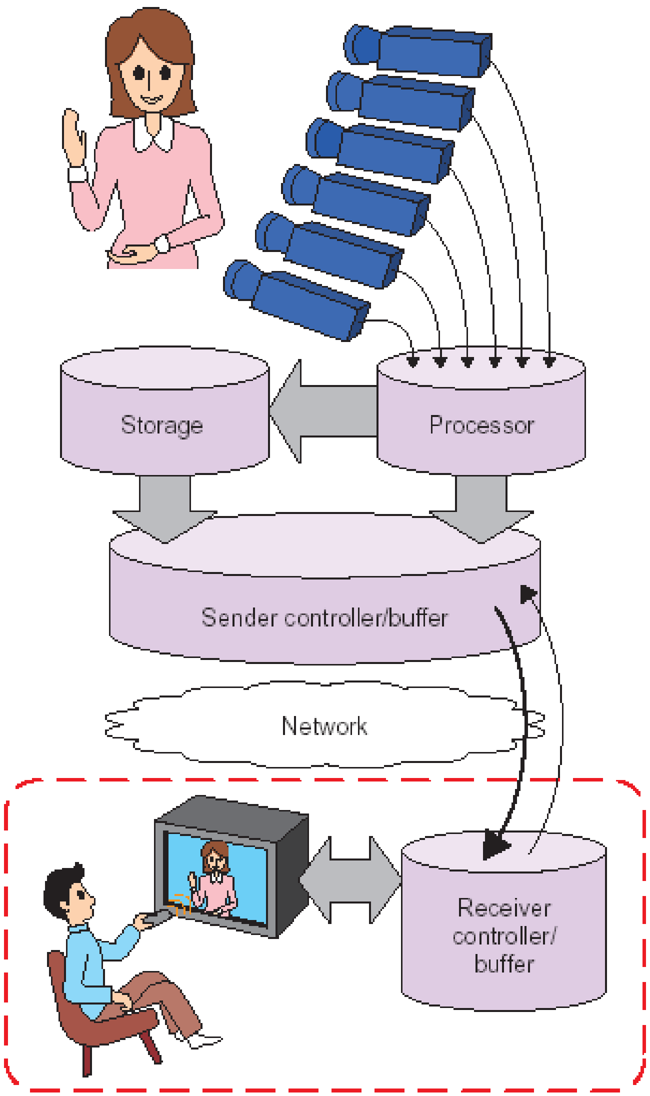
Multi-view video (MVV) refers to a set of N temporally synchronized video streams coming from cameras that capture the same real world scenery from different viewpoints.

- **Provides the ability to change viewpoint freely with multiple views available**
- **Renders one view (real or virtual) to legacy 2D display**
- **Most important case is stereo video (N = 2), with each view derived for projection into one eye, in order to generate a depth impression**

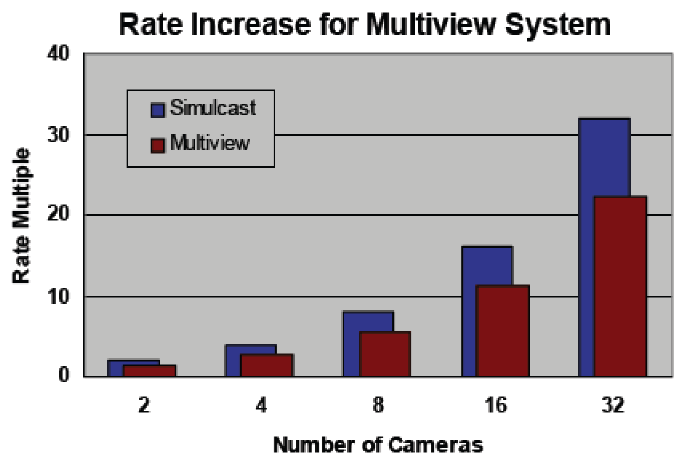
Multi-View Video Data

- Most test sequences have 8-16 views
  - But, several 100 camera arrays exist!
- Redundancy reduction between camera views
  - Need to cope with color/illumination mismatch problems
  - Alignment may not always be perfect either

Multi-View Video Coding (MVC)

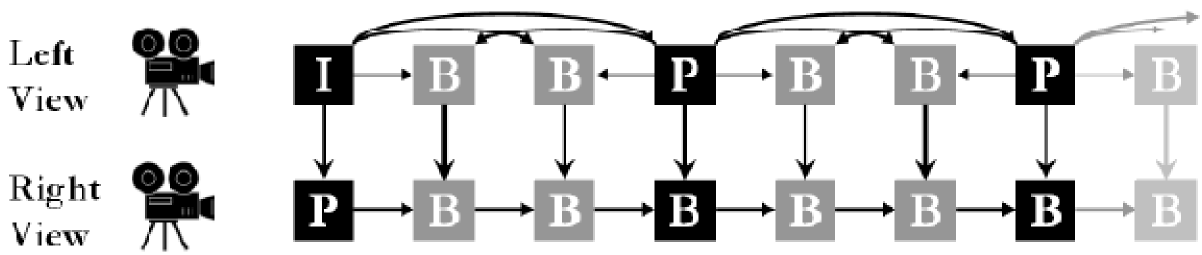


- **Direct coding of multiple views (stereo to multi-view)**
- **Exploits redundancy between views using inter-camera prediction to reduce required bit-rate**
- **Without any changes at H.264/AVC slice layer and below, bitrate reductions around 20-50% can be achieved by allowing interview predictions.**



