



International Telecommunication Union

Overview of International Video Coding Standards (preceding H.264/AVC)

Gary J. Sullivan, Ph.D.

ITU-T VCEG Rapporteur | Chair
ISO/IEC MPEG Video Rapporteur | Co-Chair
ITU/ISO/IEC JVT Rapporteur | Co-Chair

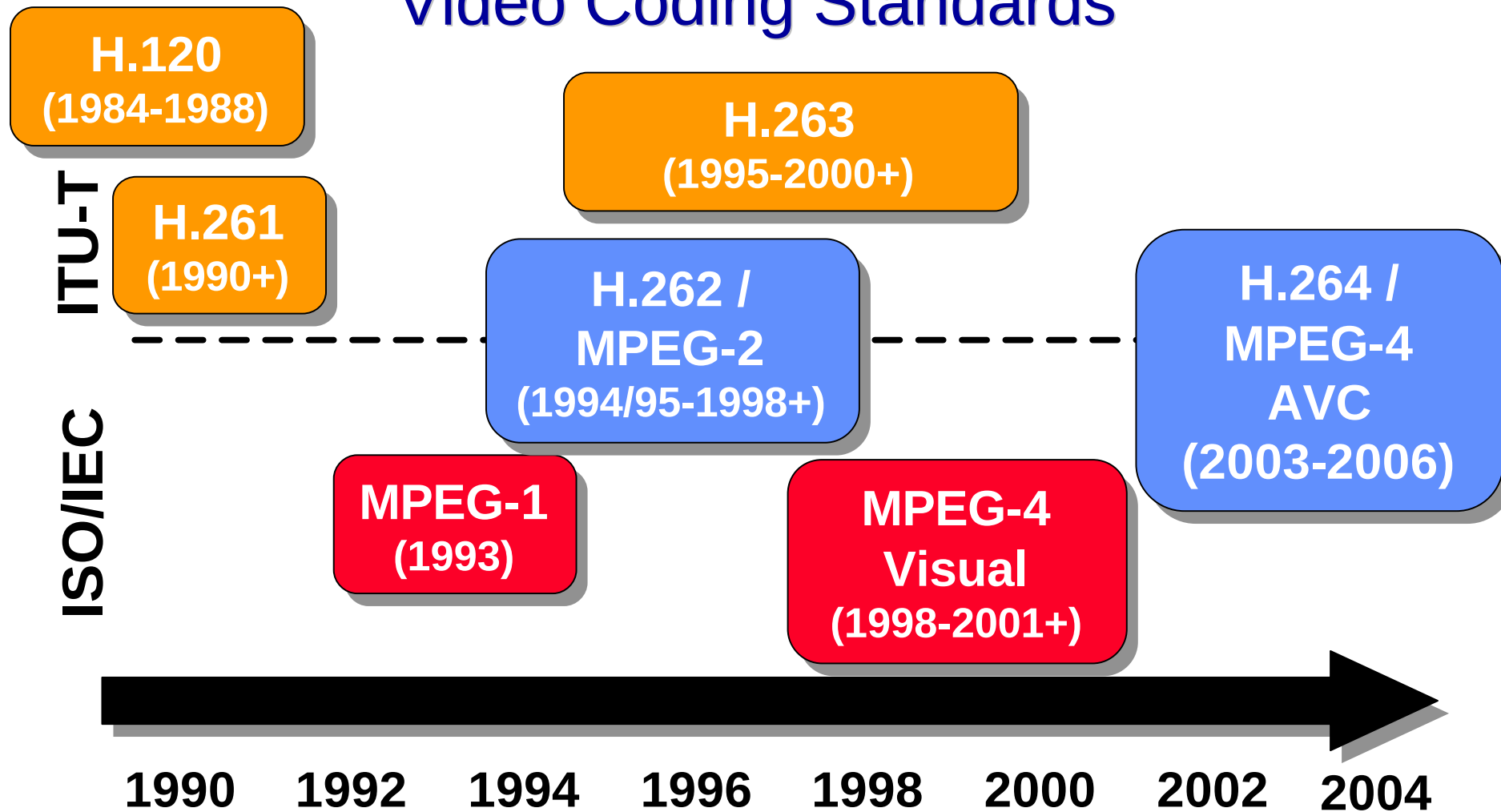
July 2005

**ITU-T VICA Workshop
22-23 July 2005, ITU Headquarter, Geneva**

Video Coding Standardization Organizations

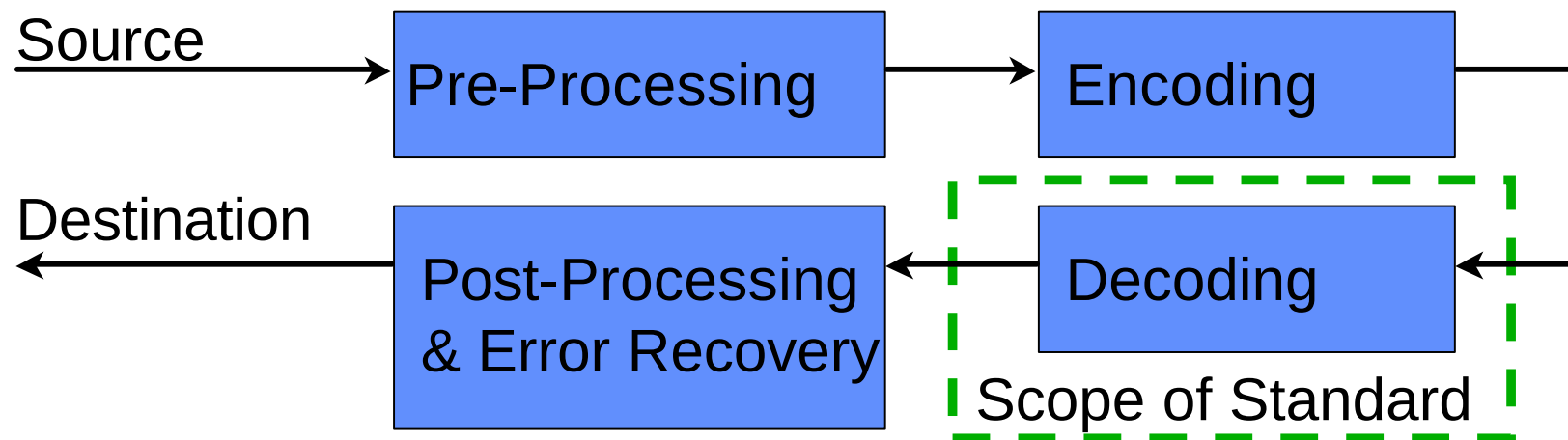
- Two organizations have dominated video compression standardization:
 - ITU-T Video Coding Experts Group (VCEG)
International Telecommunications Union –
Telecommunications Standardization Sector (ITU-T,
a United Nations Organization, formerly CCITT),
Study Group 16, Question 6
 - ISO/IEC Moving Picture Experts Group (MPEG)
