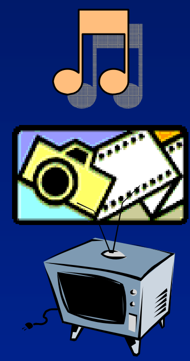
# New coding techniques, standardisation, and quality metrics

Touradj Ebrahimi
Touradj.Ebrahimi@epfl.ch

# Background

# Multimedia

- Means of Expression aimed at being consumed by humans
- Modalities
  - Audio (music, song, radio, …)
  - Image (paintings, photos, press, …)
  - Video (cinema, TV, …)
  - …

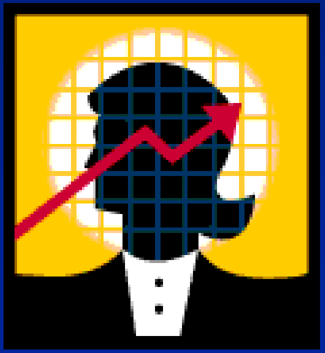


Closely linked to content creation

# Digital revolution



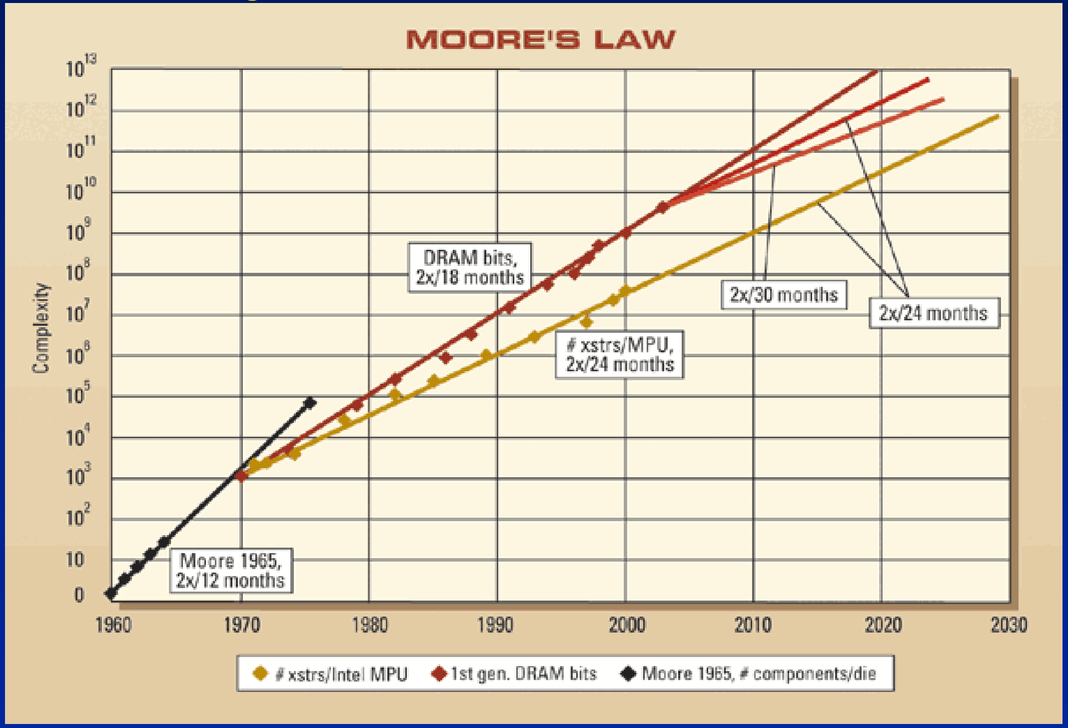
- Representation of media in digital form
  - Ease of manipulation by computers
  - Efficient transmission through various communication channels
  - Robustness to errors
  - Infinitely reproducible
  - Low cost
  - Easily searchable and accessible
  - …
- Moore's law (exponential growth)
  - Computing power
  - Storage
  - Communication channels bitrates

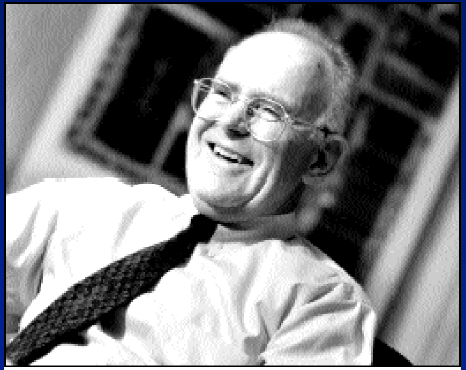


# Moore's law



- Number of transistors on a chip doubles every 18 months





Gordon Moore, Intel co-founder

# Multimedia challenges

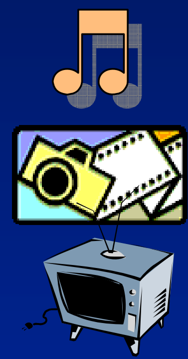


- Applications
- Compression as an enabler and essential tool in many applications
- Importance of international standards
- Seamless exchange of information
- Media Security, ownership protection, privacy
- Information management
- Quality
- New information modalities

# Digital Media Roadmap

# Multimedia

- Means of Expression aimed at being consumed by humans
- Modalities
  - Audio (music, song, radio, …)
  - Image (paintings, photos, press, …)
  - Video (cinema, TV, …)
  - …



Closely linked to content creation

# Multimedia anywhere, anytime, accessible to everybody



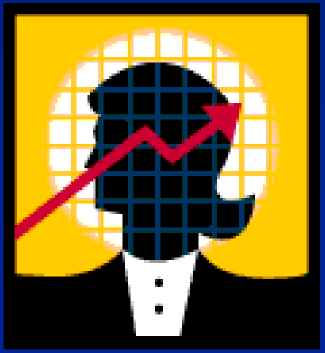
# In many forms, applications and services



# Digital revolution

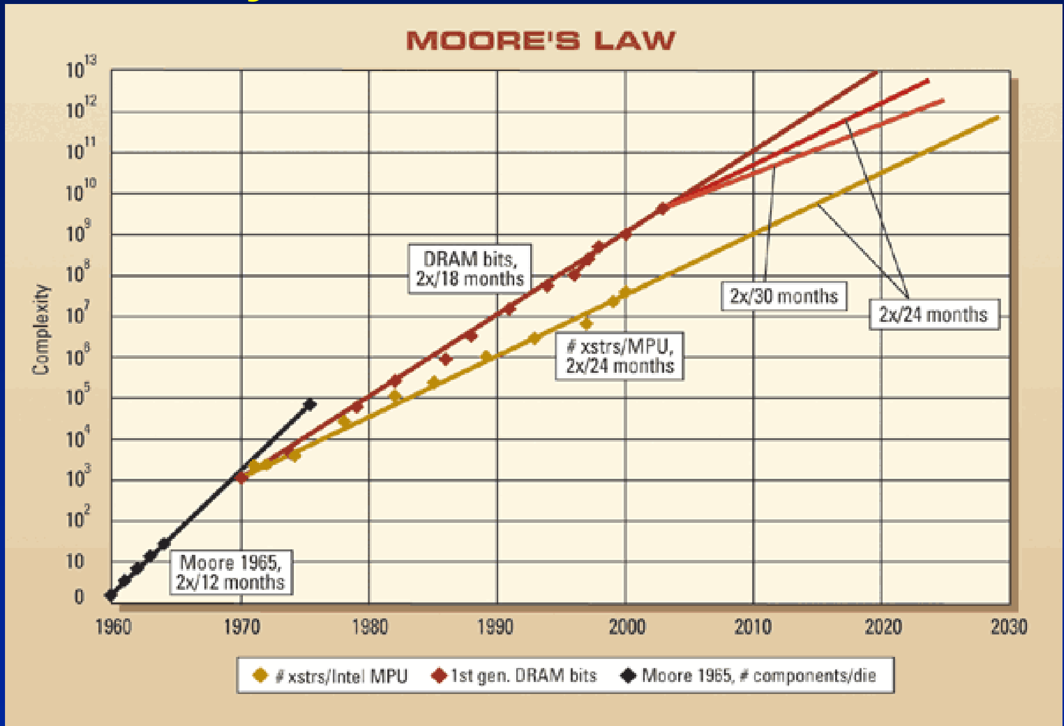


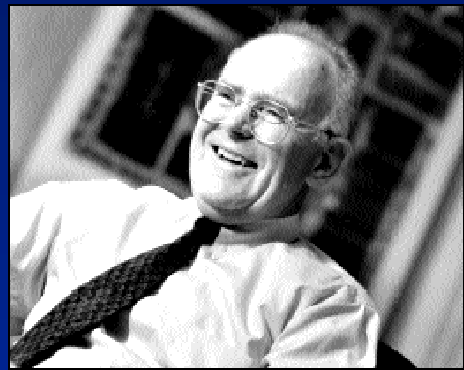
- Representation of media in digital form
  - Ease of manipulation by computers
  - Efficient transmission through various communication channels
  - Robustness to errors
  - Infinitely reproducible
  - Low cost
  - Easily searchable and accessible
  - …
- Moore's law (exponential growth)
  - Computing power
  - Storage
  - Communication channels bitrates



# Moore's law

- Number of transistors on a chip doubles every 18 months



Gordon Moore, Intel co-founder

# Digital Media Roadmap

# Phase 1 (yesterday): Era of all digital

- World of acronyms
  - CD (Digital music)
  - DVD (Digital video)
  - DAB (Digital radio)
  - DVB (Digital TV)
  - ISDN (Digital phone)
  - GSM/UMTS (Digital mobile phone)
- Media in bits!
  - Essentially same experience as before, but more efficiently deployed

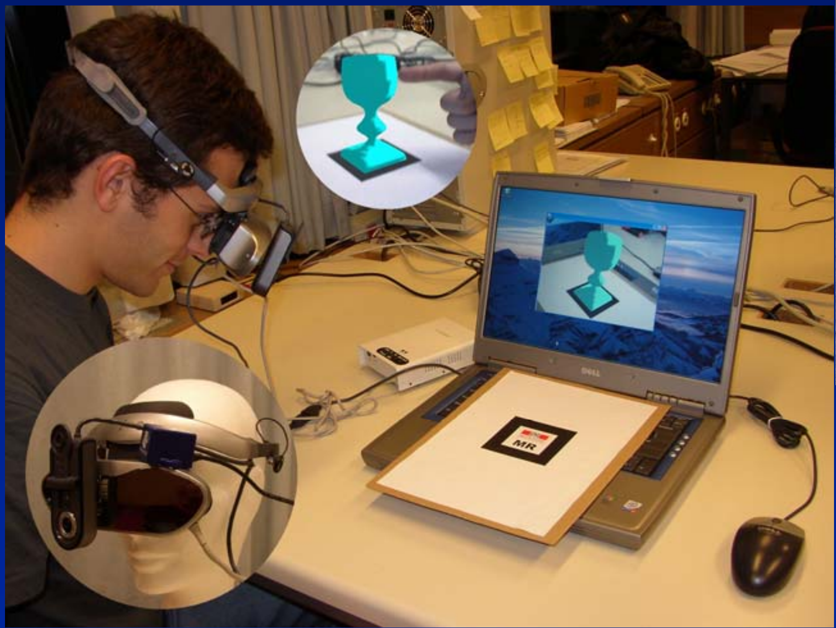
# Phase 2 (today): Era of multimedia

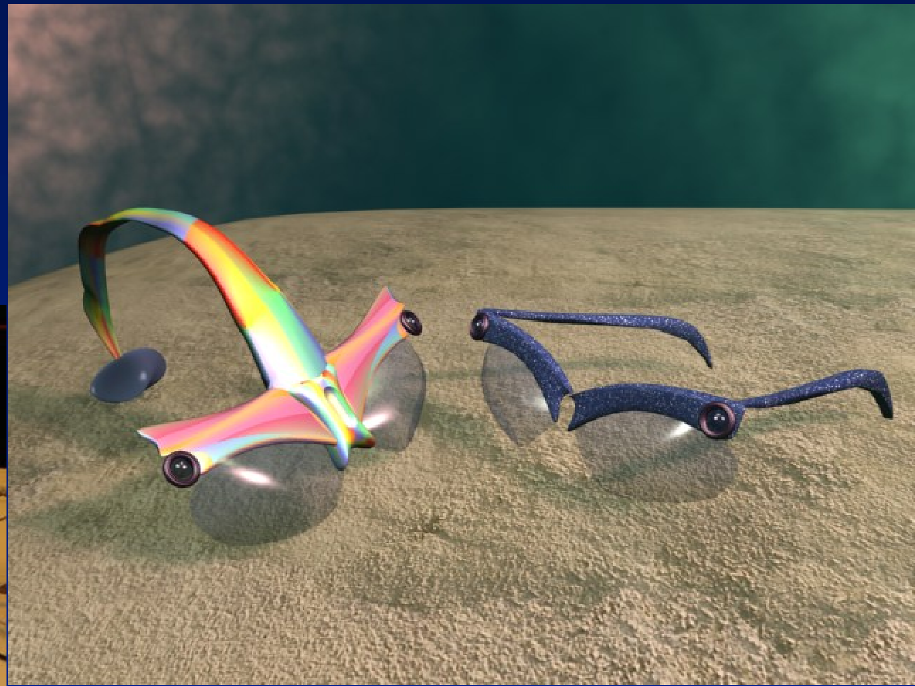


- Synthetic and natural media
  - Natural content (AV) can be acquired easily and cost efficiently
  - Content is not anymore only the result of acquisition of real world through various
  - Virtual reality
  - Augmented reality
  - Mixed reality