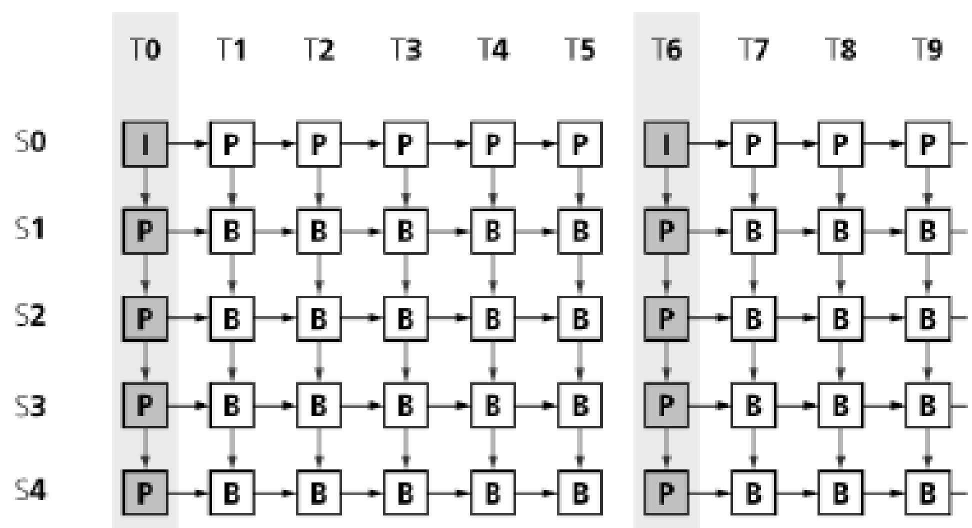
MVC: Prediction Structures

Many prediction structures possible to exploit inter-camera redundancy: trade-off in memory, delay, computation and coding efficiency.