International Standardization Organization and
International Electrotechnical Commission, Joint
Technical Committee Number 1, Subcommittee 29,
Working Group 11

Chronology of International Video Coding Standards



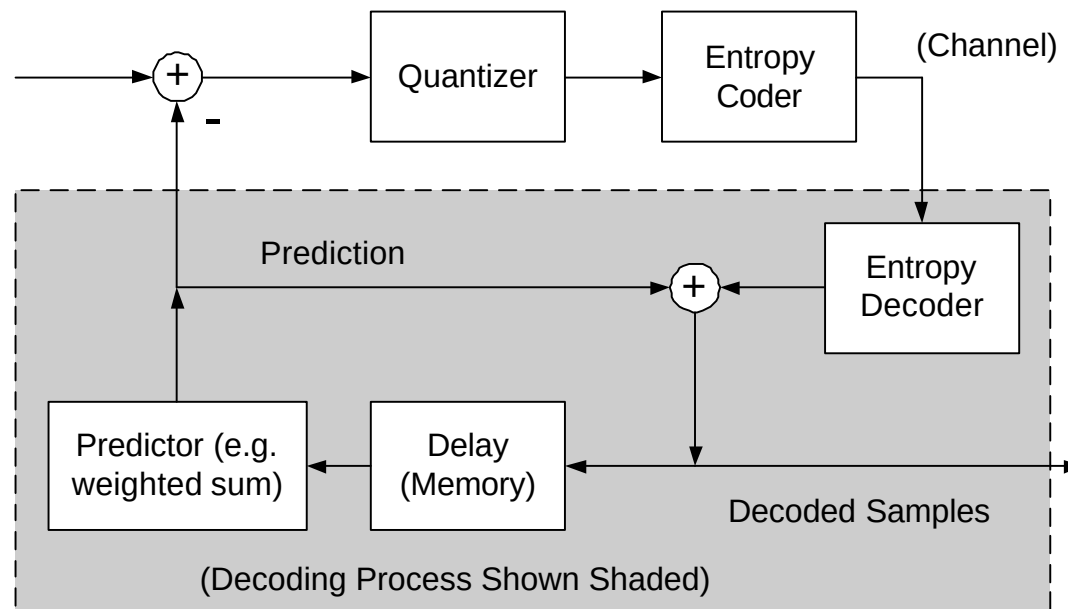
The Scope of Picture and Video Coding Standardization

- Only the *Syntax* and *Decoder* are standardized:
 - Permits optimization beyond the obvious
 - Permits complexity reduction for implementability
 - Provides *no* guarantees of Quality



Predictive Coding and DPCM

- Separate quantization of each sample is known as pulse-code modulation (PCM)
- *Predictive Coding or Differential PCM:* Generate an estimate for the value of the input data, and encode only the remaining difference.



