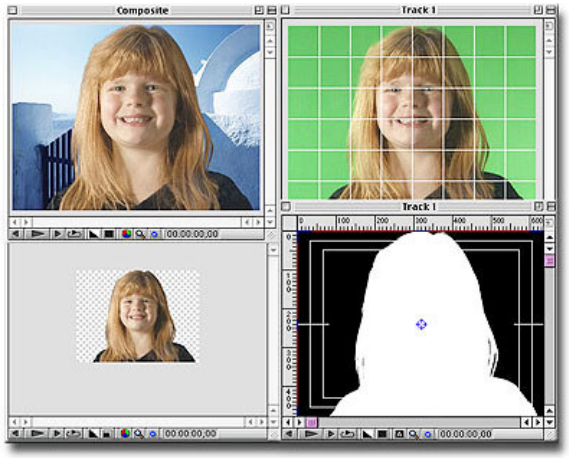
ADVANCED MULTIMEDIA CODING

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 …

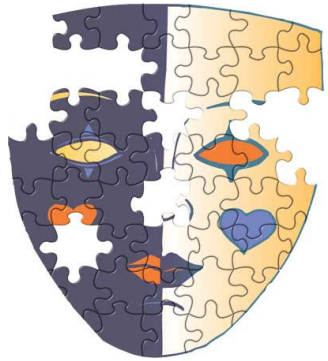
New Technologies, New Needs …

- **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 !

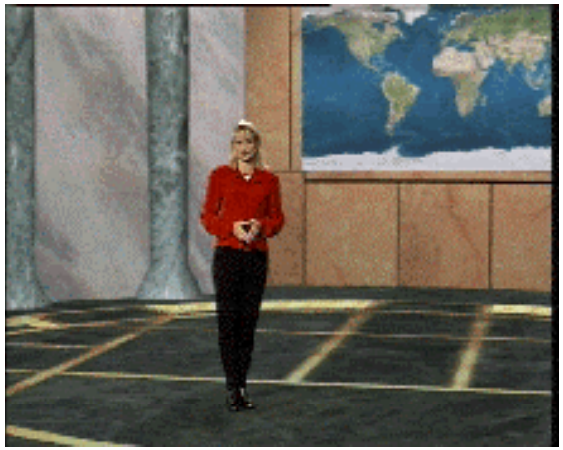
We and the World around us …

Towards the Real World: The Object-based Representation Model



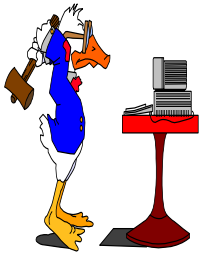
- Audiovisual scene represented as a composition of objects
- Integration of objects from different nature: A&V, natural and synthetic, text & graphics, animated faces, arbitrary and rectangular video shapes, generic 3D, speech and music, ...
- Object-based hyperlinking, processing, coding and description
- 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 …

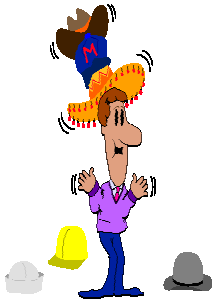
Object-based Content …



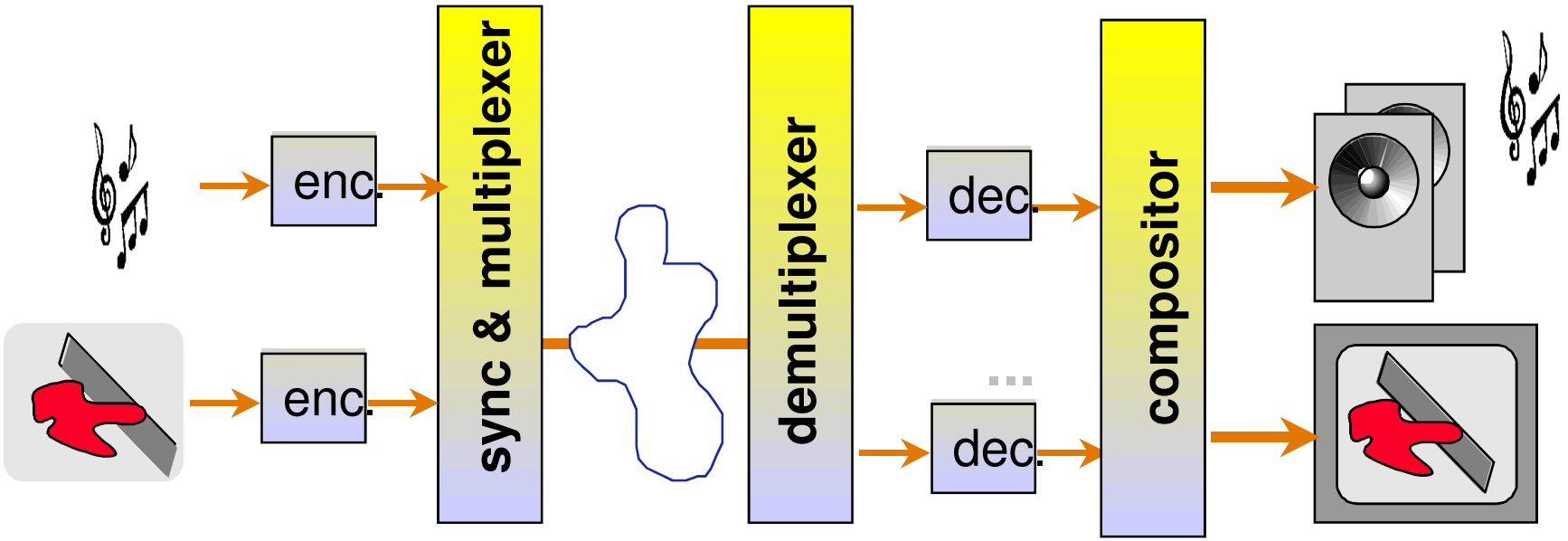
Sports results: Benfica - Sporting

Stock information ...

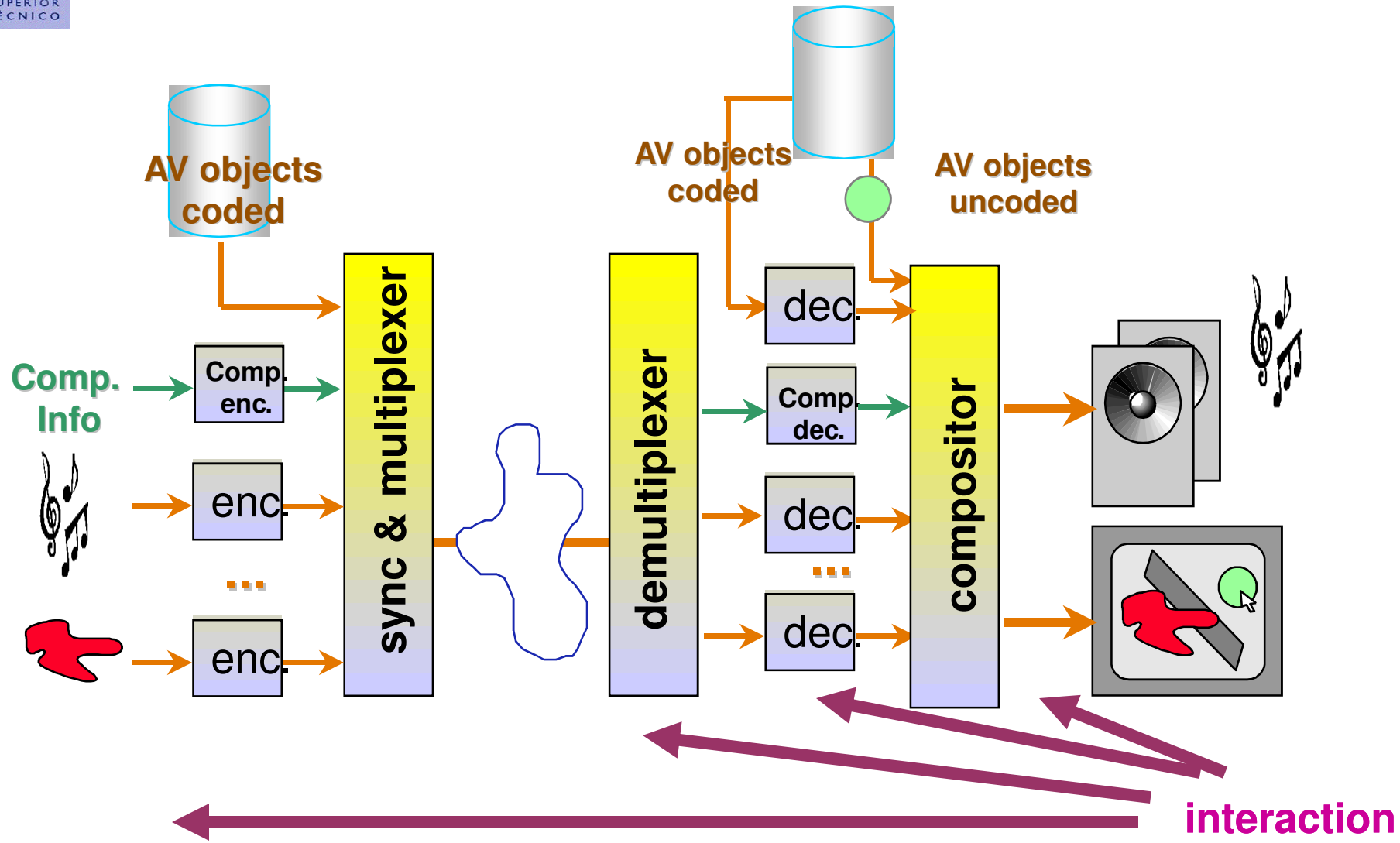




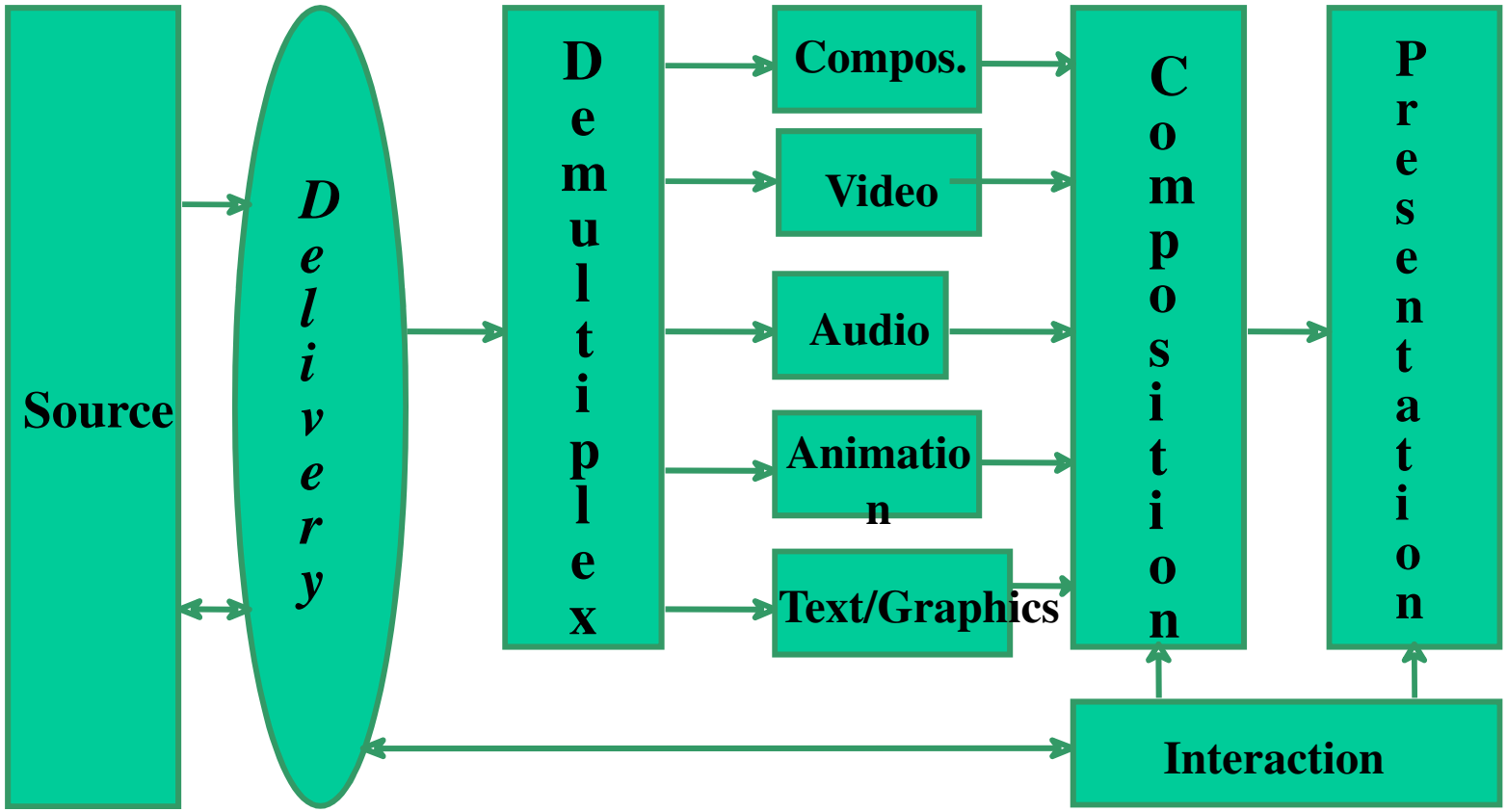
Conventional Audiovisual System



Object-based Audiovisual System

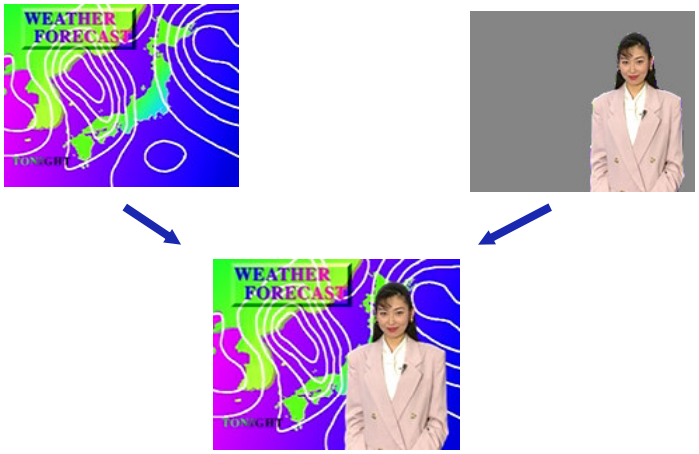


MPEG-4: the New Service Model

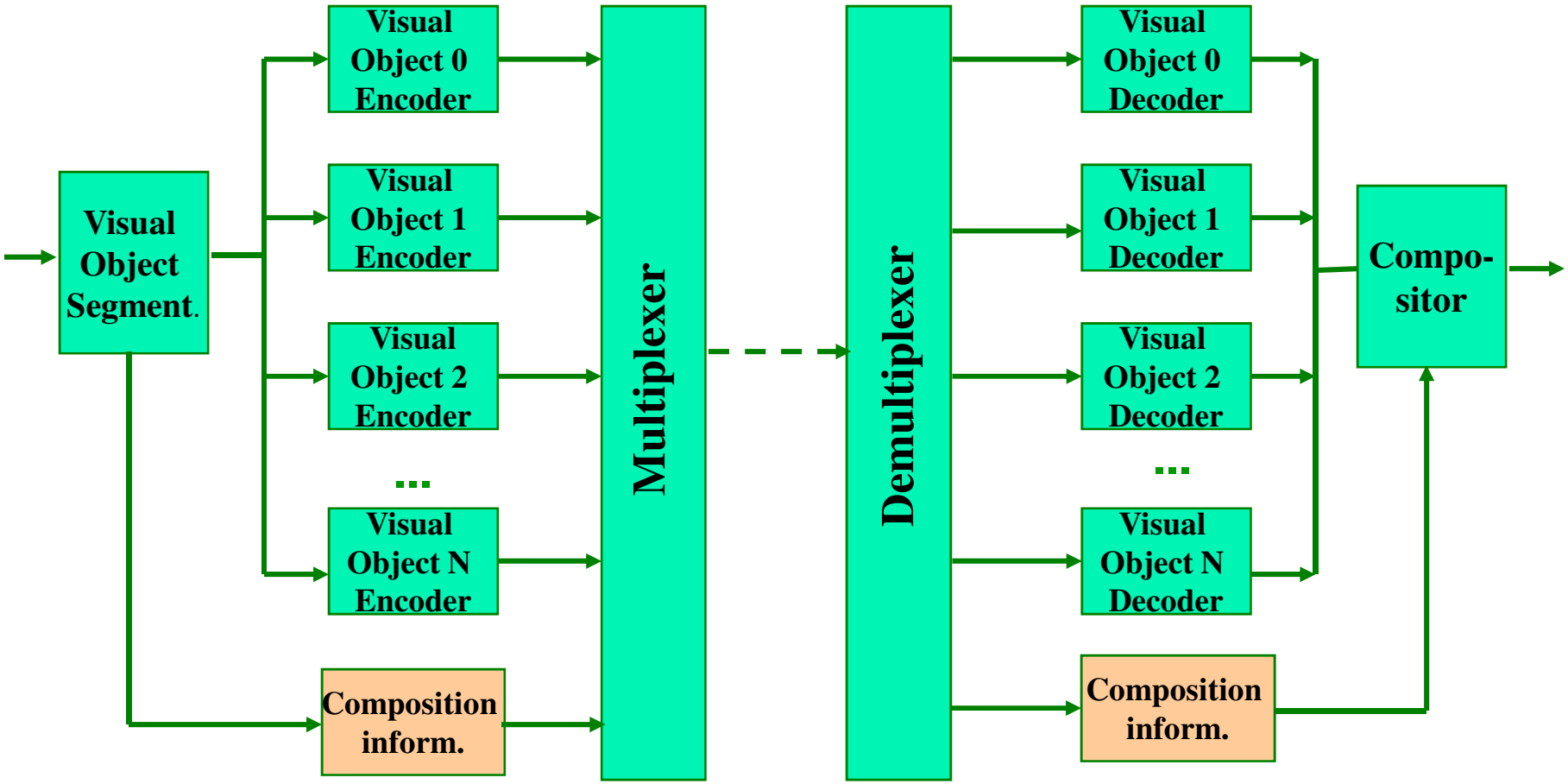


MPEG-4: Object-Based Coding Standard

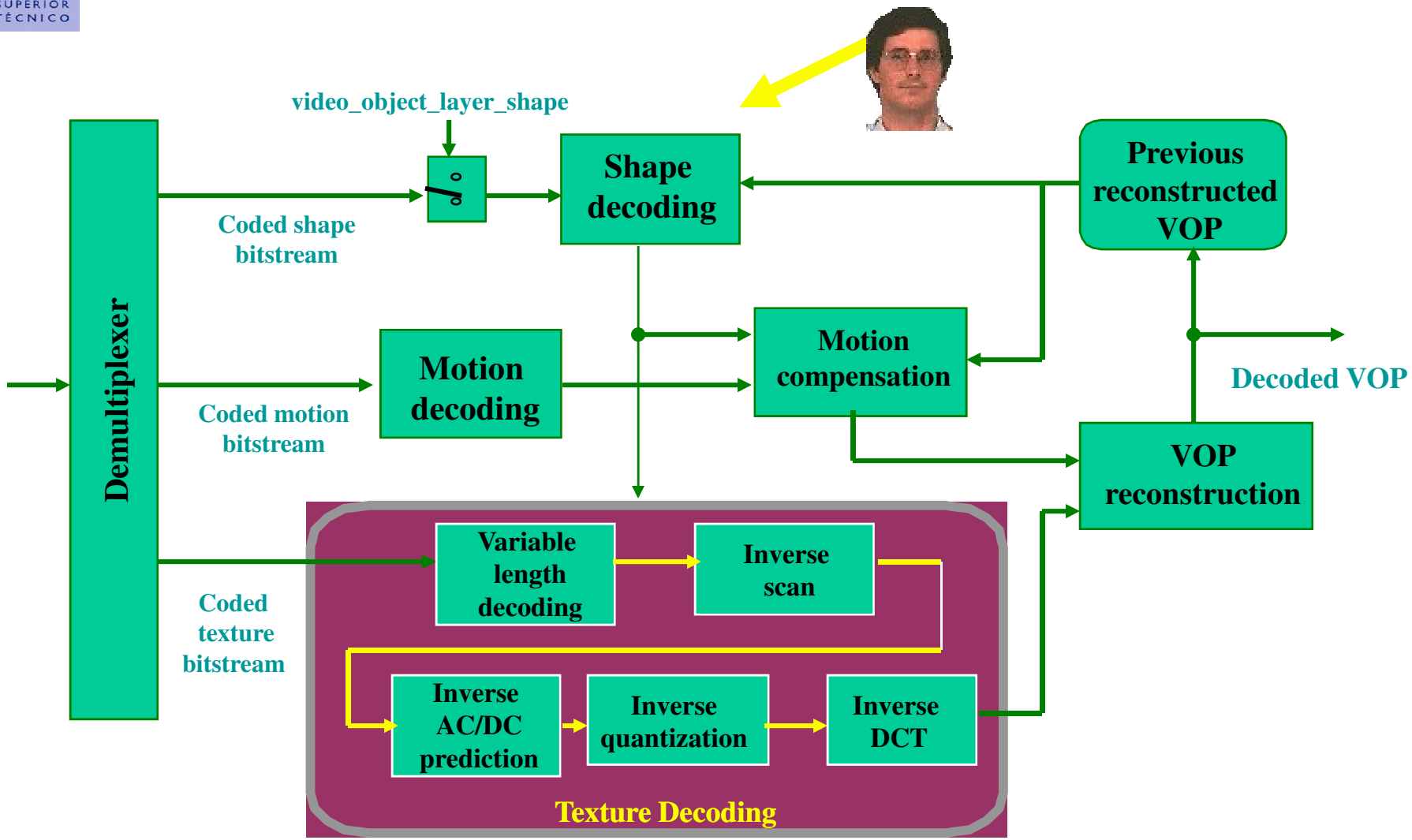
- Adopts the object-based model giving a semantic value to the data structure
- Integration of natural and synthetic content, both aural and visual
- Object-based functionalities, e.g., re-using and manipulation capabilities
- Powerful data model for interaction and personalisation
- Exploitation of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



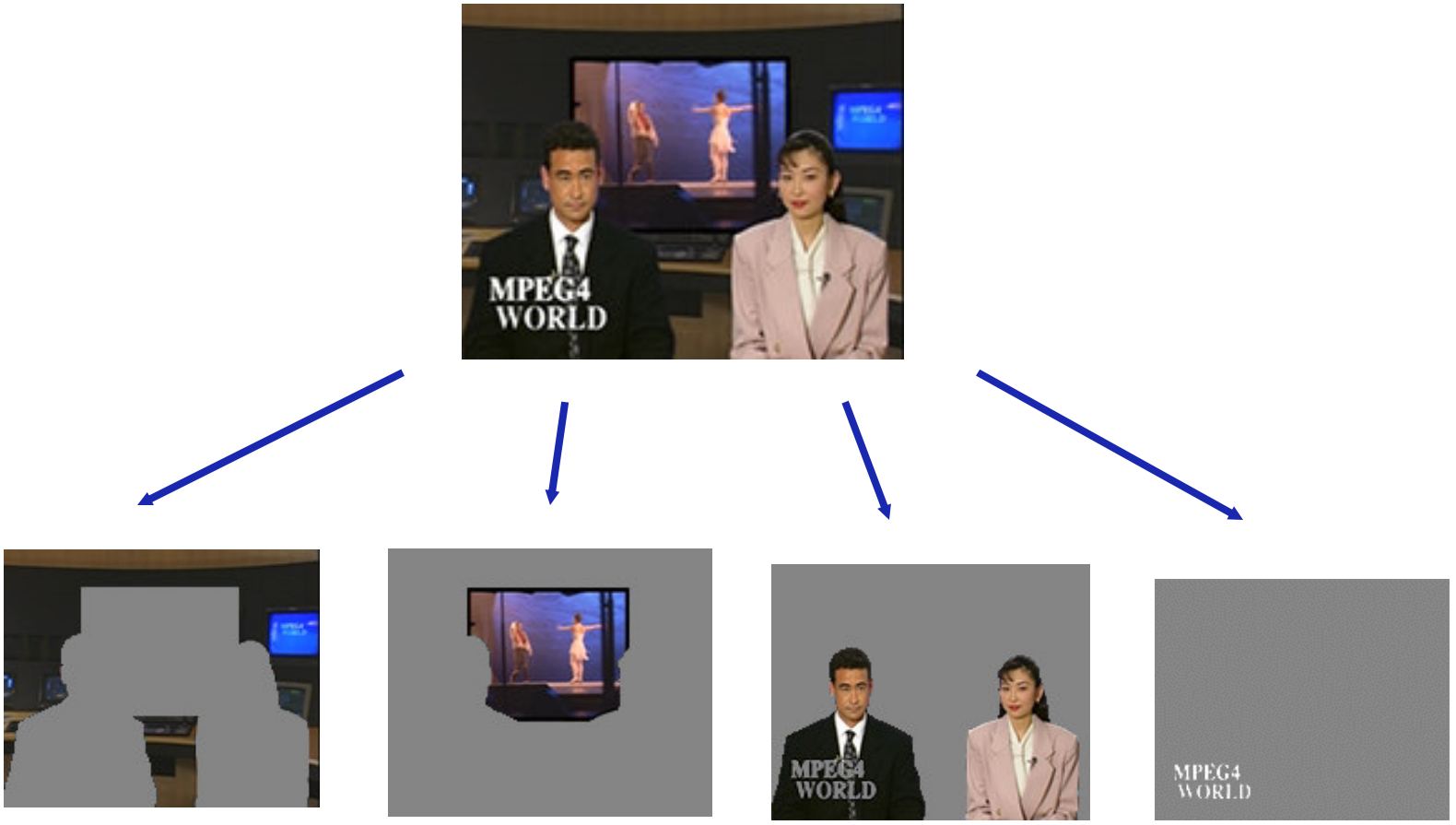
MPEG-4: Visual Coding Architecture



Basic MPEG-4 Video Decoding



Segmentation: a Limitation or not so Much ?

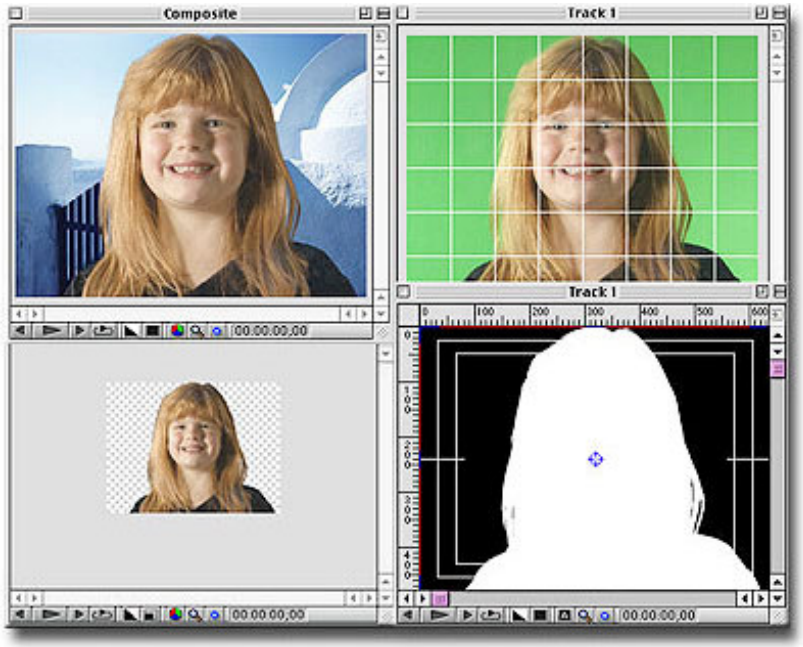


The ‘Weather’ Girl ...

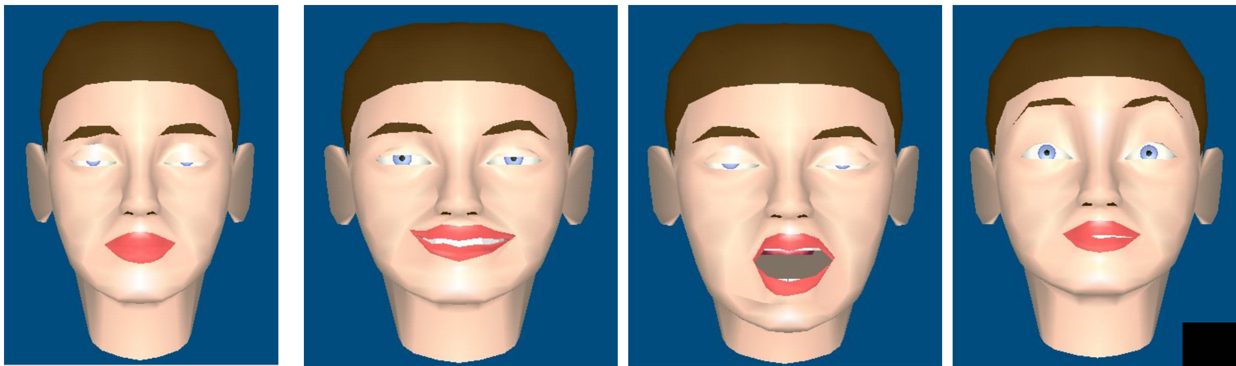
Segmentation: the Problem that Sometimes does not Exist ...