# Cyber glasses





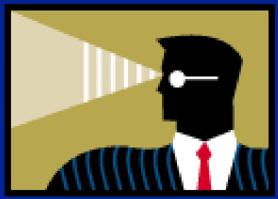


# Phase 2 (today): Era of multimedia

- New media, rich media
  - New types of content (media) and experience start appearing thanks to digital revolution
  - Content is increasingly rich (associated metadata) and accessible
  - Content democratization (consumer-producer)

# Phase 3 (tomorrow): Era of interface

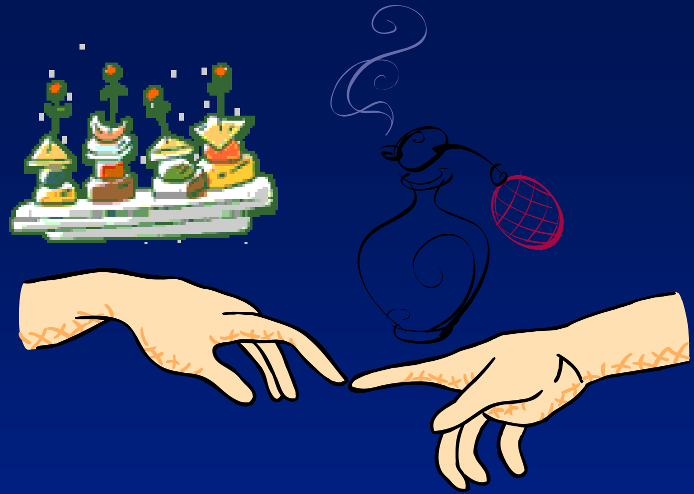
- Moore's law does not apply to the interface between man and machine
  - We basically use the same interface devices in the last 50 years (keyboard, mouse, screen, …)
  - The communication bitrate between man and machine remains very low
  - Interfaces remain unnatural
- New man-machine interfaces
  - Computers/devices that can see and hear
  - Speech recognition
  - Gesture recognition
  - …

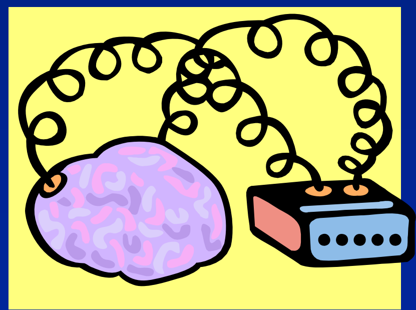


# Phase 4 (day after tomorrow): Era of extended media and interface

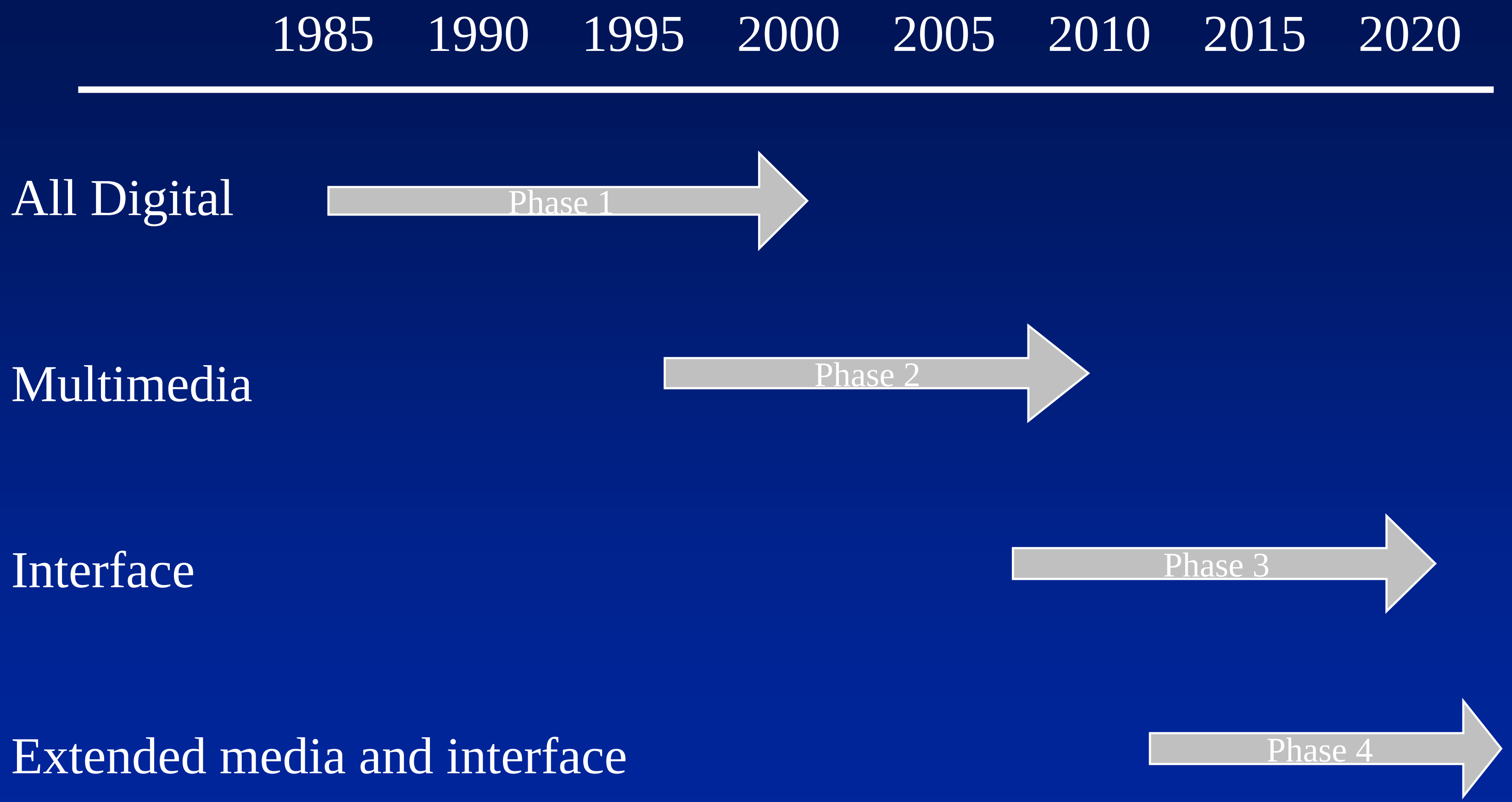


- Multimedia beyond audiovisual
  - Olfaction (smell)
  - Haptics (touch)
  - Gustation (taste)
- New interfaces
  - Emotional computers
  - Brain-computer interface



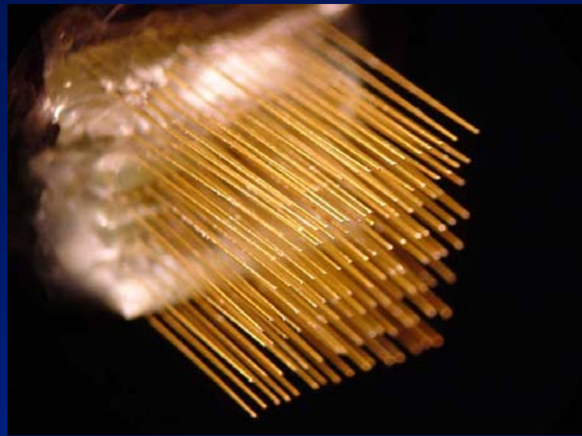


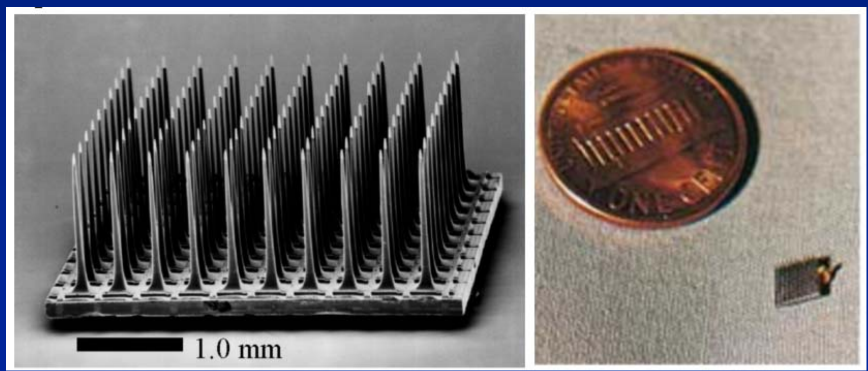
# Digital Media Roadmap

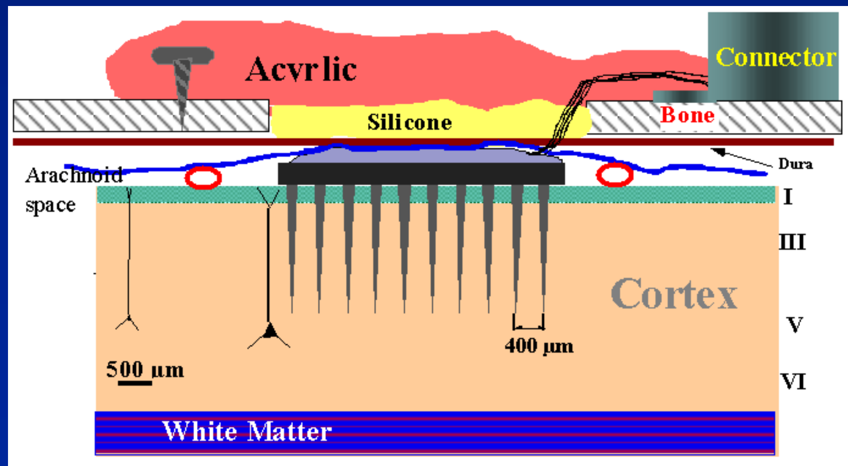


# Phase 5 : Era of symbiosis

- Neural implants
- Brain co-processors
- Communication beyond five senses

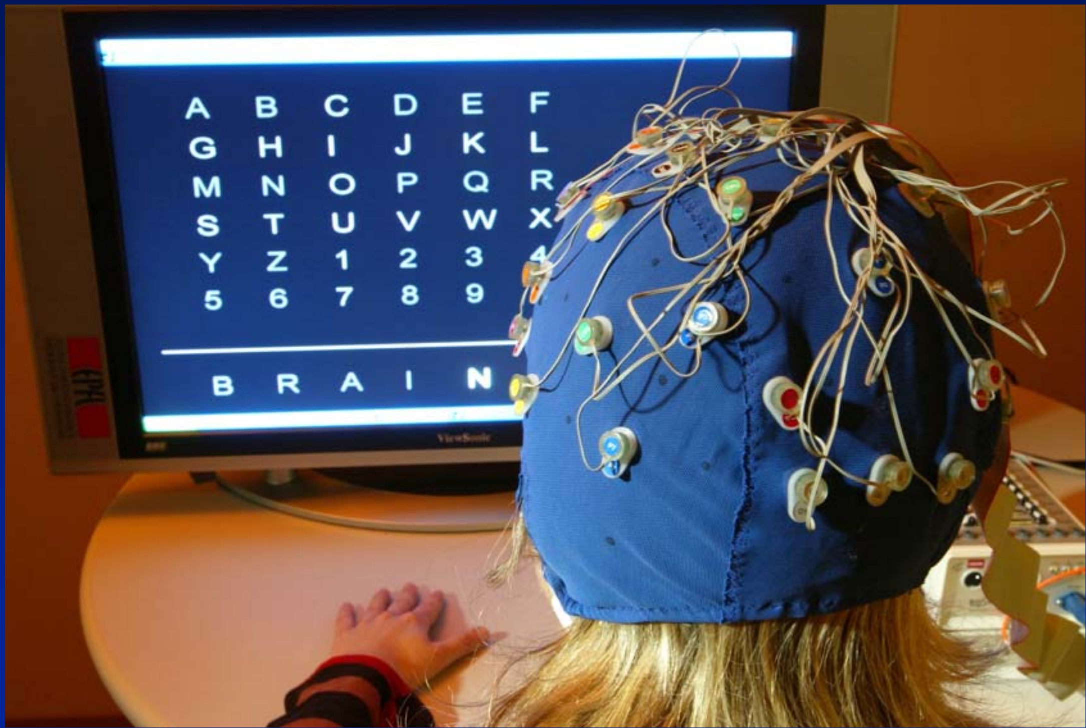


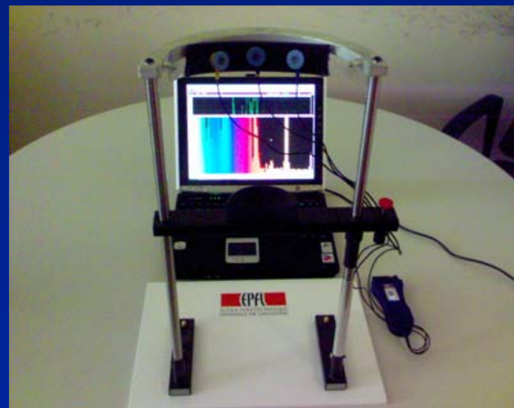


# Phase 5 : Era of symbiosis





# Phase 5 : Era of symbiosis

- Quote from Murray Gell-Mann, Nobel Prize and founder of Santa Fe institute:

*" One day, for the good or for the bad of it, such interconnections will be possible. Human beings will be able to connect to computers, without need to go through the interface of spoken language or a display, and through these computers to one or more other human beings. There will be a total sharing of thoughts and feelings, without any of the limitations of the language."*

# Media Standards

# Standards

- Communication convention (interoperability)
  - ISO/IEC
  - ITU-T
- Increasingly important in the global village
- Often long and tedious process
- Often not the best technology around by the time it is agreed upon
- IP issues

## Why compression?



- Essential in most applications because:
  - Bandwidth is limited
  - Content is bandwidth hungry
- One of the most confusing and hyped components in any multimedia system
- Intimately linked to content representation
- Intimately linked to application requirements

## Content representation

- Pixel-based
- Object-based
- Content-based
- Model-based

# Application requirements



- Functionalities
  - Compression efficiency
  - Error resiliency
  - Complexity
  - Scalability
  - Power consumption
  - Delay
  - …

# H.26x approach

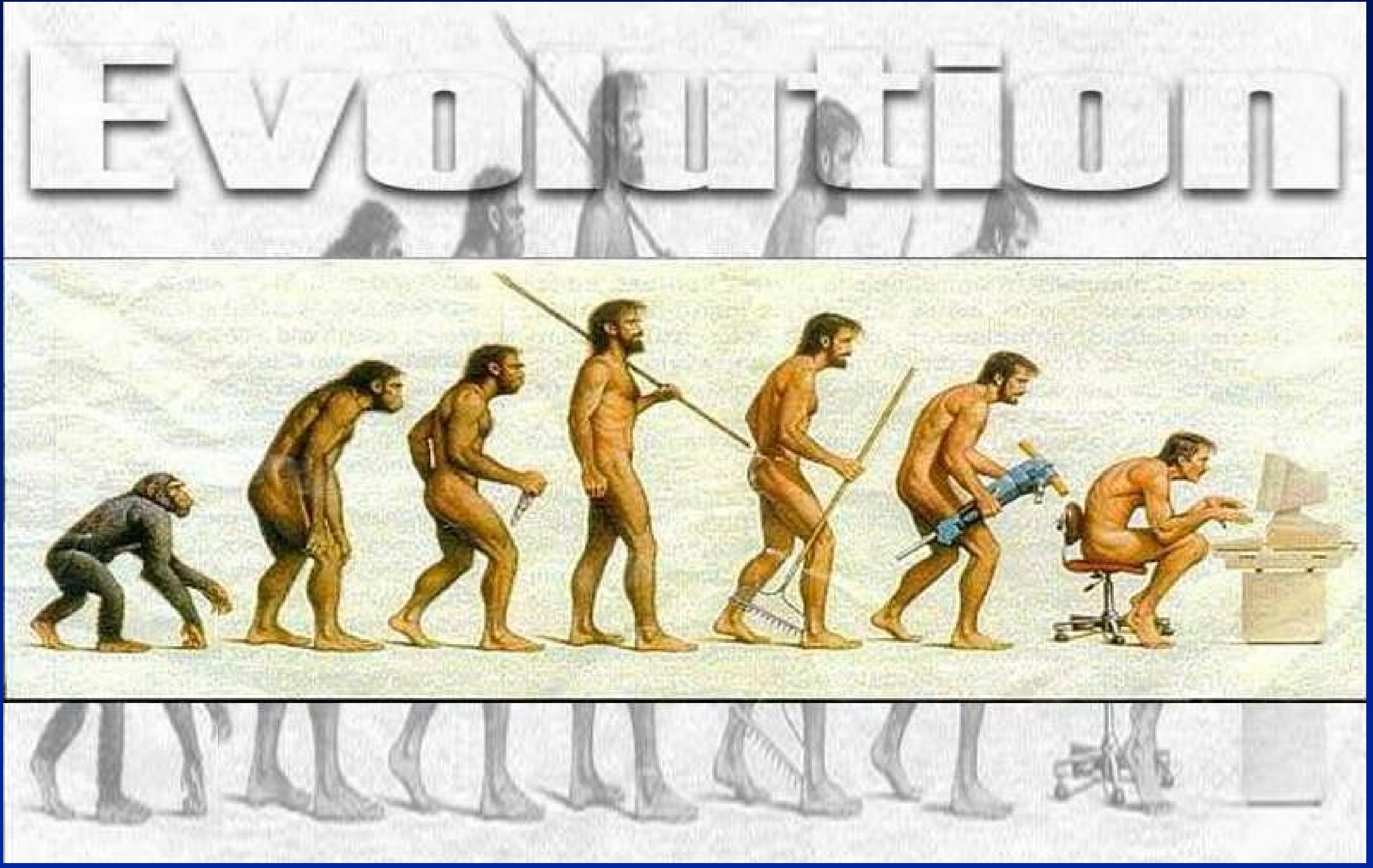
- Compression efficiency is very important
- Other functionalities are also important but not as much as compression efficiency
- Pixel-based approach is good enough to cover all requirements in today and tomorrow's applications

# MPEG-x approach

- Compression efficiency is important
- But other functionalities are as important
- Pixel-based approach is not always enough to cover well all requirements in today and tomorrow's applications

# Media Coding Standards Evolution …





# Frame-based Video/Image Coding Standards up to AVC



- JPEG (1988) - Still pictures
- H.261 (1990) - Videotelephony
- MPEG-1 (1991) - CD-ROM storage
- MPEG-2/H.262 (1993) - Digital TV
- H.263 (1995) - Video over PSTN and mobile networks

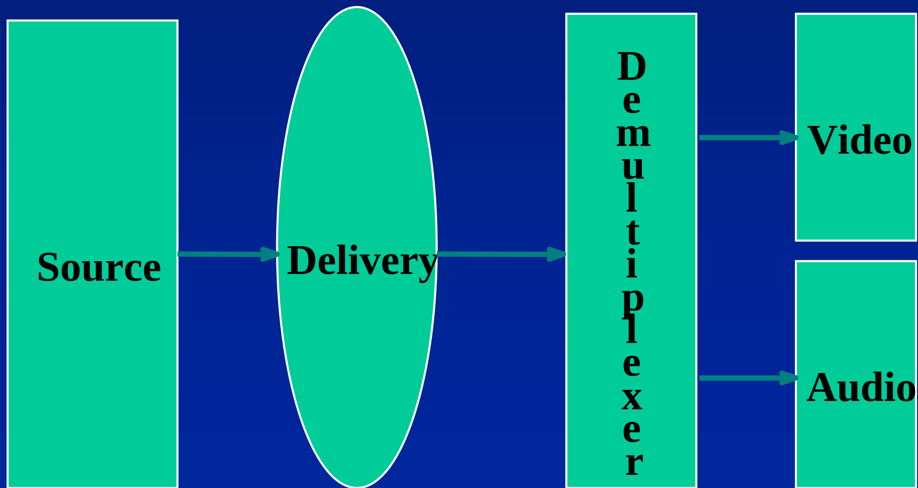
- MPEG-4 Visual (1998) – All applications
- JPEG 2000 (2000) – Still pictures and video

Standards continuously 'buy' quality and functionality with technological novelty and complexity ...

# MPEG-1: The First Entertainment Bits 1991 !



- Based on the DCT-motion compensation cocktail for video and perceptual coding for audio
- MPEG-1 is one of the most common AV format for PC environments
- Video CDs and MP3 (MPEG-1 Audio Layer III) players are big successes
- MPEG-1 content is nowadays everybody's stuff

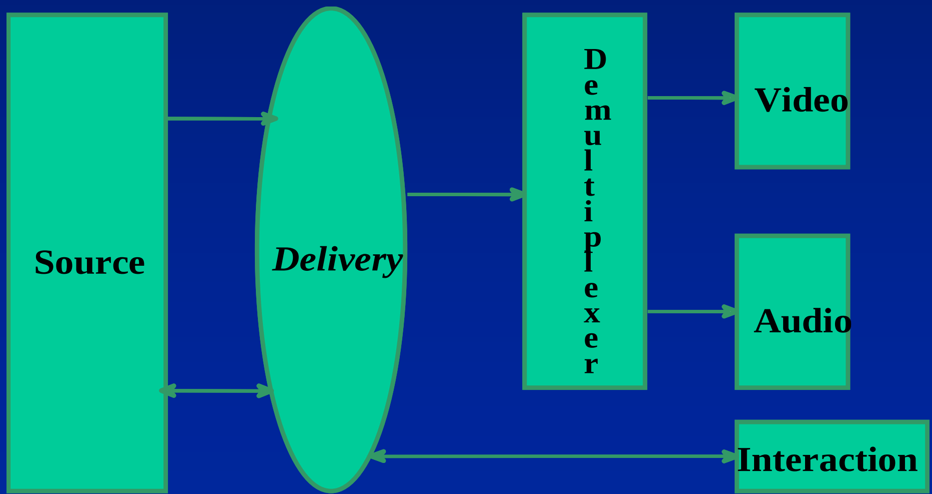



# MPEG-2: Digital TV

**1993 !**



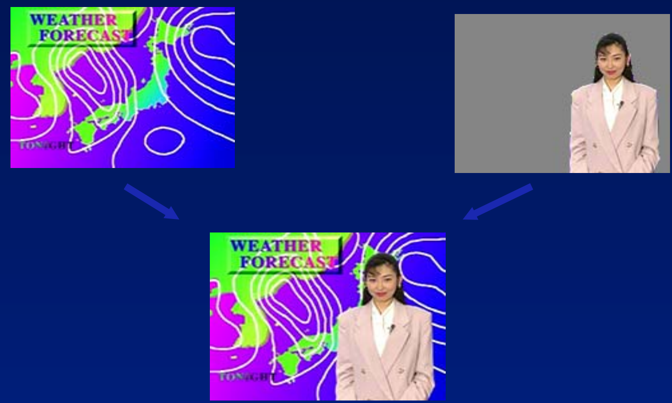
- Based on the DCT-motion compensation cocktail for video and perceptual coding for audio (multi-channel)
- Adopted in DVB, ATSC and ISBN digital TV solutions; adopted also for DVD storage
- Hundreds of millions of MPEG-2 set tops boxes in use in the US, Europe and Japan




# 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
- 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




# Video Coding in MPEG-4

- Two Parts in the MPEG-4 standard deal 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 video coding tools; this Part has been jointly developed by ISO/IEC MPEG and ITU-T through the Joint Video Team (JVT): H.264/MPEG-4 AVC.