H.120 : The First Digital Video Coding Standard

- ITU-T (ex-CCITT) Rec. H.120: The first digital video coding standard (1984)
 - v1 (1984) had conditional replenishment, DPCM, scalar quantization, variable-length coding, switch for quincunx sampling
 - v2 (1988) added motion compensation and background prediction
 - Operated at 1544 (NTSC) and 2048 (PAL) kbps
 - Few units made, essentially not in use today

“Intra” Picture Coding by DCT

Basic “intra” image representation: Discrete Cosine Transform (DCT) (early ‘70s, ITU+ISO JPEG approved ‘92):

- Analyze 8x8 blocks of image according to DCT frequency content (images tend to be smooth)
- Find magnitude of each discrete frequency within the block
- Round off (“quantize”) the amounts to scaled integer values (‘50s, ‘60s, ...)
- Send integer approximations to decoder using “Huffman” variable-length codes (VLC, early ‘50s)

The Discrete Cosine Transform

- The DCT (unitary type II DCT):

$$F_{m,n}(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \left(c_u \sqrt{\frac{2}{M}} \right) \left(c_v \sqrt{\frac{2}{N}} \right) f(mM + x, nN + y) \cdot \cos \left[\frac{(2x+1)up}{2M} \right] \cdot \cos \left[\frac{(2y+1)vp}{2N} \right]$$

- The Inverse DCT (unitary type III DCT):

$$\hat{f}(mM + x, nN + y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \left(c_u \sqrt{\frac{2}{M}} \right) \left(c_v \sqrt{\frac{2}{N}} \right) \hat{F}_{m,n}(u,v) \cdot \cos \left[\frac{(2x+1)up}{2M} \right] \cdot \cos \left[\frac{(2y+1)vp}{2N} \right]$$

- Definition of Constants

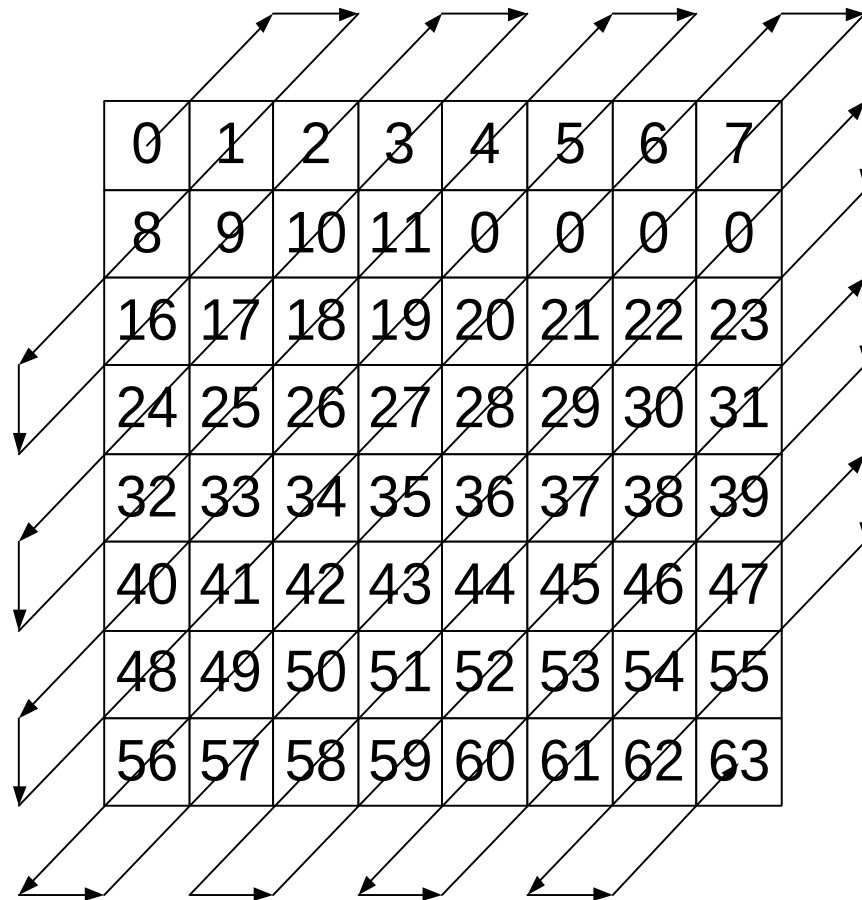
$c_u = 1 / \sqrt{2}$ for $u = 0$, otherwise 1.

$M = 8$ in current visual standards

$c_v = 1 / \sqrt{2}$ for $v = 0$, otherwise 1.