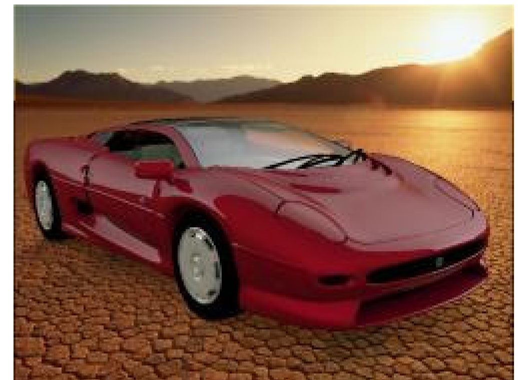
Segmentation: Automatic and Real-Time ?

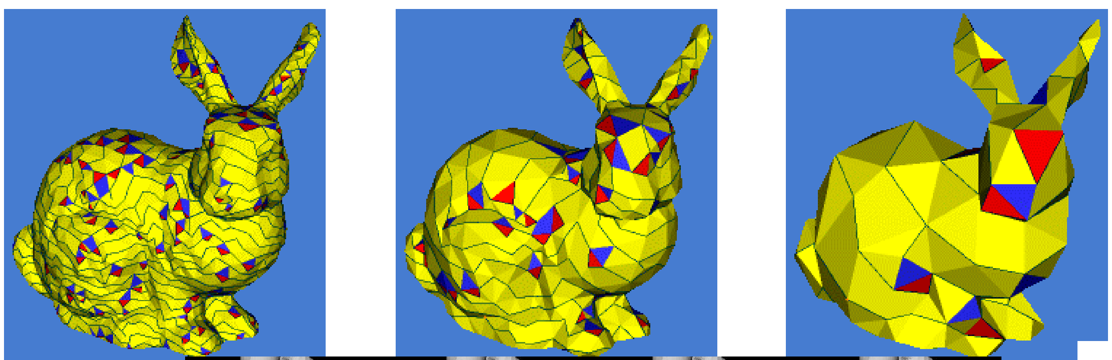
Segmentation by Chroma-Keying …

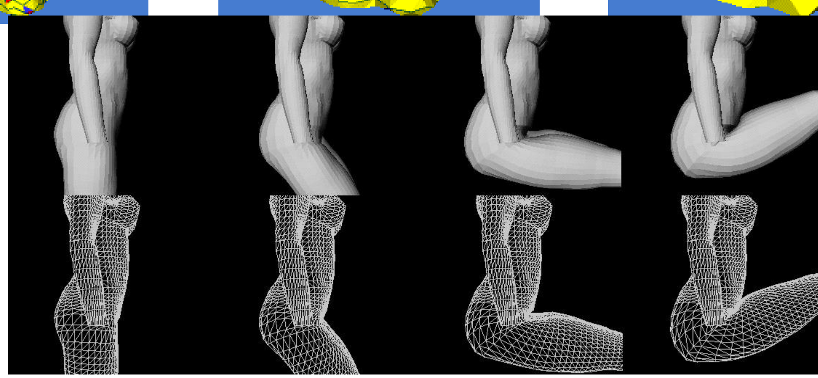


Synthetic Content: Facial Animation and More …









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.**

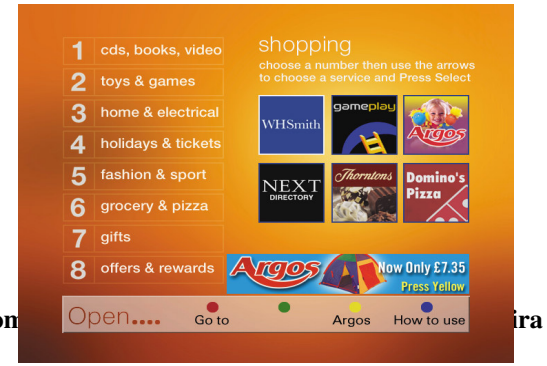
MPEG-4 Application Examples

- 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
- ...

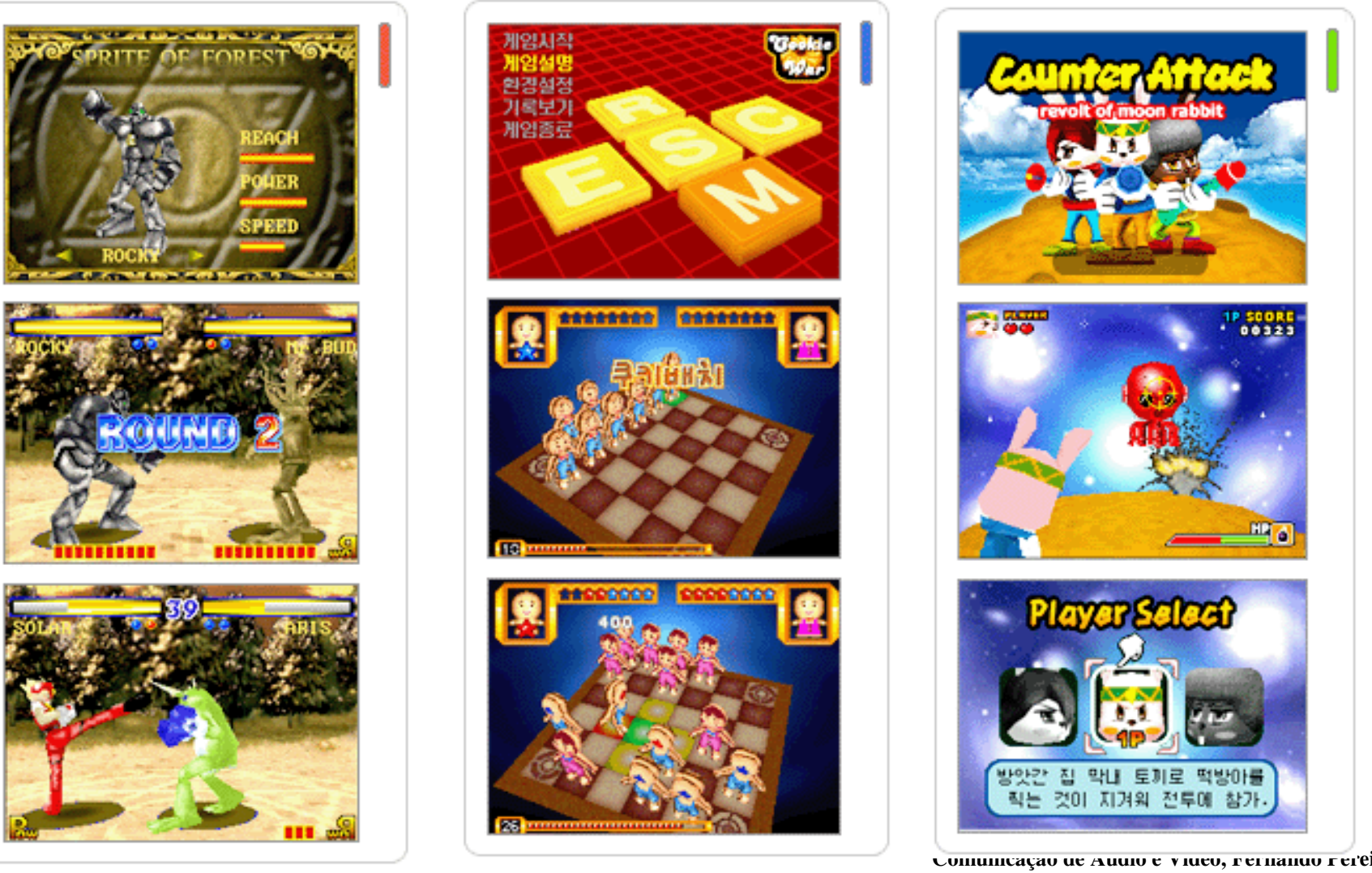
The Bloomberg Case … Today !

- **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 Standard Organisation

- **Part 1: Systems - Specifies scene description, multiplexing and synchronization**
- **Part 2: Visual - Specifies the coding of natural, and synthetic (mostly moving) images**
- **Part 3: Audio - Specifies the coding of natural and synthetic sounds**
- **Part 4: Conformance Testing - defines conformance conditions for bitstreams and terminals**
- **Part 5: Reference Software - Includes software regarding most parts of MPEG-4 (normative and non-normative)**
- **Part 6: Delivery MM Integration Framework (DMIF) - Defines a session protocol for the management of multimedia streaming over generic delivery technologies**
- **Parte 10: Advanced Video Coding (AVC) – Specifies advanced coding of rectangular video (jointly with ITU-T, H.264/AVC)**

MPEG-4 Profiling: Interoperability versus Complexity

- The MPEG-4 standard offers many tools for each Part.
- There are many tools in each Part that are useless or too expensive for certain classes of applications.
- It is essential to offer coding solutions appropriate for each application class in terms of complexity.

The specification of Profiles and Levels has the objective to offer coding solutions appropriate for relevant application classes, maximization interoperability while minimizing complexity.

MPEG-4 Profiling Basic Elements

- **Object Type - Defines the syntax of the bitstream for one single object that can represent a meaningful entity in the (Audio or Visual) scene.**
- **Profile - Defines the set of a certain type of tools that can be used in a certain MPEG-4 terminal. There are Audio, Visual, Graphics Media, Scene Description, Object Descriptor and MPEG-J profiles. Audio and Visual profiles are defined in terms of the audio and visual object types.**
- **Level – Defines the constraints and performance criteria on an Audio, Visual, Graphics Scene Description Object Descriptor or MPEG-J profile, and thus on the corresponding tools.**

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).

Detailed Goals

- **Improved Coding Efficiency**
  - Average bitrate reduction of 50% given fixed fidelity compared to any other standard
  - Complexity vs. coding efficiency scalability
- **Improved Network Friendliness**
  - Issues examined in H.263 and MPEG-4 are further improved
  - Anticipate error-prone transport over mobile networks and the wired and wireless Internet
- **Simple Syntax Specification**
  - Targeting simple and clean solutions
  - Avoiding any excessive quantity of optional features or profile configurations

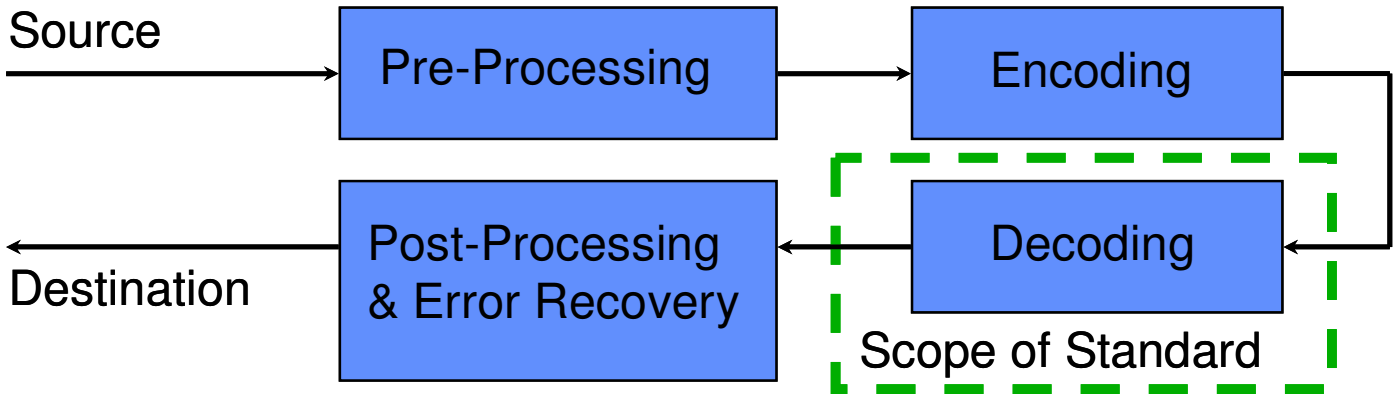
Applications

- **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**

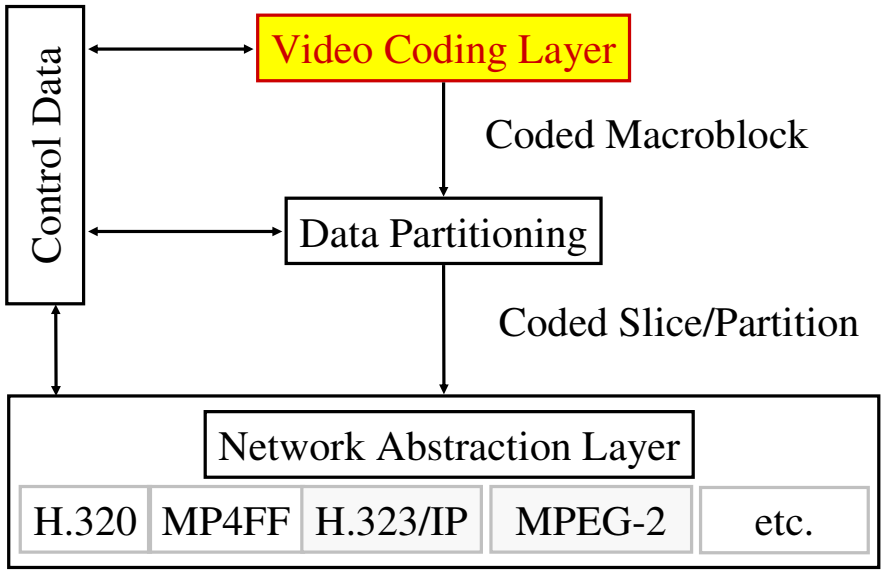
The Scope of the Standard

The standard specifies only the bitstream syntax and semantics as well as the decoding process:

- **Allows several types of encoding optimizations**
- **Allows to reduce the encoding implementation complexity (at the cost of some quality)**
- **Does NOT allow to guarantee any minimum level of quality !**



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**

NAL Basics

- **The coded video data are organized into NAL units, which are packets that each contains an integer number of bytes.**
- **A NAL unit starts with a one-byte header, which signals the type of data it contains. The remaining bytes represent payload data.**
- **NAL units are classified into VCL NAL units, which contain coded slices or coded slice data partitions, and non-VCL NAL units, which contain associated additional information.**
- **The most important non-VCL NAL units are parameter sets and Supplemental Enhancement Information (SEI).**
  - **The sequence and picture parameter sets contain infrequently changing information for a video sequence.**
  - **SEI messages are not required for decoding the samples of a video sequence; they provide additional information which can assist the decoding process or related processes like bit stream manipulation or display.**
- **A set of consecutive NAL units with specific properties is referred to as an access unit. The decoding of an access unit results in exactly one decoded picture.**
- **A set of consecutive access units with certain properties is referred to as a coded video sequence.**

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:

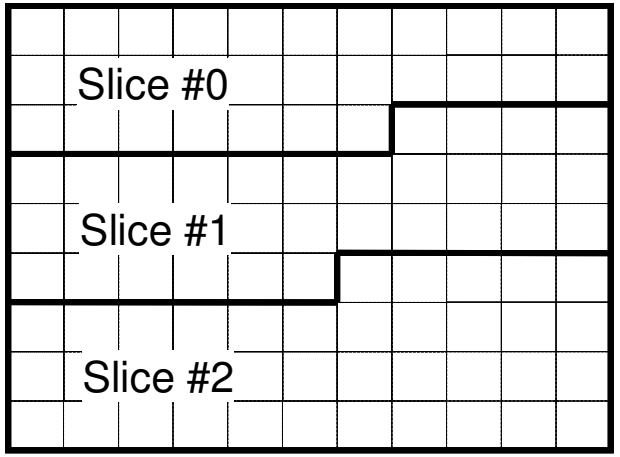
- **Variable (and smaller) block size motion compensation**
- **Multiple reference frames**
- **Hierarchical transform with smaller block sizes**
- **Deblocking filter in the prediction loop**
- **Improved, adaptive entropy coding**

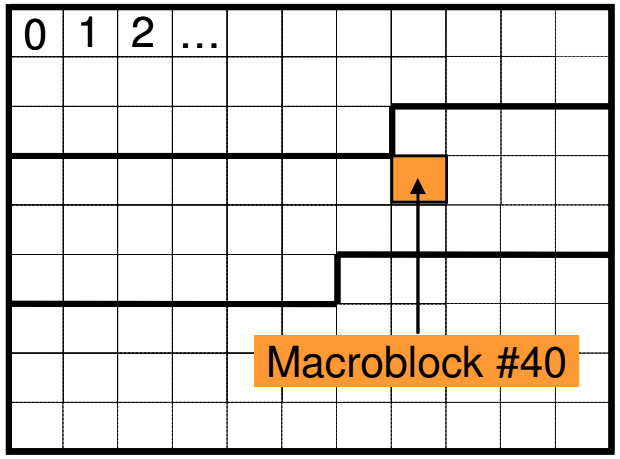
which all together allow achieving substantial gains regarding the bitrate needed to reach a certain quality level.

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

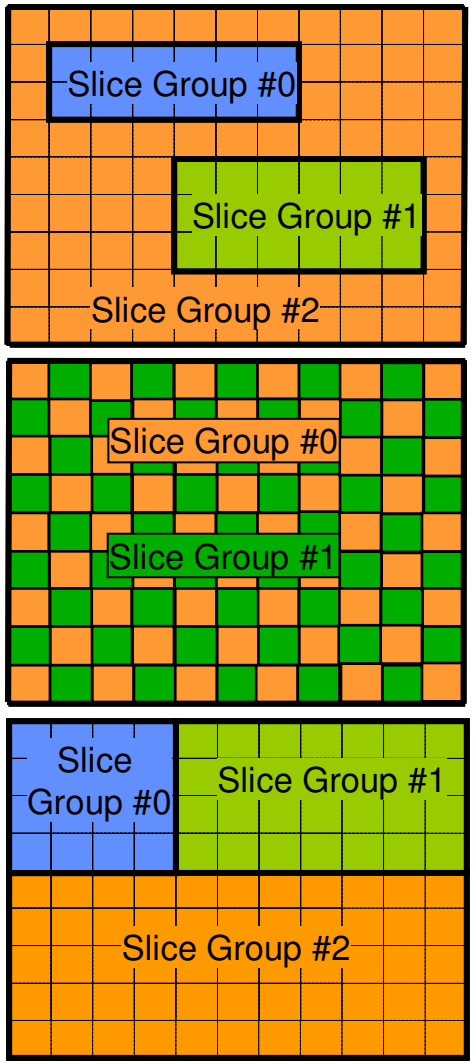
- **Picture (Y,Cr,Cb; 4:2:0 and later more; 8 bit/sample):**
  - A picture (frame or field) is split into 1 or several slices
- **Slice:**
  - Slices are self-contained
  - Slices are a sequence of macroblocks
- **Macroblock:**
  - Basic syntax & processing unit
  - Contains 16×16 luminance samples and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





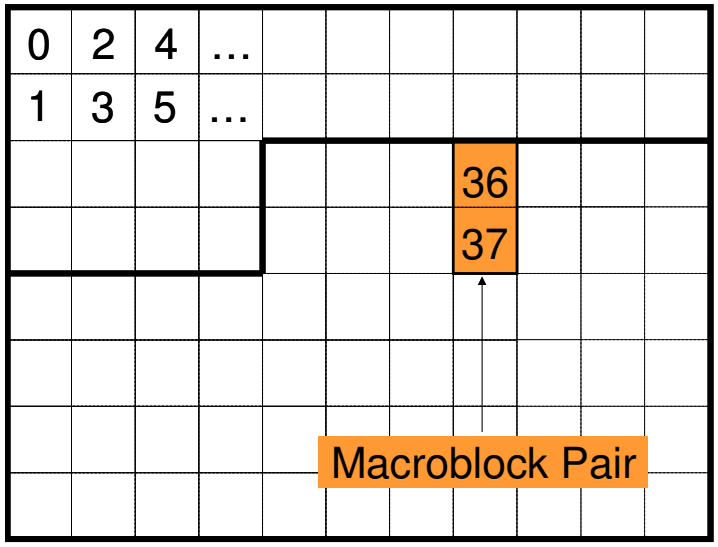
Slices and Slice Groups

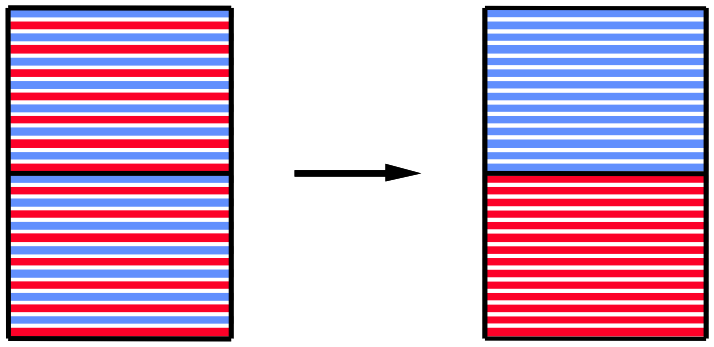
- **Slice Group:**
  - Pattern of macroblocks defined by a Macroblock Allocation Map
  - A slice group may contain 1 to several slices
- **Macroblock Allocation Map Types:**
  - Interleaved slices
  - Dispersed macroblock allocation
  - Explicitly assign a slice group to each macroblock location in raster 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



Interlaced Processing

- **Field coding**
  - each field is coded as a separate picture using fields for motion compensation
- **Frame coding**
  - Type 1: the complete frame is coded as a separate picture
  - Type 2: the frame is scanned as macroblock pairs, for each macroblock pair: switch between frame and field coding

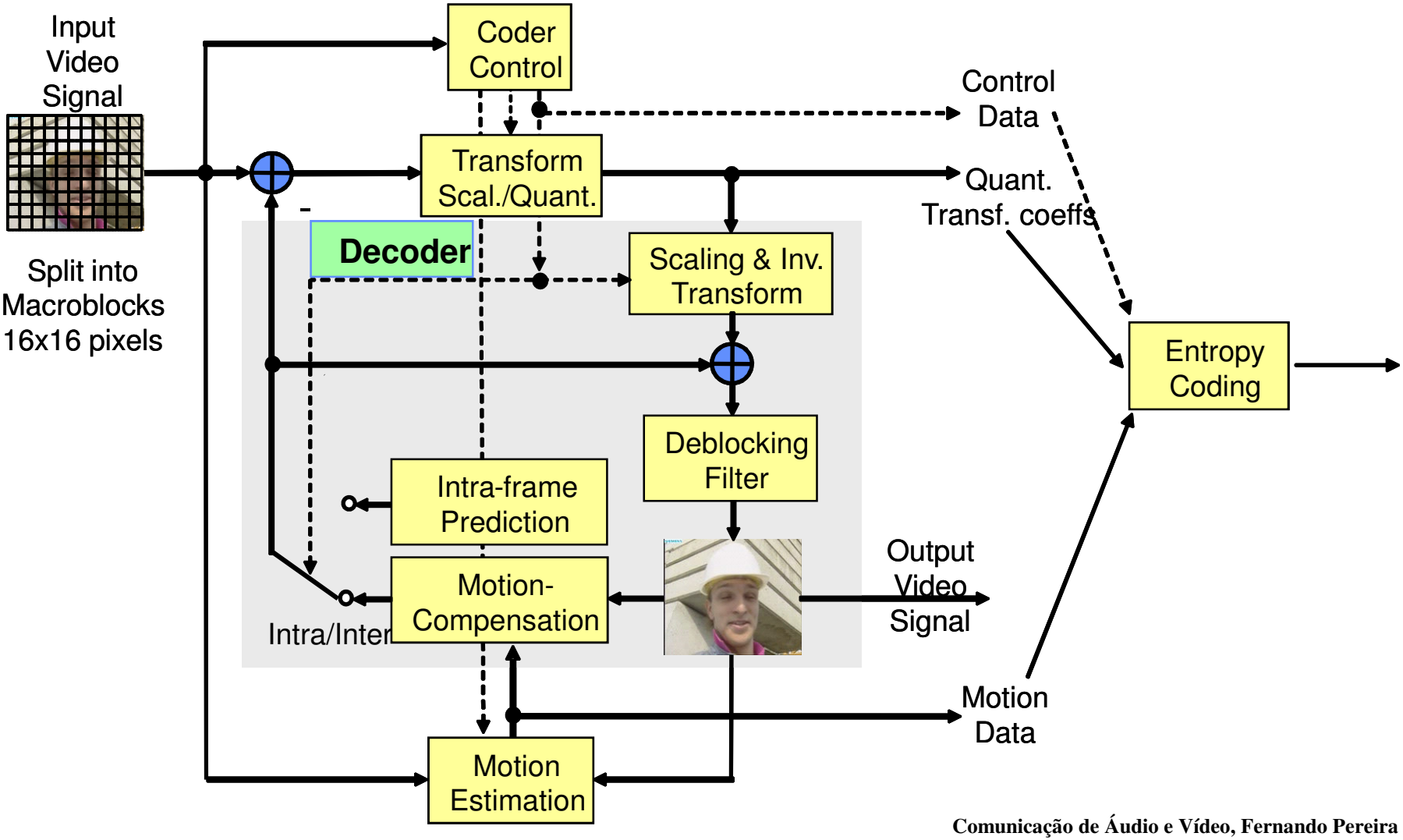




A Pair of Macroblocks in Frame Mode

Top/Bottom Macroblocks in Field Mode

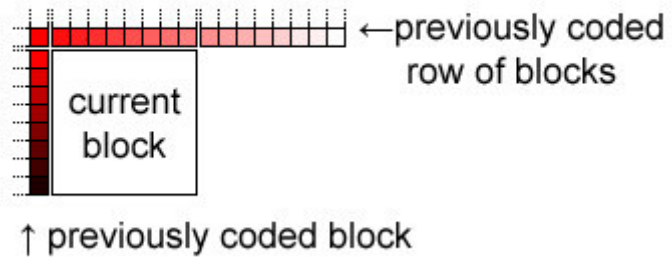
H.264/AVC Encoding Architecture



Common Elements with other Standards

- Original data: Luminance and two chrominances
- Macroblocks: 16×16 luminance + $2 \times 8 \times 8$ 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.