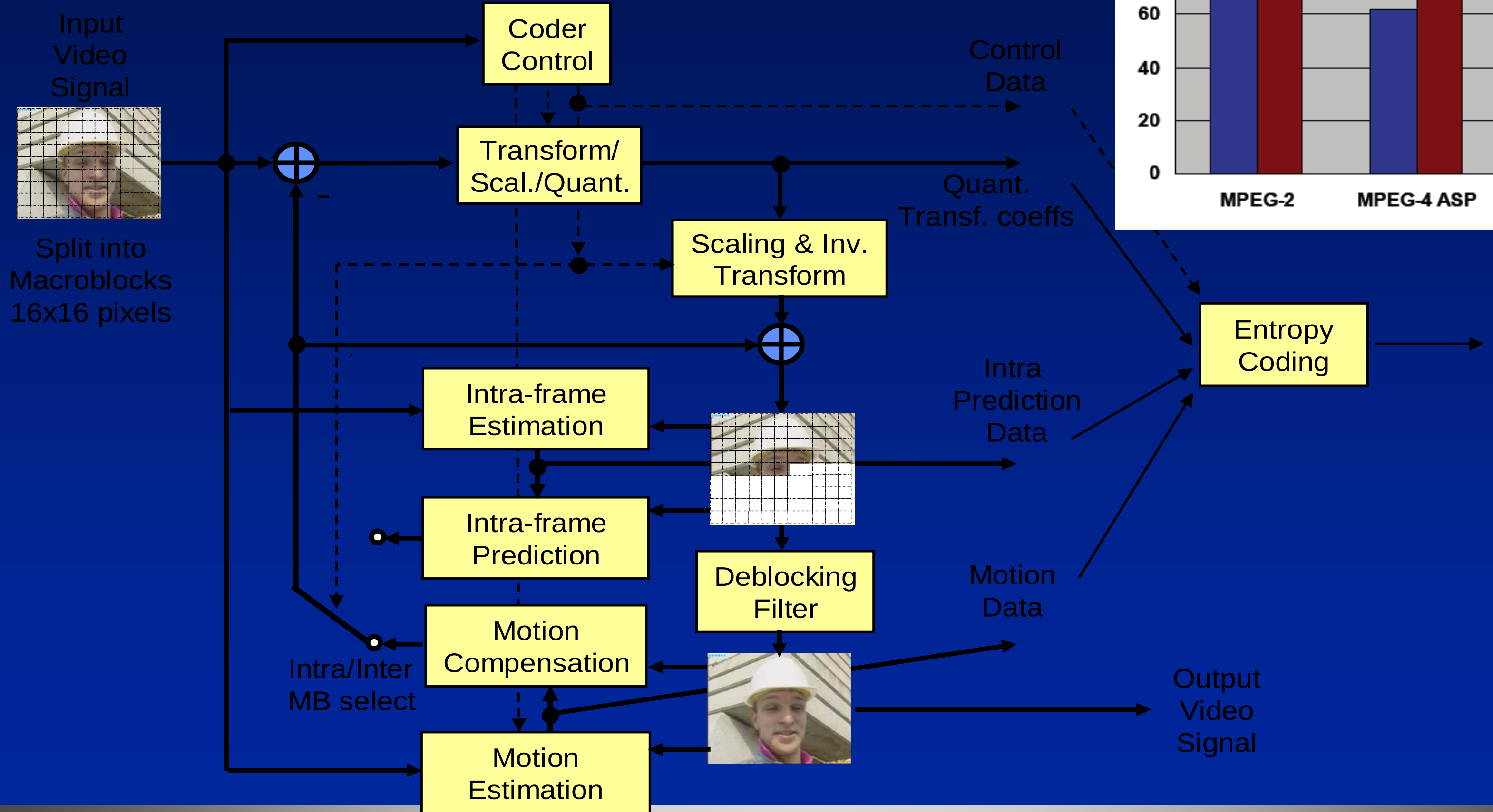
- While Part 2 addresses object and frame-based coding, Part 10 only addresses rectangular frames.

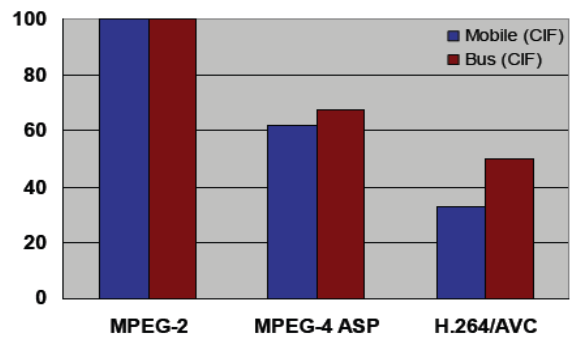
# H.264/AVC: Main Functionalities



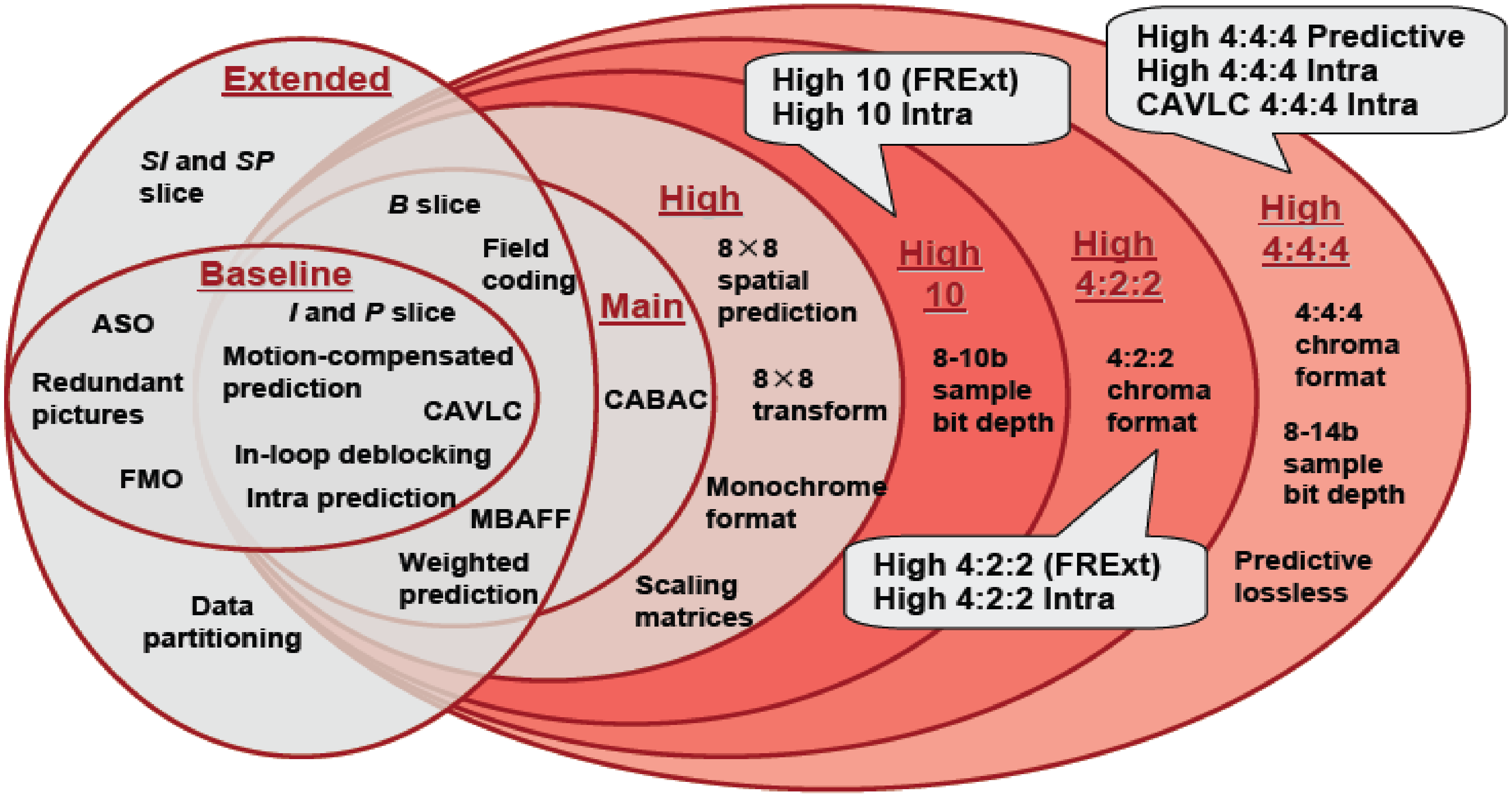
- **Coding efficiency (about 50% compared to previous standards)**
  - Enhanced motion compensation, multiple reference frames, hierarchical B frames, improved de-blocking filter, smaller blocks for transform, enhanced entropy coding
- **Network friendliness**
  - Enabled through the Network Adaptation Layer (NAL) design
- **Error resilience**
  - Data partitioning, flexible macroblock ordering (FMO), slices, redundant slices, resynchronization markers, multiple reference pictures, parameter sets, etc.
- **Temporal scalability**
  - Enabled by flexible reference picture management via hierarchical predictions

# Basic Coding Architecture



# 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
- 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)
- US DoD MISB (US government preferred codec up to 1080p)
- (and, of course, MPEG and the ITU-T)

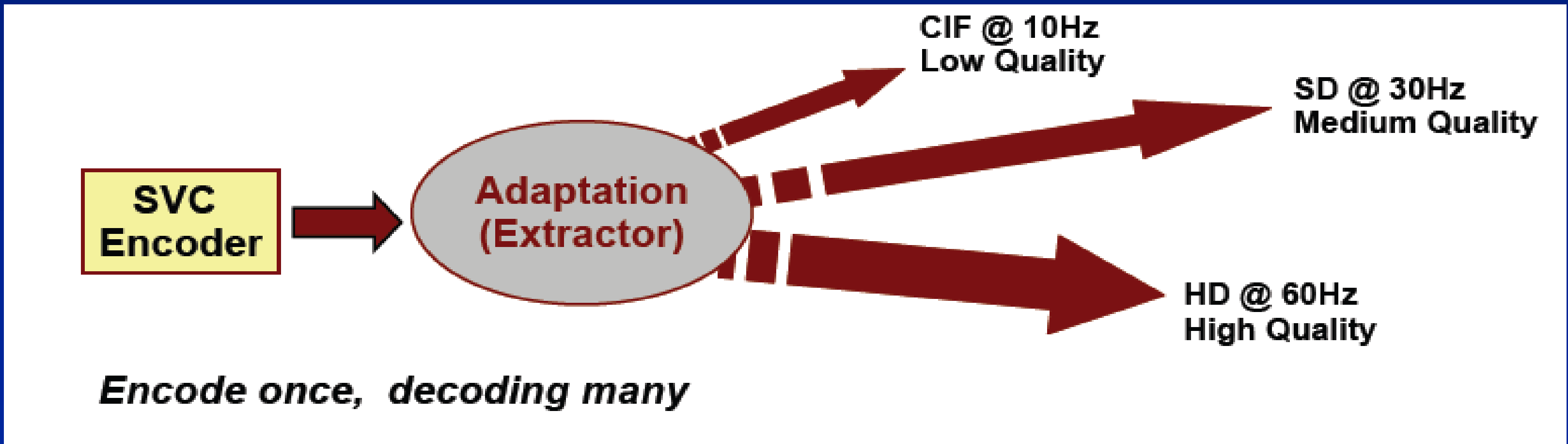
- Every 4 years, the video compression factors have been roughly doubled
  - H.263/MPEG-4 SP as good as H.261, at half the bitrate
  - H.264/MPEG-4 Part10 as good as MPEG-2/MPEG-4 ASP, at half the bitrate
  - …
- How long before H.264 is replaced by an alternative twice better ?

# Scalable Video Coding: Objectives



Scalability is a functionality regarding the useful 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.