$N = 8$ in current visual standards

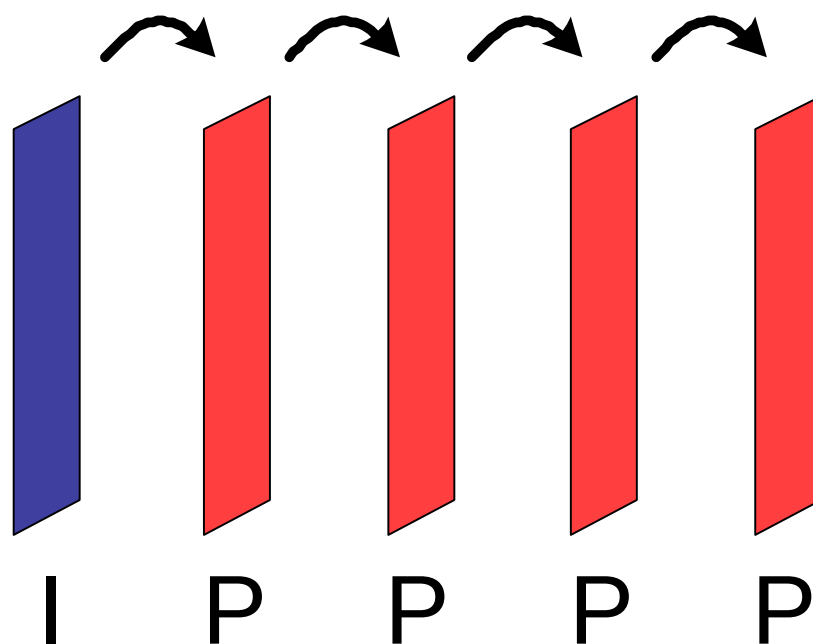
Coefficient Scan Order: The Zig-Zag Scan



Interframe Motion Prediction

- Large areas of images stay the same from frame to frame, changing mostly due to motion
- *Conditional Replenishment*: Can signal to leave a block area of the image unchanged, or replace it with new data
- *Interframe Difference Coding*: Could encode a refinement to the value of an area
- *Displaced Frame Difference Coding*: Can predict an image area by copying some nearby part of the previous image (motion compensation) and optionally adding some refinement