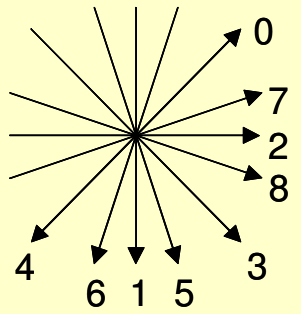
Intra Prediction Types

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)**

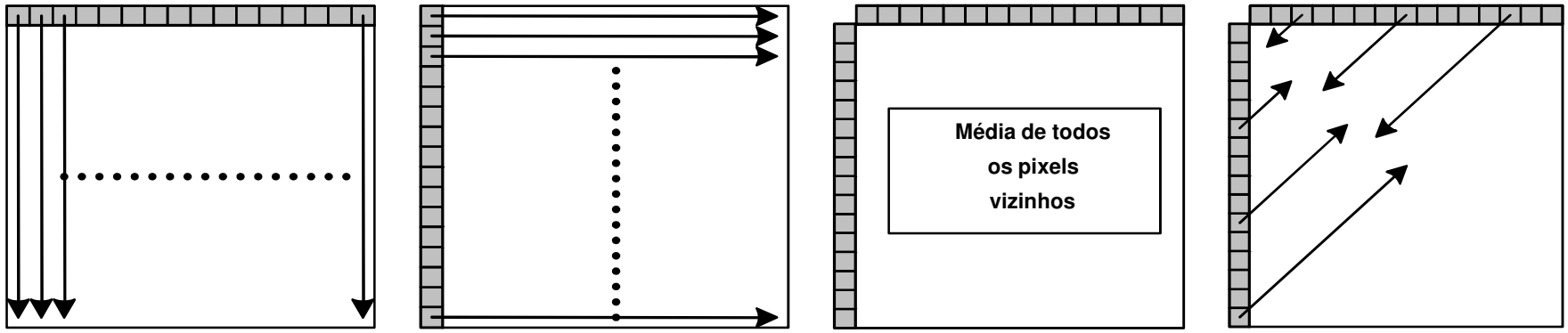
- Directional spatial prediction (9 types for luma, 1 chroma)

	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				



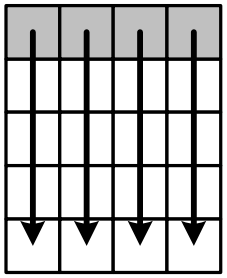
- e.g., Mode 3: diagonal down/right prediction a, f, k, p are predicted by (A + 2Q + I + 2) >> 2

16×16 Blocks Intra Prediction Modes

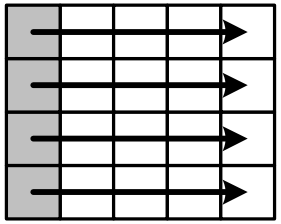


- **The luminance is predicted in the same way for all samples of a 16×16 MB (Intra16×16 modes).**
- **This coding mode is adequate for the image areas which have a smooth variation.**

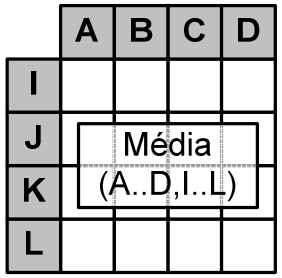
4×4 Intra Prediction Directions



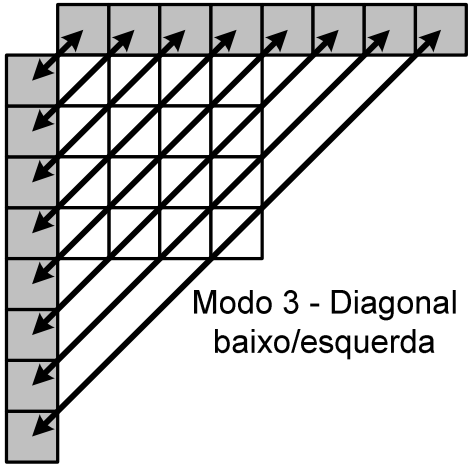
Modo 0 - Vertical



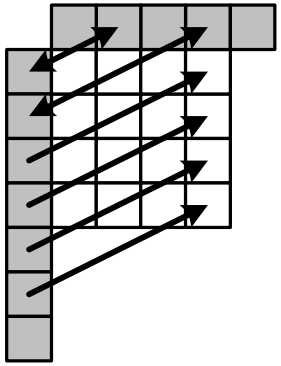
Modo 1 - Horizontal



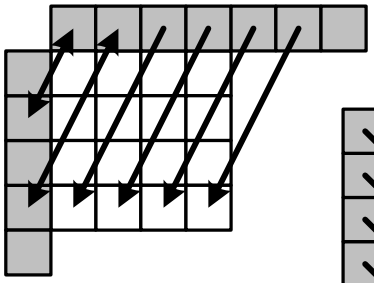
Modo 2 - DC



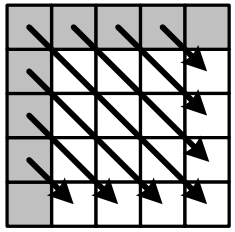
Modo 3 - Diagonal baixo/esquerda



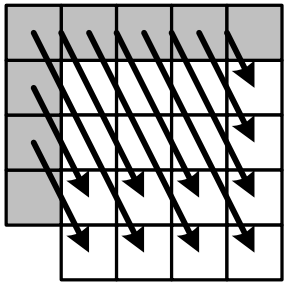
Modo 8 - Horizontal cima



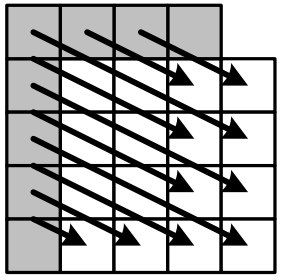
Modo 7 - Vertical/ esquerda



Modo 4 - Diagonal baixo/direita

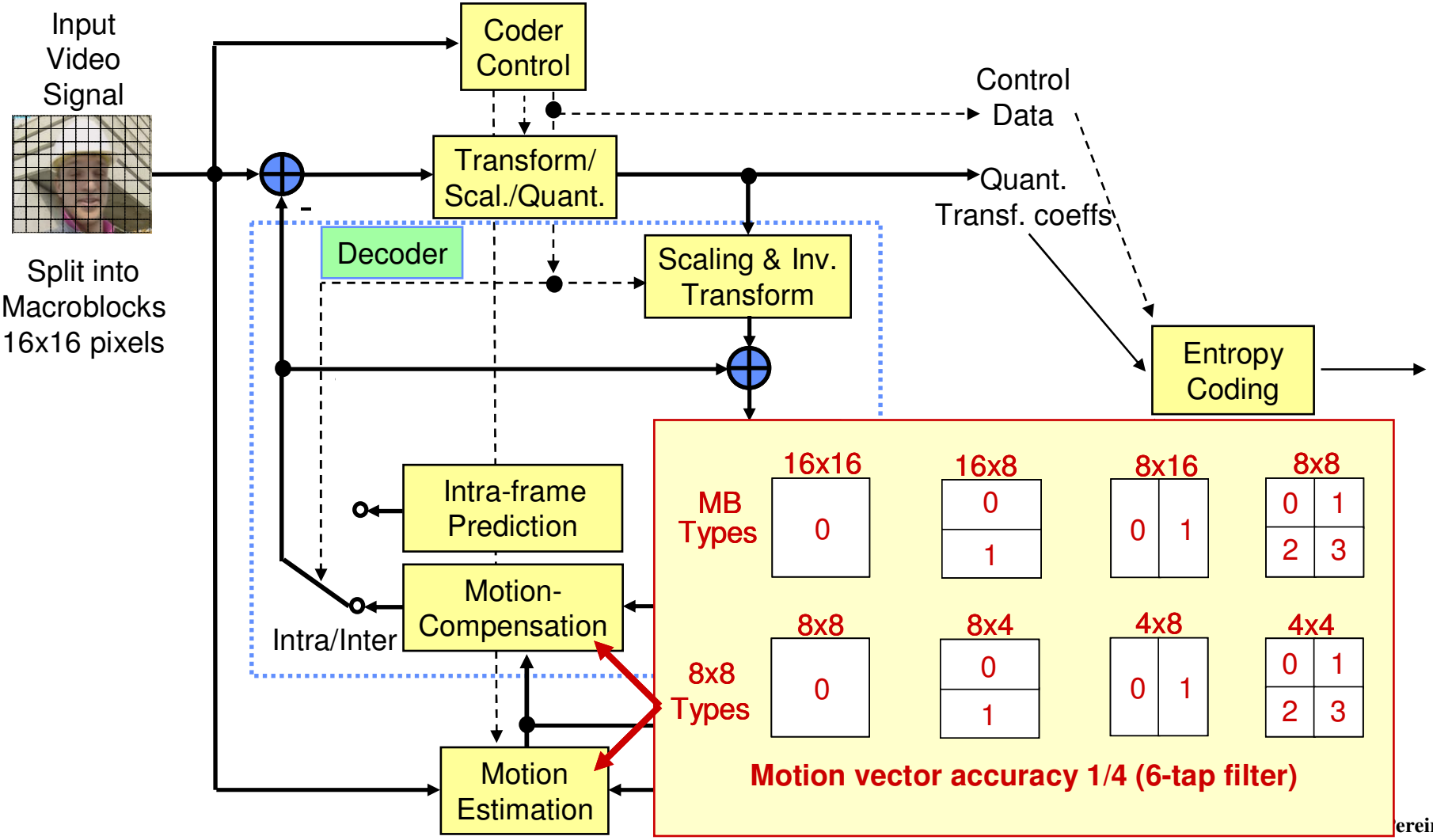


Modo 5 - Vertical/ direita



Modo 6 - Horizontal/ baixo

Variable Block- Size Motion Compensation

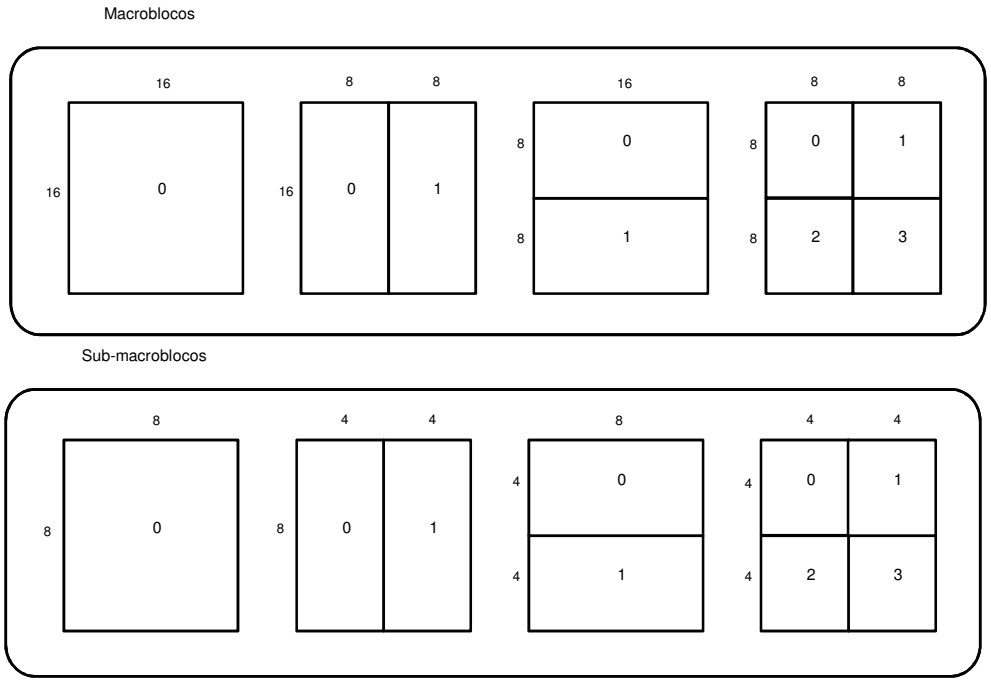


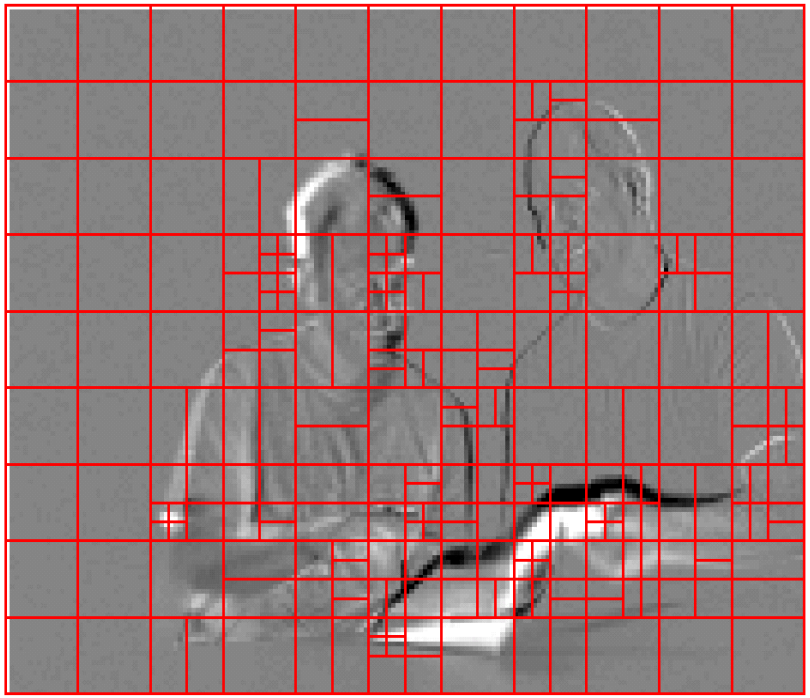
Flexible Motion Compensation

- 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.

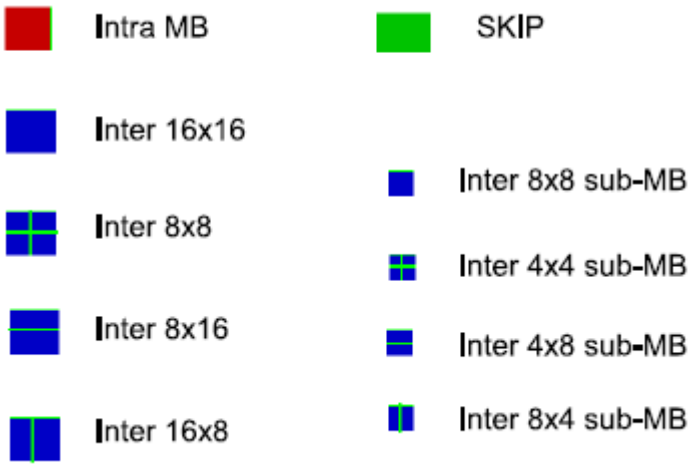
MBs and sub-MBs Partitioning for Motion Compensation



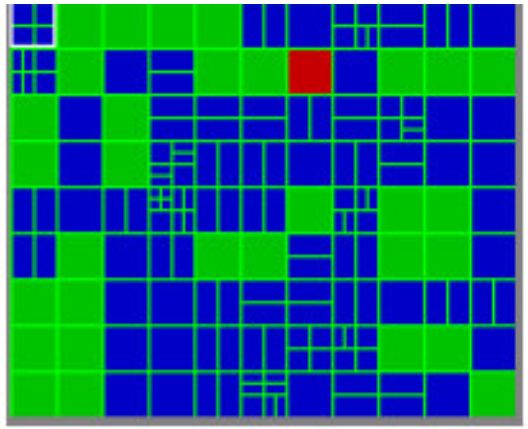


Motion vectors are differentially coded but not across slices.

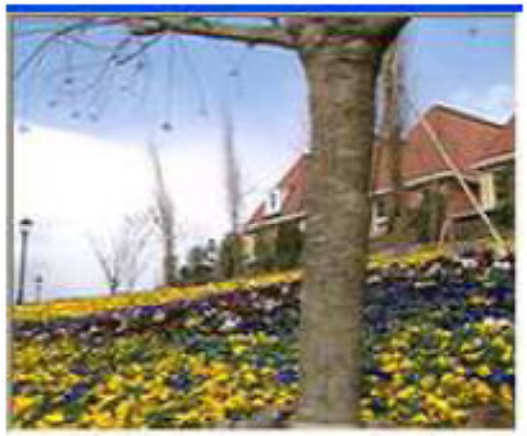
INSTITUTO SUPERIOR TÉCNICO



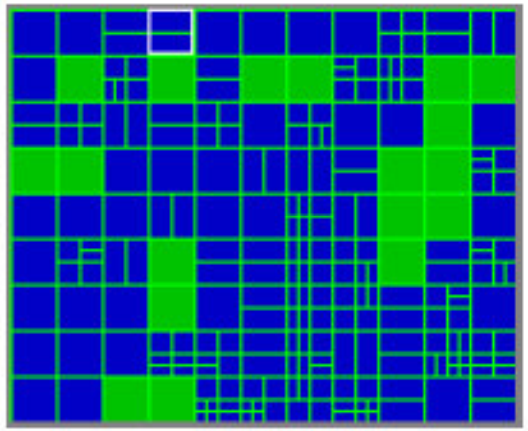
(a) Foreman second frame



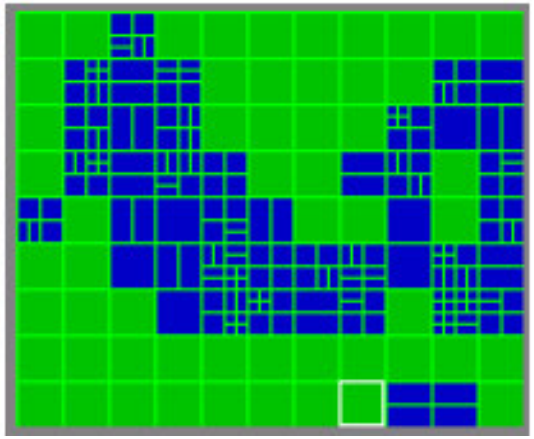
(b) Foreman second frame mode decision



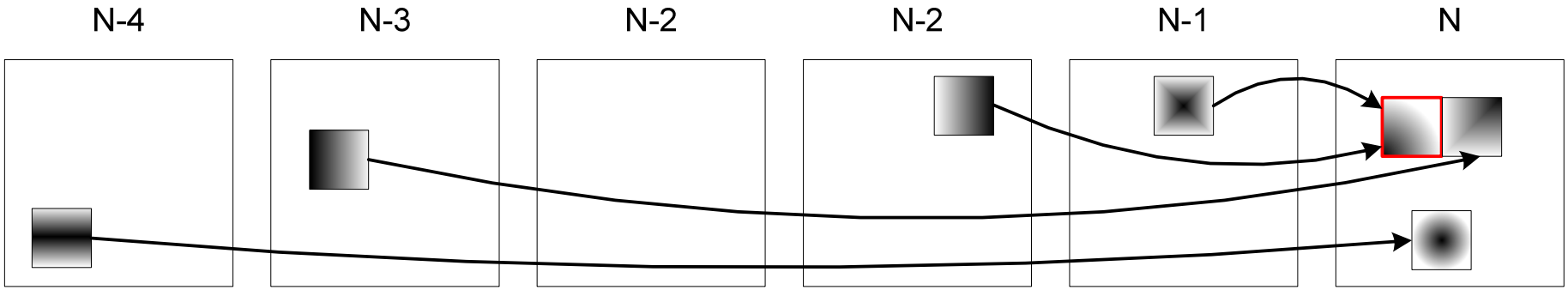
(c) Flower second frame



(d) Flower second frame mode decision



Multiple Reference Frames



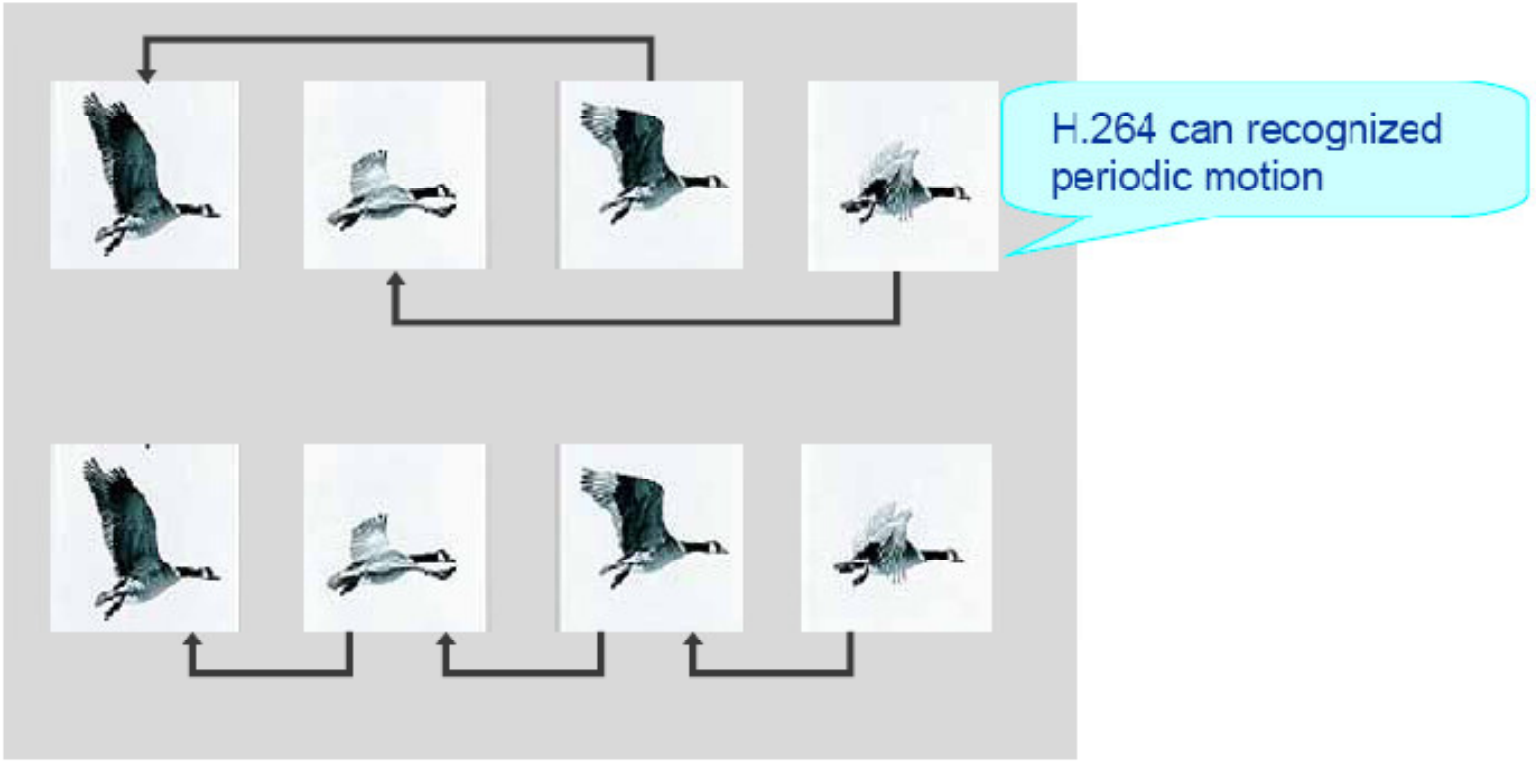
The H.264/AVC standard supports motion compensation with multiple reference frames this means that more than one previously coded picture may be simultaneously used as prediction reference for the motion compensation of the MBs in a picture (at the cost of memory and computation).

- **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

H.264/AVC

Other standards



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.

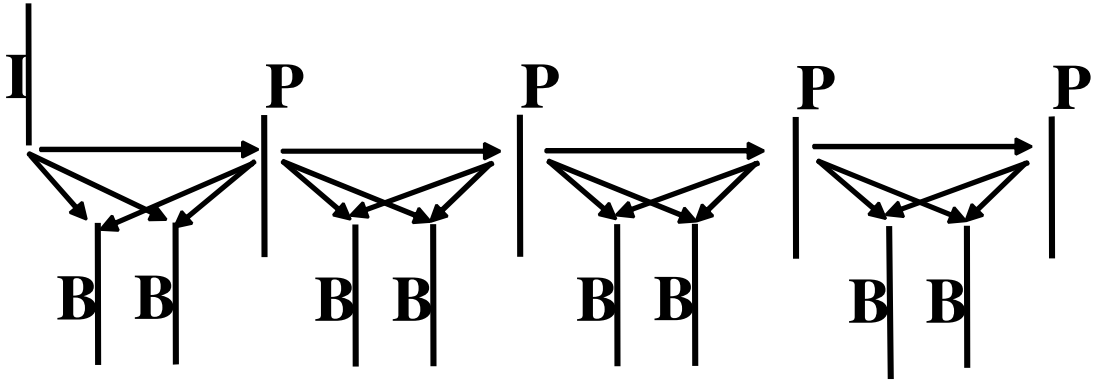
- **For B slices, some blocks or MBs are coded using a weighted prediction of two blocks or MBs in two reference frames, both in the past, both in the future, or one in the past and another in the future.**
- **B type frames use two reference frames, referred as the first and second reference frames.**
- **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.**

Weighted Prediction for P and B Slices

- For each MB partition, it is possible to use a weighted prediction obtained from one or two reference frames.
- In addition to shifting in spatial position, and selecting from among multiple reference pictures, each region's prediction sample values can be multiplied by a weight, and given an additive offset.
- For B-MBs, the weighted prediction may consist in performing motion compensation from the two reference frames and compute the prediction using a set weights w_1 and w_2 .
- Some key uses: improved efficiency for B coding, e.g., accelerating motion, illumination variations; excels at representation of fades: fade-in, fade-out, cross-fade from scene-to-scene.

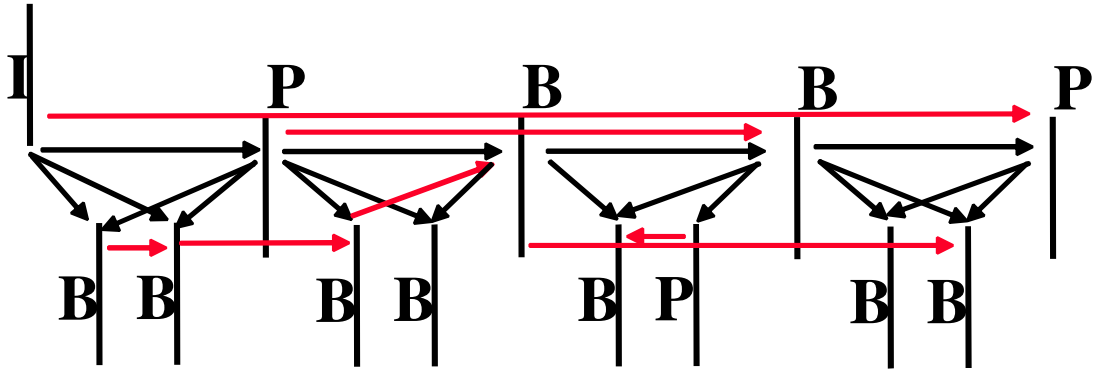
New Types of Temporal Referencing

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

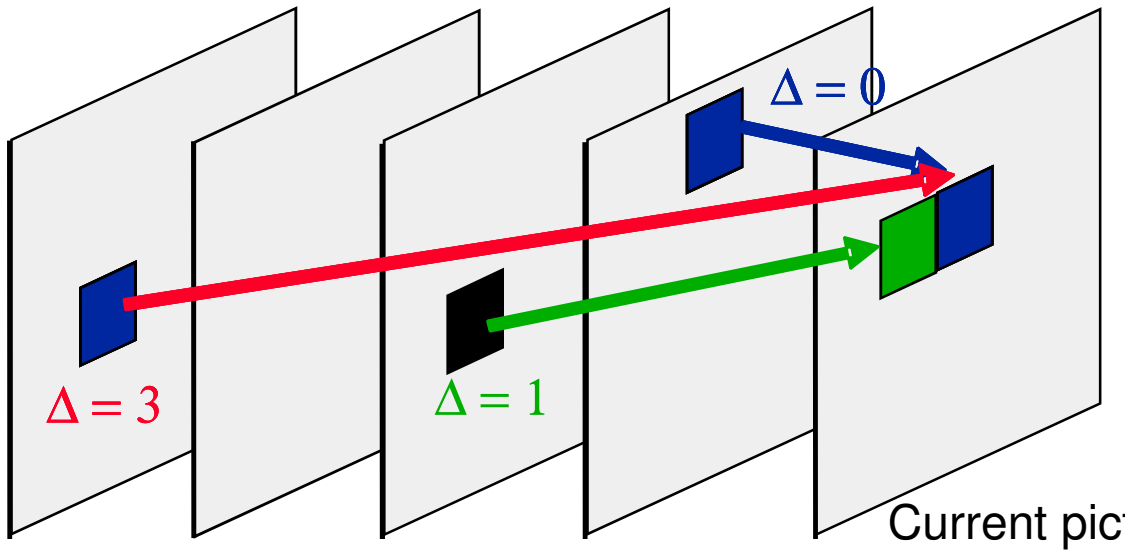


New types of dependencies:

- **Referencing order and display order are decoupled**
- **Referencing ability and picture type are decoupled, e.g. it is possible to use a B frame as reference**



Multiple Reference Frames and Generalized Bi-Predictive Frames

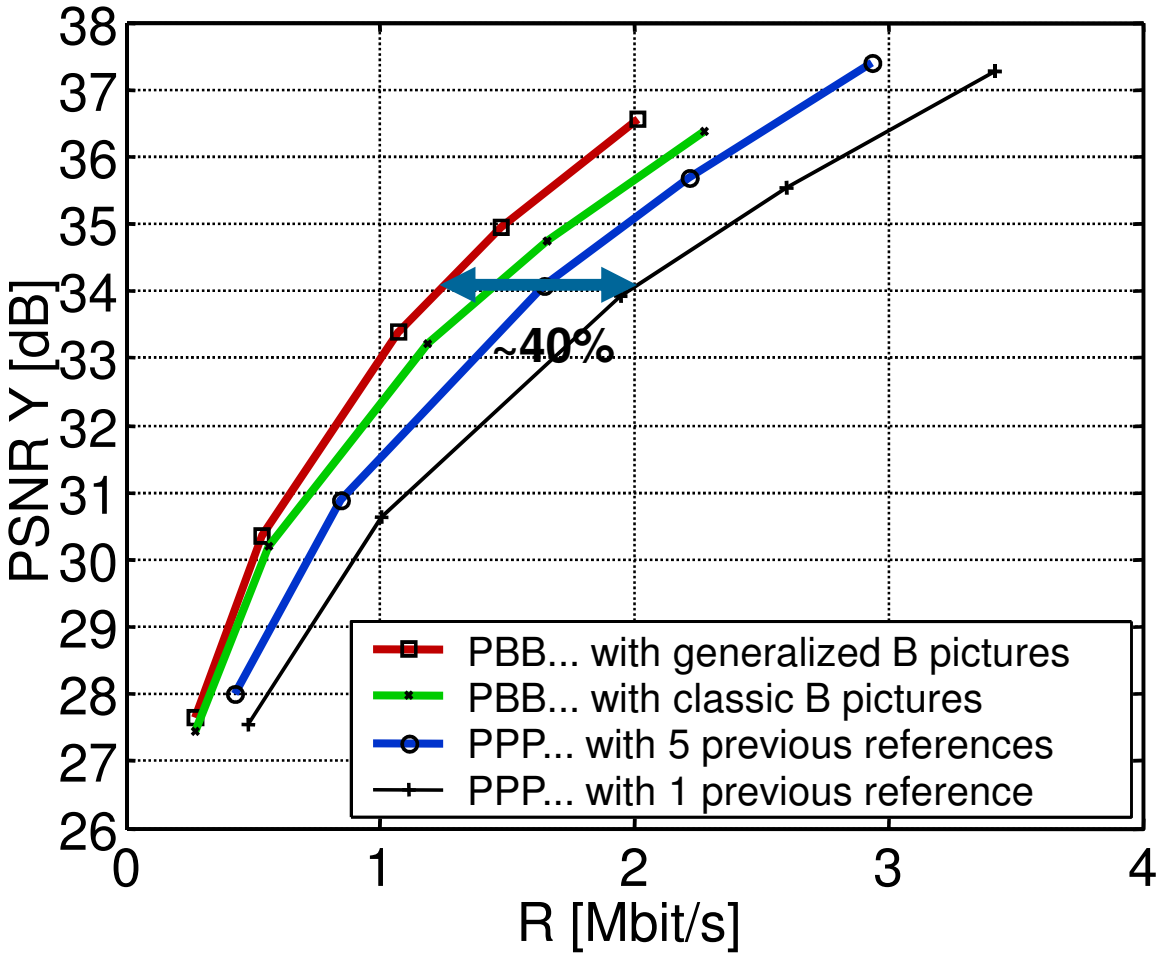


4 Prior Decoded Pictures as Reference

1. Extend motion vector by reference picture index Δ
2. Provide reference pictures at decoder side
3. In case of bi-predictive pictures: decode 2 sets of motion parameters

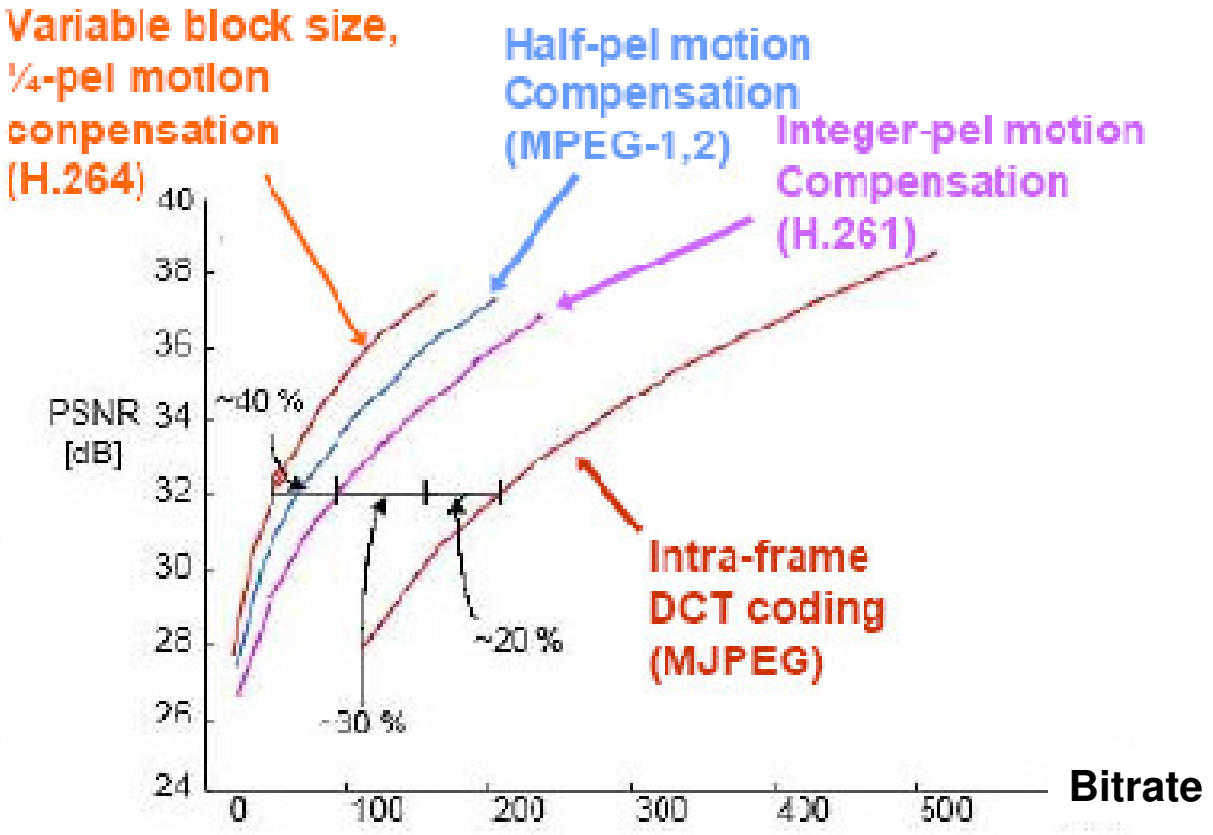
If the memory allows to store more than one picture, the reference picture index is transmitted for each 16×16, 8×16, 16×8 or 8×8 MB partition, indicating to the decoder which reference pictures should be used for that MB from those available in the memory.

