ImmediaTV Corporation

Innovative video networking

Innovations today for Broadcasters of tomorrow

H.264 Compression Review

Ciro A. Noronha, Ph.D.

Innovations today for Broadcasters of tomorrow

Terminology 101: Alphabet Soup

- **MPEG-2**
- **H.262**

The ubiquitous video codec standard used in digital television today – Terrestrial Broadcasting, Cable, DBS, DVD-V

- **MPEG-4 Part 2**
- **MPEG-4 SP/ASP**

A follow-on video codec standard – not widely used for DTV

- **MPEG-4 AVC**
- **H.264**
- **MPEG-4 Part 10**
- **AVC**
- **“JVT”**

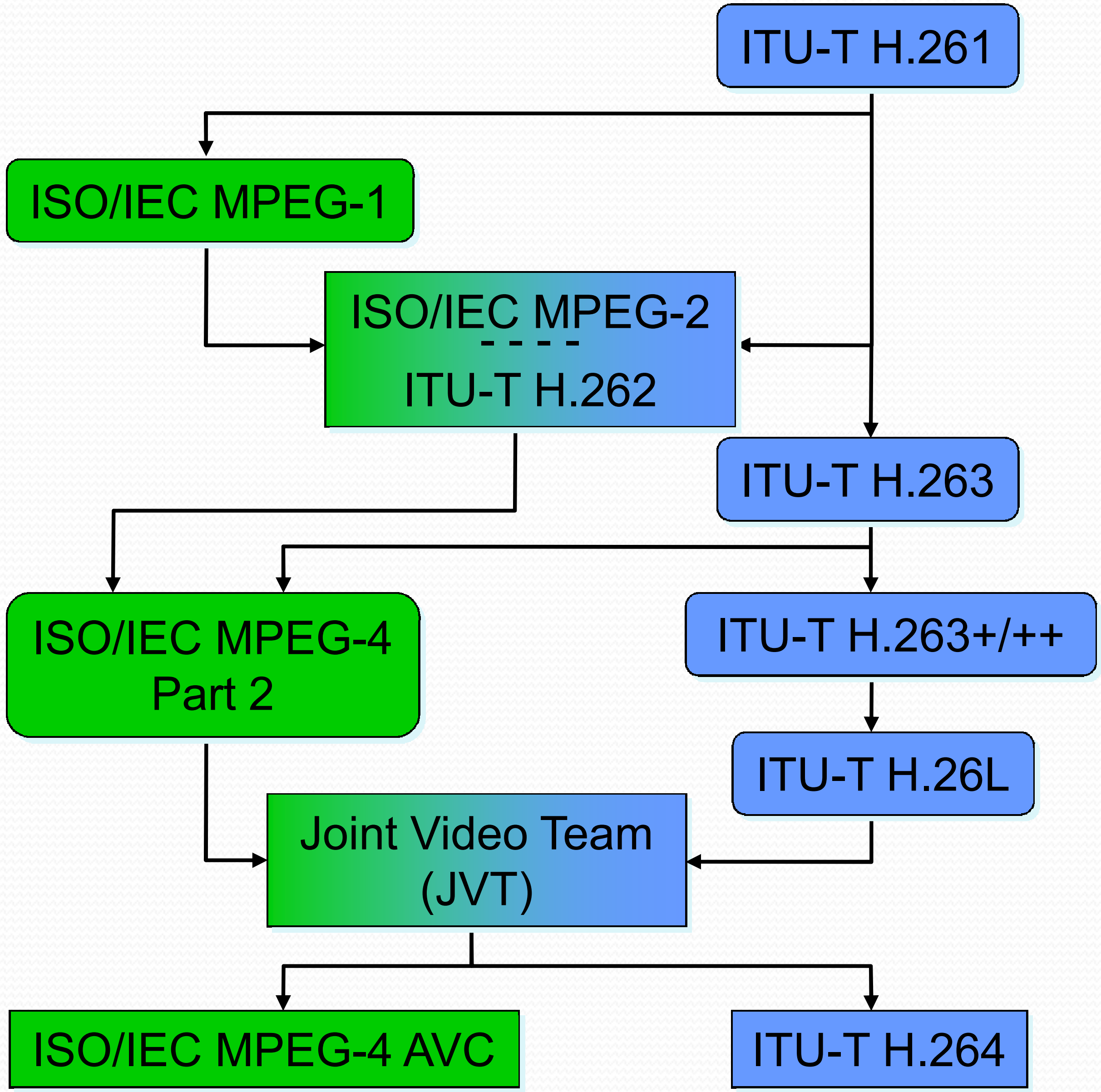
Advanced Video Codec (AVC) – A next generation video codec standard jointly developed by ISO/IEC MPEG and ITU-T VCEG