P-Picture Predictive Coding

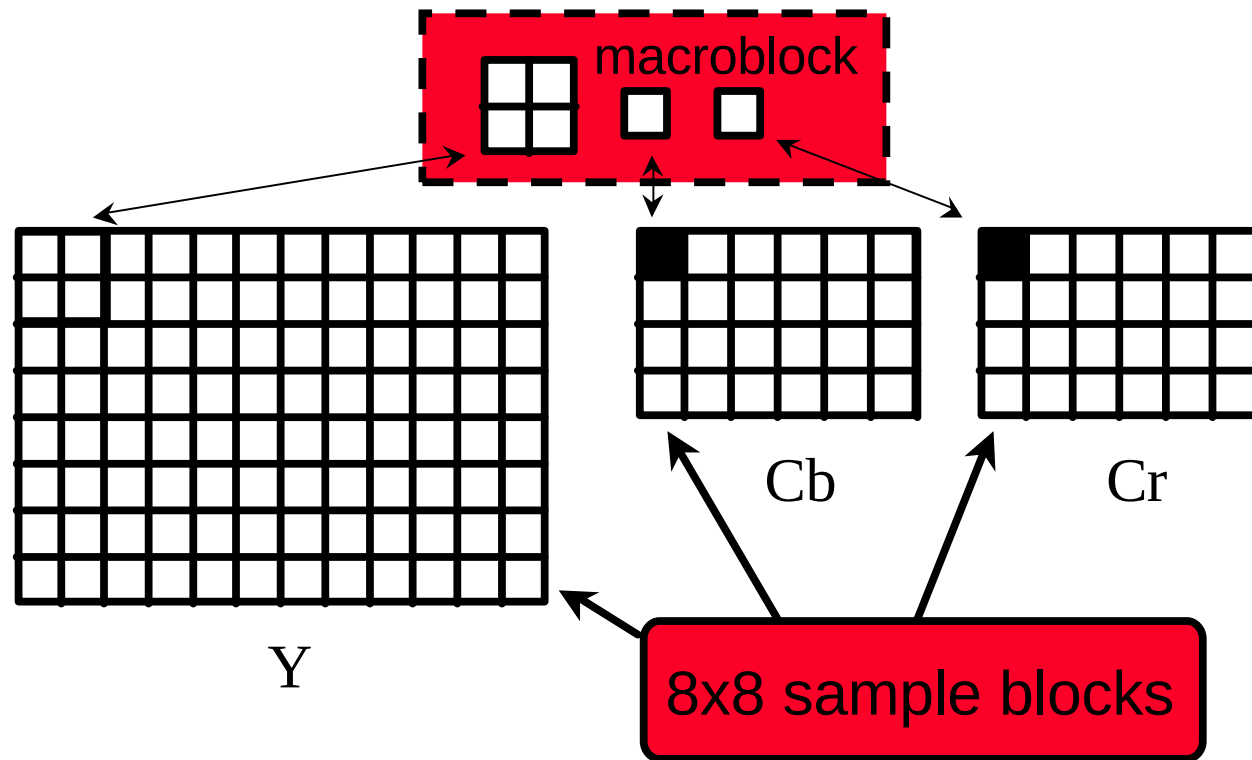


H.261: The Basis of Modern Video Compression

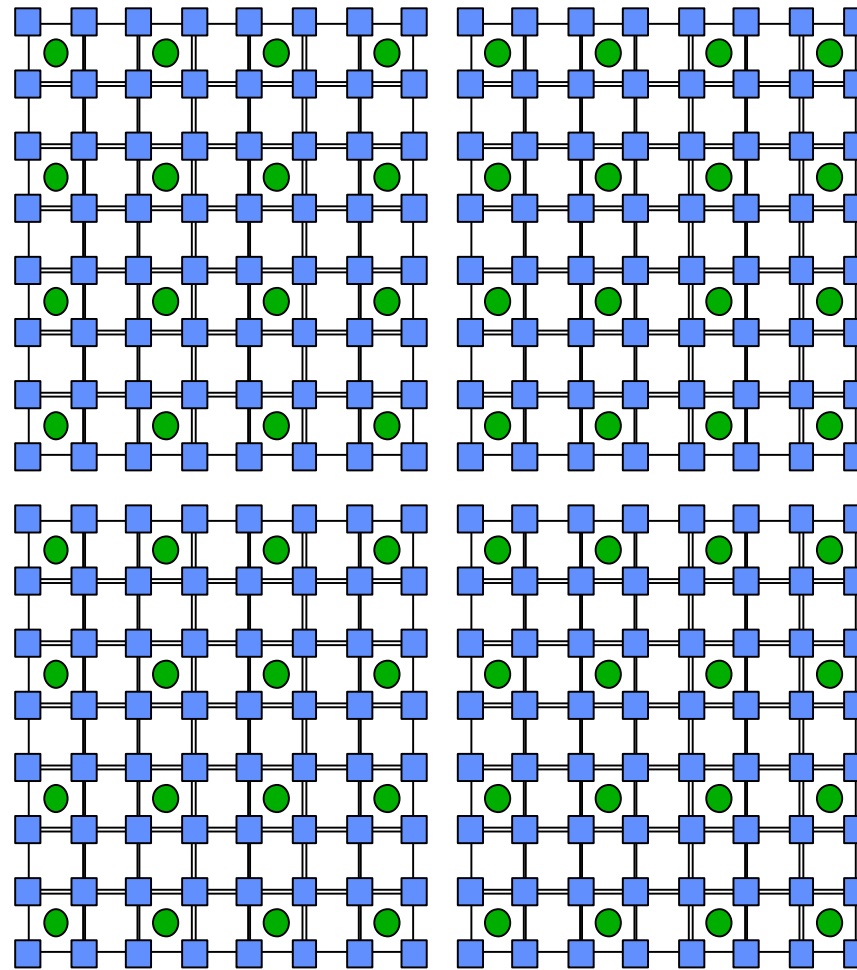
- ITU-T (ex-CCITT) Rec. H.261: The first widespread practical success
 - First design (late '90) embodying typical structure dominating today:
 - **16x16 macroblock motion compensation,**
 - **8x8 DCT,**
 - **scalar quantization,**
 - **zig-zag scan, and**
 - **run-length**
 - **variable-length coding**
 - Key aspects later dropped by other standards: loop filter, integer motion comp., 2-D VLC, header overhead
 - v2 (early '93) added a backward-compatible high-resolution graphics trick mode
 - Operated at 64-2048 kbps
 - Still in use, although mostly as a backward-compatibility feature – overtaken by H.263

Blocks and Macroblocks

The luma and chroma planes are divided into blocks. Luma blocks are associated with Cb and Cr blocks to create a macroblock.



H.261&3 Macroblock Structure



■ = luma pixel

● = chroma pixel

(two chroma fields)

Intra/Inter Decisions:

16x16 macroblock

DCT of 8x8 blocks

H.261:

16x16 1-pel motion

H.263:

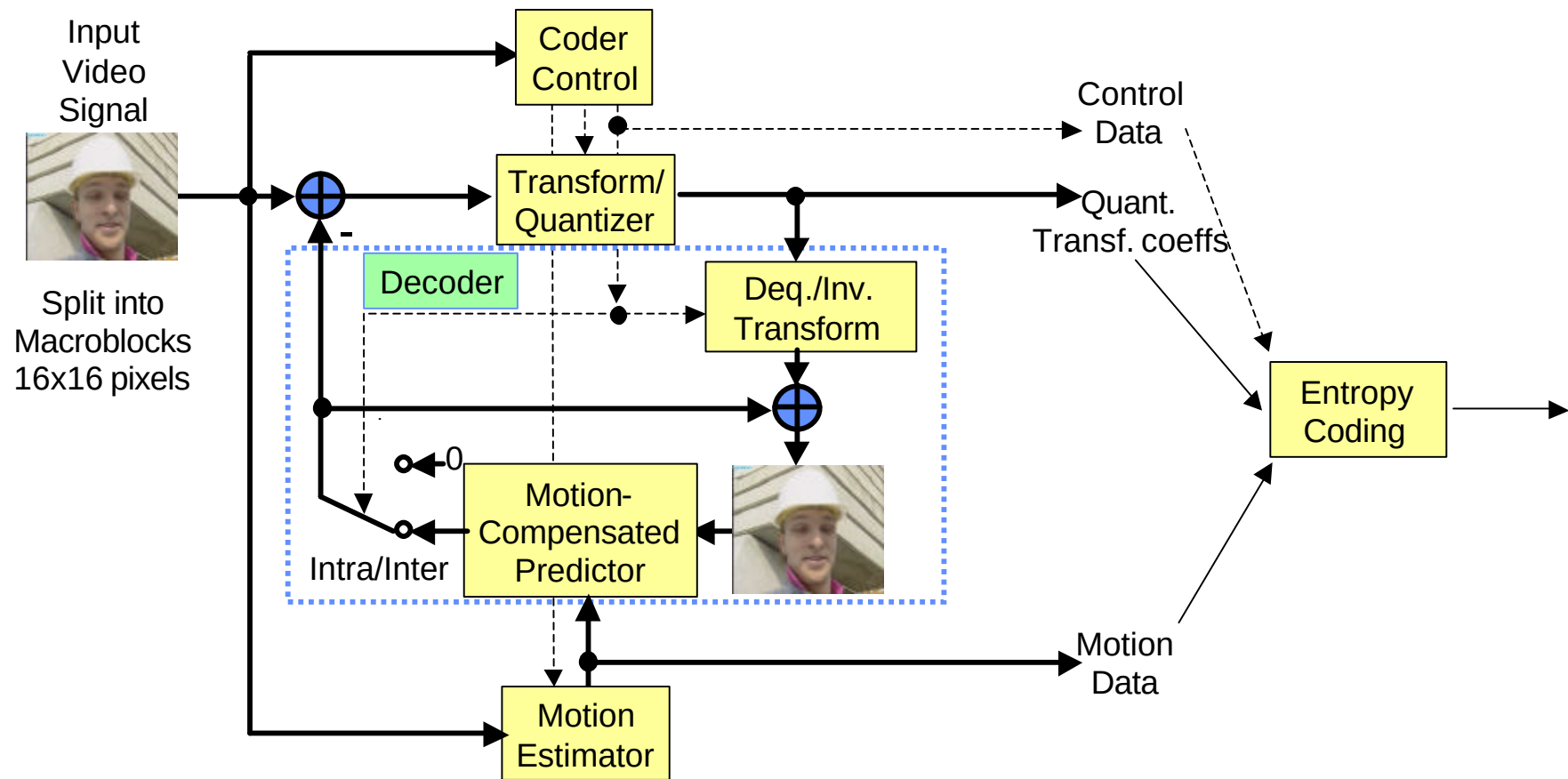
16x16 1/2-pel motion

or (AP mode)

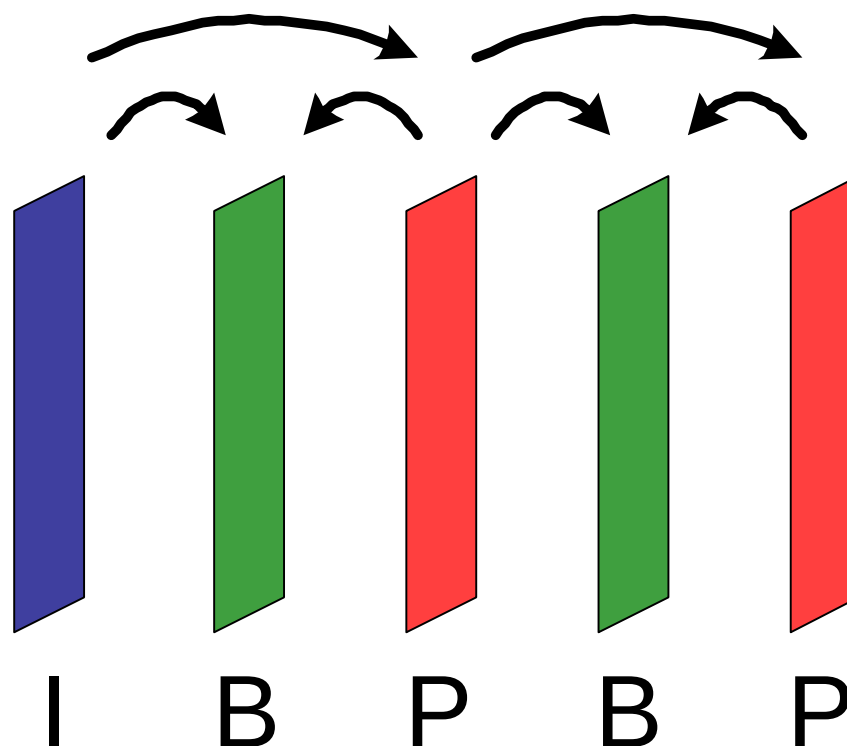
8x8 1/2-pel motion

with overlapping

Basic Hybrid Structure of H.261, etc. (late '90)