# SVC Applications

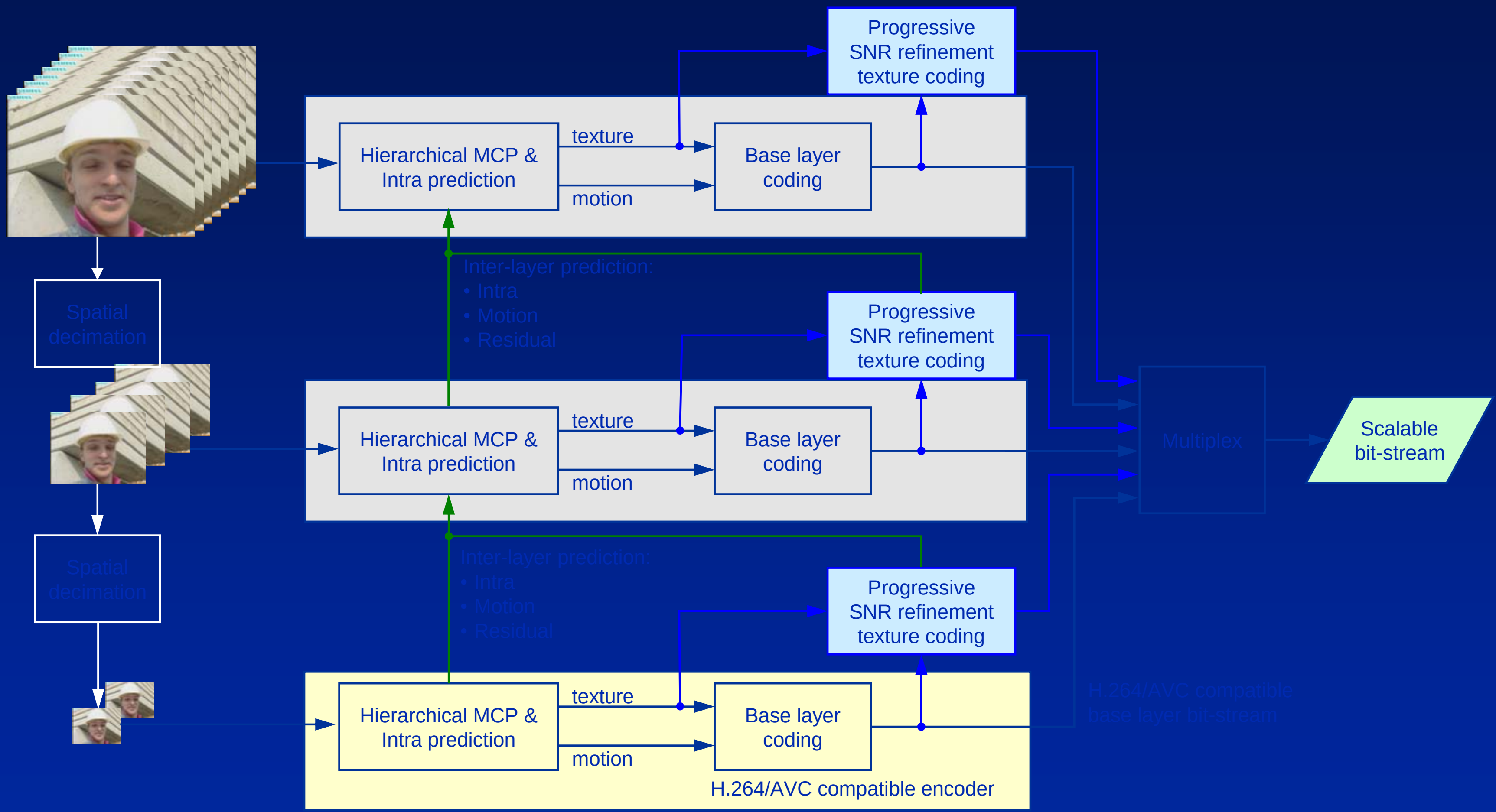


- Robust Video Delivery
  - Adaptive delivery over error-prone networks and to devices with varying capability
  - Combined with unequal error protection
  - Internet/mobile transmission
- Scalable Storage
  - Scalable export of video content
  - Graceful expiration or deletion
  - Surveillance DVR’s and Home PVR’s
- Enhancement Services
  - Upgrade delivery from 1080i/720p to 1080p
  - DTV broadcasting, optical storage devices

# SVC Coding Architecture

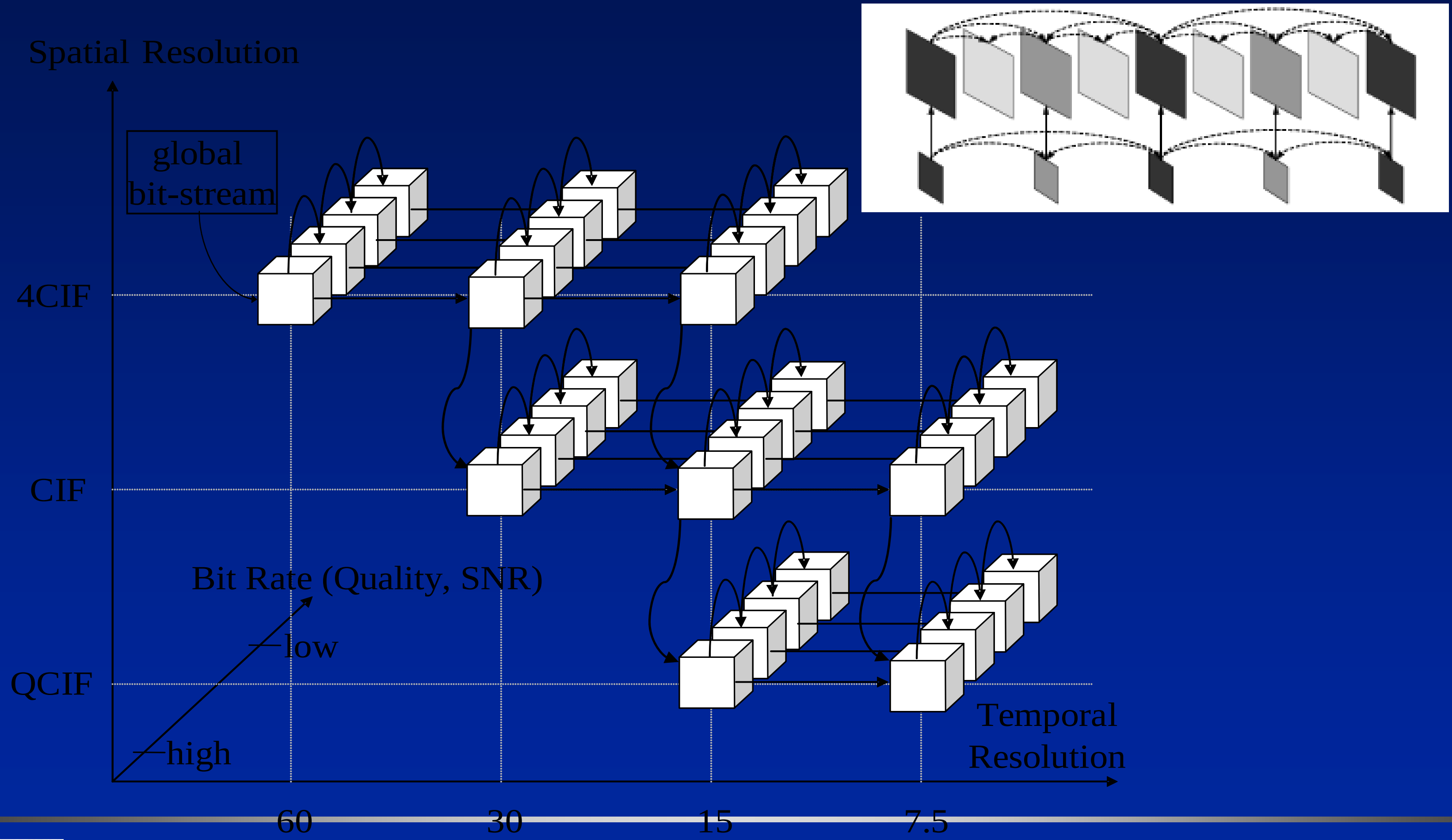
**2007 !**





# Spatio-Temporal-Quality Cube





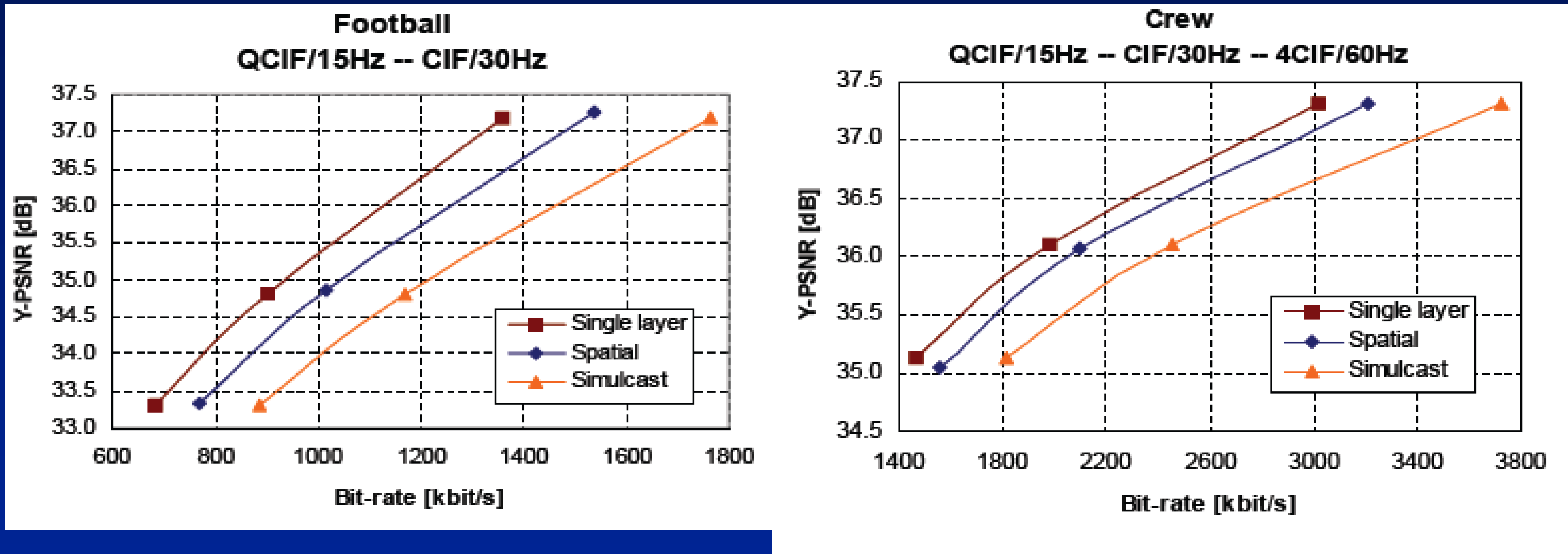
## SVC Scalability Types



- Temporal scalability - Can be typically achieved without losses in rate-distortion performance.
- Spatial scalability - When applying an optimized SVC encoder control, the bit rate 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 become worse as spatial resolution of both layers decreases and results improve as spatial resolution increases.
- SNR/Quality scalability - When applying an optimized encoder control, the bit rate 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 bit rate range with a factor of 2-3 between the lowest and highest supported rate point.

# SVC Performance: Spatial Scalability






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

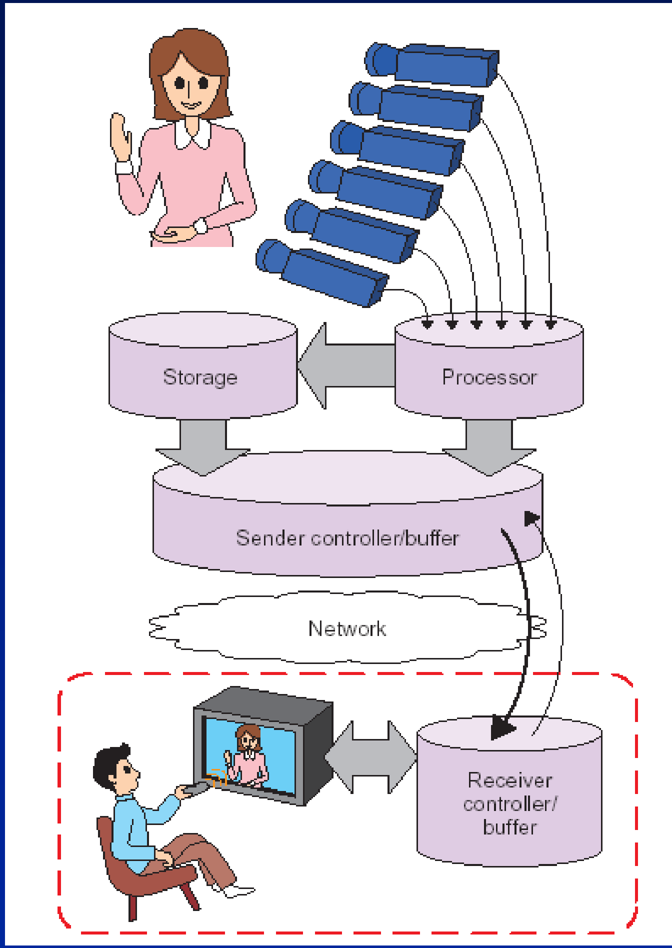
# 3D Worlds



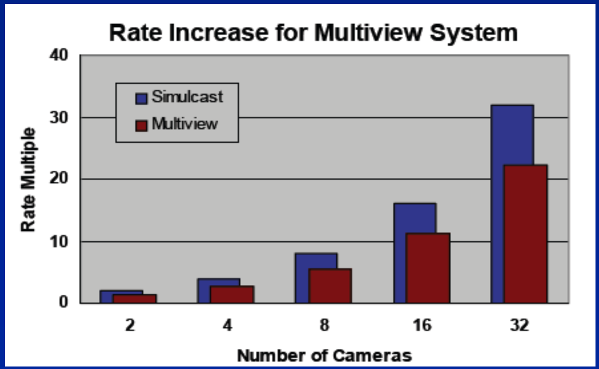
- 3D experiences may be provided through multi-view video, notably
  - 3D video (also called stereo) which brings a depth impression of a scene
  - Free viewpoint video (FVV) which allows an interactive selection of the viewpoint and direction within certain ranges.
- May require especial 3D display 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

# Multiview Video Coding (MVC)





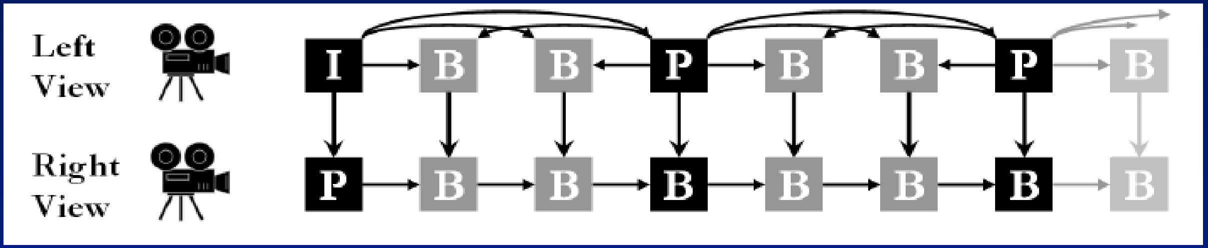

- In addition to exploiting the temporal and spatial redundancy within each view to achieve coding gains, redundancy can also be exploited across the different views.
- JVT standardization goal is to reach 50% bitrate savings over independent coding of views with same quality by defining another extension to the H.264/AVC standard.