MPEG-2 Video Multi-view profile

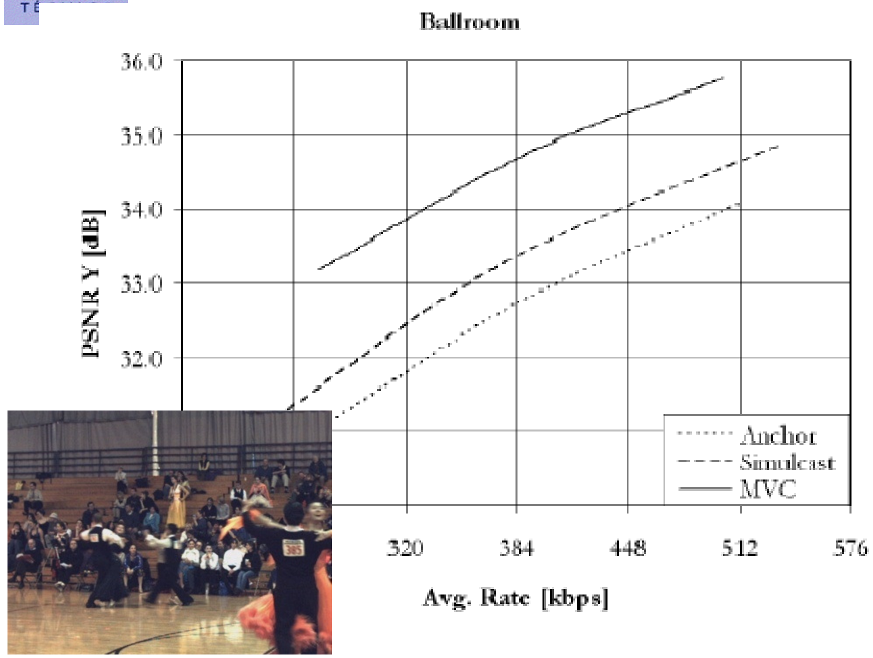
(JVT) MVC

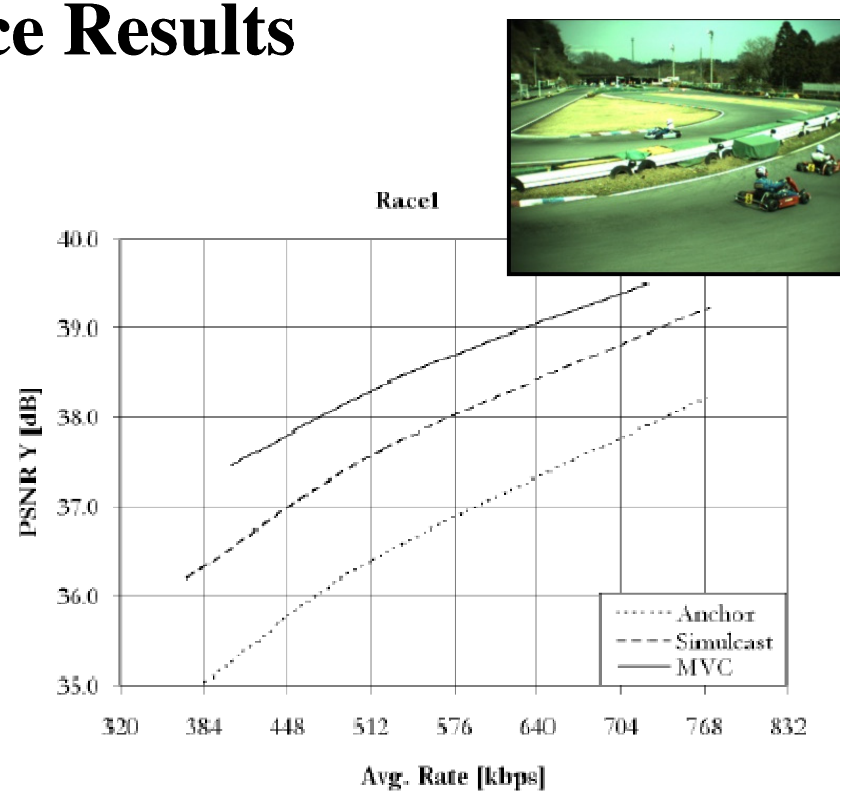


MVC: Technical Solution

- **Key elements of MVC design**
  - Does not require any changes to lower-level syntax, so it is very compatible with single-layer AVC hardware
  - Base layer required and easily extracted from video bitstream (identified by NAL unit type)
- **Inter-view prediction**
  - Enabled through flexible reference picture management
  - Allow decoded pictures from other views to be inserted and removed from reference picture buffer
  - Core decoding modules do not need to be aware of whether reference picture is a time reference or multiview reference
- **Small changes to high-level syntax, e.g. specify view dependency**
- **MPEG-2 based transport and MP4 file format specs to follow**