- **SMPTE VC-1**
- **SMPTE 421M**
- **Windows Media™ Video 9 (WMV-9)**

A next generation video codec originally developed by Microsoft – now a proposed standard in the SMPTE

Due to ambiguity, “MPEG-4” should never be used without a qualifier, i.e. “MPEG-4 AVC”

Video Compression Family Tree



Principles of Compression

- Encoders use the advantage of redundancy in moving images
  - Similarity within a frame (spatial redundancy)
  - Similarity between frames (temporal redundancy)
- Encode the smallest amount of information for a given "quality" level
  - Take advantage of the deficiencies of the human visual/aural system to remove information that "won't be noticed"
- Encoders create a prediction for the decoder to use
  - Then send the error signal giving the difference between the prediction and the actual
- Minimising the error signal minimises the data to be sent
  - Maximises coding performance

Video Compression Basics: Exploiting Spatial Correlation

- Points to consider:
  - An image pixel is likely to be similar to neighboring pixels
  - Fine detail is “less important” to the human eye as major features
- How to do it:
  - Represent the image in the frequency domain using a cosine transform or equivalent
    - This captures the similarity
    - Most of the image information is in the low frequency components
  - Quantize the frequency domain coefficients (this removes “imperceptible” detail), using coarser quantization for the higher frequencies
- Well-known technique; used in JPEG still image compression

Video Compression Basics: Exploiting Temporal Correlation

- Successive images are typically similar
- Results can be improved by coding the ***difference*** between this image and the previous (and/or next) image
- Dealing with motion:
  - A block may move from one frame to another (example: camera pan)
  - Simple difference won't catch that
  - Encoder needs to search the "best match" for a block prior to coding the difference
  - Variables: block size, search radius, number of frames involved

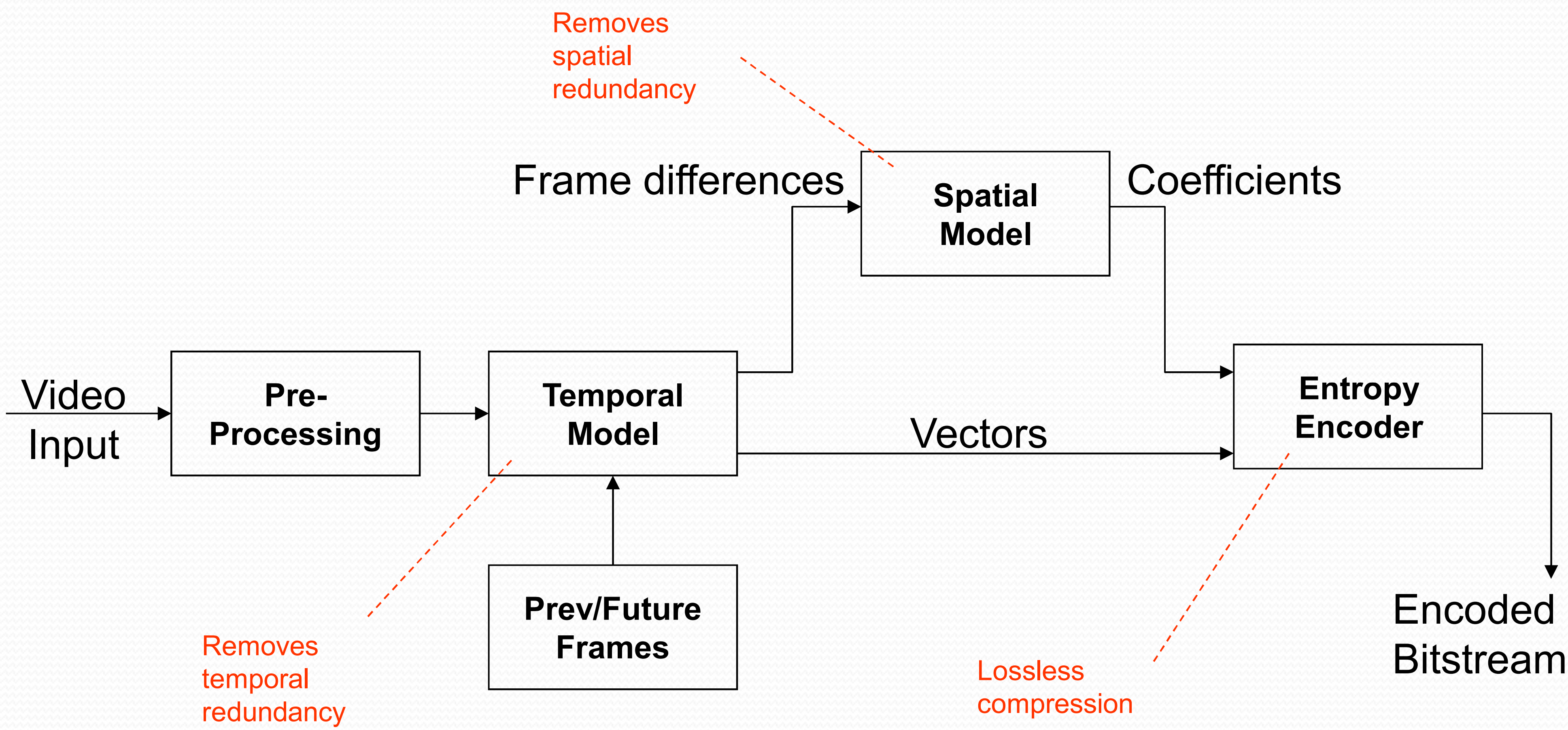
Video Compression Basics: Entropy Coding, the Last Stage