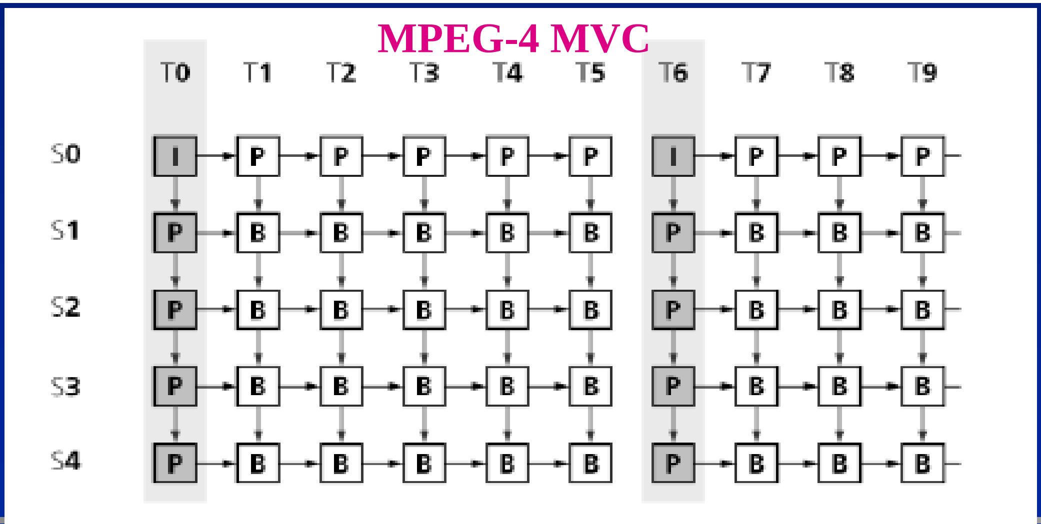


# MVC Prediction Structures

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

MPEG-2 Video Multiview profile



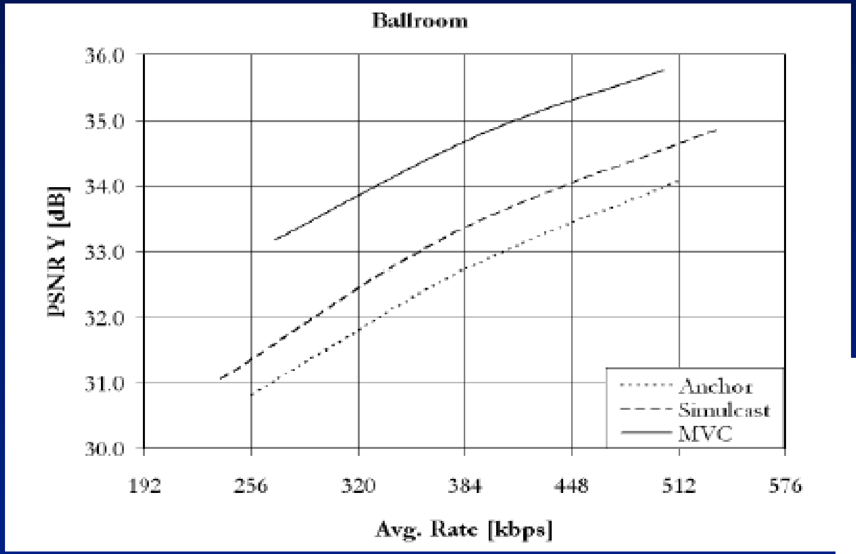
# MVC: Technical Solution

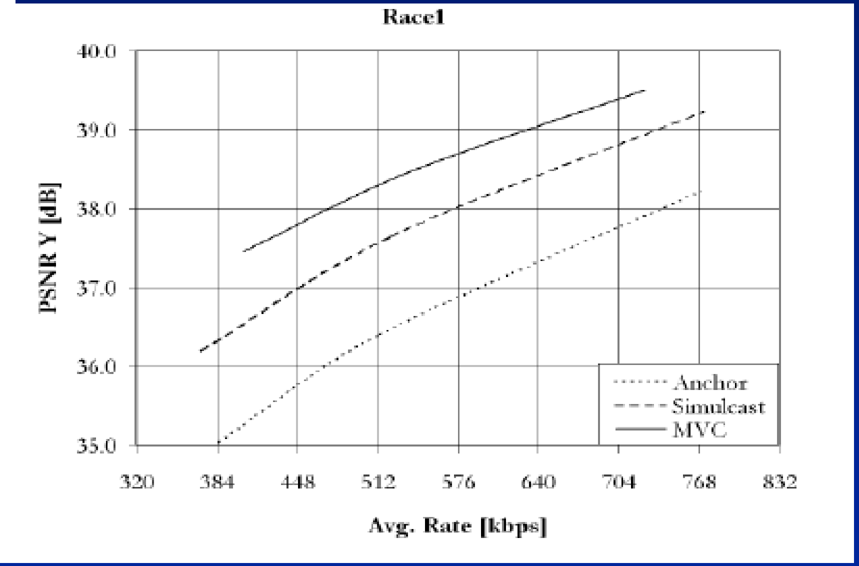


- Current H.264/AVC multiview extension does not require any changes to lower-level syntax
  - Very compatible with single-layer AVC hardware
- Inter-view prediction
  - Enabled through flexible design of decoded reference picture management
  - Allow decoded pictures from other views to be inserted and removed from reference picture buffer
- Small changes to high-level syntax
  - E.g., specify view dependency

# Some MVC Performance Results



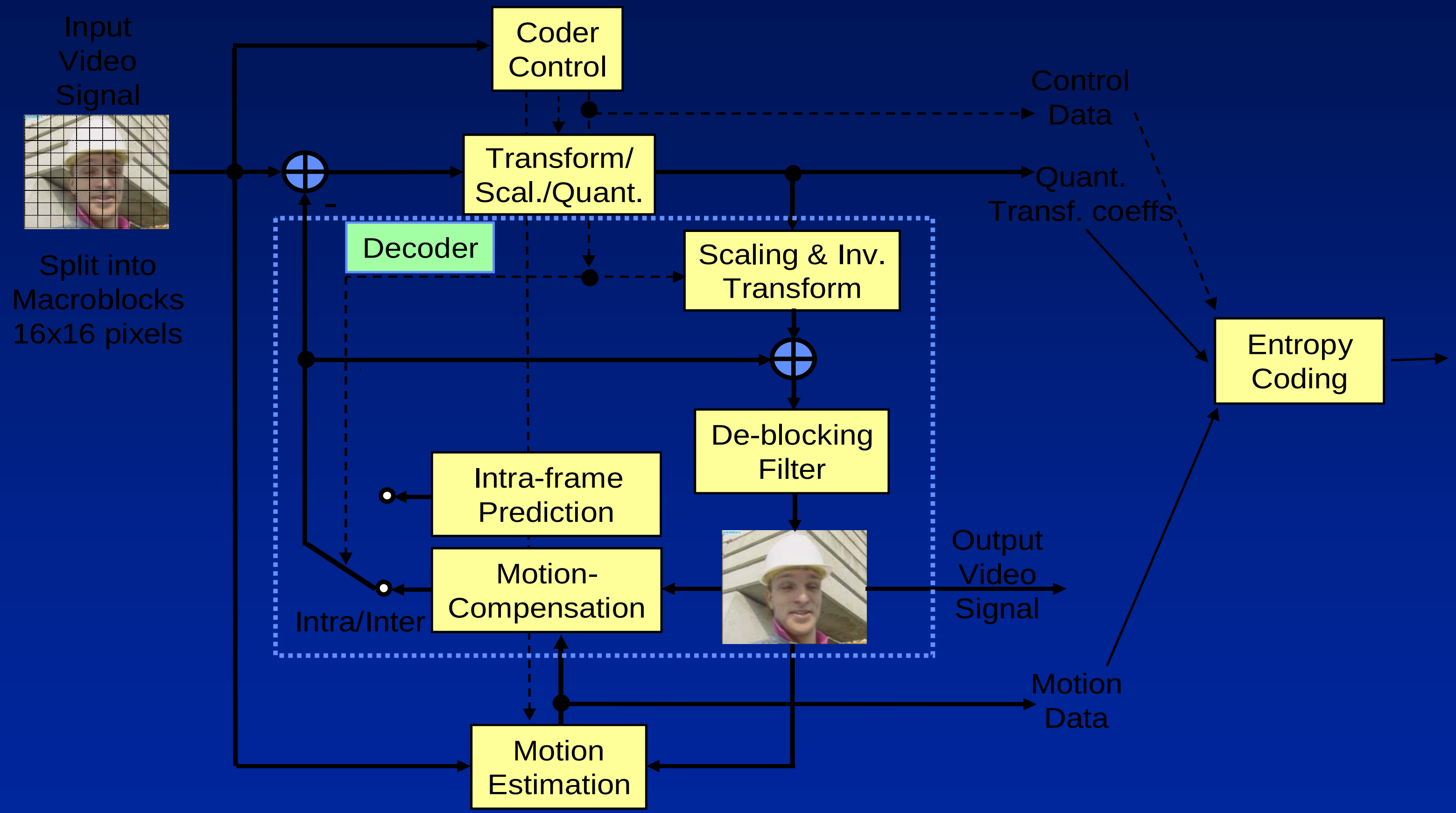



**Anchor is H.264/AVC without hierarchical B pictures; however, Simulcast already includes hierarchical B pictures.**

# State-of-the-Art: H.264/AVC Encoding Architecture

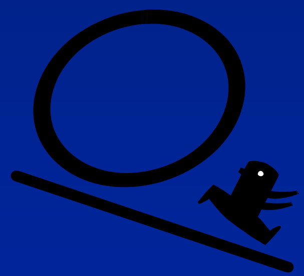




# The Current Video Coding Paradigm



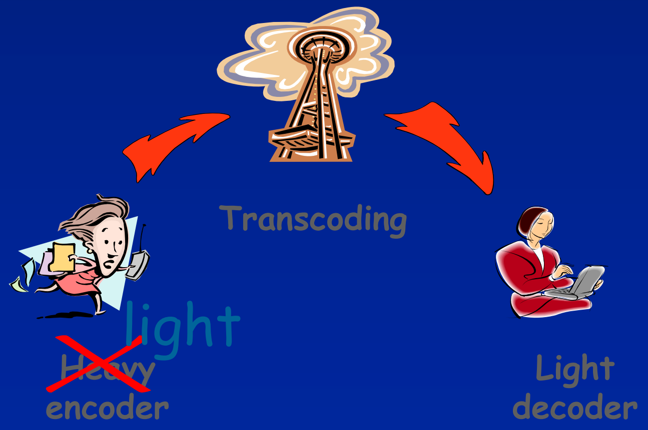
- Coding efficiency: a smart-dummy relation
- Complexity budget: a rigid master-slave relation
- Error robustness: the snow ball effect
- Fitting better one-to-many topologies