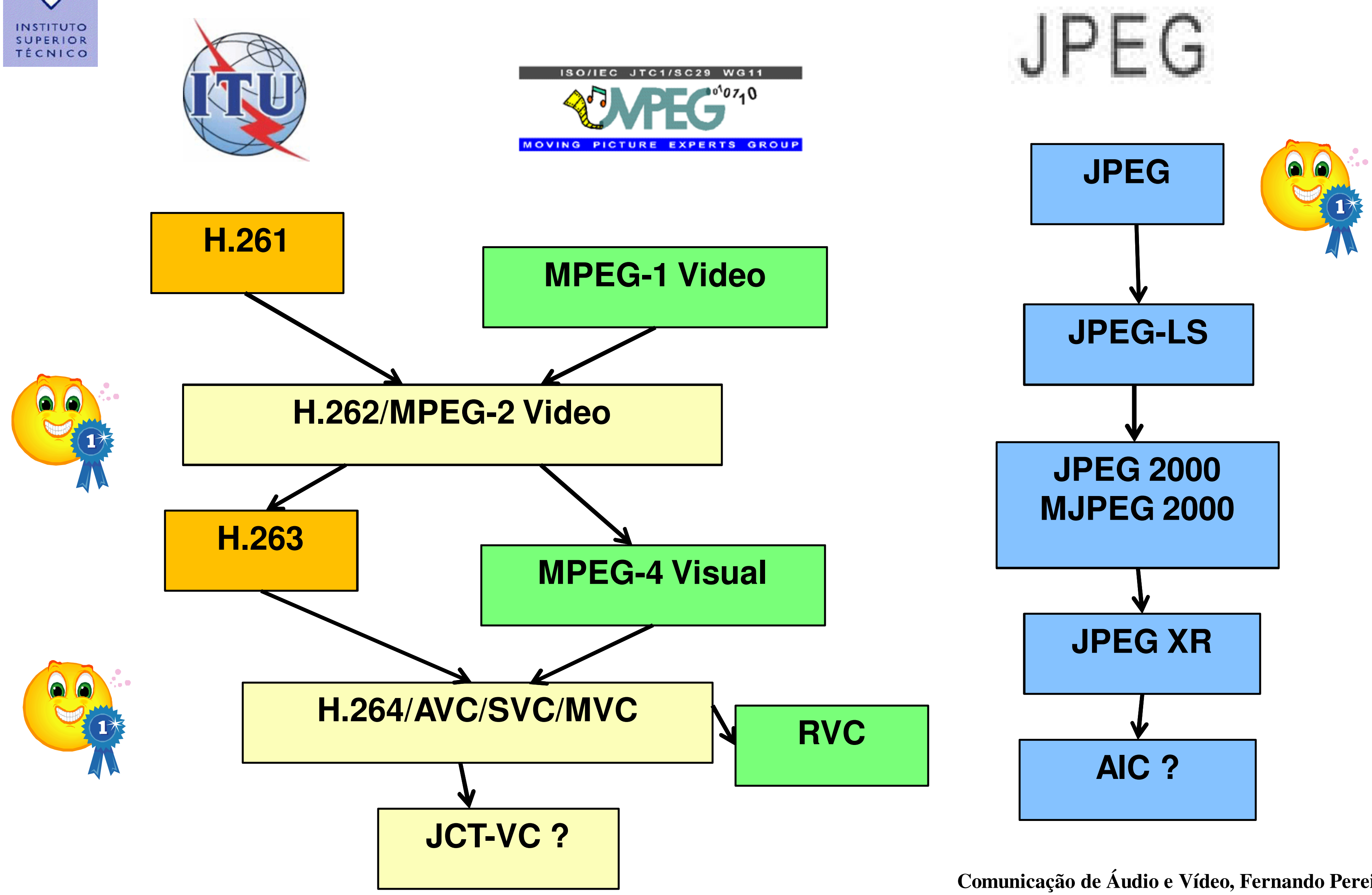
Some MVC Performance Results





- Anchor is H.264/AVC without hierarchical B pictures
- Simulcast already includes hierarchical B pictures
- Majority of gains due to inter-view prediction at I-picture locations
- Although more efficient than simulcast, rate of MVC is still proportional to the number of views (varies with scene, camera arrangement, etc.)

The Standardization Path …



Video Coding Standards: a Summary

Standard	Year	Main Applications	Profiles	Main Bitrates	Frame Types	Ref. Frames	Transform	Number Motion Vectors (if any)	Motion Vectors Precision	Entropy Coding	Deblocking Filter
H.261	1988	Videotelephony and videoconference	No	p×64 kbit/s	-	1	DCT	1 per MB	Integer pel	Huffman based	In loop
MPEG-1 Video	1991	Digital storage in CD-ROM	No	Around 1-1.2 Mbit/s	I, P, B, and D	0-2	DCT	1 or 2 per MB (P and B)	Half pel	Huffman based	Out of the loop
H.262/MPEG-2 Video	1994	Digital TV and DVD	Yes, most used is Main Profile	From 2 to 10 Mbit/s	I, P and B	0-2	DCT	1 or 2 per MB (2 to 4 for interlaced video)	Half pel	Huffman based	Out of the loop
H.263	1995	Videotelephony and videoconference and more	Only in extensions	From very low rates to around 1 Mbit/s	I, P and B	0-2	DCT	1 or 2 per MB (4 in the optional modes)	Half pel	Huffman based	Out of the loop
MPEG-4 Visual	1998	Large range with objects	Yes, most used are Simple and Advanced Simple	Very large range using levels	I, P and B	0-2	DCT	1 or 2 per MB (4 in the optional modes); also global motion vectors	1/4 pel	Huffman based; arithmetic coding for the shape	Out of the loop
H.264/AVC	2004	Large range, from mobile to Blu-ray	Yes, most used are Baseline, Main and High	Very large range using levels	I, P, generalized B, SP and SI	Up to 16	Integer DCT	1 to 16 per MB (P slices) and 1 to 32 (B slices)	1/4 pel	CAVLC and CABAC	Out of the loop
SVC	2007	Robust delivery, graceful deletion, broadcasting,	Yes	Very large range using layers	I, P and generalized B,	Up to 16	Integer DCT	1 to 16 per MB (?)	1/4 pel	CAVLC and CABAC	In loop
MVC	2009	Stereo TV, Free viewpoint TV	Yes	Very large range using levels	I, P, B,	Up to 16	Integer DCT	1 to 16 per MB (?)	1/4 pel	CAVLC and CABAC	In loop

Final Remarks on AVC and Extensions

- The H.264/AVC standard builds on previous coding standards to achieve a typical compression gain of about 50%, largely at the cost of increased encoder and decoder complexity.
- The compression gains are mainly related to the variable (and smaller) block size motion compensation, multiple reference frames, smaller blocks transform, deblocking filter in the prediction loop, and improved entropy coding.
- The H.264/AVC standard represents nowadays the state-of-the-art in video coding and it is currently being adopted by a growing number of organizations, companies and consortia.
- The SVC and MVC extensions are technically powerful but their market relevance has still to be fully checked ...