Predictive Coding with B Pictures



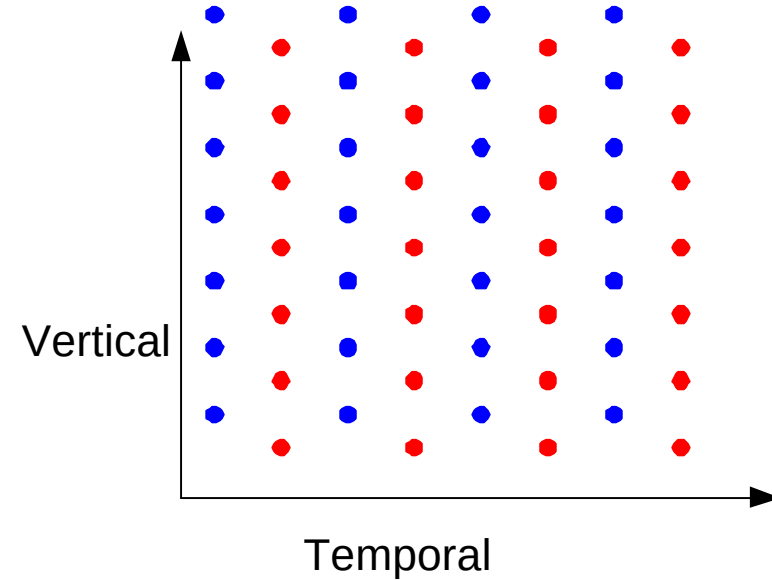
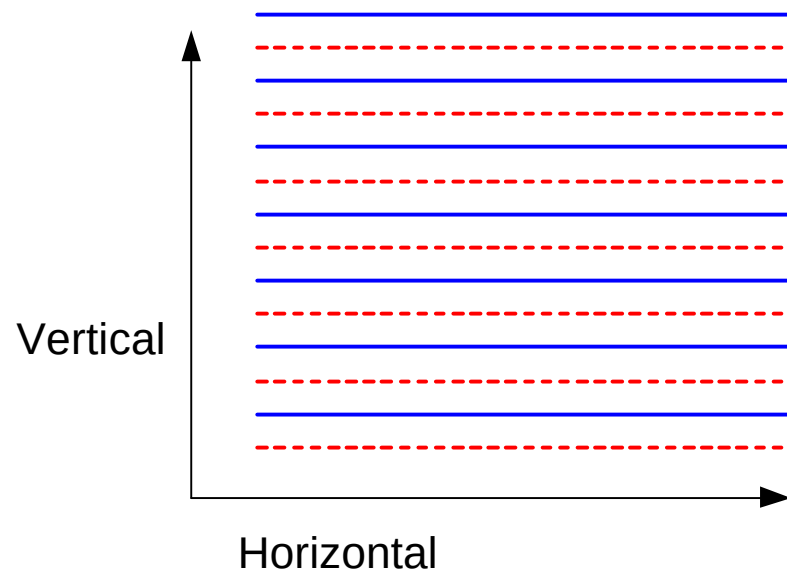
MPEG-1:

Practical Video at Higher Rates than H.261

- Formally ISO/IEC 11172-2 ('93), developed by ISO/IEC JTC1 SC29 WG11 (MPEG) – use is fairly widespread (esp. Video CD in Asia), but mostly overtaken by MPEG-2
 - Superior quality to H.261 when operated at higher bit rates (≥ 1 Mbps for CIF 352x288 resolution)
 - Can provide approximately VHS quality between 1-2 Mbps using SIF 352x240/288 resolution
 - Technical features inherited from H.261
 - **16x16 macroblocks**
 - **16x16 motion compensation,**
 - **8x8 DCT,**
 - **scalar quantization,**
 - **zig-zag scan, and**
 - **run-length**
 - **variable-length coding**
 - Technical features added:
 - **Bi-directional motion prediction**
 - **Half-pixel motion**
 - **Slice-structured coding**
 - **DC-only “D” pictures**
 - **Quantization weighting matrices**

Interlaced Video

(Welcome to the 1940 Analog World)



MPEG-2/H.262: Even Higher Bit Rates and Interlace

- Formally ISO/IEC 13818-2 & ITU-T H.262, developed ('94) jointly by ITU-T and ISO/IEC SC29 WG11 (MPEG) – Now in very wide use for DVD and standard and high-definition DTV (the most commonly used video coding standard)
 - Primary new technical features:
 - **Support for interlaced-scan pictures**
 - **Increased DC quantization precision**
 - Also
 - **Various forms of scalability (SNR, Spatial, breakpoint)**
 - **I-picture concealment motion vectors**
 - Essentially the same as MPEG-1 for progressive-scan pictures, and MPEG-1 forward compatibility required
 - Not especially useful below 2-3 Mbps (range of use normally 2-5 Mbps SDTV broadcast, 6-8 DVD, 20 HDTV)
 - Essentially fixed frame rate

H.263: The Next Generation

- ITU-T Rec. H.263 (v1: 1995): The next generation of video coding performance, developed by ITU-T – the current premier ITU-T video standard (has overtaken H.261 as dominant videoconferencing codec)
 - **Superior quality to prior standards at all bit rates (except perhaps for interlaced video)**
 - **Better by a factor of two at very low rates**
 - **Versions 2 (late 1997/early 1998) & v3 (2000) later developed with a large number of new features**
 - **Profiles defined early 2001**
 - A somewhat tangled relationship with MPEG-4

What Was in H.263 Version 1?