# Emerging Challenges

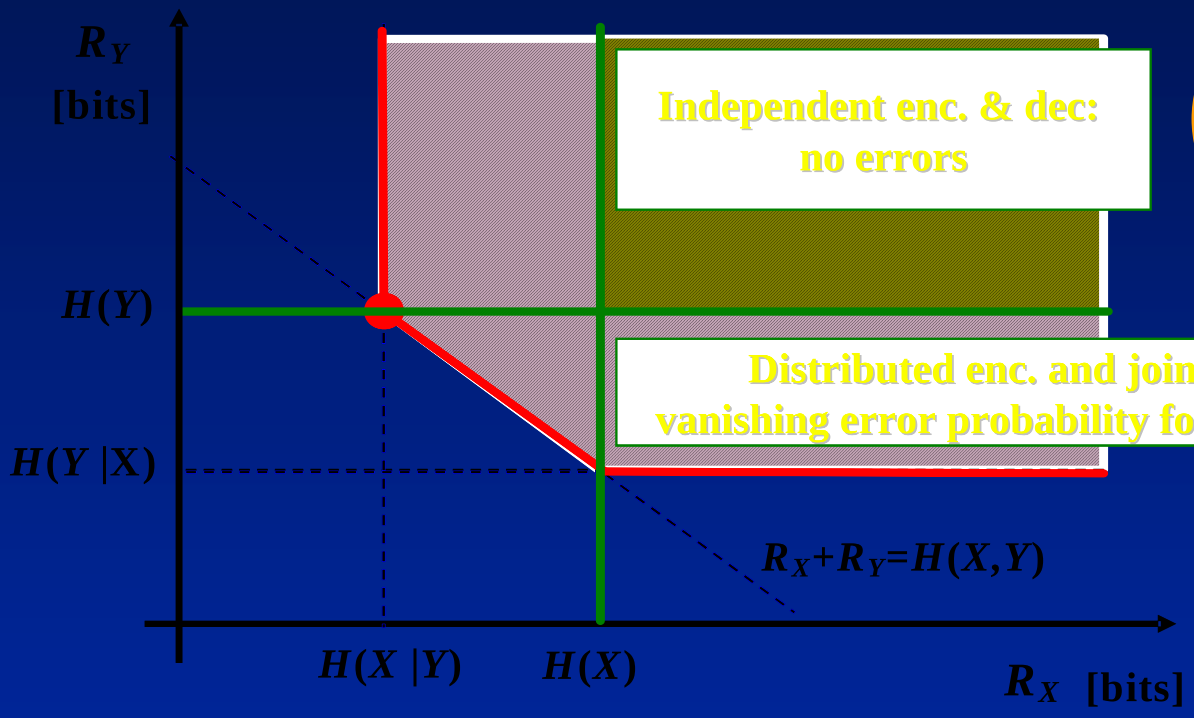
- Applications (from down-link to up-link)
  - Wireless digital video cameras
  - Multimedia mobile phones and PDAs
  - Low-power video sensors and surveillance cameras
  - Wireless video teleconferencing systems

- Requirements
  - Light and flexible distribution codec complexity
  - Robustness to packet/frame losses
  - High compression efficiency
  - Low latency
- Target
  - Inter coding efficiency
  - Intra coding complexity (encoder)
  - Intra coding robustness



# Distributed Coding of Dependent Sources: The Slepian-Wolf Theorem

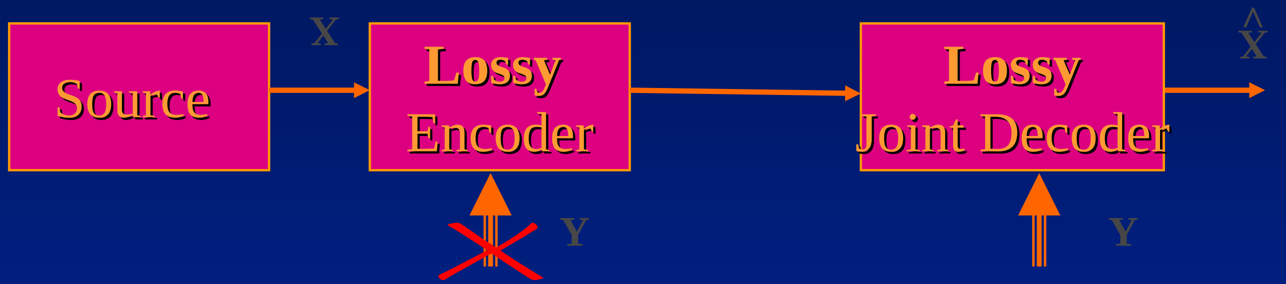





$$R_X \geq H(X \mid Y)$$
$$R_Y \geq H(Y \mid X)$$
$$(R_X + R_Y) \geq H(X,Y)$$

Red zone: Combinations of $R_X$ and $R_Y$ needed to reconstruct $X$ and $Y$, with an arbitrarily small probability of error

# Lossy Coding with Side Information: The Wyner-Ziv Theorem

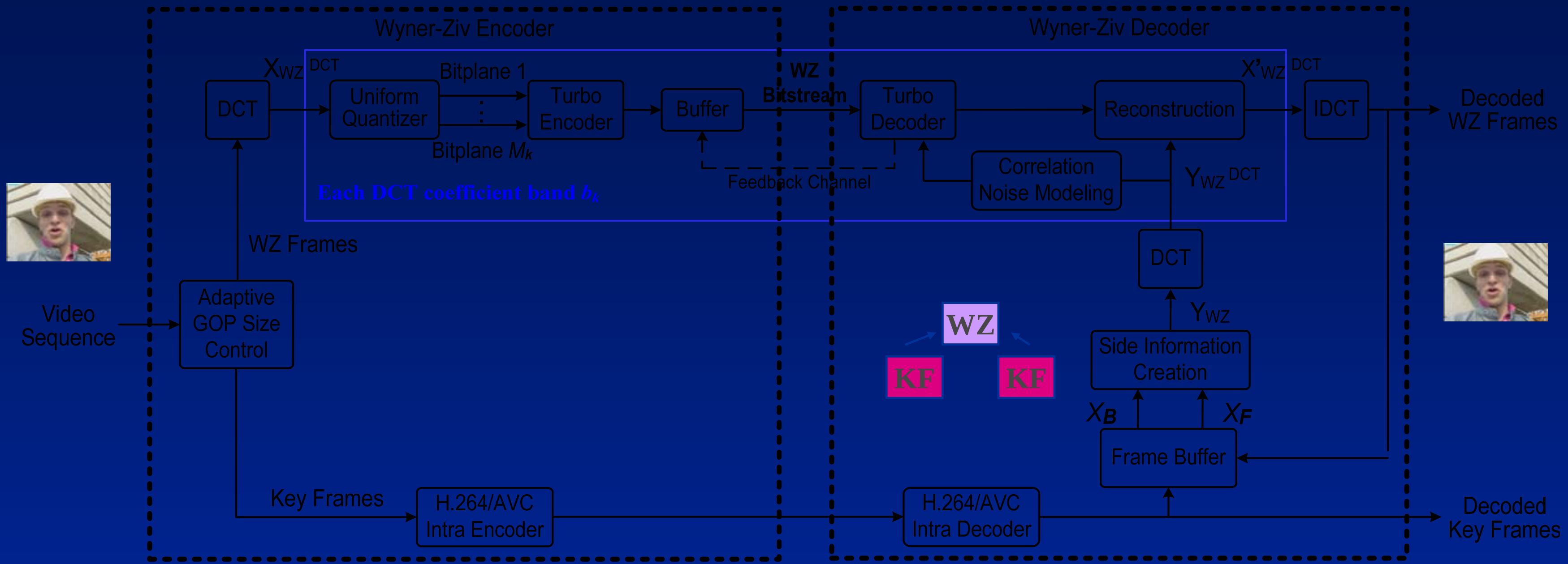



The Wyner-Ziv Theorem states that when X and Y are only jointly decoded, there is no coding efficiency loss with respect to the case when they are jointly encoded and decoded, if X and Y are jointly Gaussian and a mean squared error distortion measure is considered.

Later, it was proved that only the 'innovation', X–Y, needs to be Gaussian.

# Feedback Channel Based DVC Architecture



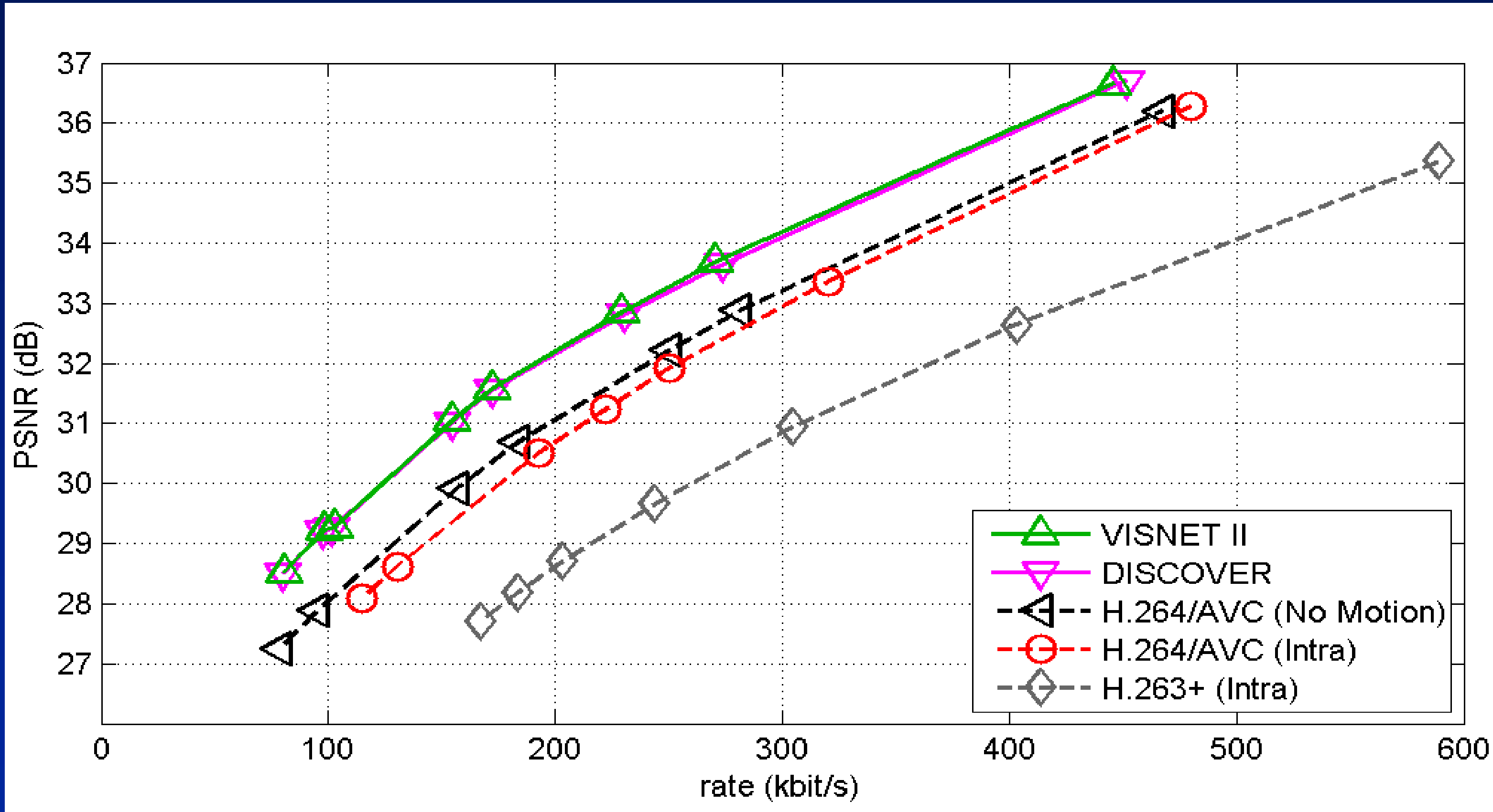



- Based on a split between Wyner-Ziv (WZ) and key frames.
- Key frames used with a regular (GOP size) or dynamic periodicity.
- Key frames coded with H.264/AVC Intra.
- Turbo codes or LDPC codes may be used.