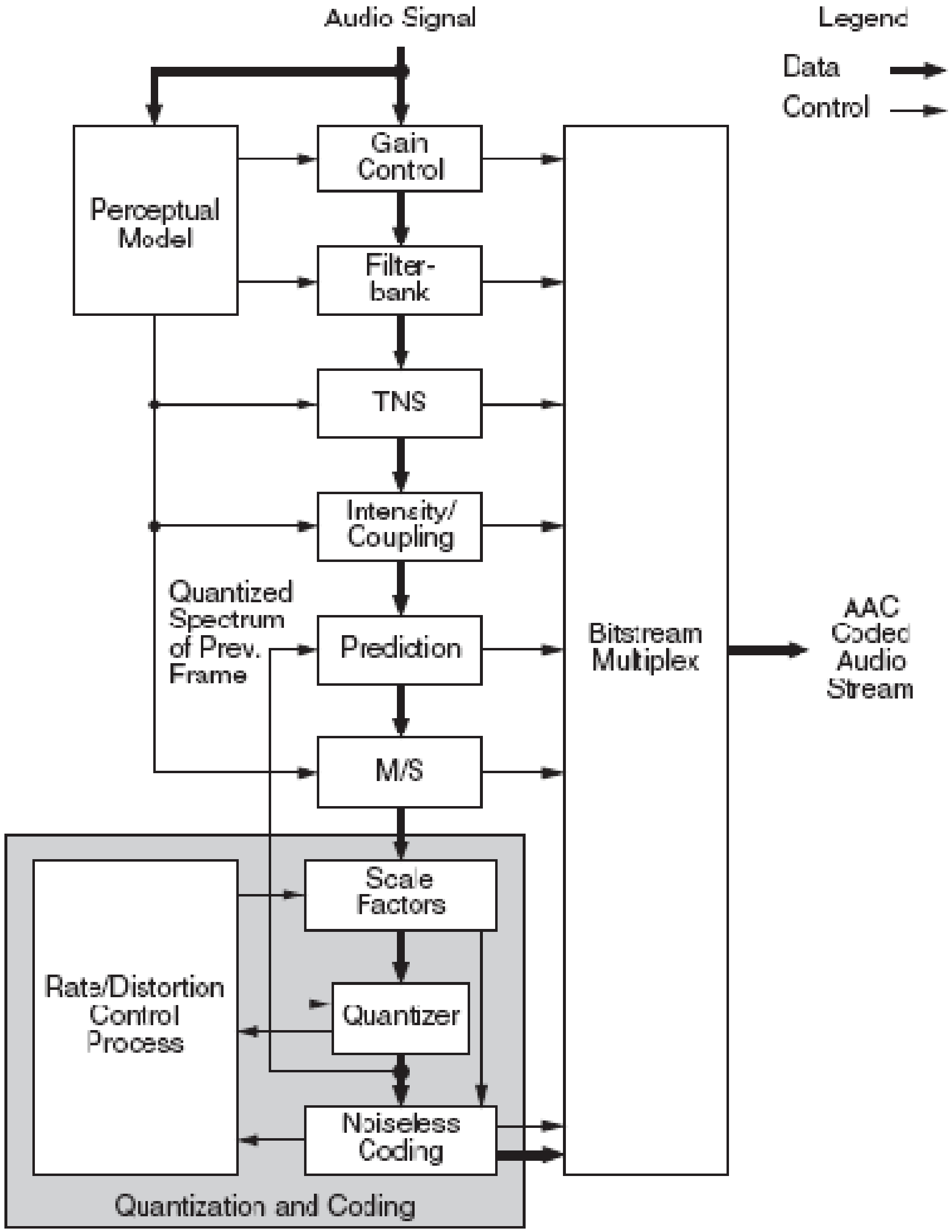
Advanced Audio Coding (MPEG-2 e MPEG-4)

AAC: Objectives

To provide a substantial increase of coding efficiency regarding previous audio coding standards, notably indistinguishable quality at 384 kbit/s or lower for five full bandwidth channels.

Advanced Audio Coding (AAC) - initially called Non- Backward Compatible (NBC) - is defined in two MPEG standards:

- **MPEG-2 AAC (Part 7) – Defines the core AAC codec;**
- **MPEG-4 Audio (Part 3) - Building on the MPEG-2 AAC core technology, MPEG-4 defines a number of extensions, notably to enhance compression performance (perceptual noise substitution, long-term prediction) and enable operation at very low delays (low-delay AAC).**



AAC is based on the Time-Frequency paradigm (T/F) of perceptual audio coding where a spectral (frequency domain) representation of the input signal rather than the time domain signal itself is coded. This paradigm was already adopted in MPEG-1 Audio.

MPEG-2 AAC Encoder Architecture

MPEG-2 AAC Compression Performance

- **MPEG-2 AAC demonstrated near-transparent subjective audio quality at a bitrate of 256 to 320 kbit/s for five channels and at 96 to 128 kbit/s for stereophonic signals.**
- **Although originally designed for near-transparent audio coding, testing inside MPEG revealed that the coder exhibits excellent performance also at very low bitrates down to 16 kbit/s.**
- **As a result, MPEG-2 AAC was adopted as the core of the MPEG-4 General Audio (T/F) coder, now called MPEG-4 AAC or simply AAC.**

MPEG-4 AAC Tools

MPEG-4 AAC Additions to MP3

- More sample frequencies (from 8 kHz to 96 kHz) than MP3 (16 kHz to 48 kHz)
- Up to 48 channels (MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode)
- Arbitrary bitrates and variable frame length
- Higher efficiency and simpler filterbank (hybrid → pure MDCT)
- Higher coding efficiency for stationary signals (blocksize: 576 → 1024 samples)
- Higher coding efficiency for transient signals (blocksize: 192 → 128 samples)
- Much better handling of audio frequencies above 16 kHz
- More flexible joint stereo (separate for every scale band)
- Adds additional modules (tools) to increase compression efficiency: TNS, Backwards Prediction, PNS, etc... These modules can be combined to constitute different profiles.