Comparative Performance: Mobile & Calendar, CIF, 30 Hz



Motion Estimation is King …

e.g. Foreman
10Hz, QCIF

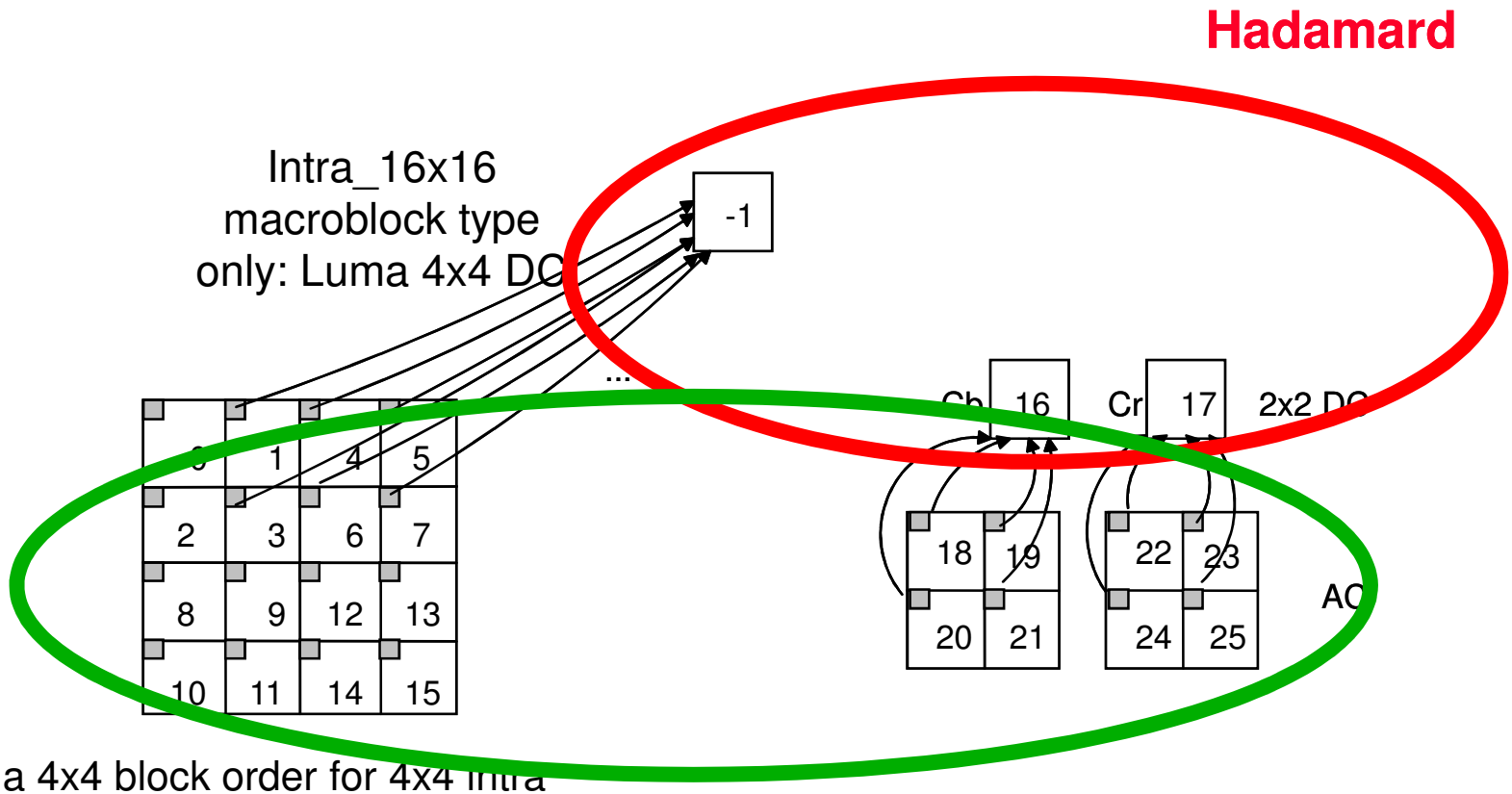


Multiple Transforms

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 ?



Luma 4x4 block order for 4x4 intra prediction and 4x4 residual coding

Chroma 4x4 block order for 4x4 residual coding, shown as 16-25, and Intra4x4 prediction, shown as 18-21 and 22-25

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.
- One 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% on the bitrate for an increment of 1 in the quantization step value, Q_{step}.

Deblocking Filter in the Loop (1)

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

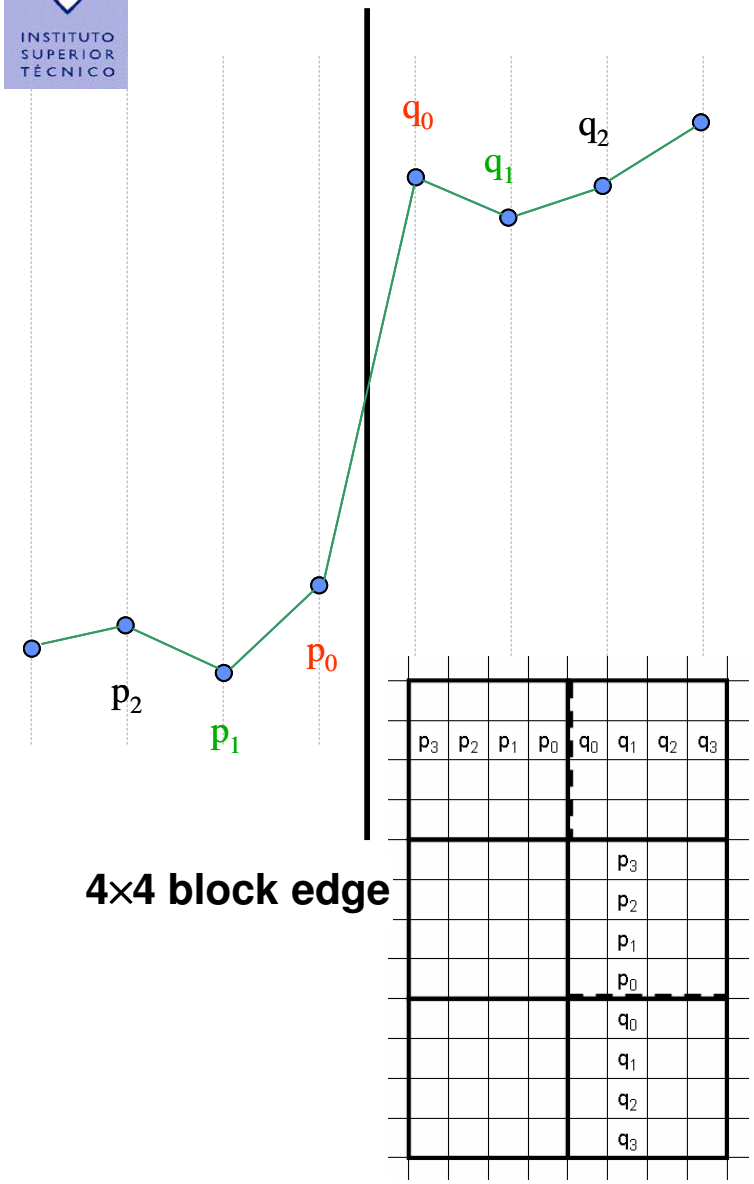
- **This filter needs to be present at the encoder and decoder (normative at decoder) since the filtered blocks are after used for motion estimation (filter in the loop). This filter has a superior performance to a post-processing filter (not in the loop and thus not normative).**

- **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. $|q_1 - q_0| < \beta(QP)$

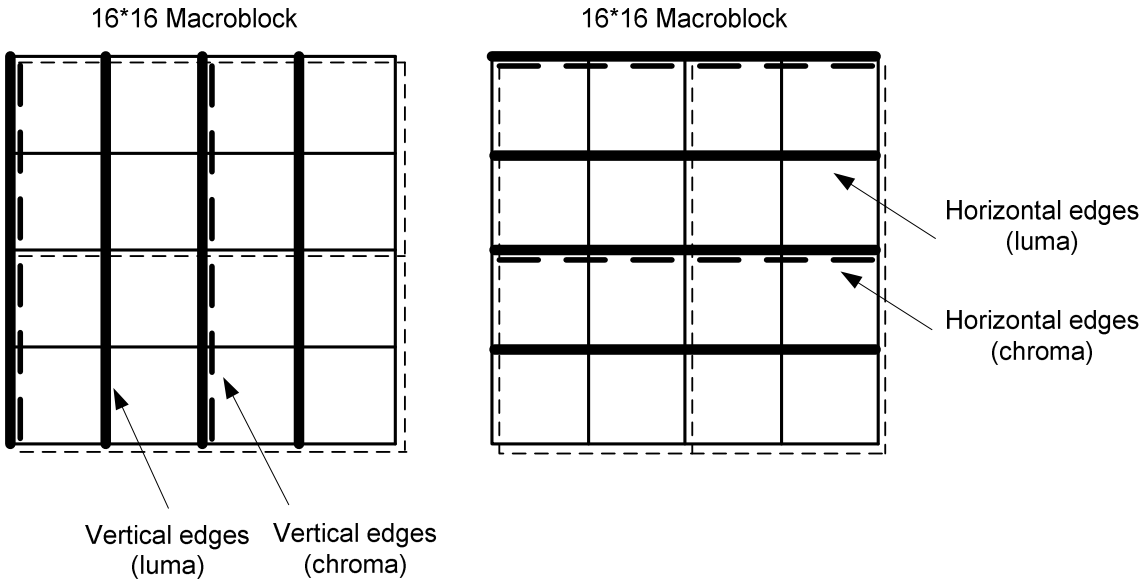
Where $\beta(QP)$ is considerably smaller than $\alpha(QP)$

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

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

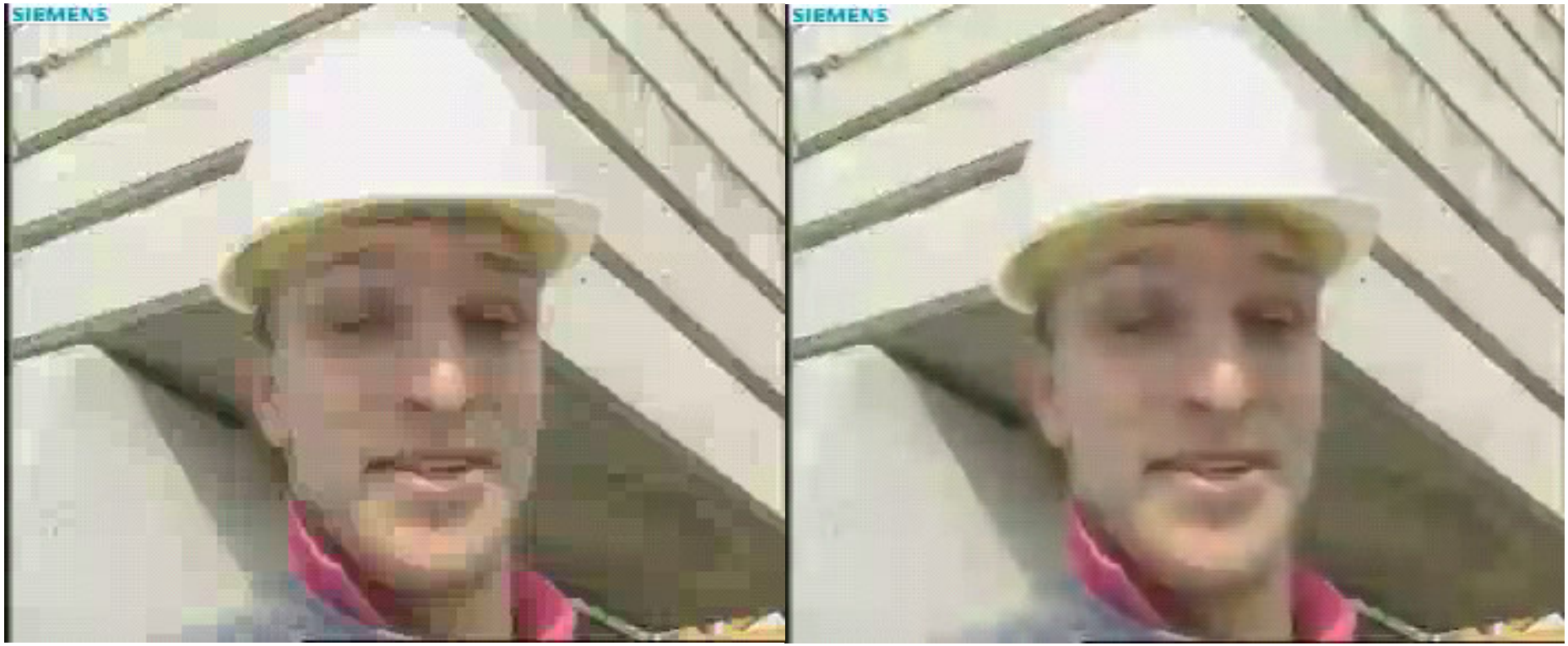
(QP = quantization parameter)

Order of Filtering



- **Filtering can be done on a macroblock basis that is, immediately after a macroblock is decoded.**
- **First, the vertical edges are filtered then the horizontal edges.**
- **The bottom row and right column of a macroblock are filtered when decoding the corresponding adjacent macroblocks.**

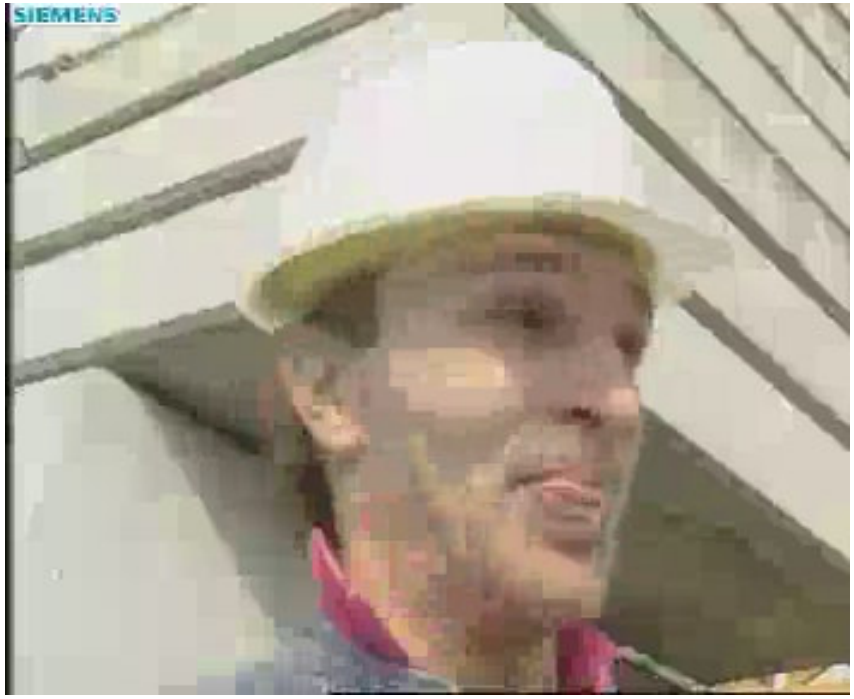
Deblocking: Subjective Result for Intra Coding at 0.28 bit/sample



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

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)**

Non-Intra H.264/AVC Profiles …

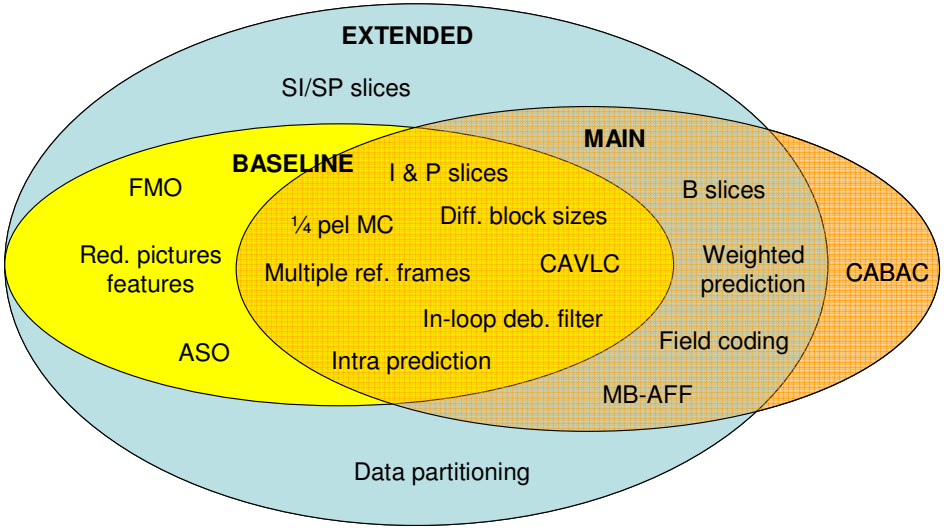
- **Baseline Profile (BP):** Primarily for lower-cost applications with limited computing resources, this profile is used widely in videoconferencing and mobile applications.
- **Main Profile (MP):** Originally intended as the mainstream consumer profile for broadcast and storage applications, the importance of this profile faded when the High profile was developed for those applications.
- **Extended Profile (XP):** Intended as the streaming video profile, this profile has relatively high compression capability and some extra tricks for robustness to data losses and server stream switching.
- **High Profile (HiP):** The primary profile for broadcast and disc storage applications, particularly for high-definition television applications (this is the profile adopted into HD DVD and Blu-ray Disc, for example).
- **High 10 Profile (Hi10P):** Going beyond today's mainstream consumer product capabilities, this profile builds on top of the High Profile — adding support for up to 10 bits per sample of decoded picture precision.
- **High 4:2:2 Profile (Hi422P):** Primarily targeting professional applications that use interlaced video, this profile builds on top of the High 10 Profile — adding support for the 4:2:2 chroma sampling format while using up to 10 bits per sample of decoded picture precision.
- **High 4:4:4 Predictive Profile (Hi444PP):** This profile builds on top of the High 4:2:2 Profile — supporting up to 4:4:4 chroma sampling, up to 14 bits per sample, and additionally supporting efficient lossless region coding and the coding of each picture as three separate color planes.

H.264/AVC Intra Profiles

In addition, the standard defines four additional all-Intra profiles, which are defined as simple subsets of other corresponding profiles. These are mostly for professional (e.g., camera and editing system) applications:

- **High 10 Intra Profile: The High 10 Profile constrained to all-Intra use.**
- **High 4:2:2 Intra Profile: The High 4:2:2 Profile constrained to all-Intra use.**
- **High 4:4:4 Intra Profile: The High 4:4:4 Profile constrained to all-Intra use.**
- **CAVLC 4:4:4 Intra Profile: The High 4:4:4 Profile constrained to all-Intra use and to CAVLC entropy coding (i.e., not supporting CABAC).**

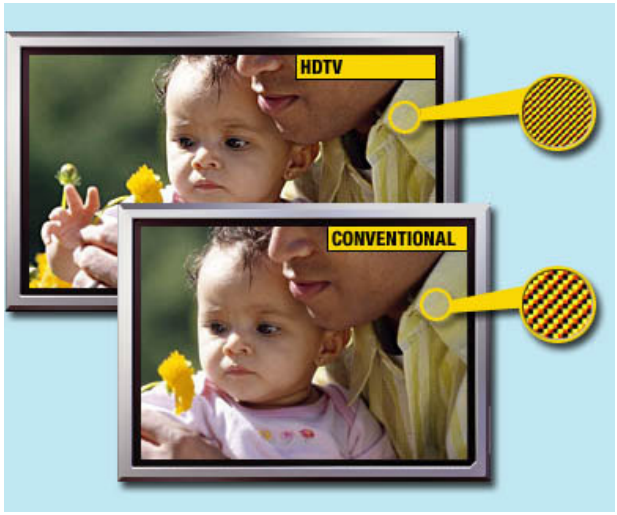
First H.264/MPEG-4 AVC Profiles …



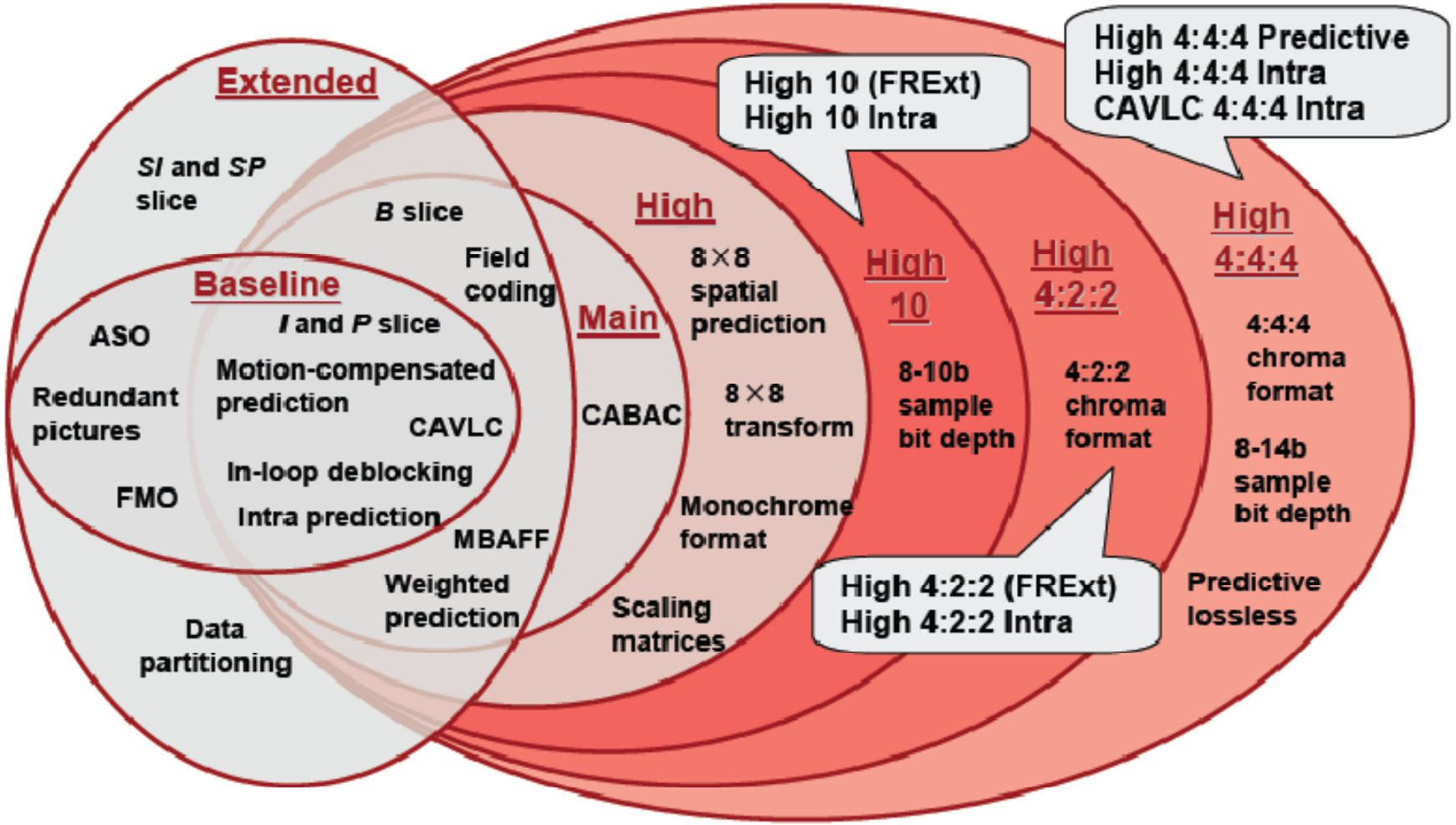
- **Baseline Profile** is targeted towards real-time encoding and decoding for CE devices. Supports progressive video, uses I and P slices, CAVLC entropy coding.
- **Main Profile** is targeted mainly towards the broadcast market. Supports both interlaced and progressive video with macroblock or picture level field/frame mode selection. Uses I, P, B slices, weighted prediction, both CAVLC and CABAC for entropy coding.
- **Extended Profile** is targeted towards error prone channels (such as mobile communication). Uses I, P, B, SP, SI slices, supports both interlaced and progressive video, allows CAVLC coding only.

The Fidelity Range Extensions (FREXT) Profiles

- **High Profile** extends functionality of main profile for effective coding of high definition content. Uses adaptive 8×8 or 4×4 transform, enables perceptual quantization matrices.
- **High 10 Profile** is an extension of High profile for 10 bit component resolution.
- **High 4:2:2 Profile** supports 4:2:2 chroma format and up to 10 bit component resolution. Suitable for video production and editing.
- **High 4:4:4 Profile** supports 4:4:4 chroma format and up to 12 bit component resolution. In addition, it enables lossless mode of operation and direct coding of RGB signal. Targeted for professional production and graphics.