# DVC RD Performance

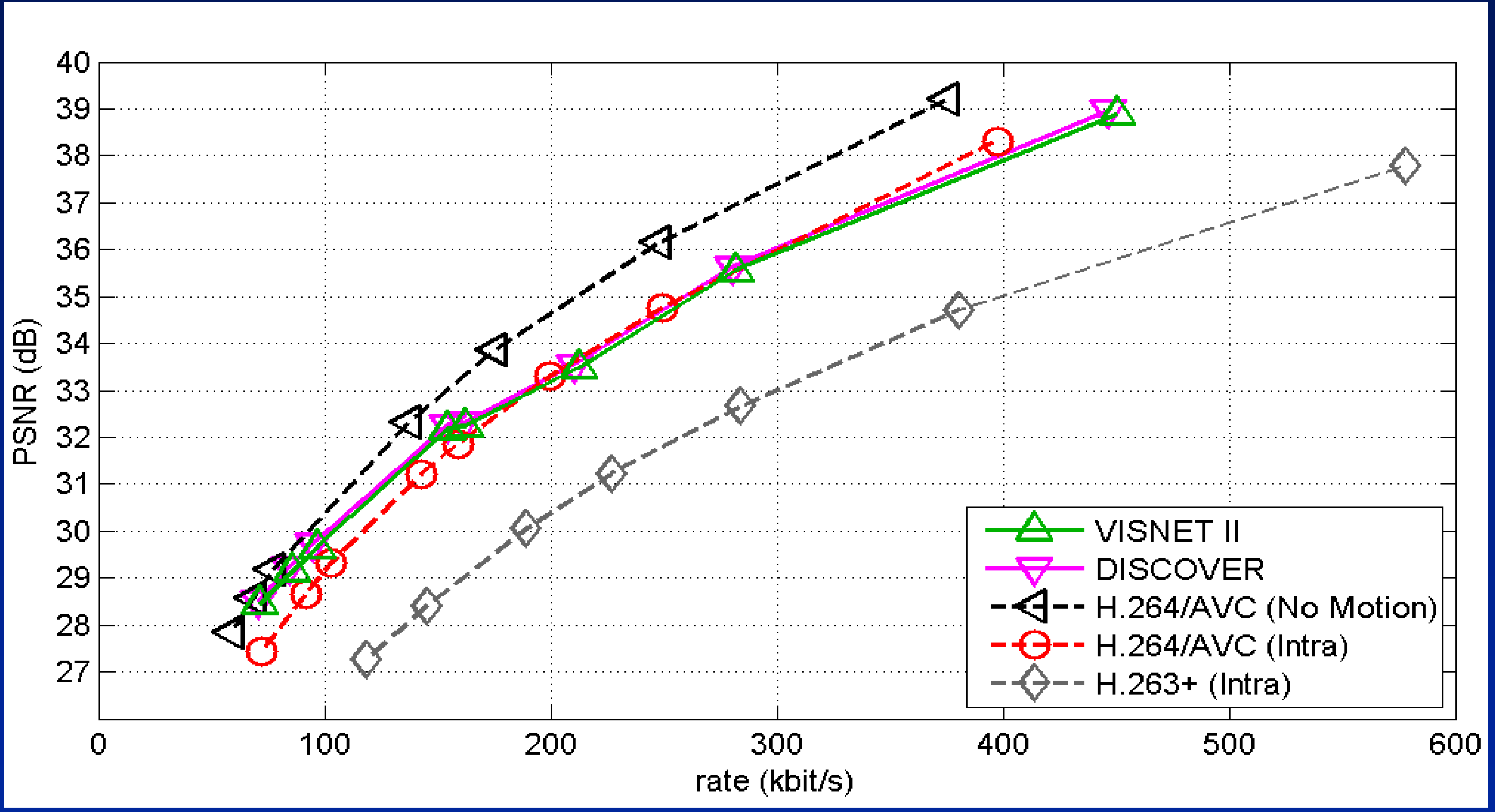


Coastguard QCIF@15Hz, GOP2



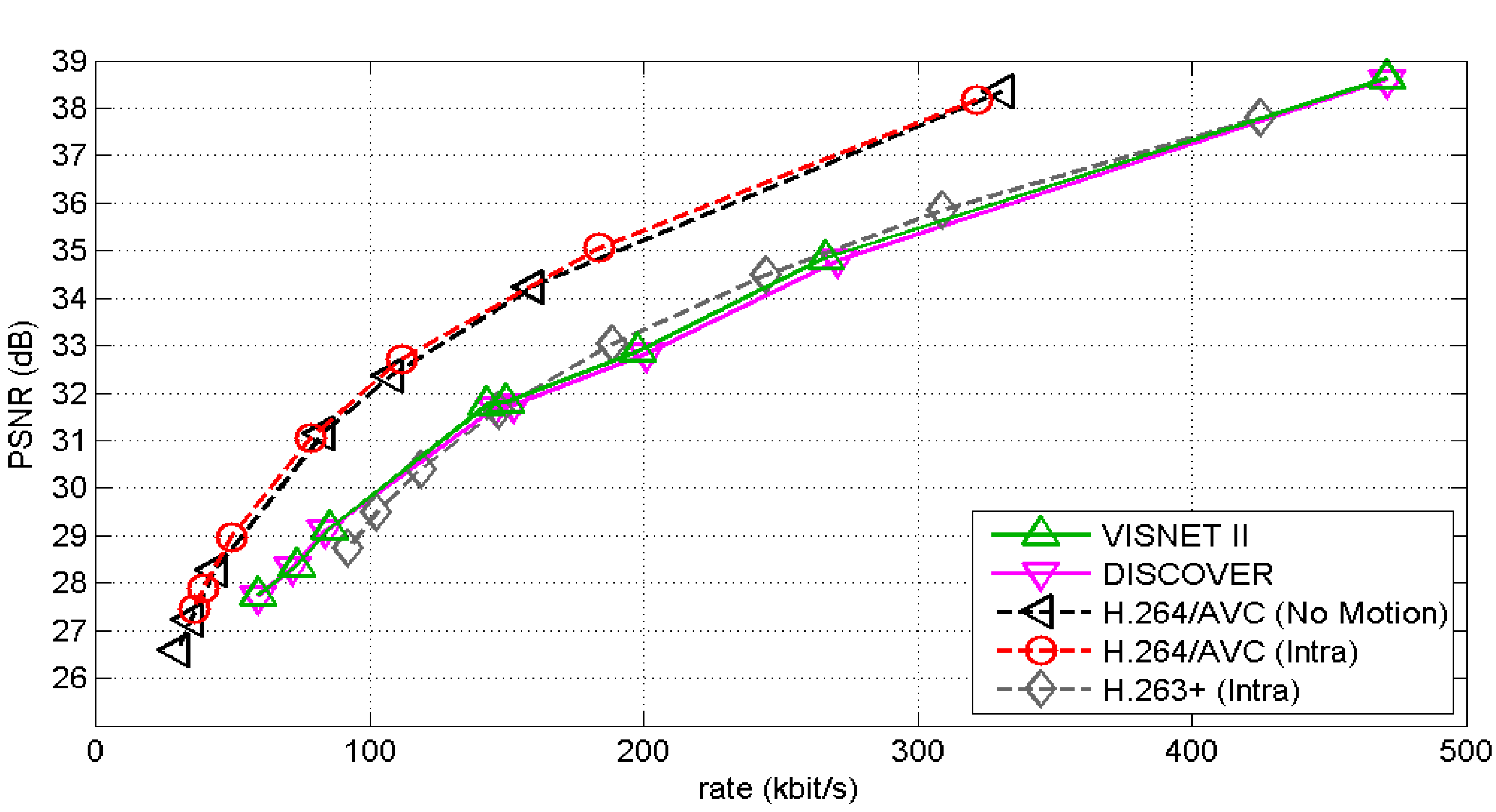
# VISNET II DVC RD Performance

Foreman QCIF@15Hz, GOP2



# VISNET II DVC RD Performance

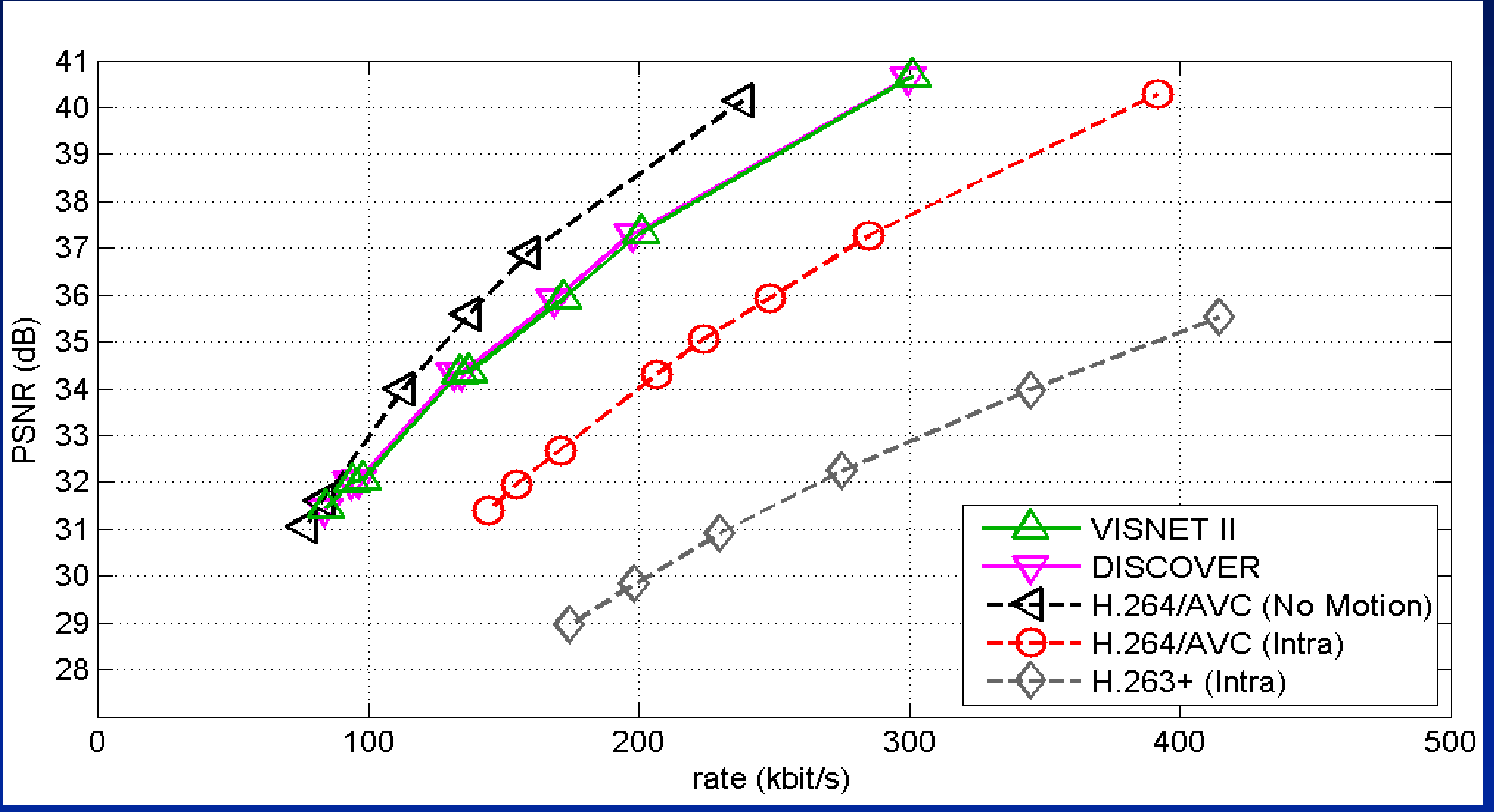
Soccer QCIF@15Hz, GOP2



# VISNET II DVC RD Performance



Hall&Monitor QCIF@15Hz, GOP2



# Looking over the Horizon …

- From middle 80s until recently, major video compression gains have been obtained in an almost continuous way: about 50% gains every 5 years.
- Current hybrid coding schemes seem to be close to saturation in terms of RD performance ... Other approaches may follow …
- Besides compression efficiency, there are other important functionalities and scenarios that still need much attention, notably scalability, multiview, etc.
- DVC is a new interesting way to look to video coding … still not clear what 'real' will come out of it …
- New breakthroughs in video coding are not emerging … is there more to learn from the Human Visual System (HVS) ?

# ‘rugby’ 4.7 Mbit/s AVC



# ‘rugby’ 4.7 Mbit/s  AVC + texture synthesis



# Quality of Experience



- What matters is the impact of the content on the end-user!

- Metrics to estimate the perceived value and impact of the content on the end-user, encompassing:
  - Inherent quality of the signal
  - Quality of interaction
  - Multimodal nature of the quality
  - Context
  - Environment (device, surrounding, etc.)
  - …

# Quality of Experience



- Quality of Service: Estimated value by provider about user's experience
- Quality of Experience: Value (estimated or measured) of the actual user's experience

- Quality of Experience is the dual (and extended) view of QoS problem
  - QoS=provider-centric
  - QoE=user-centric

# Challenges in visual quality assessment



- Primary focus:
  - Reliable and widely accepted perceived quality metrics for pictures
  - Reliable perceived quality metrics for video
  - Reliable perceived quality metrics without reference or with reduced reference

- Should be extended to:
  - Stereoscopic and multi-view quality metrics
  - 3D quality metrics
  - Multi-modal quality metrics
  - Segmentation and tracking quality metrics
  - …