AAC Licensing and Patents

- **No licenses or payments are required to be able to stream or distribute content in AAC format. This reason alone makes AAC a much more attractive format to distribute content than MP3, particularly for streaming content (such as Internet radio).**

- **However, a patent license is required for all manufacturers or developers of AAC codecs, that require encoding or decoding. It is for this reason that some implementations are distributed in source form only, in order to avoid patent infringement.**

- **AAC requires patent licensing, and thus uses proprietary technology. But contrary to popular belief, it is not the property of a single company, having been developed in a standards-making organization, MPEG; the same is true for MP3.**

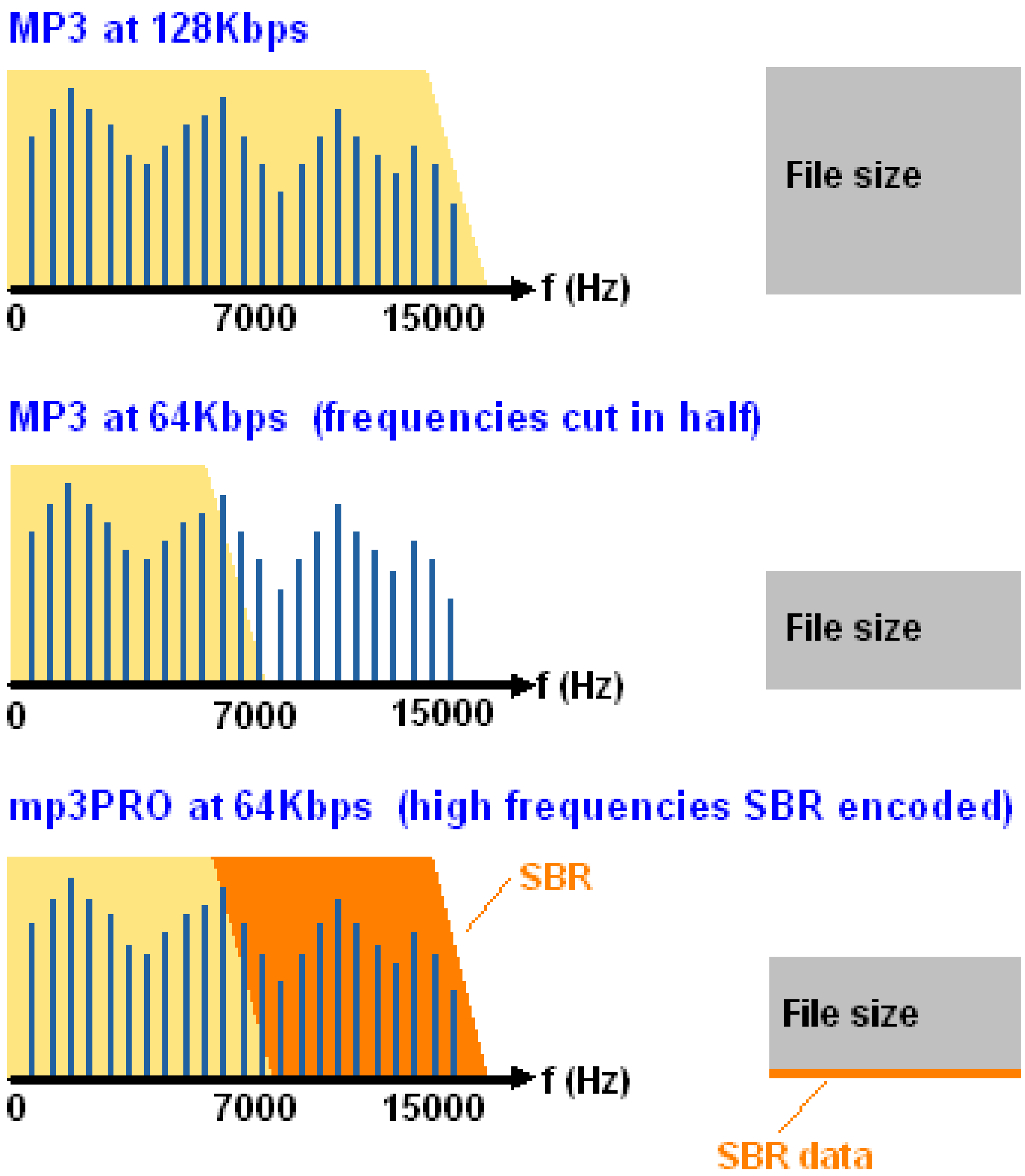
MPEG-4 High-Efficiency AAC (HE-AAC)