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.**

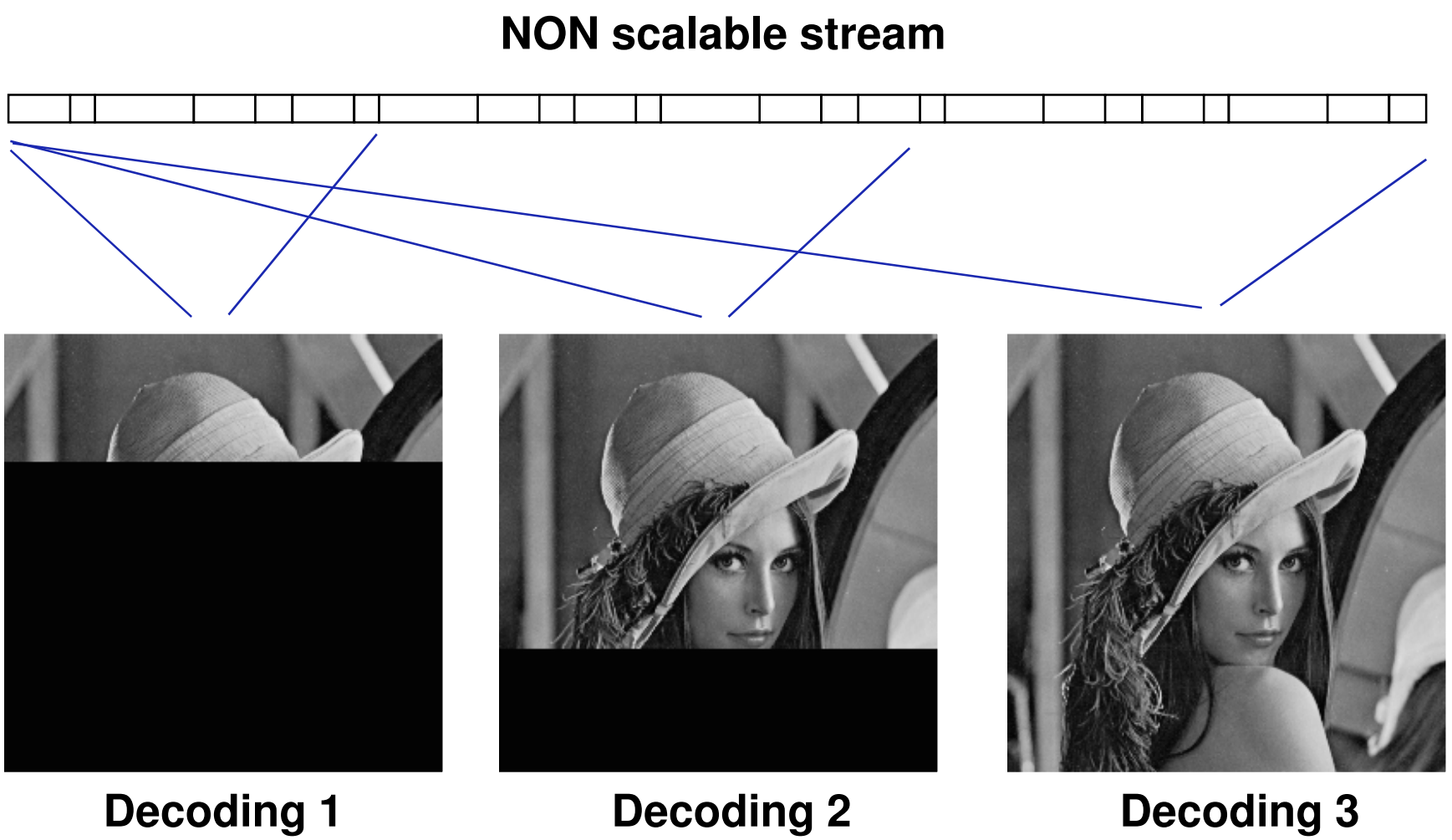
Participation Fees (2)

- **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.**

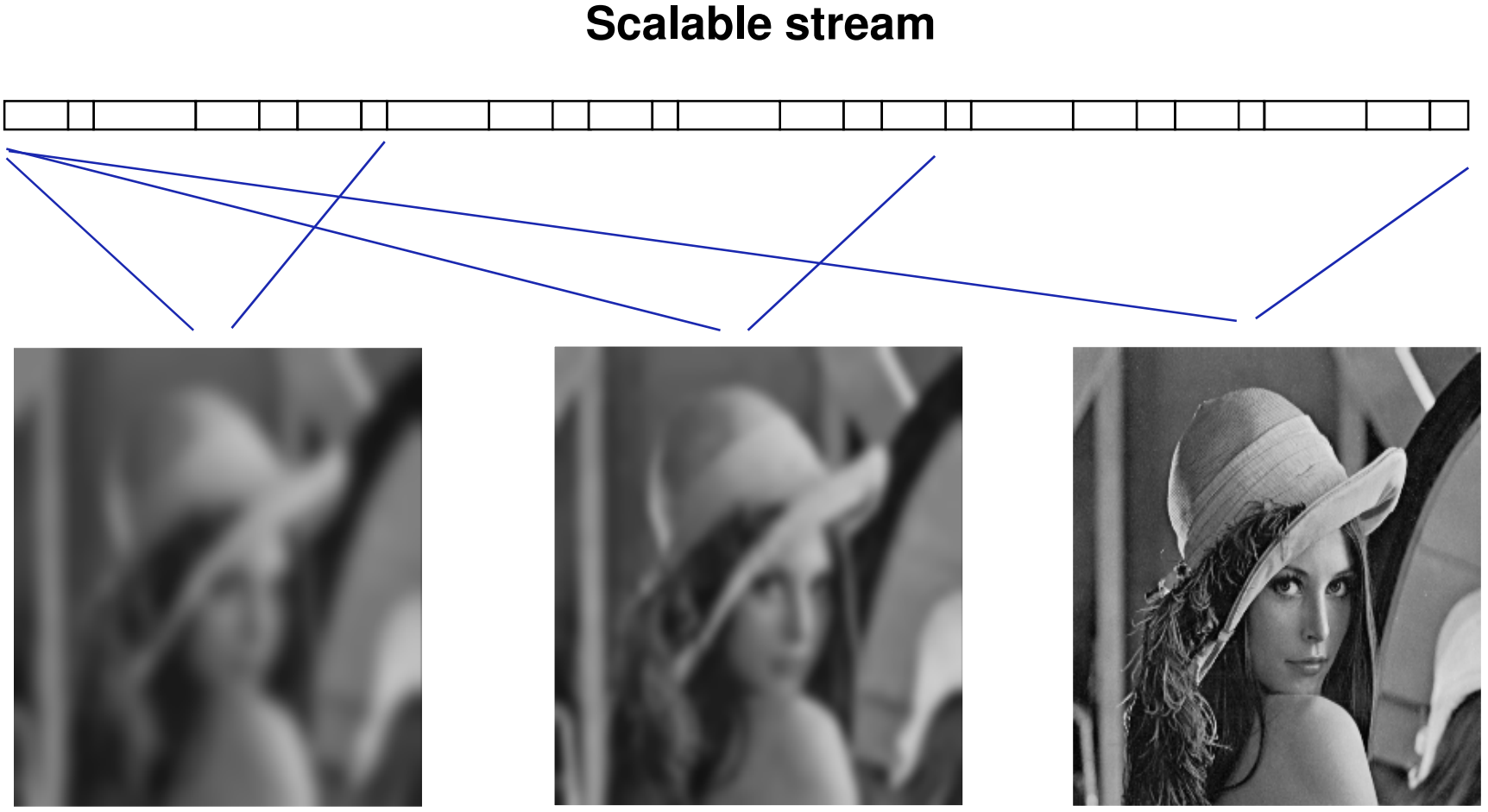
Scalable Video Coding (SVC)

An H.264/AVC Extension

Non-Scalable Coding ...



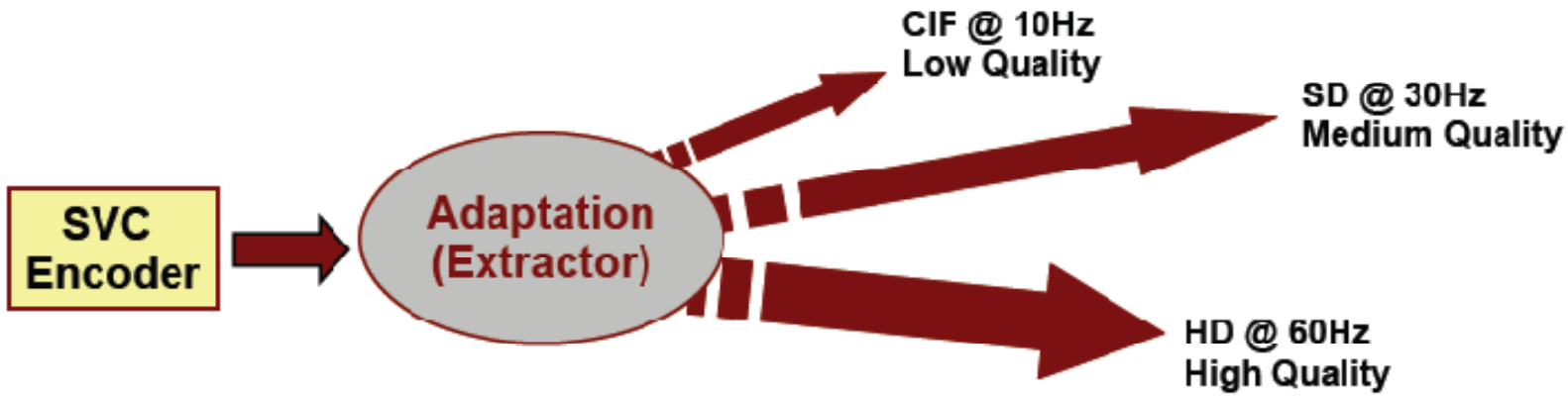
Quality or SNR Scalable Coding …



Scalable Video Coding: Objectives

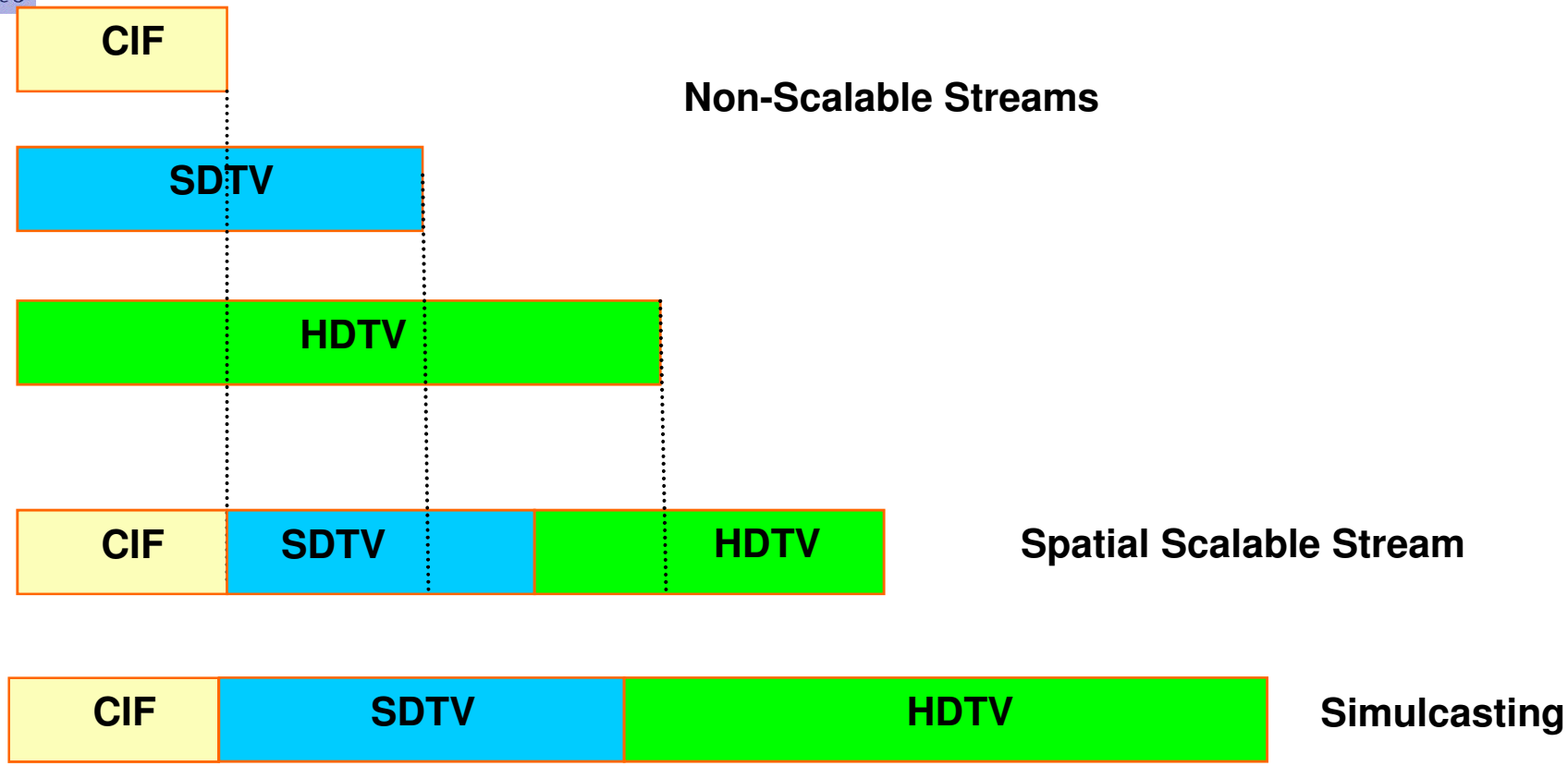
Scalability is a functionality regarding the decoding of parts of the coded bitstream, ideally

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.**



Encode once, decoding many

The Price of Scalability …



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.

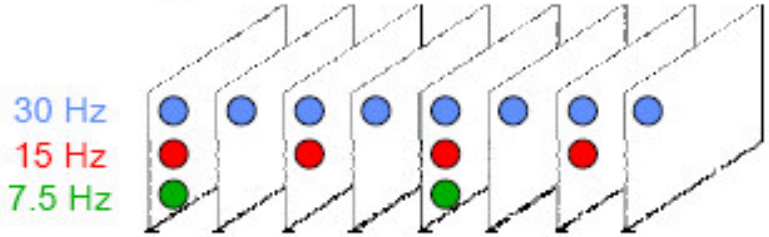
- **SVC should provide functionalities such as graceful degradation in lossy transmission environments as well as bitrate, format, and power adaptation; this should provide enhancements to transmission and storage applications.**
- **Previous video coding standards, e.g. MPEG-2 Video and MPEG-4 Visual, 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)

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

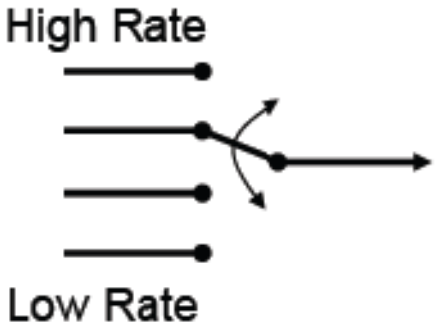
- **Scalable Storage**
  - Scalable export of video content
  - Graceful expiration or deletion
  - Surveillance DVR's and Home PVR's

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

SVC Alternatives

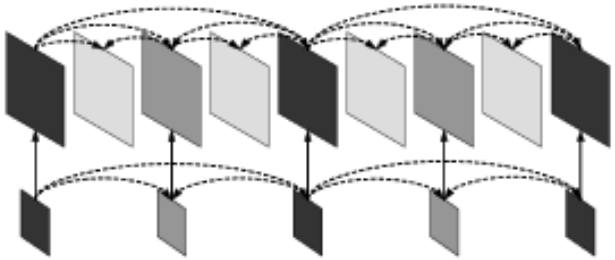
- **Simulcast**
  - Simplest solution
  - Code each layer as an independent stream
  - Incurs increase of rate

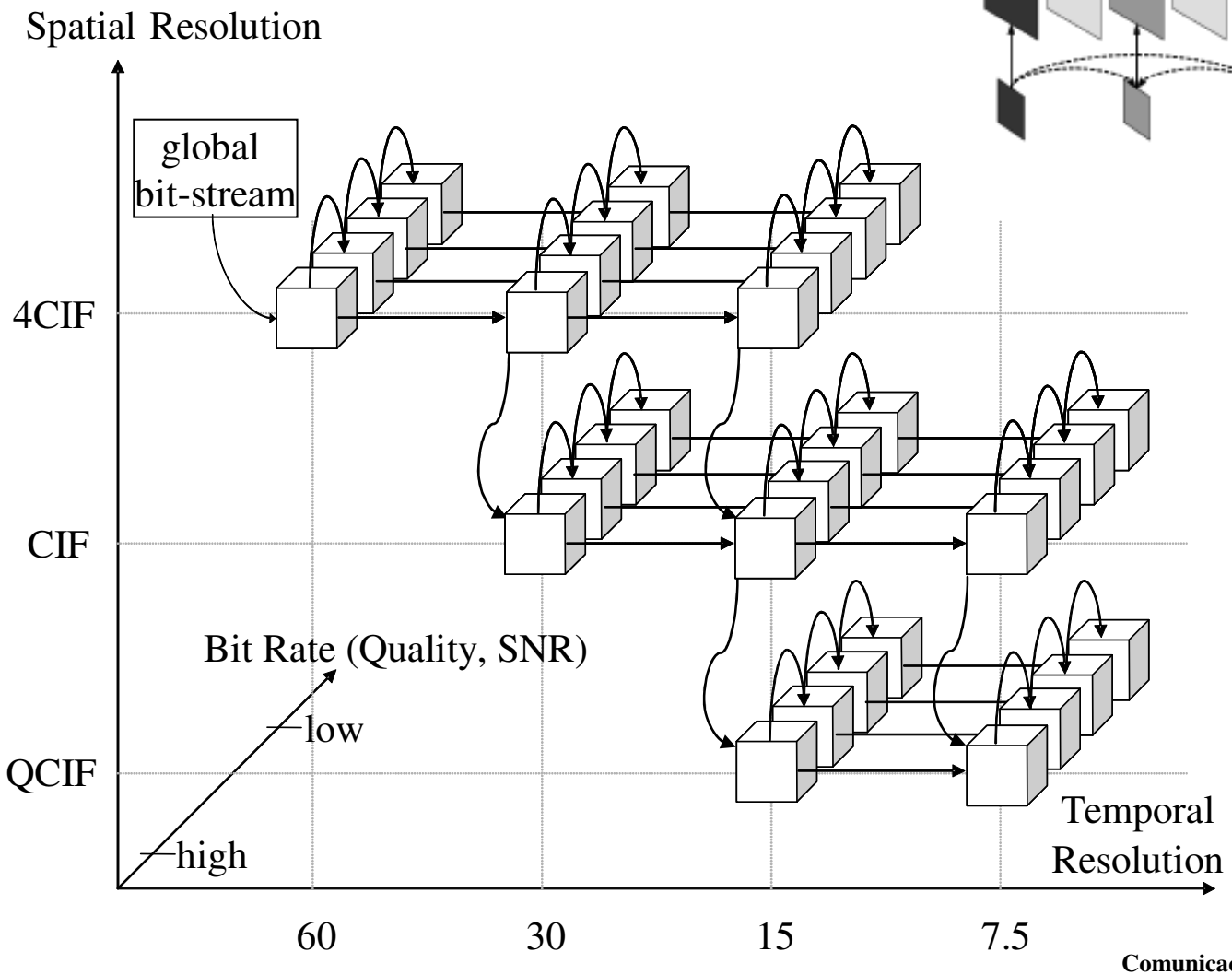
- **Stream Switching**
  - Viable for some application scenarios
  - Lacks flexibility within the network
  - Requires more storage/complexity at server



- **Transcoding**
  - Low cost, designed for specific application needs
  - Already deployed in many application domains

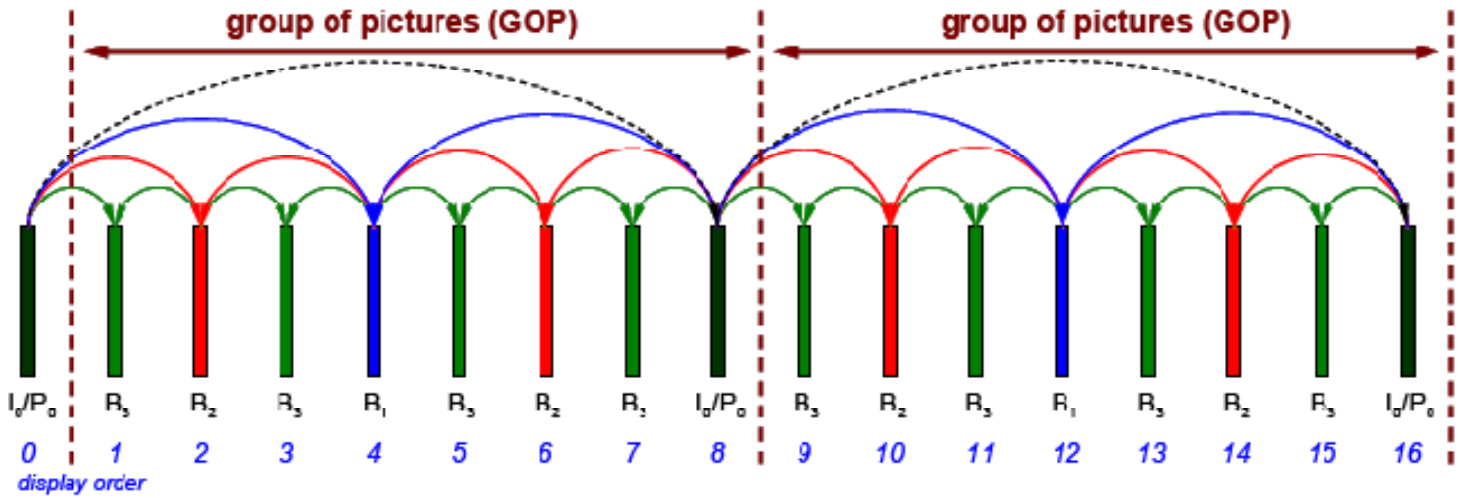
Spatio-Temporal-Quality Cube



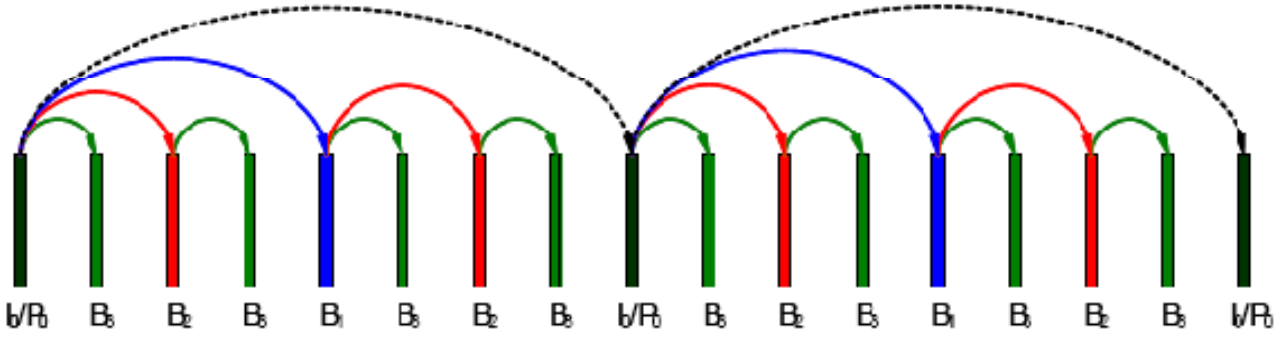


Hierarchical Prediction Structures

- Dyadic temporal scalability

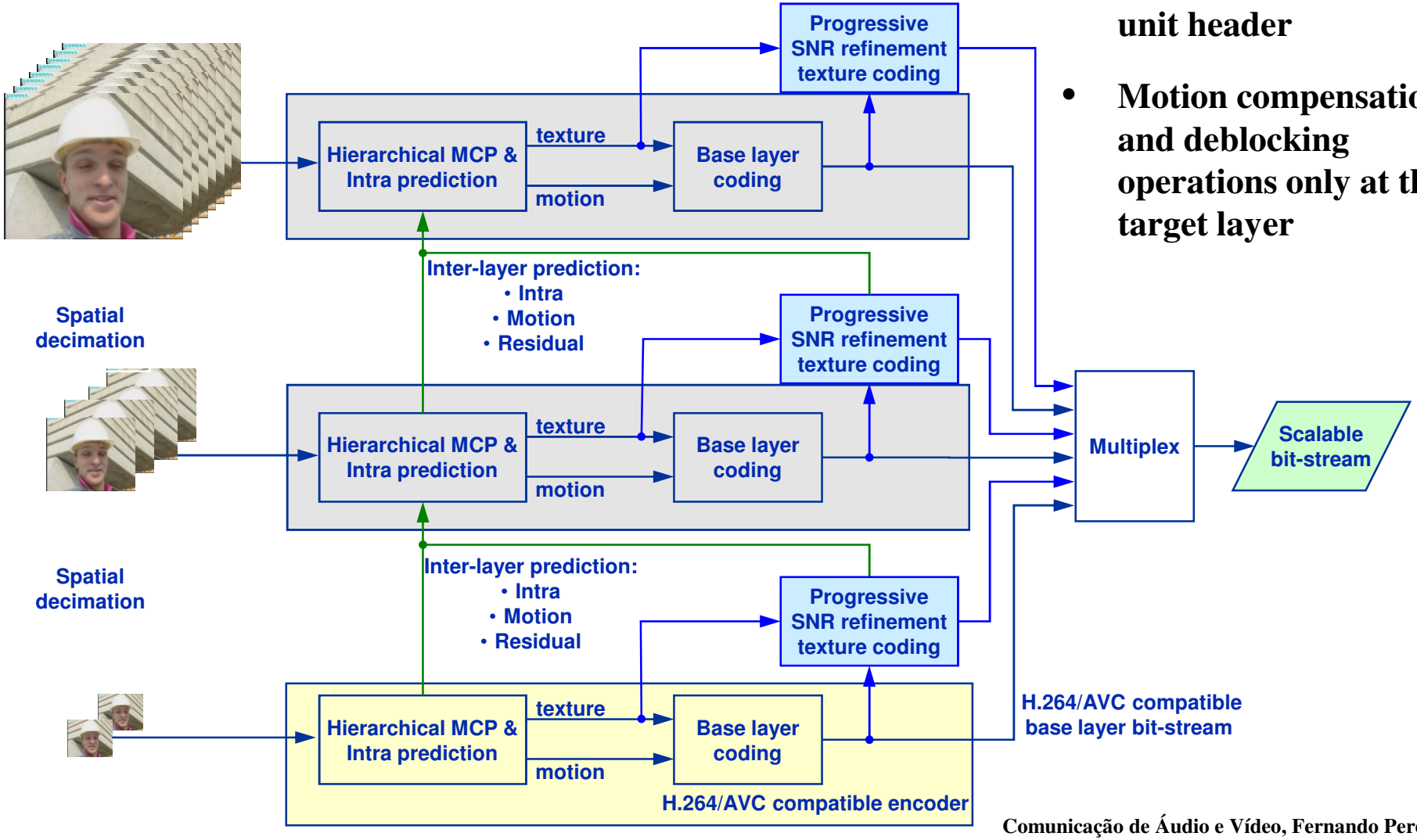


- Low-delay prediction structure (structural delay is 0)

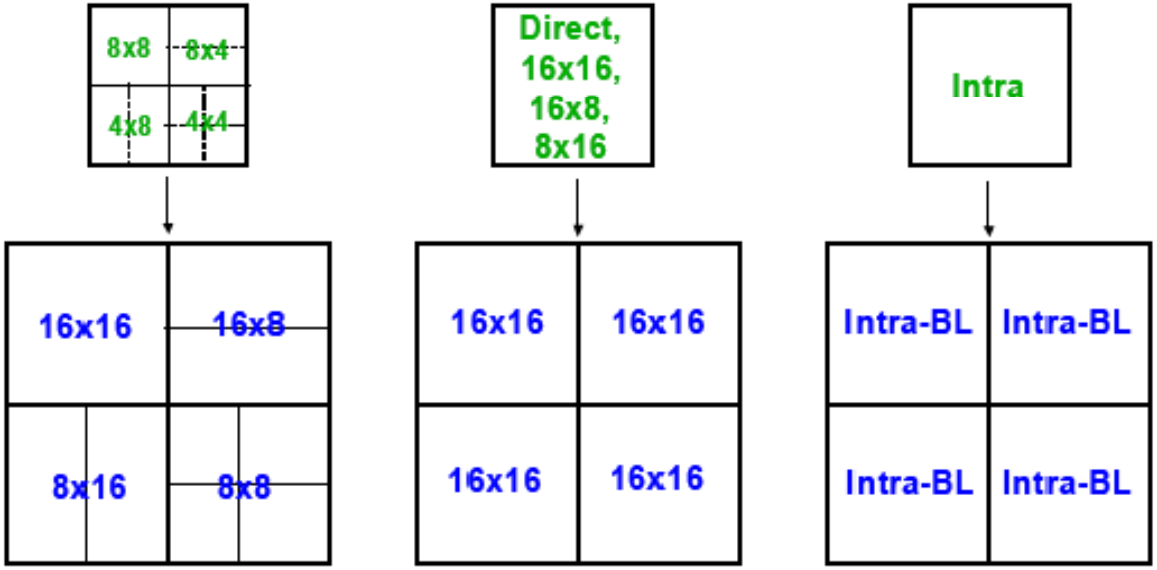


SVC Coding Architecture

- 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.

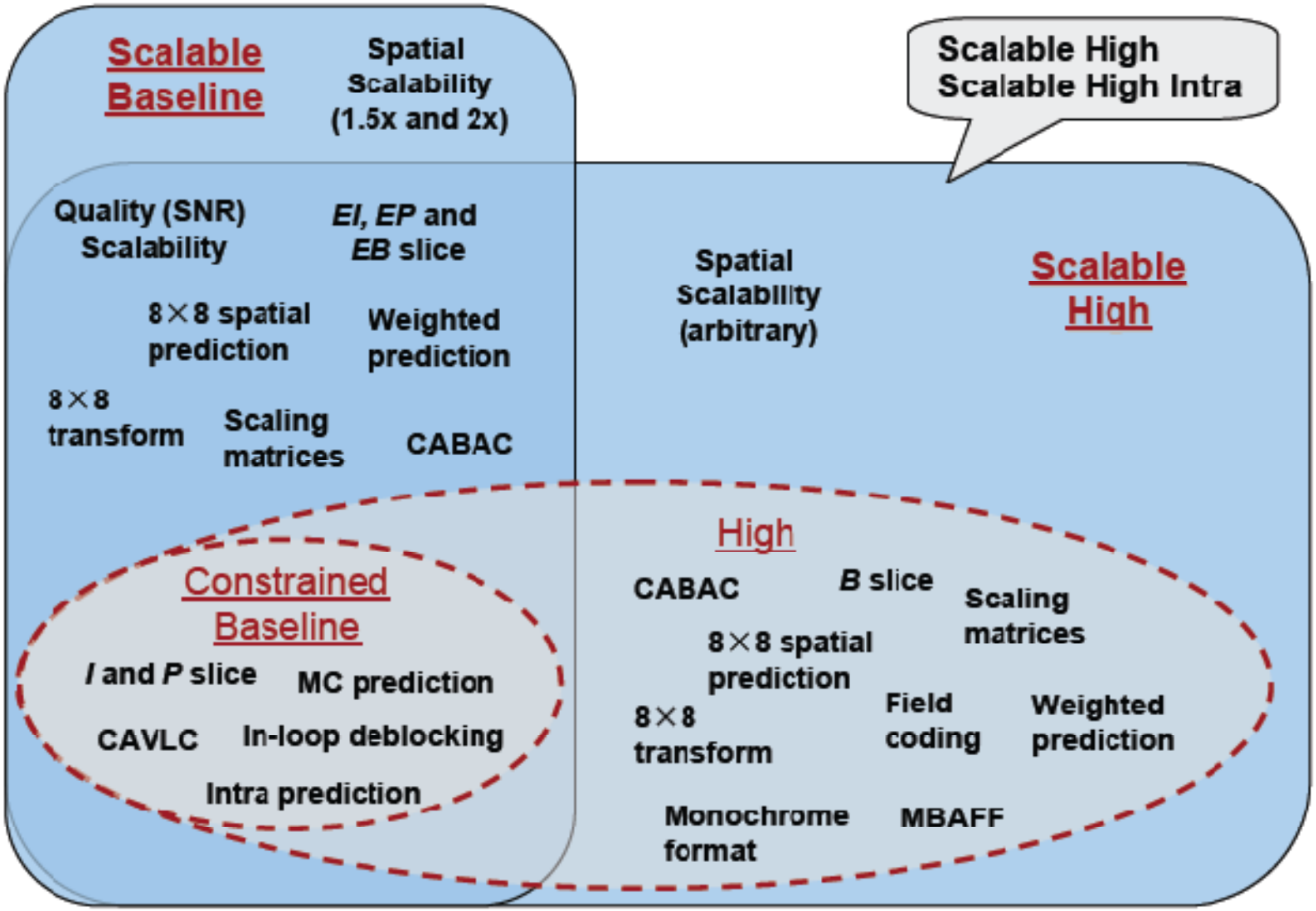
SVC Novelty Regarding Previous Scalable Standards

2007 !