- “Baseline” Algorithm Features beat H.261
 - Half-pel motion compensation (also in MPEG-1)
 - 3-D variable length coding of DCT coefficients
 - Median motion vector prediction
 - More efficient coding pattern signaling (?)
 - Deletable GOB header overhead (also in MPEG-1, but not 2?)
- Optional Enhanced Modes
 - Increased motion vector range with picture extrapolation
 - Variable-size, overlapped motion with picture extrapolation
 - PB-frames (bi-directional prediction)
 - Arithmetic entropy coding
 - Continuous-presence multipoint / video mux



H.263+ Feature Categories

- Error resilience
- Improved compression efficiency (e.g., 15-25% overall improvement over H.263v1)
- Custom and Flexible Video Formats
- Scalability for resilience and multipoint
- Supplemental enhancement information

H.263++ Version 3 Features

- **Annex U:** Fidelity enhancement by **macroblock and block-level reference picture selection** – a significant improvement in picture quality
- **Annex V:** Packet Loss & Error Resilience using data partitioning with reversible VLCs (roughly similar to MPEG-4 data partitioning, but improved by using **reversible coding of motion vectors** rather than coefficients)
- **Annex W:** Additional Supplemental Enhancement Information
 - **IDCT Mismatch Elimination (specific fixed-point fast IDCT)**
 - Arbitrary binary user data
 - Text messages (arbitrary, copyright, caption, video description, and URI)
 - Error Resilience:
 - **Picture header repetition (current, previous, next+TR, next-TR)**
 - **Spare reference pictures for error concealment**
 - Interlaced field indications (top & bottom)

MPEG-4 “Visual”: Baseline H.263 and Many Creative Extras

- MPEG-4 **part 2** (v1: early 1999), formally ISO/IEC 14496-2
- Contains the H.263 baseline design
 - **coding efficiency enhancements (esp. at low rates)**
- Adds many creative new extras:
 - **more coding efficiency enhancements**
 - **error resilience / packet loss enhancements**
 - **segmented coding of shapes**
 - **zero-tree wavelet coding of still textures**
 - **coding of synthetic and semi-synthetic content,**
 - **10 & 12-bit sampling,**
 - **more**
 - **v2 (early 2000) & v3 (early 2001) later added**

MPEG-4 Visual Focus: Simple Profile

- The most basic video coding profile of MPEG-4
- No shape coding
- Progressive-scan video only
- Most popular in low cost / low rate / low resolution apps (e.g., mobile) – top bit rate & resolution limited
- Basic contents
 - H.263 baseline
 - Motion vectors over picture boundaries
 - Variable block-size motion compensation
 - Intra DCT coefficient prediction
 - Handling of four streams in most levels
 - Error / packet-loss features – data partitioning, RVLC

MPEG-4 **Visual** Focus: Advanced Simple Profile

- Target goal: General rectangular video with improved coding efficiency
- Progressive-scan and interlaced video support
- Up to SDTV resolution
- Basic contents
 - All of Simple profile
 - B pictures
 - Global motion compensation
 - Quarter-sample motion compensation
 - Interlace handling

MPEG-4 Visual Focus: Studio Profile

- Target goal: studio & professional use
- Progressive-scan and interlaced video support
- Up to very high resolution and bit rate
- Basic contents
 - Enhanced-accuracy IDCT
 - B pictures
 - 10 & 12 bit sample accuracy
 - 4:2:2 & 4:4:4 chroma sampling structures

The Advanced Video Coding Project

AVC = ITU-T H.264 / MPEG-4 part 10

- History: ITU-T Q.6/SG16 (**VCEG - Video Coding Experts Group**) “H.26L” standardization activity (where the “L” stood for “long-term”)
- **August 1999**: 1st test model (TML-1)
- **July 2001**: MPEG open call for technology: H.26L demo'ed
- **December 2001**: Formation of the **Joint Video Team (JVT)** between VCEG and MPEG to finalize H.26L as a new joint project (similar to MPEG-2/H.262)
- **July 2002**: Final Committee Draft status in MPEG
- **Dec '02** technical freeze, FCD ballot approved
- **May '03** completed in both orgs
- **July '04** Fidelity Range Extensions (FRExt) completed
- **January '05** Scalable Video Coding launched

AVC Objectives

- **Primary technical objectives:**
 - Significant improvement in coding efficiency
 - High loss/error robustness
 - “Network Friendliness” (carry it well on MPEG-2 or RTP or H.32x or in MPEG-4 file format or MPEG-4 systems or ...)
 - Low latency capability (better quality for higher latency)
 - Exact match decoding
- **Additional version 2 objectives (in FRExt):**
 - Professional applications (more than 8 bits per sample, 4:4:4 color sampling, etc.)
 - Higher-quality high-resolution video
 - Alpha plane support (a degree of “object” functionality)

AVC Structure