- After all the processing, a bitstream is created
- Some additional compression gains can be realized by using some sort of lossless compression on the resulting bitstream
  - Just like doing a "zip file" of the bitstream
  - This is called "entropy coding"
- Both MPEG-2 and AVC include an entropy coding stage at the end
  - Different compression algorithms

Video Compression Basics: Pre-Processing

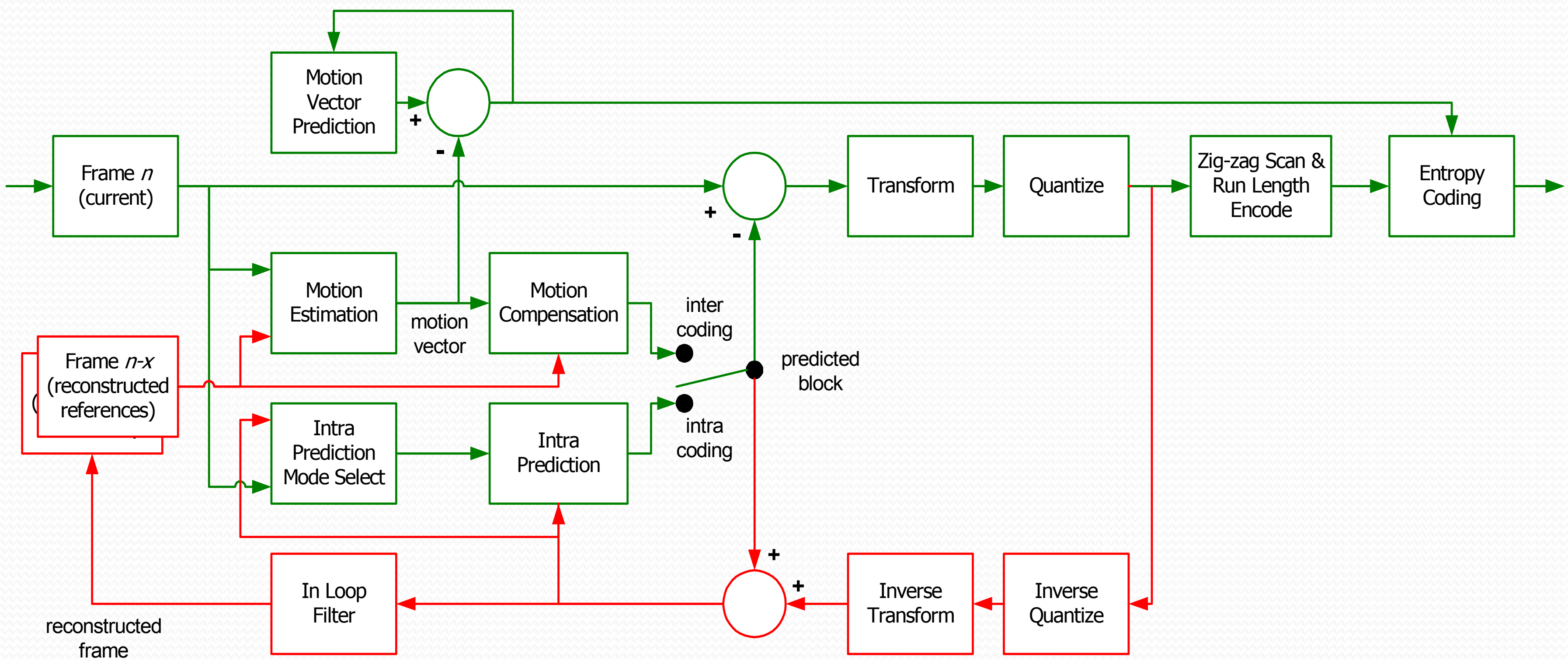
- All modern encoders include some sort of pre-processing
- Pre-processing is filtering the image **before** it goes into the encoder so it is easier to encode and looks better
- Types of pre-processing:
  - Spatial Filtering (reduces a bit spatial noise)
  - Temporal Filtering (can even be motion-compensated)
  - Noise reduction (MPEG encoders do not like noise – they tend to spend a lot of bits coding it)
  - Other filters (deblocking filters, flash/fade detection, scene cut detection, inverse telecine, etc.)
- Pre-processing can make a big difference in quality