- **Possibility to employ hierarchical prediction structures for providing temporal scalability with several layers while improving the coding efficiency and increasing the effectiveness of quality and spatial scalable coding.**
- **New methods for inter-layer prediction of motion and residual improving the coding efficiency of spatial scalable and quality scalable coding.**
- **Concept of key pictures for efficiently controlling the drift for packet-based quality scalable coding with hierarchical prediction structures.**
- **Single motion compensation loop decoding for spatial and quality scalable coding providing a decoder complexity close to that of single-layer coding.**
- **Support of a modified decoding process that allows a lossless and low-complexity rewriting of a quality scalable bit stream into a bit stream that conforms to a non-scalable H.264/AVC profile.**

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

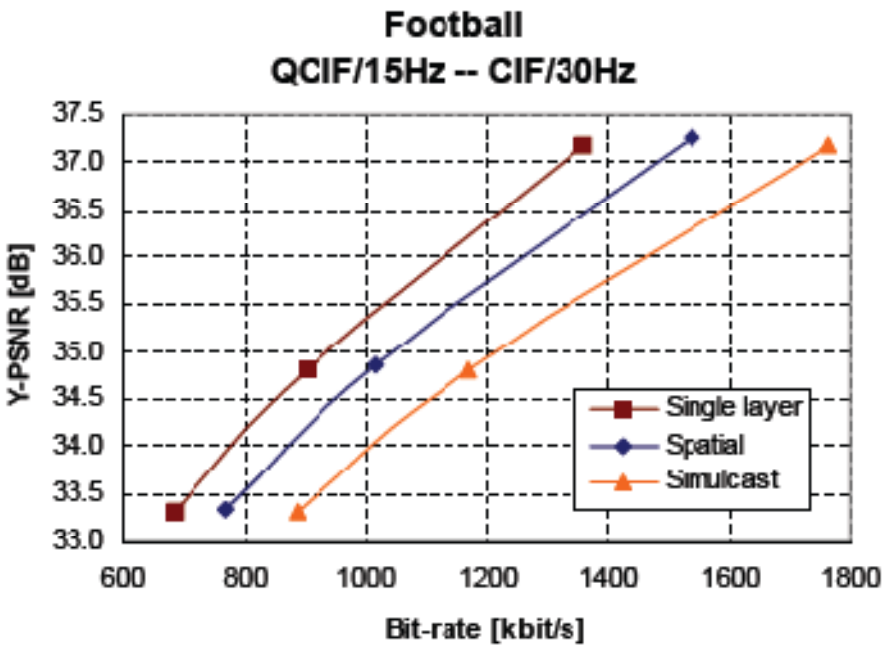
SVC Profiles

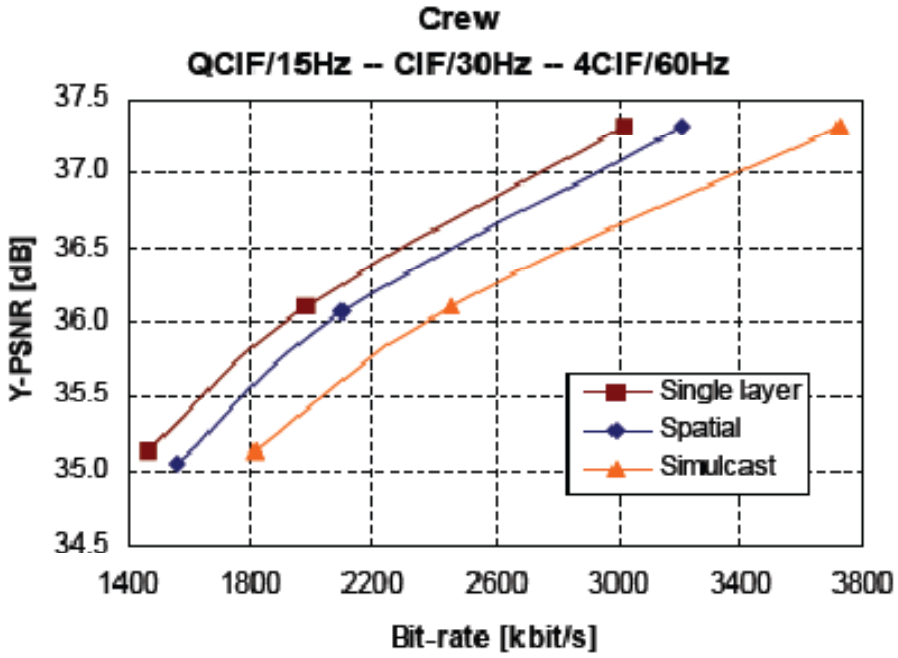


SVC Profiles

- **Scalable Baseline Profile**
  - Base layer: Restricted H.264/AVC Baseline Profile
  - Conversational and surveillance applications
  - Low decoding complexity
  - Restrictions on scaling ratios between spatial layers
- **Scalable High Profile**
  - Base layer: H.264/AVC High Profile
  - Broadcast, streaming, and storage applications
  - Arbitrary scaling rations between spatial layers
- **Scalable High Intra Profile**
  - Base layer: H.264/AVC High Profile, Intra only
  - Professional applications: video post processing / editing
  - Only IDR pictures (IDR = Instantaneous Decoding Refresh)

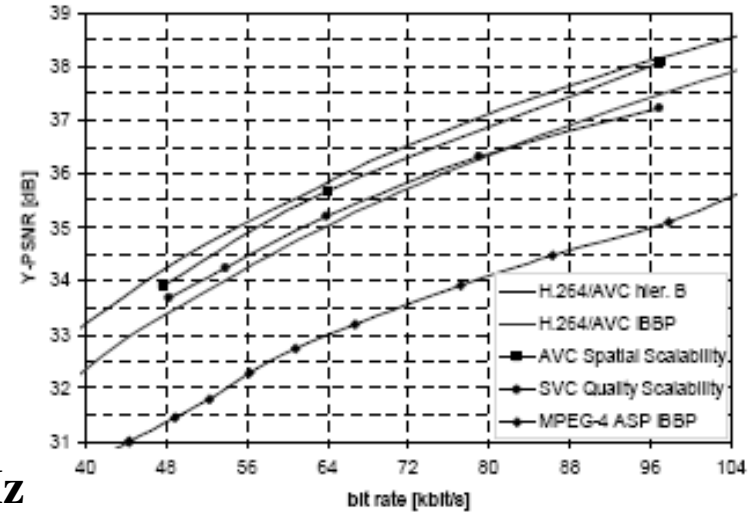
SVC Performance: Spatial Scalability



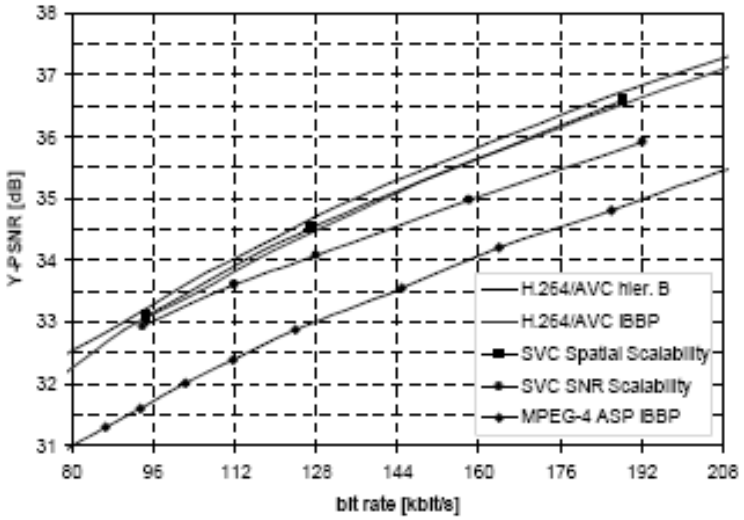


- **10~15% gains over simulcast**
- **Performs within 10% of single layer coding**

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



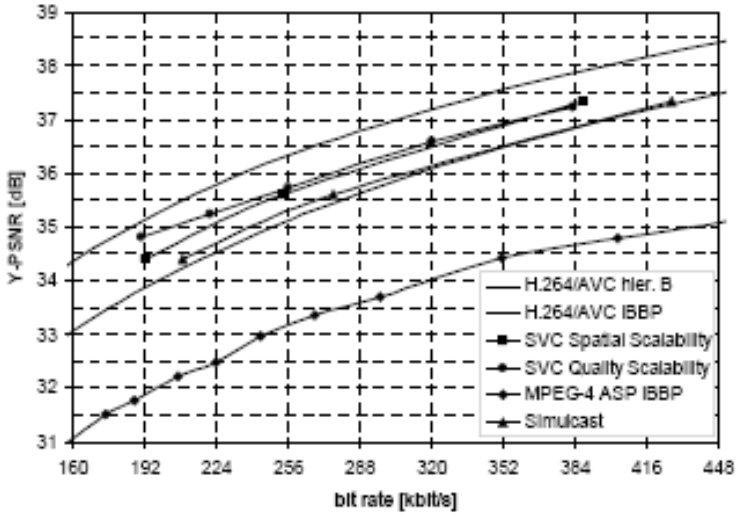
(a)



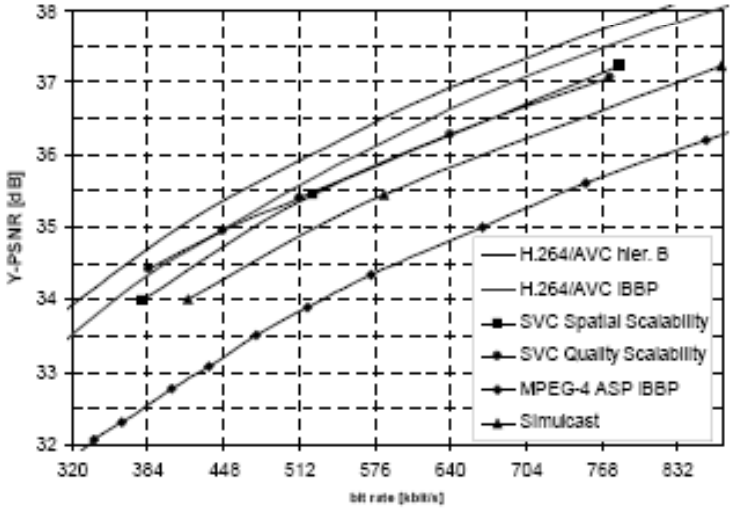
(a)

QCIF@15 Hz

CIF@30 Hz



(b)



(b)

SVC Performance: Foreman and Crew

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

SVC: Conclusions

- **Scalable extension of H.264/AVC**
  - Layered approach
  - Additional tools for fidelity and spatial scalability
  - Single-loop decoding

- **Performance evaluation results**
  - SVC enables scalability at not more than 10% bitrate overhead without compromising the visual quality
  - About 40% bitrate savings in HD broadcast compared to simulcasting single layer streams
  - About 20% bitrate savings in mobile broadcast and conversational compared to simulcasting single layer streams
  - About 20-30% bitrate savings in three-layer intra scenario compared to simulcasting single layer streams

- **SVC provides a coding efficiency that is similar to single-layer coding for simple scalability configurations at only a slightly increased decoder complexity**

SVC: What Future ?

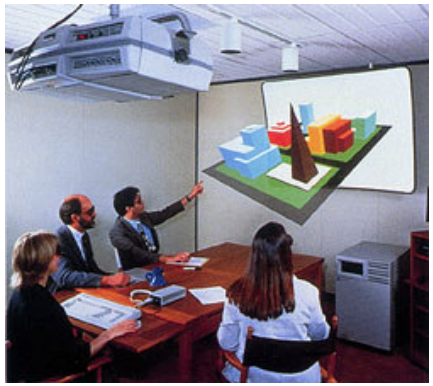
- **Technically, the standard is a great success already with some adoption**
  - *Google* Gmail service
  - *Vidyo* video conferencing for the Internet
  - Industry appears to be open towards embracing SVC for DTV broadcast services
  - Specifically, enhancement of 720p to 1080p

- **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**

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

Coming 3D Displays ….

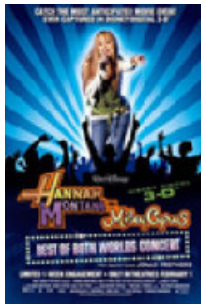
- **Where several years ago the latest television sets were hailed as "HDTV ready", a handful of TV makers are now touting models that are "3D ready".**
- **Mitsubishi's latest 3D ready HDTV set, a 65-inch model called LaserVue, goes on sale this summer, with a 73-inch version following later in the year. Meanwhile, Panasonic exhibited a 103-inch plasma display showing 3D video from a Blu-ray player.**
- **And the Philips Quad Full Autosteroscopic 3D HD TV increases a display's screen resolution to a 8.29 million pixels, four times the number of pixels of the highest HDTV standard.**
- **Autostereoscopic displays that don't call for special glasses tend to have a fairly small "sweet spot", and therefore a limited viewing angle.**

partly from

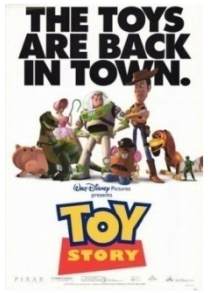
The Economist, January 2009

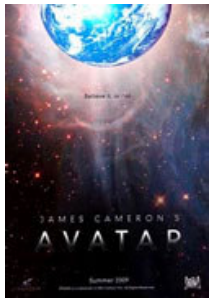
Coming 3D Content …

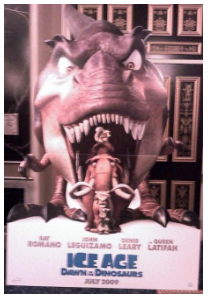
- ✶ **Nine 3D title releases to date since 2005**
  - **Recent: Beowulf, Hannah Montana, U23D**



- ✶ **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**

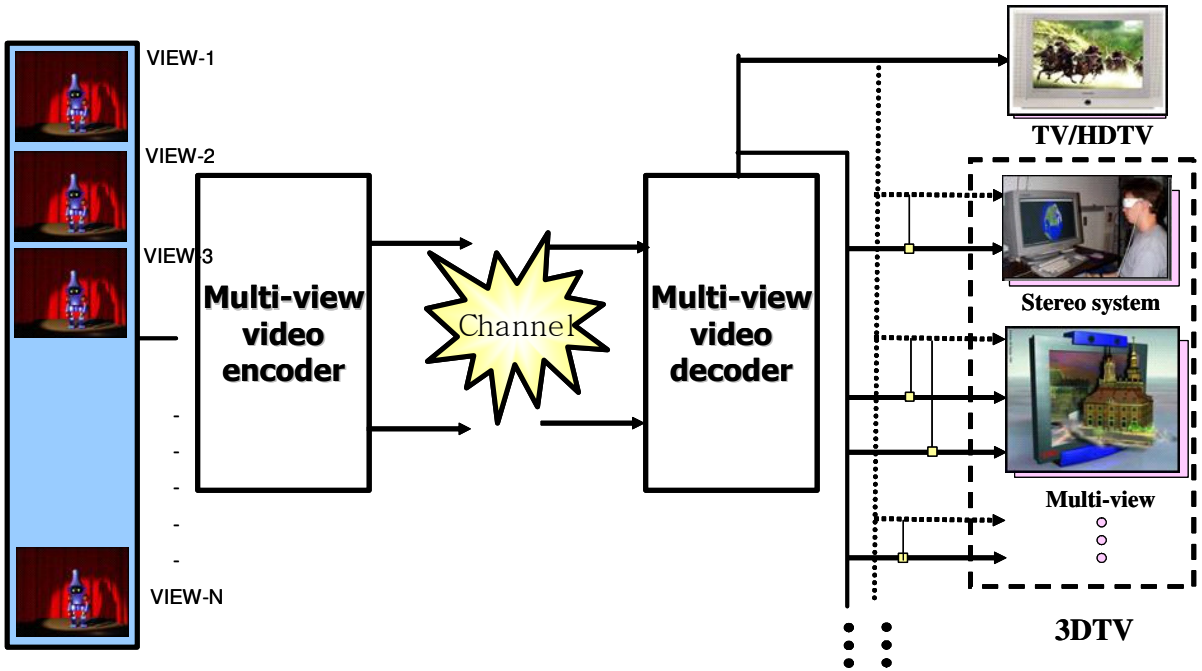
Productions Committed to be Released in 3D (Feb. 2009)

- Monsters vs. Aliens (March 27, 2009)
- Up (May 29, 2009)
- Ice Age: Dawn of the Dinosaurs (July 1, 2009)
- G-Force (July 24, 2009)
- Piranha 3-D (July 24, 2009)
- Final Destination: Death Trip 3D (August 21, 2009)
- Toy Story (October 2, 2009)
- Astro Boy (October 23, 2009)
- Horrorween (October 30, 2009)
- A Christmas Carol (November 6, 2009)
- Crood Awakening (November 2009)
- Planet 51 (November 20, 2009)
- Avatar (December 18, 2009)
- The Princess and the Frog (December 25, 2009)
- Cloudy with a Chance of Meatballs (January 15, 2010)
- Toy Story 2 (February 12, 2010)
- Alice in Wonderland (March 5, 2010)
- How to Train Your Dragon (March 26, 2010)
- Alpha and Omega (April 16, 2010)
- Shrek Goes Fourth (May 21, 2010)
- Toy Story 3 (June 18, 2010)
- Ghostbusters III (2010)
- Beauty and the Beast (2010)
- Despicable Me (July 16, 2010)
- Guardians of Ga'Hoole (July 23, 2010)
- Master Mind (November 5, 2010)
- Rapunzel (December 25, 2010)
- Kung Fu Panda 2 (June 3, 2011)
- Puss in Boots: The Story of an Ogre Killer (2011)
- The Bear and the Bow (2011)
- Newt (film) (2012)
- Cars 2 (June 24, 2011)
- Madagascar 3 (2012)
- Shrek 5 (2013)

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



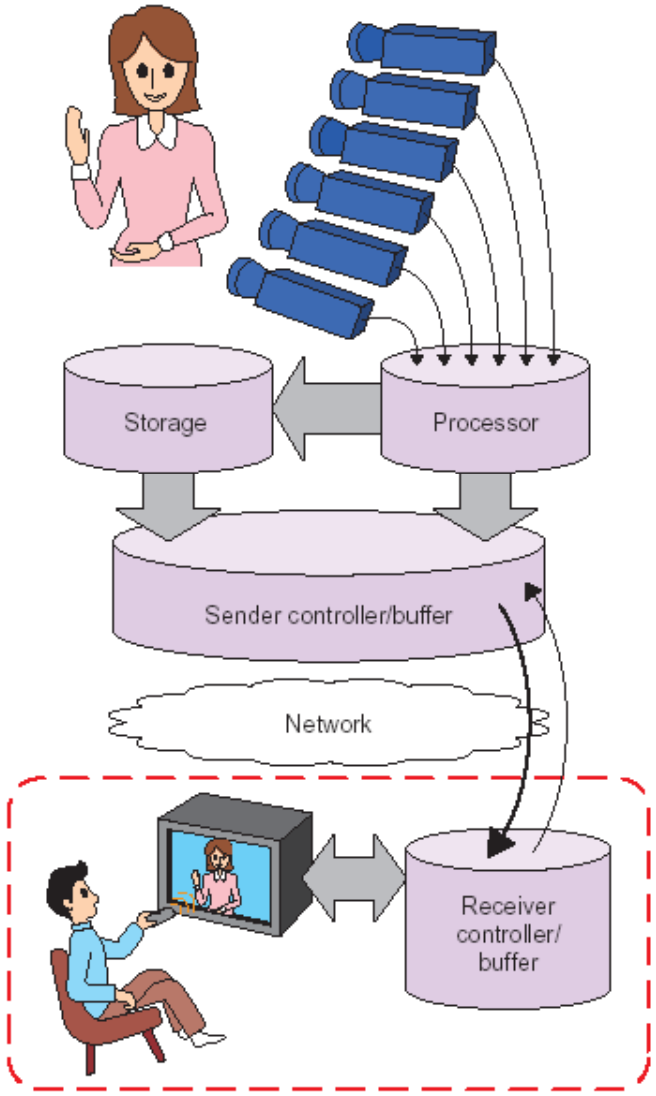
Multi-view video (MVV) refers to a set of N temporally synchronized video streams coming from cameras that capture the same real world scenery from different viewpoints.

- **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**

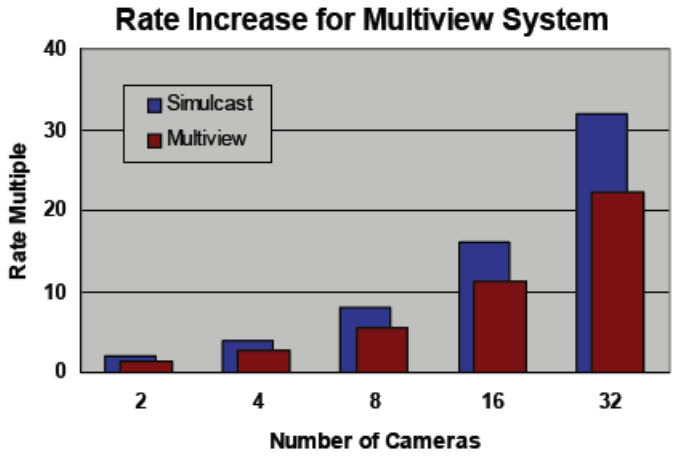
Multi-View Video Data

- 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)

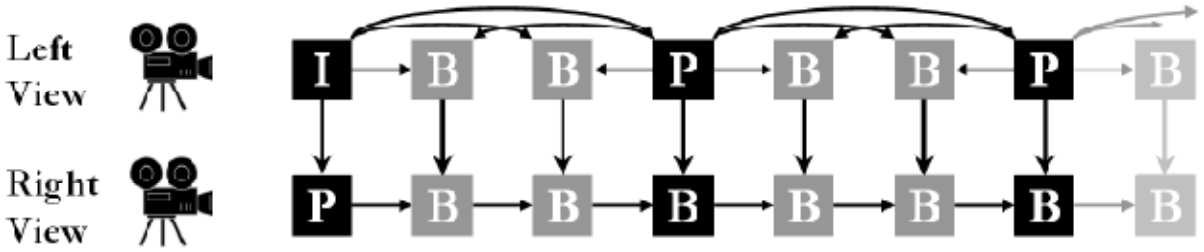


- **Direct coding of multiple views (stereo to multi-view)**
- **Exploits redundancy between views using inter-camera prediction to reduce required bit-rate**
- **Without any changes at H.264/AVC slice layer and below, bitrate reductions around 20-50% can be achieved by allowing interview predictions.**



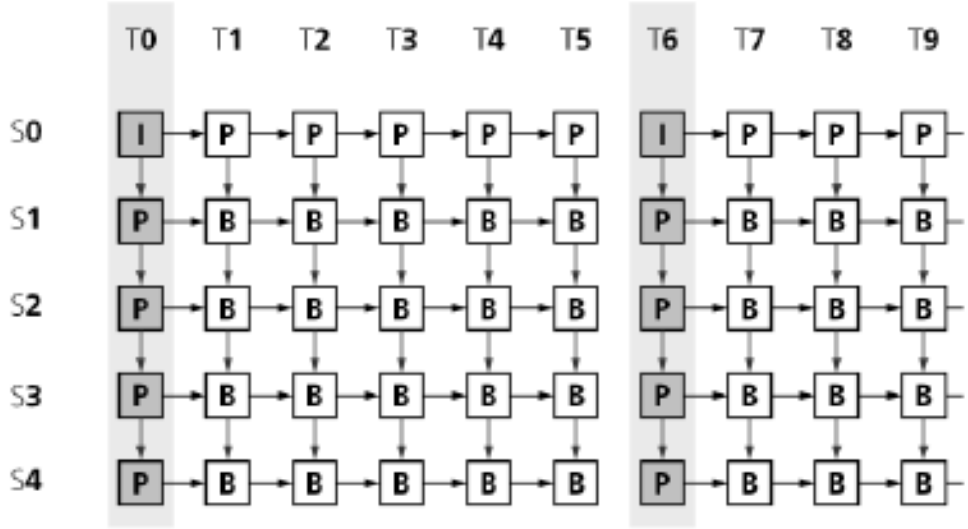
MVC: Prediction Structures

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



MPEG-2 Video Multi-view profile

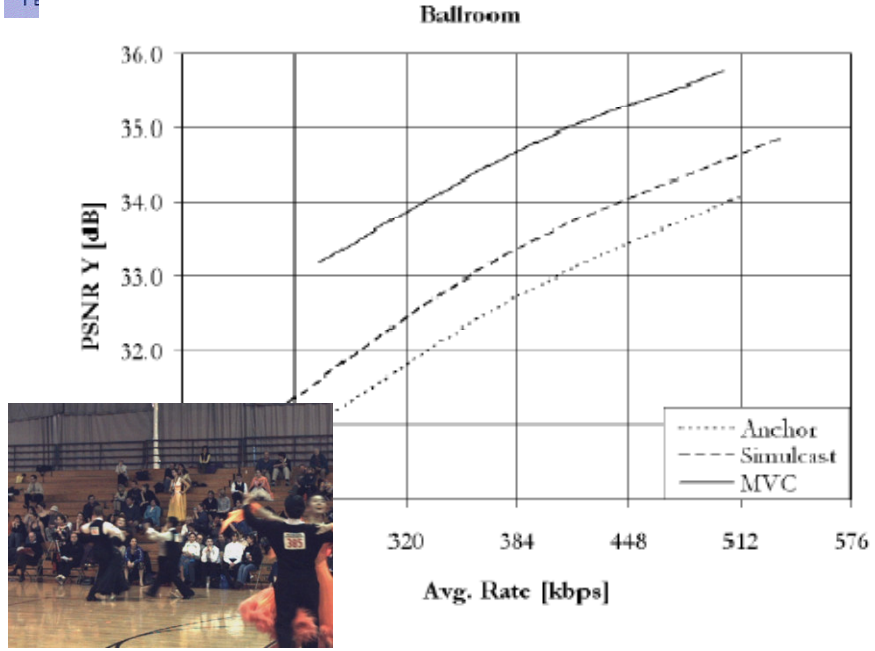
(JVT) MVC

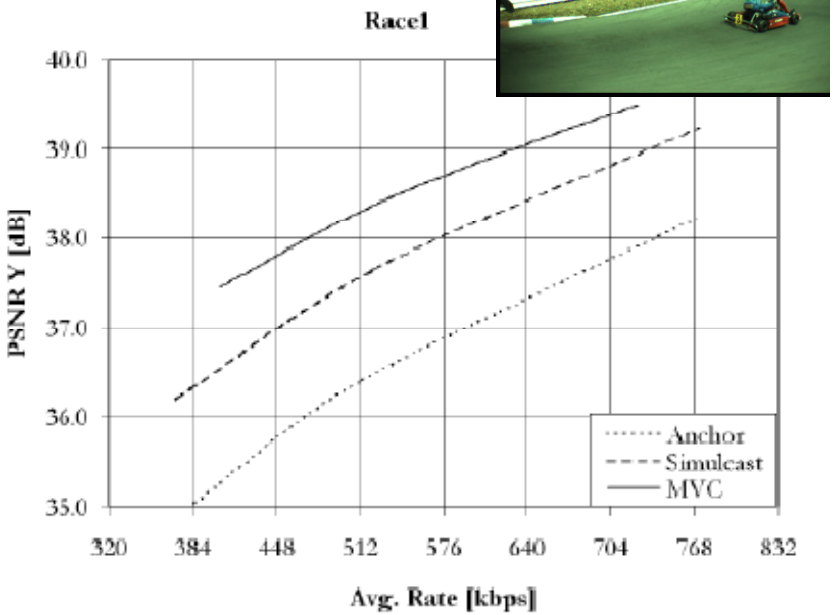


MVC: Technical Solution

- **Key elements of MVC design**
  - **Does not require any changes to lower-level syntax, so 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