HE-AAC: Objectives

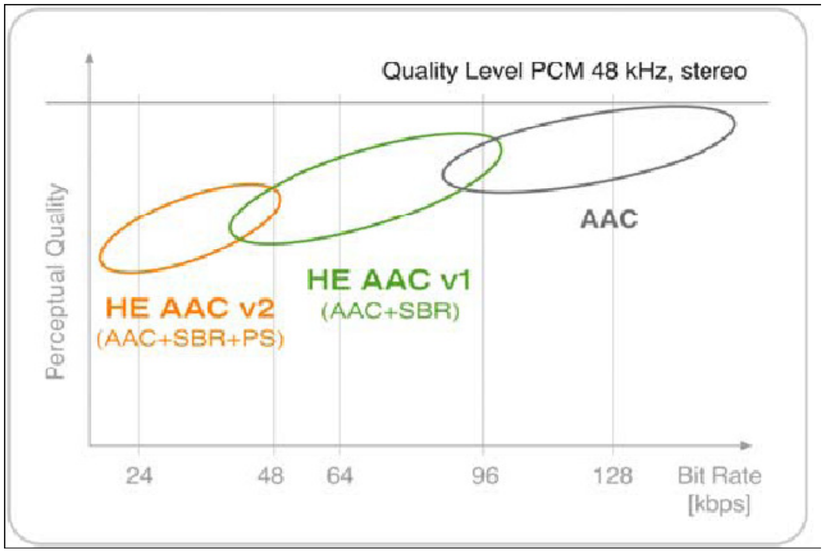
To enable audio and music delivery for very low bitrate applications, a substantial increase of coding efficiency is required compared to the performance offered by regular AAC at such rates.

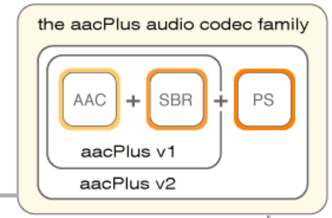
- **Extension of the established MPEG-4 Advanced Audio Coding (AAC) architecture.**
- **Compression format for generic audio signals offering high audio quality also to applications limited in transmission bandwidth or storage capacity.**
- **Targets applications that cannot be served well using regular AAC to deliver high audio quality and full audio bandwidth even at very low data rates, e.g. 24 kbit/s and below per audio channel.**

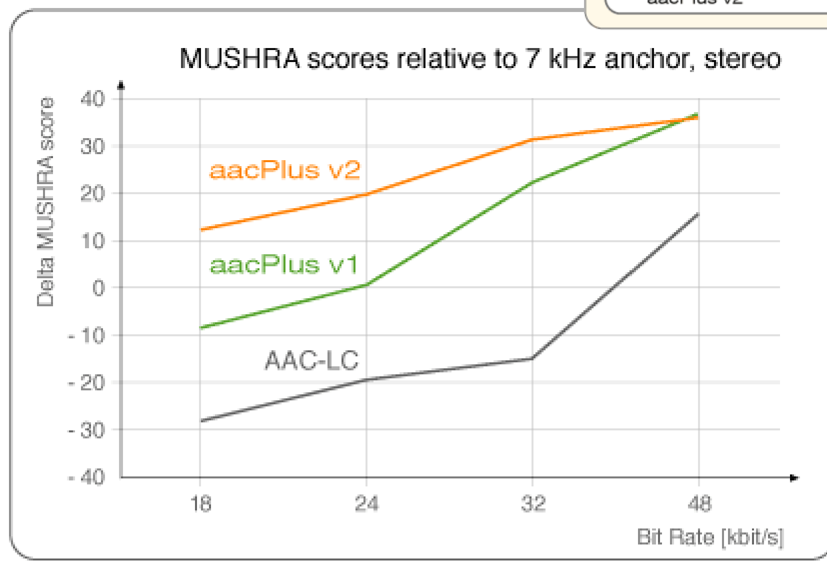
The SBR Principle and Benefit



HE-AAC Family: Compression Performance







- **HE-AAC v1 offers an increase in coding efficiency by more than 25% over AAC, when operated at or near 24 kb/s per audio channel.**

- **With the inclusion of parametric stereo coding, a further increase in coding efficiency is achieved; HE-AAC v2 typically performs as well as HE-AAC v1 when the latter is operating at a 33% higher bitrate (up to 40 kbit/s stereo, according to MPEG verification tests).**