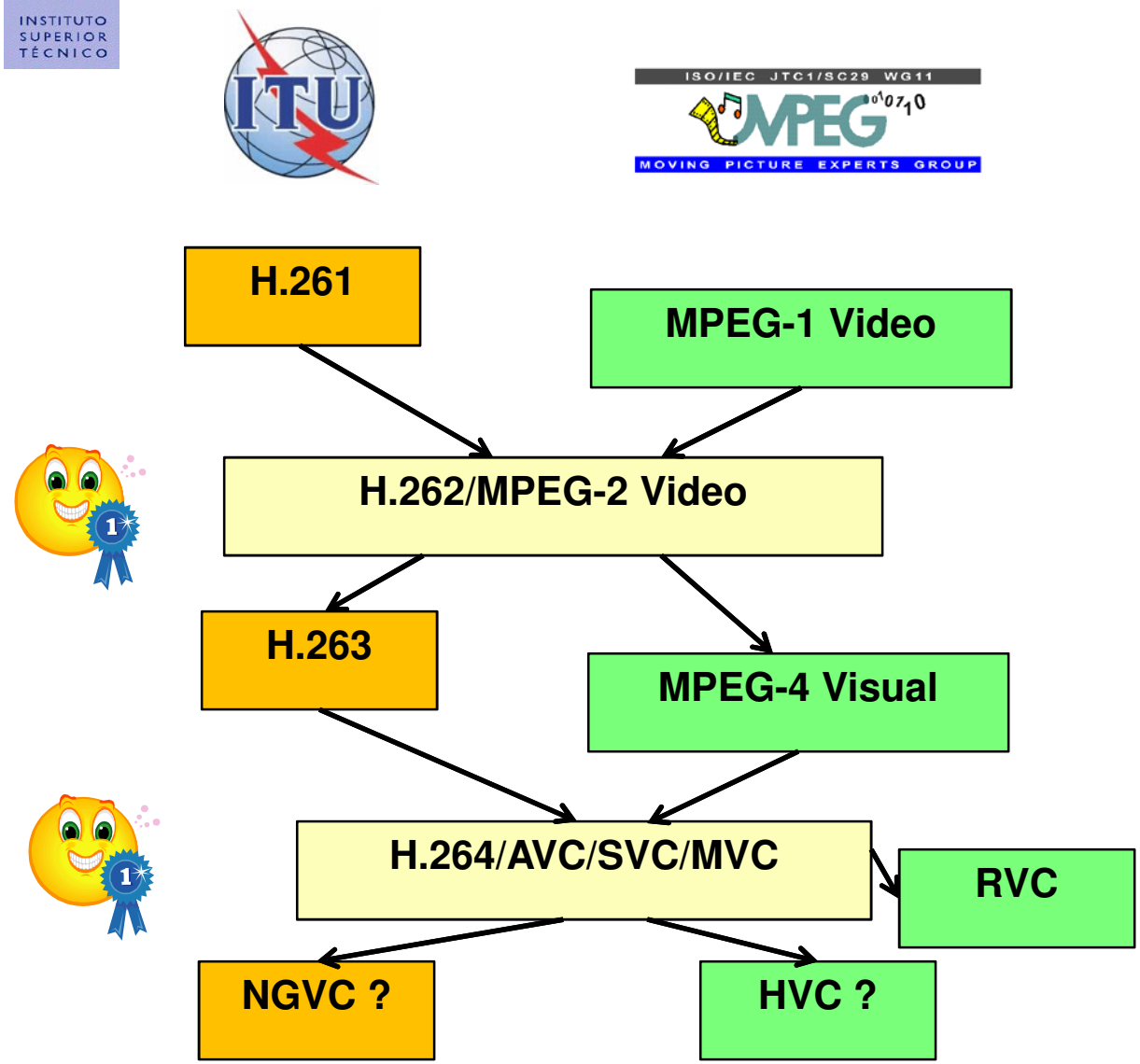


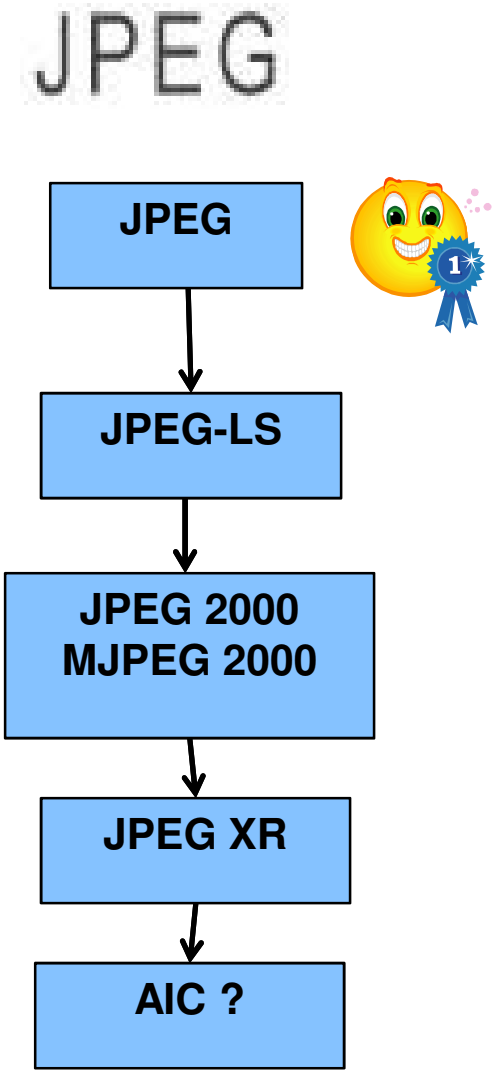
- Anchor is H.264/AVC without hierarchical B pictures
- Simulcast already includes hierarchical B pictures
- Majority of gains due to inter-view prediction at I-picture locations
- Although more efficient than simulcast, rate of MVC is still proportional to the number of views (varies with scene, camera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

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 checked ...

The Standardization Path …





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.

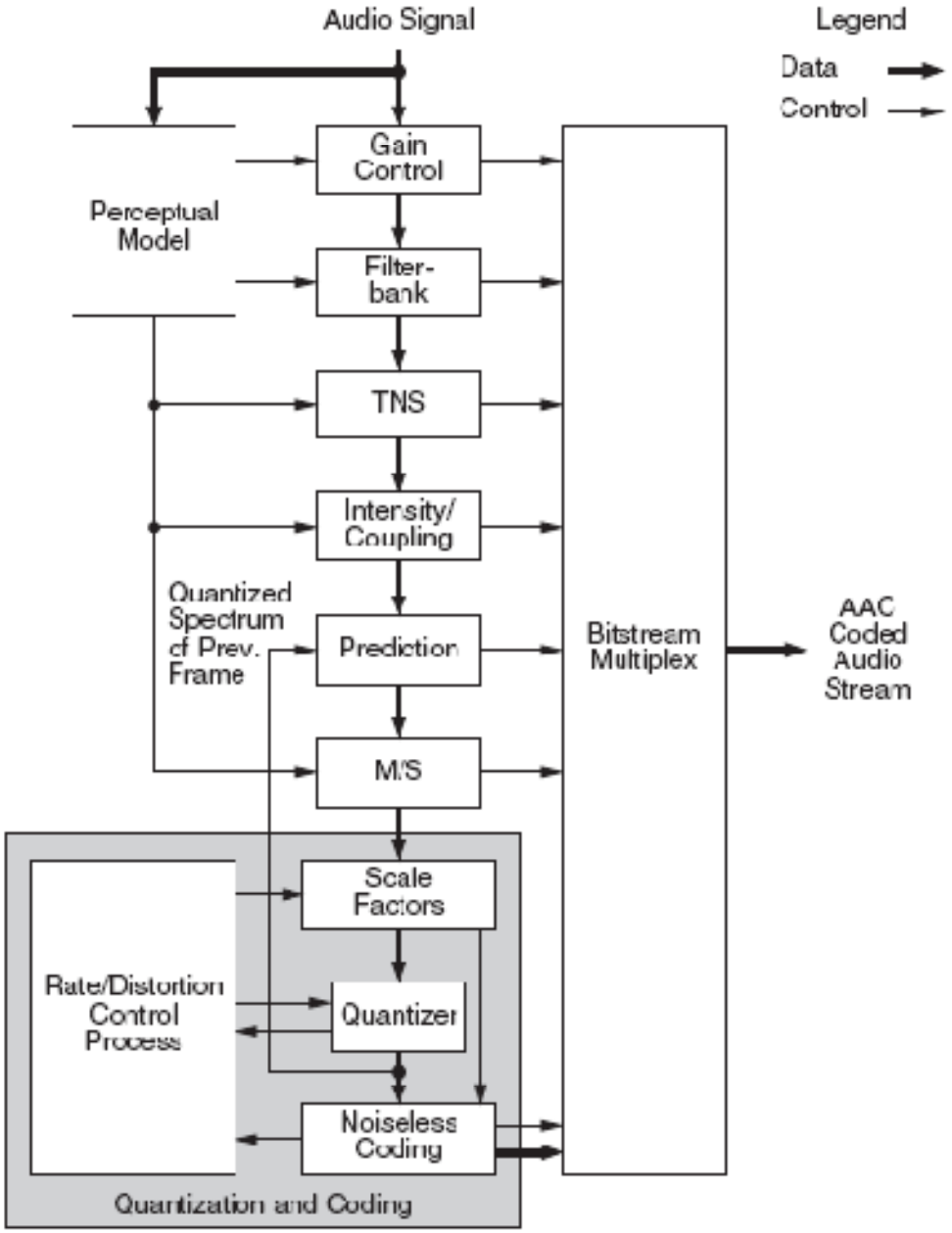
Advanced Audio Coding (AAC) - initially called Non- Backward Compatible (NBC) - is defined in two MPEG standards:

- **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).**

MPEG-2 AAC versus MP3

- **In terms of overall approach and structure, some commonalities between the MPEG-2 AAC coder and the MPEG-1/2 Layer 3 coder can be observed in that**
  - **both schemes employ a switched filterbank providing a high-frequency resolution,**
  - **a nonuniform power-law quantizer, and**
  - **a Huffman code–based entropy coding.**

- **Beyond these commonalities, the MPEG-2 AAC codec includes a considerable number of novel coding tools to increase the codec flexibility and performance.**



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 Tools: Gain Control

The pre/postprocessing stage is designed to reduce the temporal spread of the quantization noise for transient input signals (pre-echo).

- The gain control (preprocessing) module is used exclusively by the MPEG-2 AAC SSR profile as an additional block of the input stage of the encoder.
- The module includes a polyphase quadrature filterbank (PQF), gain detectors, and gain modifiers.
- Each audio channel input is split into four frequency bands of equal bandwidth (for a sampling rate of 48 kHz, this corresponds to the bands of 0–6 kHz, 6–12 kHz, 12–18 kHz, and 18–24 kHz). The signals in these bands are examined for rapid changes in signal energy by the gain detectors.
- Based on the result of this analysis, adjustments of the signal amplitude over time are conducted by the gain modifiers in order to compress the dynamics of the signal.
- Each preprocessed signal is subsequently passed on to an MDCT filterbank to produce 256 spectral coefficients, resulting in a total of 1024 spectral coefficients for each input frame of 1024 samples.
- The postprocessing (inverse gain control) in the AAC SSR decoder uses the same components as the encoder preprocessing but arranged in reverse order.

MPEG-2 AAC Tools: Filterbank

The MPEG-2 AAC encoder employs a high-frequency resolution filterbank to map the time domain input samples to a subsampled spectral representation.

- An MDCT is used which is a perfect reconstruction filterbank.
- There is an overlap of 50% of the window size between subsequent analysis windows.
- In standard operation mode, the AAC encoder analyzes input windows of 2048 samples with a shift length of 1024 samples between subsequent windows. As a result, the filterbank produces 1024 spectral coefficients, representing 1024 uniformly spaced filterbank channels with a frequency resolution of 23.4 Hz (assuming a sampling rate of 48 kHz).
- This high-frequency resolution allows for a very fine spectral shaping of the quantization noise, which is particularly important in the lower frequency range, where the critical bands are narrower.

MPEG-2 AAC Tools: Temporal Noise Shaping – The Problem

The TNS tool allows a fine temporal shaping of the coder's quantization noise.

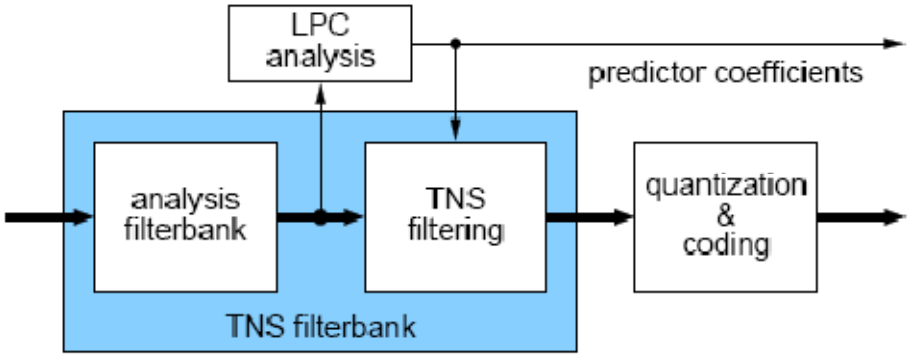
- Conventional transform coding schemes often encounter problems with signals that vary heavily over time, such as castanets, glockenspiel, or certain types of speech signals.
- The main reason for this is that the distribution of quantization noise can be controlled over frequency but is constant over a complete transform block. If the signal characteristic changes drastically within such a block without leading to a switch to shorter transform lengths, this equal distribution of quantization noise can lead to audible artifacts.
- Using a spectral signal decomposition for quantization and coding implies that a quantization error introduced in this domain will be spread out in time after reconstruction by the synthesis filterbank (time/frequency uncertainty principle).
- For commonly used filterbank designs (e.g. a 1024 lines MDCT), this means that the quantization noise may be spread over a period of more than 40 ms (for a sampling rate of 48 kHz). This will lead to problems when the signal contains strong signal components only in parts of the analysis filterbank window, i. e. for transient signals.

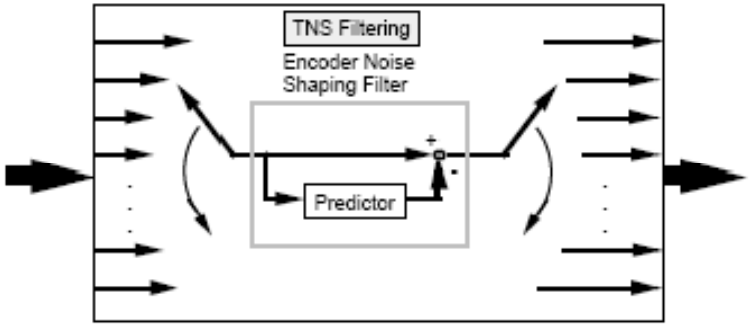
MPEG-2 AAC Tools: Temporal Noise Shaping – The Solution

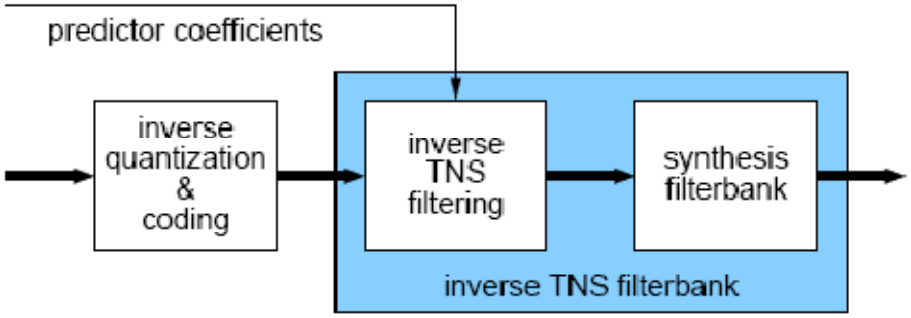
The TNS tool allows a fine temporal shaping of the coder's quantization noise.

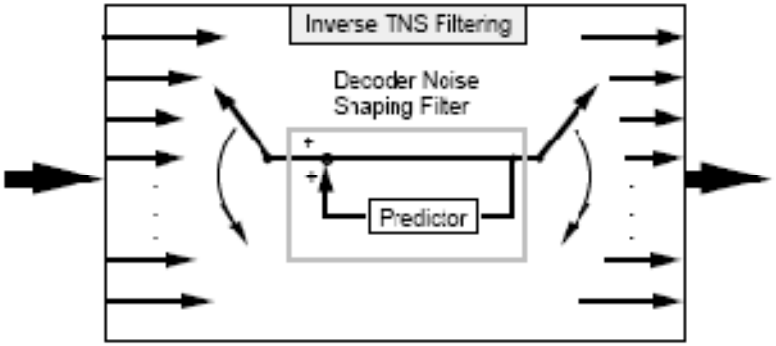
- The basic idea of TNS relies on the duality of time and frequency domain. TNS uses a prediction approach in the frequency domain to shape the quantization noise over time.
- It applies a filter to the original spectrum and quantizes this filtered signal. Additionally, quantized filter coefficients are transmitted in the bitstream. These are used in the decoder to undo the filtering performed in the encoder, leading to a temporally shaped distribution of quantization noise in the decoded audio signal.
- TNS can be viewed as a post-processing step of the transform, creating a continuous signal adaptive filter bank instead of the conventional two step switched filter bank approach.
- The actual implementation of the TNS approach within MPEG-2 AAC and MPEG-4 AAC allows for up to three distinct filters applied to different spectral regions of the input signal, further improving the flexibility of this novel approach.

Temporal Noise Shaping Encoder and Decoder



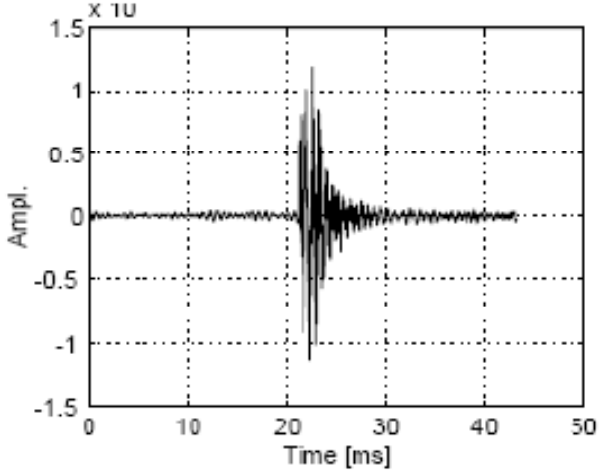




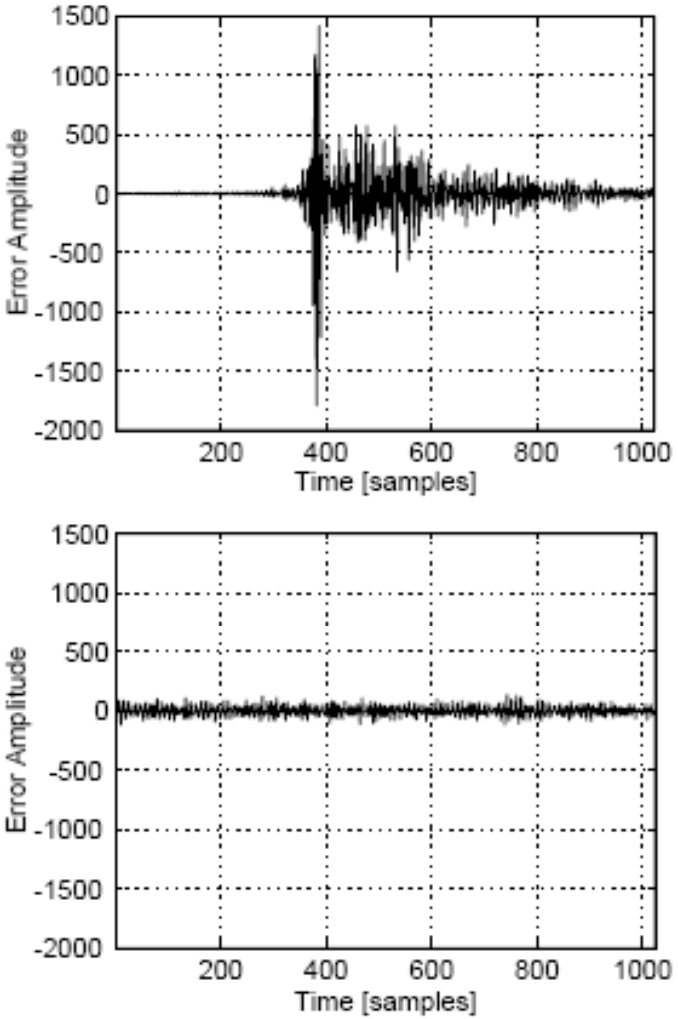


Forward prediction in the frequency domain shapes noise in the time domain.

Using TNS …



Transient signal (castanets, uncoded).



Coding noise in decoded castanets signal with (above) and without (below) TNS.

MPEG-2 AAC Tools: Prediction

The frequency domain prediction improves redundancy reduction of stationary signal segments.

- Since stationary signals can nearly always be found in long transform blocks, it is not supported in short blocks.
- The actual implementation of the predictor is a second order backwards adaptive lattice structure, independently calculated for every frequency line.
- The use of the predicted values instead of the original ones can be controlled on a scalefactor band basis and is decided based on the achieved prediction gain in that band.
- To improve stability of the predictors, a cyclic reset mechanism is applied which is synchronized between encoder and decoder via a dedicated bitstream element.
- The required processing power of the frequency domain prediction and the sensitivity to numerical imperfections make this tool hard to use on fixed point platforms. Additionally, the backwards adaptive structure of the predictor makes such bitstreams quite sensitive to transmission errors.

MPEG-2 AAC Tools: Scalefactors

- To improve the subjective quality of the coded signal, the noise is further shaped via scalefactors.
- The way the scalefactors are working is the following: Scalefactors are used to amplify the signal in certain spectral regions (the scalefactor bands) to increase the signal-to-noise ratio in these bands. Thus, they implicitly modify the bit-allocation over frequency since higher spectral values usually need more bits to be coded afterwards.
- Like the global quantizer, the stepsize of the scalefactors is 1.5 dB.
- To properly reconstruct the original spectral values in the decoder the scalefactors have to be transmitted within the bitstream.
- MPEG-4 AAC uses an advanced technique to code the scalefactors as efficiently as possible. First, it exploits the fact that scalefactors usually do not change too much from one scalefactor band to another. Thus, a differential encoding already provides some advantage. Second, it uses a Huffman code to further reduce the redundancy within the scalefactor data.

MPEG-2 AAC Tools: Quantization

Adaptive quantization of the spectral values is the main source of the bitrate reduction in all transform coders.

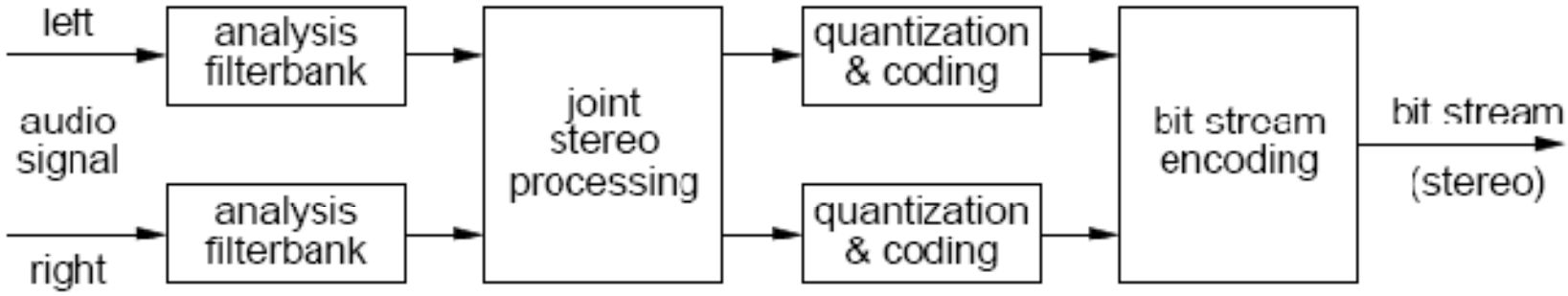
- It assigns a bit allocation to the spectral values according to the accuracy demands determined by the perceptual model, realizing the irrelevancy reduction.
- The key components of the quantization process are the actually used quantization function and the noise shaping that is achieved via the scalefactors.
- The quantizer used in MPEG-4 AAC has been designed similar to the one used in MPEG 1/2 Layer-3; it is a non-linear quantizer.
- The main advantage of this non-linear quantization over a conventional linear quantizer is the implicit noise shaping that this quantization creates.
- The absolute quantizer stepsize is determined via a specific bitstream element; it can be adjusted in 1.5 dB steps.

MPEG-2 AAC Tools: Noiseless Coding

Noiseless coding regards the part of the AAC codec which does not imply any losses, mainly entropy coding.

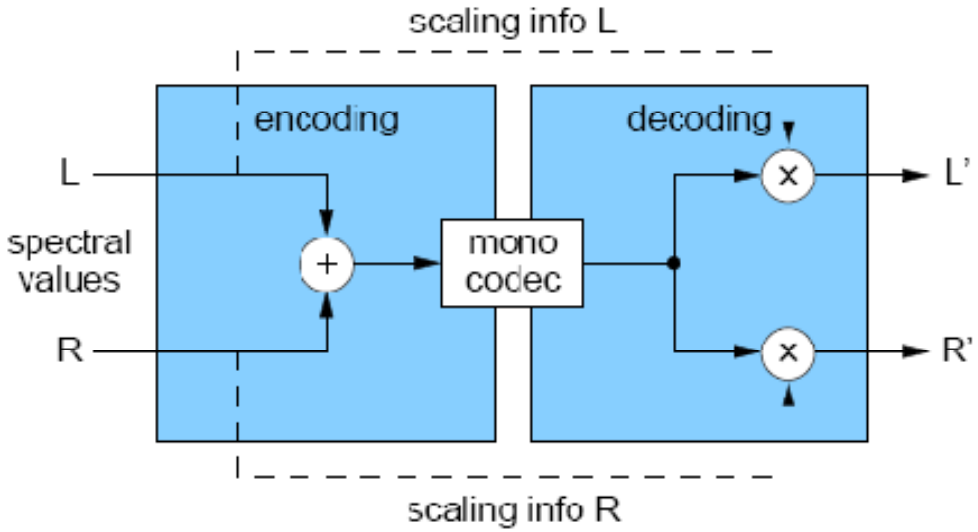
- The noiseless coding kernel within an MPEG-4 AAC encoder tries to optimize the redundancy reduction within the spectral data coding.
- The spectral data is encoded using a Huffman code which is selected from a set of available code books according to the maximum quantized value.
- The set of available codebooks includes one to signal that all spectral coefficients in the respective scalefactor band are "0", implying that there are neither spectral coefficients nor a scalefactor transmitted for that band.
- The selected table has to be transmitted inside the so-called section_data, creating a certain amount of side-information overhead. To find the optimum tradeoff between selecting the optimum table for each scalefactor band and minimizing the number of section_data elements to be transmitted, an efficient grouping algorithm is applied to the spectral data.

Joint Stereo Coding: Mid-Side (MS) Stereo Coding



- Joint stereo coding methods try to increase the coding efficiency when encoding stereo signals by exploiting commonalties between the left and right signal.
- AAC contains 2 different joint stereo coding algorithms, namely Mid-Side (MS) stereo coding and Intensity stereo coding.
- MS stereo applies a matrix to the left and right channel signals, computing sum and difference of the two original signals. Whenever a signal is concentrated in the middle of the stereo image, MS stereo can achieve a significant saving in bitrate. By applying the inverse matrix at the decoder the quantization noise becomes correlated and falls in the middle of the stereo image where it is masked by the signal.

Joint Stereo Coding: Intensity Stereo Coding



- Intensity stereo coding is a method that achieves a saving in bitrate by replacing the left and the right signal by a single representing signal plus directional information.
- This replacement is psychoacoustically justified in the higher frequency range since the human auditory system is insensitive to the signal phase at frequencies above, approximately, 2 kHz.
- Intensity stereo is by definition a lossy coding method, thus it is primarily useful at low bitrates; for coding at higher bitrates, only MS stereo is used.

MPEG-2 AAC Profiles

The MPEG-2 AAC standard defines three profiles, corresponding to different configurations of the basic coding scheme, providing different trade-off options between coding performance and complexity:

- **Low-Complexity (LC) profile -** Defines a baseline coder that is both efficient in coding and has moderate complexity (no interframe prediction is used, the maximum temporal noise shaping (TNS) filter order is limited to 12).

- **Main profile -** Does not carry the preceding restrictions and delivers somewhat higher compression performance at the expense of higher memory and computational demands. Because the Main profile is a true superset of the LC profile, all LC profile bitstreams can be decoded by a Main profile decoder.

- **Scalable Sampling Rate (SSR) profile -** Can provide decoder configurations with even lower complexity than the LC profile; if not, the entire audio bandwidth is decoded. This is achieved by using a preprocessing stage (including a first filterbank and the gain control stage) in combination with a filterbank of modified length. Only partial compatibility is achieved with the LC profile.

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.**

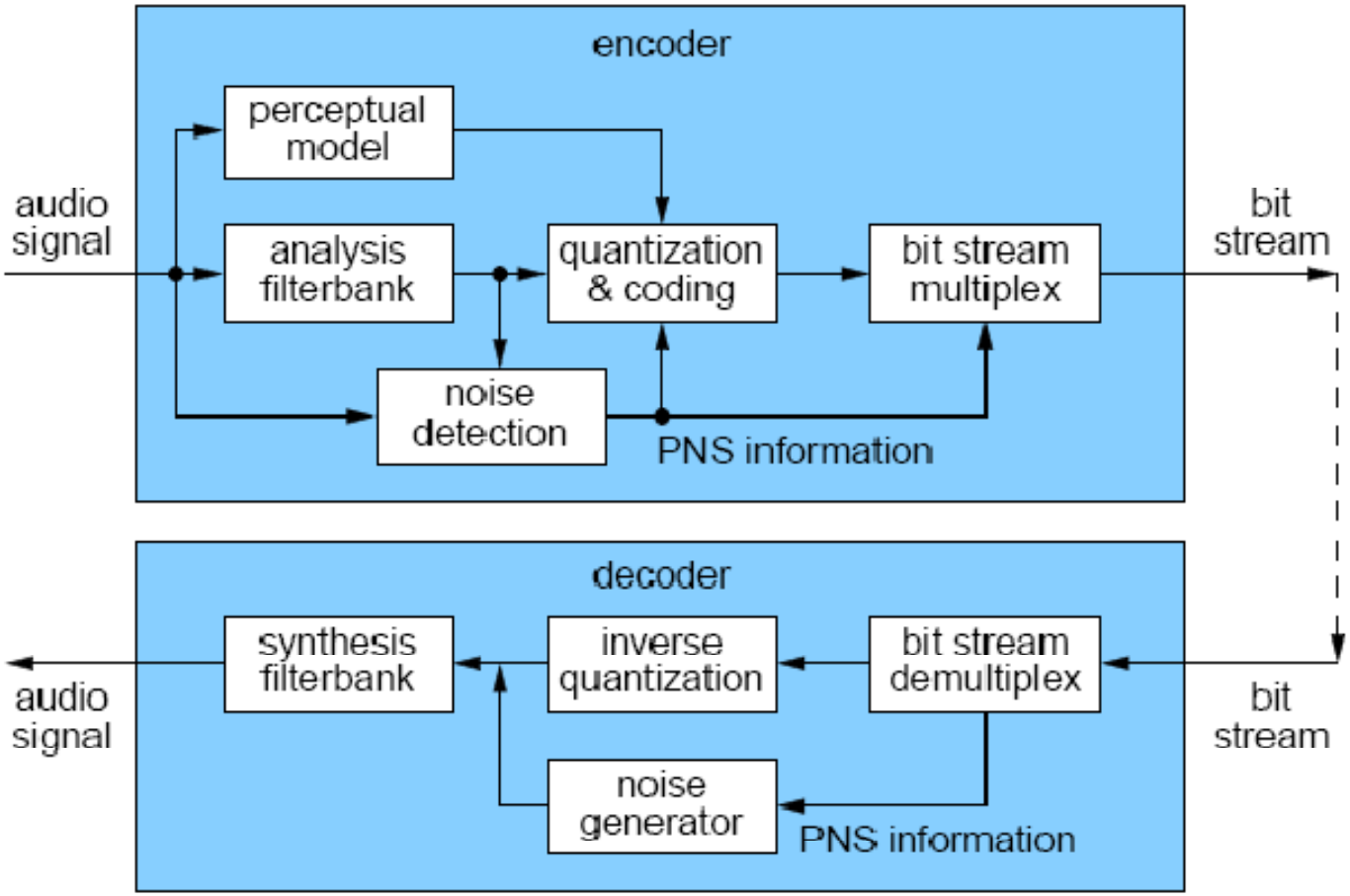
MPEG-4 AAC Tools

Perceptual Noise Substitution (PNS)

Perceptual Noise Substitution (PNS) aims at further increasing the AAC compression efficiency at lower bitrates.

- PNS is based on the observation that one noise sounds like the other. This means that the actual fine structure of a noise signal is of minor importance for the subjective perception of such a signal.
- Consequently, instead of transmitting the actual spectral components of a noisy signal, the bitstream would just signal that this frequency region is a noise-like one and give some additional information on the total power in that band.
- PNS can be switched on a scalefactor band basis so even if there just are some spectral regions with a noisy structure, PNS can be used to save bits. In the decoder, a randomly generated noise will be inserted into the appropriate spectral region, according to the power level signaled within the bitstream.
- The most challenging task in the context of PNS is not to enter the appropriate information into the bitstream but reliably determining which spectral regions may be treated as noise like and thus may be coded using PNS, without creating severe coding artifacts.

Perceptual Noise Substitution (PNS)

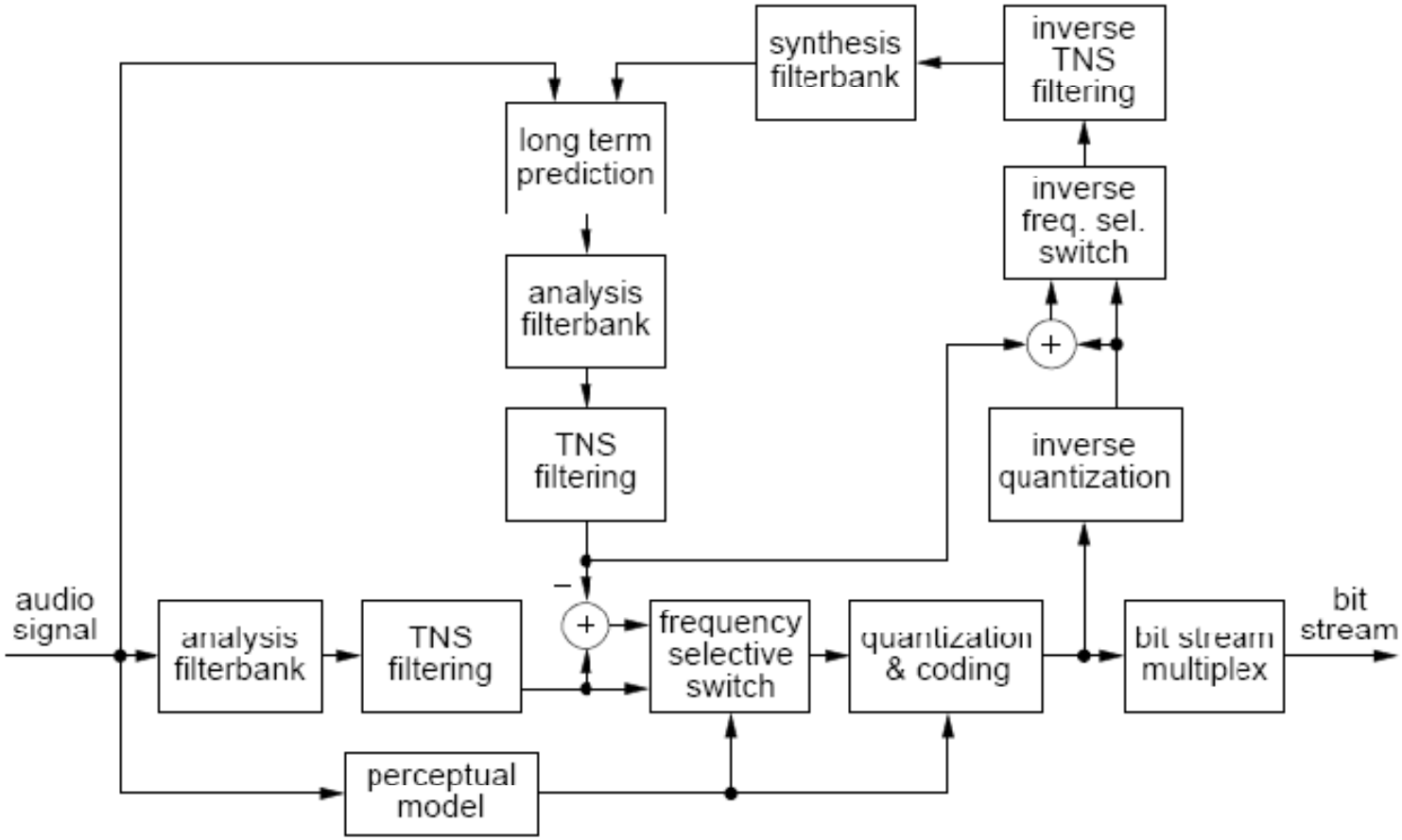


Long Term Prediction (LTP)

Long term prediction (LTP) is an efficient tool for reducing the redundancy of a signal between successive coding frames.

- **This tool is especially effective for the parts of a signal which have a clear pitch property (the pitch property is designed to set the pitch level for a given speaking voice).**
- **The implementation complexity of LTP is significantly lower than the complexity of the MPEG-2 AAC frequency domain prediction.**
- **Because the Long Term Predictor is a forward adaptive predictor (prediction coefficients are sent as side information), it is inherently less sensitive to round-off numerical errors in the decoder or bit errors in the transmitted spectral coefficients.**

Long Term Prediction (LTP)



AAC Low Delay

- AAC exhibits a minimum theoretic algorithmic delay of up to several hundred ms; thus, it is not well-suited for low delay applications, such as real-time bidirectional communications.
- In contrast to this, traditional speech coding schemes provide good quality coding at low delay only for a narrow class of signals (i.e., speech).
- To enable coding of audio signals with an algorithmic delay down to 20 ms, MPEG-4 specifies a so-called low-delay audio coding mode:
  - Operating at sampling rates up to 48 kHz and using a frame length of 512 or 480 samples (compared to the 1024 or 960 samples used in core AAC); also the size of the window used in the analysis and synthesis filterbank is reduced by a factor of two.
  - No window switching is used to avoid the look-ahead delay; to reduce pre-echo artifacts, only TNS is employed with window shape adaptation.
  - Although for the non-transient parts of the signal a sine window is used, a so-called low overlap window is applied for the case of transient signals in order to achieve optimum TNS performance, reducing the effects of temporal aliasing as a result of the MDCT filterbank.
  - Furthermore, the use of the bit reservoir is minimized at the encoder in order to reach the desired target delay; in the extreme case, no bit reservoir is used at all.

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.**

HE-AAC: Target Applications

Target applications for HE-AAC are mobile music, mobile TV, digital radio and TV broadcasting, Internet streaming, and consumer electronics.

- In the mobile music and TV market, HE-AAC is used for music downloads, music streaming, ring tones, ring-back tones and the audio part of various mobile TV broadcasting systems.
- In audio broadcasting, HE-AAC is a mandatory component of multiple existing and emerging systems.
- In TV broadcasting, the codec is of special interest in combination with H.264//AVC for video, in new systems. The first commercial services using both MPEG standards (AAC and AVC) were launched in 2007.
- In Internet streaming HE-AAC is of special interest because of the significant bandwidth savings at the server side and the capability to stream directly into the mobile environment.

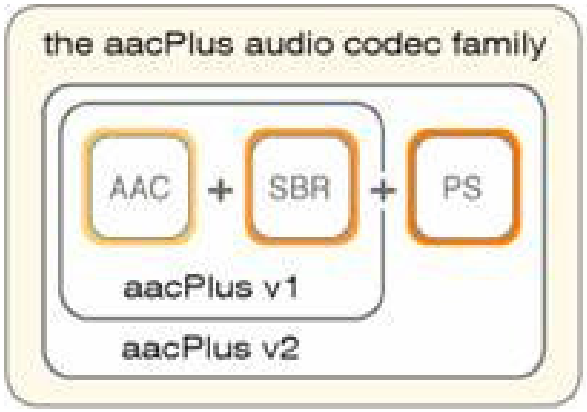
HE-AAC: Basic Targets

- **Previous audio coders typically have to reduce the transmitted audio bandwidth when operating at low bitrates (e.g. below 48 kbit/s per audio channel) in order to avoid excessive coding artifacts from being introduced in the transmitted low frequency region.**
- **HE-AAC technology was designed to overcome this obstacle by reproducing a wide audio bandwidth independently of the coding bitrate by using audio bandwidth extension.**
- **An enhanced version of the coder (HE-AAC v2) has been designed to additionally exploit models of human spatial perception to achieve a further boost in coding efficiency.**
- **In both cases (HE-AAC v1 and HE-AAC v2), the objective was to achieve this goal by means of simple extensions to the AAC architecture coming at a limited increase in complexity.**

HE-AAC: Normative Elements

- **The MPEG-4 HE-AAC audio coding standard specifies the bitstream format and the decoding process, including conformance testing methods and reference implementations.**
- **The decoding process defines how the syntax elements present in the encoded bitstream are converted into a time domain Pulse Code Modulated (PCM) digital audio signal. As a result, every decoder conforming to the standard will produce a well-defined output signal for any bitstream conforming to the standard.**
- **The encoding algorithm, on the other hand, is not normatively specified, thus e.g. allowing to balance real-time execution speed and audio quality, depending on the individual application demands.**

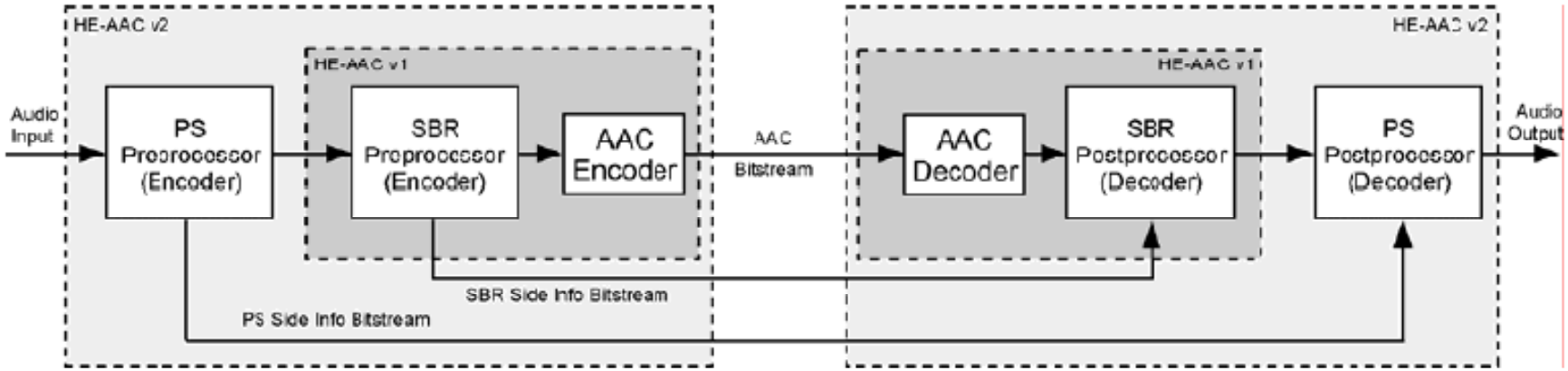
HE-AAC: Functionalities



- **High-Efficiency AAC supports a broad range of compression ratios and configurations ranging from highly efficient mono and stereo coding (typical operation point 32 kbit/s stereo with HE-AAC v2) via high quality multi-channel coding (typical operation point 160 kbit/s for 5.1 configuration) to near-transparent multi-channel compression (typical operation point 320 kbit/s using AAC-only operation).**

- **Because subsequent HE-AAC versions form a superset of their predecessors, HE-AAC v2 decoding is fully compatible with AAC-only and HE-AAC v1 content.**

HE-AAC: Architecture



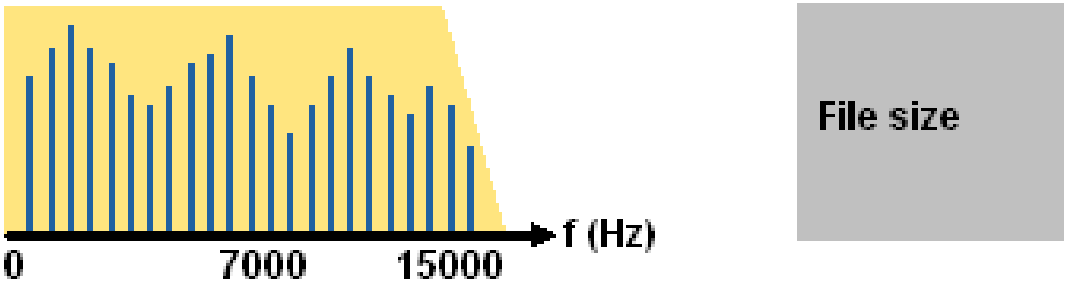
- **The core of the system is the AAC waveform codec.**
- **For increased compression efficiency, the Spectral Band Replication (SBR) bandwidth enhancement tool and the Parametric Stereo (PS) advanced stereo compression tool are added to the system.**
- **Both SBR and PS act as preprocessing blocks at the encoder side and post-processing blocks at the decoder side.**
- **The bitstream syntax of HE-AAC allows for up to 48 audio channels; in practice, mono, stereo and 5.1 multi-channel are the most commonly used configurations.**

Spectral Band Replication (SBR)

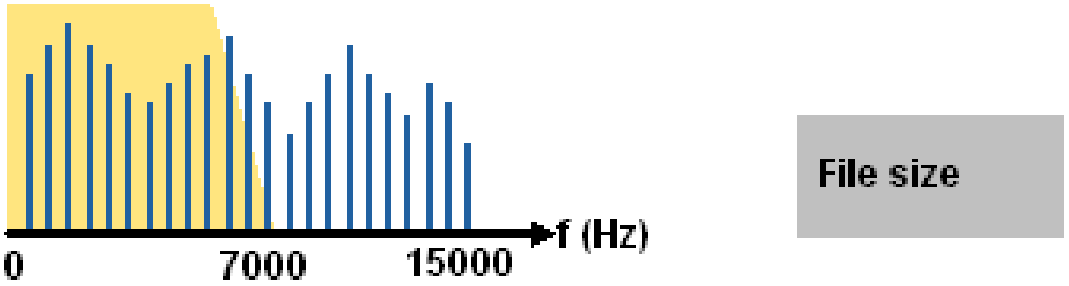
- **Bandwidth extension technology is based on the observation that usually the upper part of the spectrum of an audio signal contributes only marginally to the "perceptual information" contained in the signal, and that human auditory perception is less sensitive in the high frequency range.**
- **SBR exploits this observation for the purpose of improved compression: instead of transmitting the upper part of the spectrum with AAC, SBR regenerates it from the lower part with the help of some low-bitrate guidance data.**
- **For regenerating the missing high-frequency components, SBR operates in the frequency domain using a QMF (Quadrature Mirror Filter) filterbank analysis/synthesis system.**
- **The SBR bitstream data controls both the operation of the high-frequency reconstruction and the envelope adjustment. Depending on the specific configuration, the SBR side information rate is typically a few (e.g. 2-3) kbit/s.**

The SBR Principle and Benefit

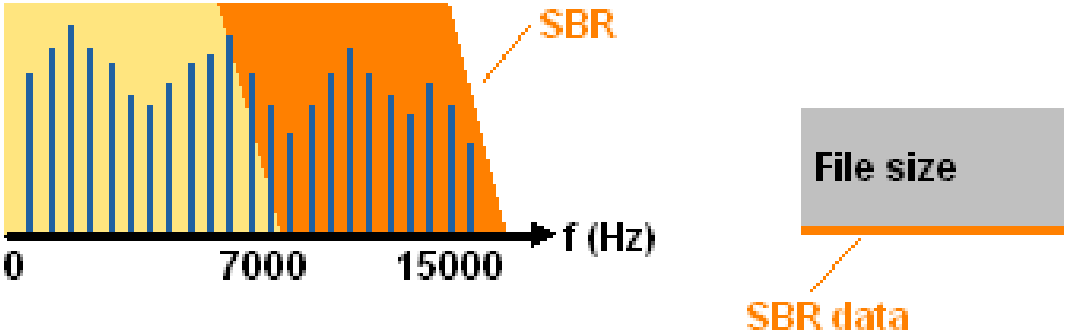
MP3 at 128Kbps



MP3 at 64Kbps (frequencies cut in half)



mp3PRO at 64Kbps (high frequencies SBR encoded)



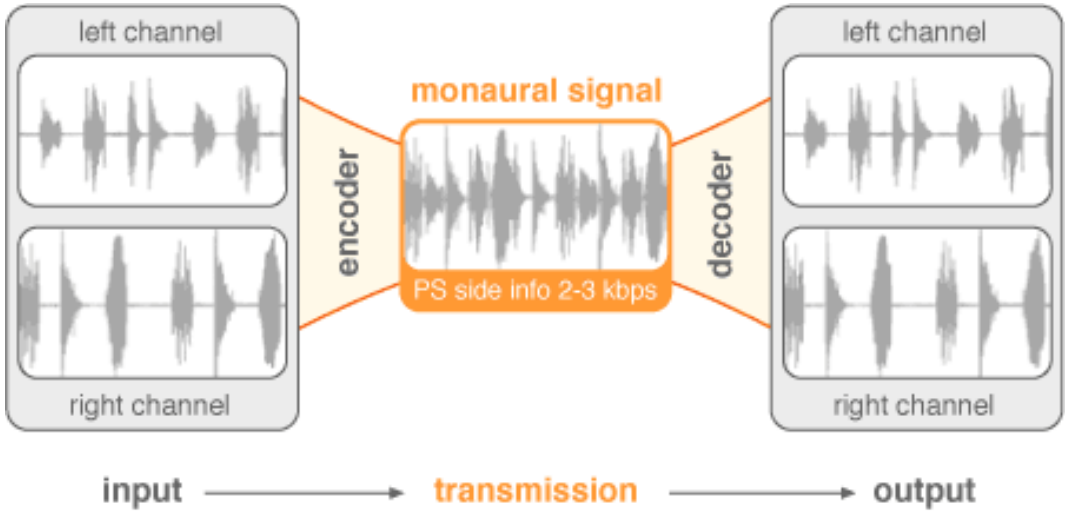
Detailing Spectral Band Replication (SBR)

The most important SBR building blocks are:

- **High Frequency Reconstruction - The so-called transposer generates a first estimate for the upper part of the spectrum by copying and shifting the lower part of the transmitted spectrum. In order to generate a high-frequency spectrum that is close to the original spectrum in its fine structure, several provisions are available including the addition of noise, the flattening of the spectral fine structure and the addition of missing sinusoids.**

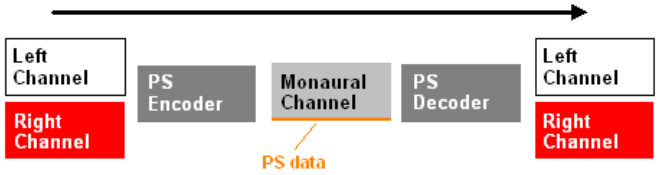
- **Envelope Adjustment - The upper spectrum generated by the transposer needs to be shaped subsequently with respect to frequency and time in order to match the original spectral envelope as closely as possible.**

Parametric Stereo (PS)



- **Parametric Stereo is an extension of a well-known principle for efficient joint coding of stereo audio: instead of the stereo signal, just a mono-downmix is transmitted, along with a small data stream describing how to upmix the signal back to stereo in the decoder. The PS technology is defined for stereo configurations only.**
- **The intensity stereo tool available in AAC and many other codecs (like MP3) is a simple implementation of this approach, whereas PS is a significantly more sophisticated variant thereof.**

Parametric Stereo (PS)



- To reproduce a high-quality stereophonic sound image, it is vital to consistently preserve the cues that determine human spatial hearing of sound, i.e. inter-aural level difference, inter-aural time/phase difference and inter-aural correlation/coherence.
- While traditional intensity stereo can only reproduce level (=intensity) differences between the stereo channels, the PS technology can also produce phase differences and decorrelation between the stereo pair to yield a convincing upmix quality.
- Most notably, PS includes a decorrelator tool that creates an adjustable degree of decorrelation between the 2 channels and is steered by coherence factors measured in the encoder and transmitted in the PS data. This is vital for modeling sound sources with a wide sound image (e.g. a choir) or room ambience.
- Since PS operates on the same spectral representation as SBR, both can be efficiently integrated to form an even more efficient compression algorithm for audio signals at relatively low additional computational complexity. Also PS coding requires only a few kbit/s transmitted as its side information data rate.

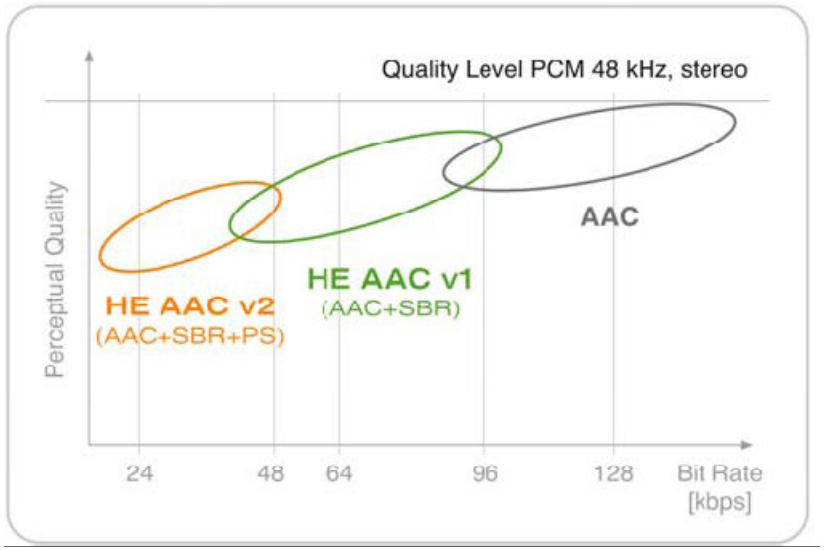
MPEG-4 AAC Profiles

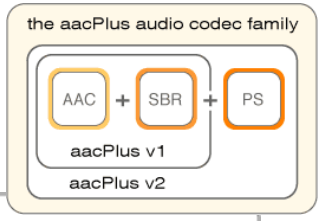
- **AAC Profile - fairly similar to the MPEG-2 AAC LC profile, but with some additional tools making MPEG-4 AAC**

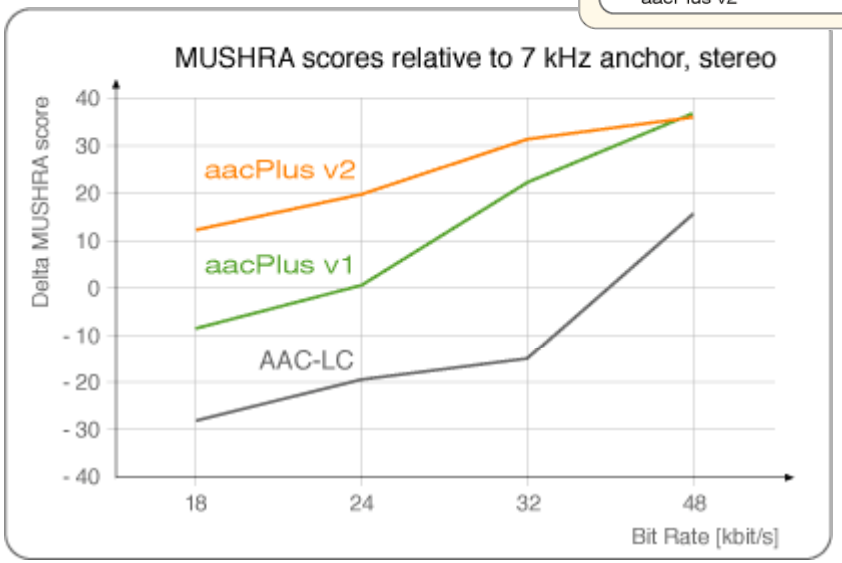
- **High Efficiency AAC v1 Profile - MPEG-4 AAC and SBR**

- **High Efficiency AAC v2 Profile – MPEG-4 AAC, SBR, and PS**

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).**

HE-AAC Family: Complexity Performance

- The significant increase in coding efficiency of HE-AAC over MPEG-4 AAC comes at moderate additional computational complexity.
- While both the SBR and the PS tool consume additional calculations, this increase is partially compensated by running the AAC core at half the original sampling rate and just for one channel (in case of PS). As a consequence, the approximate computational complexity of the decoder is increased by a factor of 1.5 and 2, when comparing HE-AAC v1 and HE-AAC v2 to AAC, respectively.
- The encoder complexity is roughly similar for all three variants.

HE-AAC Family: Profiles and Levels

Level	AAC profile		HE-AAC v1 profile				HE-AAC v2 profile			
	Max. channels	Max. sampling rate [kHz]	Max. channels	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate [kHz]	Max. channels	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate [kHz]
1	2	24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	2	48	2	48	24	48	2	48	24	48
3	N/A	N/A	2	48	48	48	2	48	48	48
4	5	48	5	48	48	48	5	48	48	48
5	5	96	5	96	48	96	5	96	48	96

- **The profiles and levels have been designed in a strictly hierarchical way such that the HE-AAC v2 profile is a superset of the HE-AAC v1 Profile, which in turn is a superset of the AAC profile.**
- **Also, within all profiles each higher level is a superset of the lower levels.**
- **In practice, the most relevant levels are Level 2 for stereo devices (e.g. cell phones, broadcasting receivers) and Level 4 for multi-channel systems (e.g. digital television).**

Final Remarks on AAC

- The AAC standard and its variations builds on previous audio coding standards to achieve high compression for a wide range of bitrates.
- The compression gains are mainly related to additional tools such as frequency prediction, temporal noise shaping, perceptual noise shaping, long term prediction and spectral band replication.
- The AAC standard represents nowadays the state-of-the-art in audio coding and it is currently being adopted by a growing number of organizations, companies and consortia.
- Next development regards Surround Sound
  - The MPEG Surround standard supports very efficient parametric coding of multi-channel audio, to permit transmission of such signals over channels that typically support only the transmission of stereo (or even mono) signals. Moreover, it is able to provide backward compatibility with non-multi-channel audio systems: while legacy receivers decode an MPEG Surround bitstream as stereo, enhanced receivers provide multi-channel output.

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 …


2001


2003


2004


2004


2004


2005


2005


2005


2006


2006


2006


2006

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