Recent and Emerging Advanced Coding Successes

iPod Classic and nano

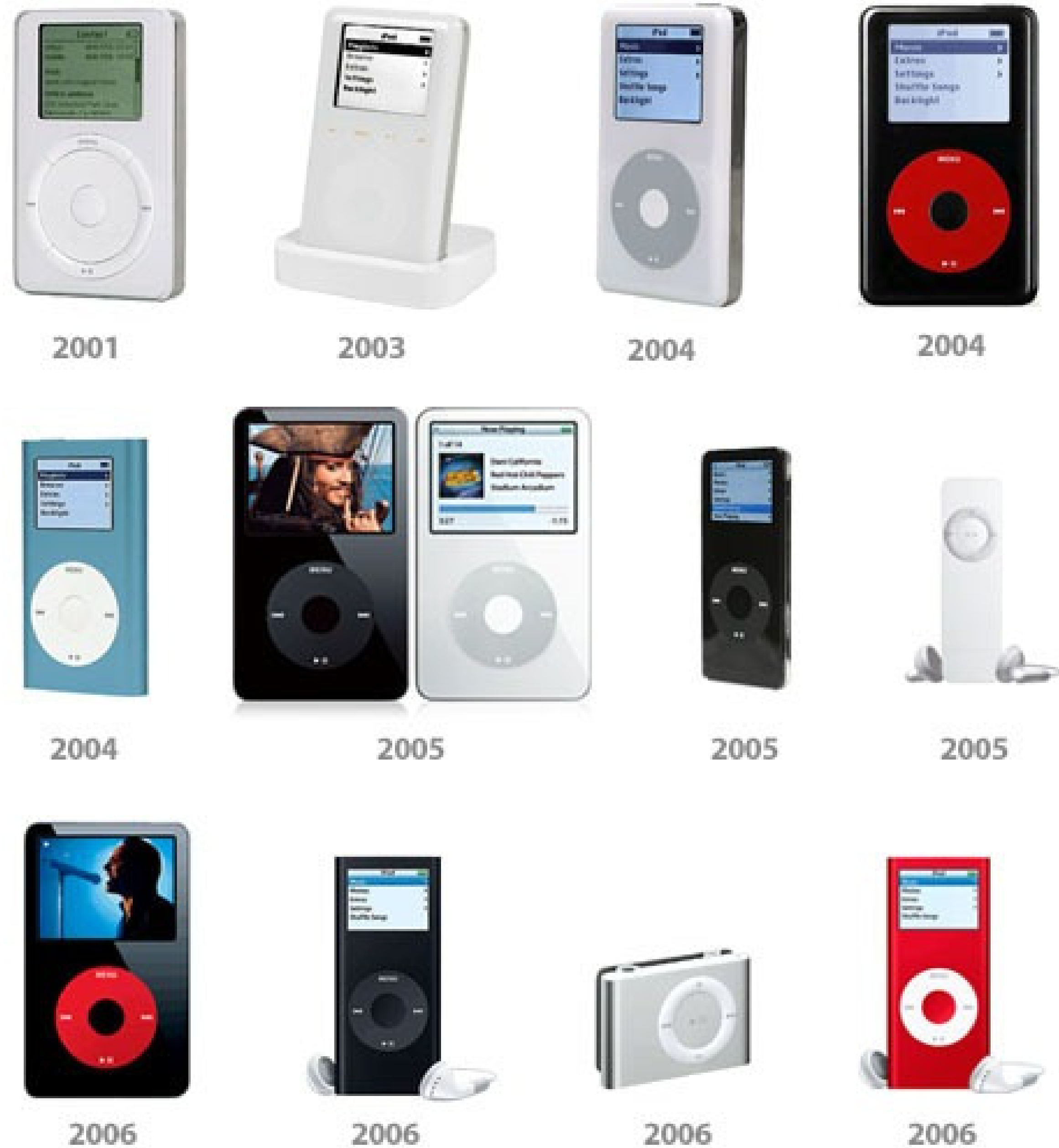
Audio

- Frequency response: 20 Hz to 20000 Hz
- Audio formats supported: AAC (16 to 320 Kbps), Protected AAC (from iTunes Store), MP3 (16 to 320 Kbps), MP3 VBR, Audible (formats 2, 3, and 4), Apple Lossless, WAV, and AIFF

Video

- H.264/AVC video, up to 1.5 Mbps, 640 by 480 pixels, 30 frames per second, Low-Complexity version of the H.264/AV Baseline Profile with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats;
- H.264/AVC video, up to 2.5 Mbps, 640 by 480 pixels, 30 frames per second, Baseline Profile up to Level 3.0 with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats;
- MPEG-4 video, up to 2.5 Mbps, 640 by 480 pixels, 30 frames per second, Simple Profile with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats

iPods for All Tastes …



First iPod ?

"Amplifiers at Bolling Field, 1921." Two giant horns with ear tubes, evidently designed to listen for approaching aircraft.

iPhone

Audio

- Frequency response: 20 Hz to 20000 Hz
- Audio formats supported: AAC, Protected AAC, MP3, MP3 VBR, Audible (formats 1, 2, and 3), Apple Lossless, AIFF, and WAV

Video

- H.264/AVC video, up to 1.5 Mbps, 640 by 480 pixels, 30 frames per second, Low-Complexity version of the H.264 Baseline Profile with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats;

- H.264/AVC video, up to 768 Kbps, 320 by 240 pixels, 30 frames per second, Baseline Profile up to Level 1.3 with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats;

- MPEG-4 video, up to 2.5 Mbps, 640 by 480 pixels, 30 frames per second, Simple Profile with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio in .m4v, .mp4, and .mov file formats

Bibliography

- **The MPEG-4 Book,** F. Pereira, T. Ebrahimi, Prentice Hall, 2002
- **H.264 and MPEG-4 Video Compression**, I. Richardson, John Wiley & Sons, 2003
- **Introduction to Digital Audio Coding and Standards**, M. Bosi and R. Goldberg, Klewer Academic Publishers, 2003