Generic Encoder Block Diagram

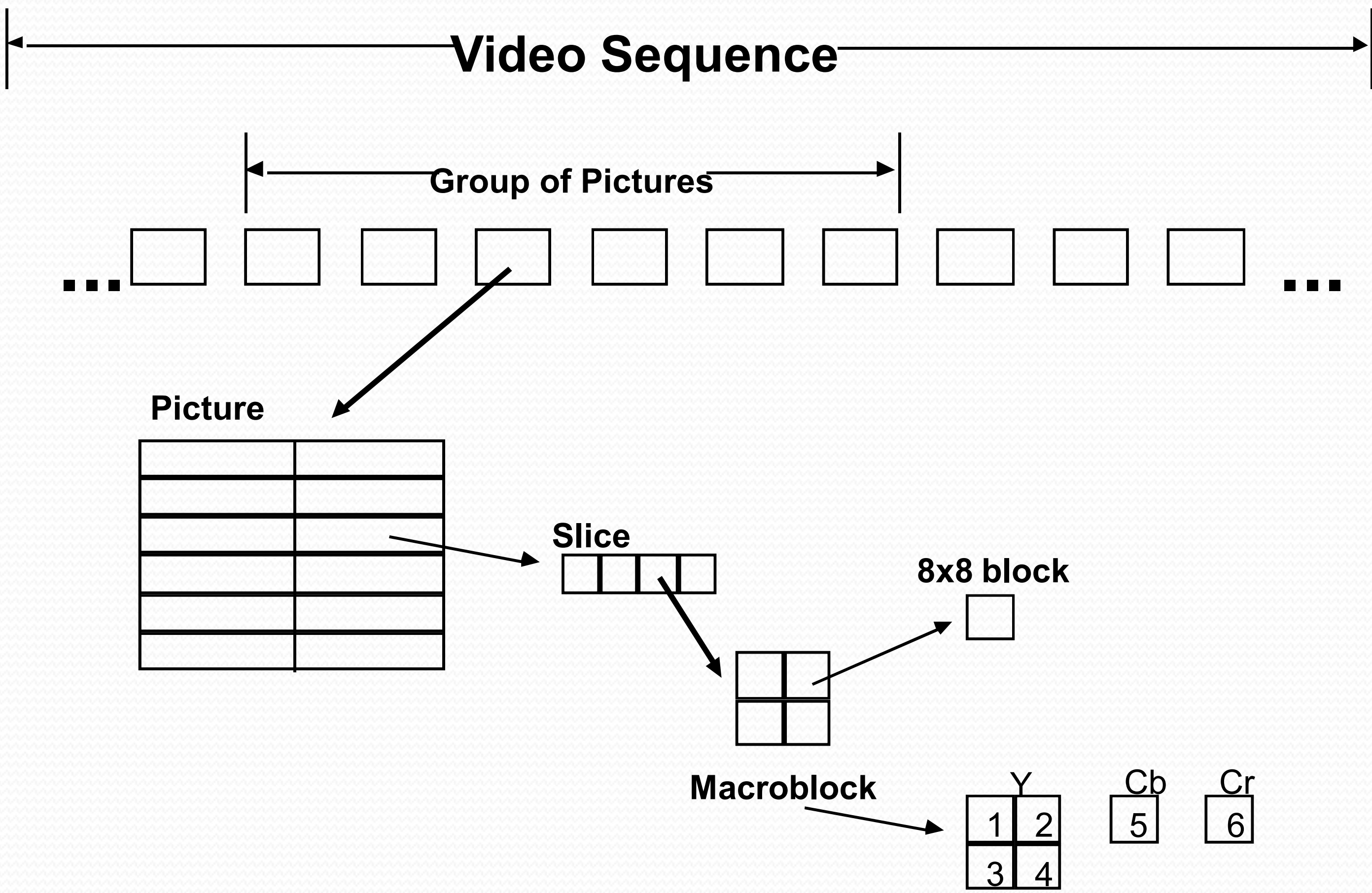


Adapted from "H.264 and MPEG-4 Video Compression", Iain E.G. Richardson

AVC Encoder


Motion
Vector
Prediction
+
-
Frame n
(current)
+
-
Transform
Quantize
Zig-zag Scan &
Run Length
Encode
Entropy
Coding
Motion
Estimation
motion
vector
Motion
Compensation
inter
coding
predicted
block
Frame n-x
(reconstructed
references)
Intra
Prediction
Mode Select
Intra
Prediction
intra
coding
+
+
In Loop
Filter
Inverse
Transform
Inverse
Quantize
reconstructed
frame

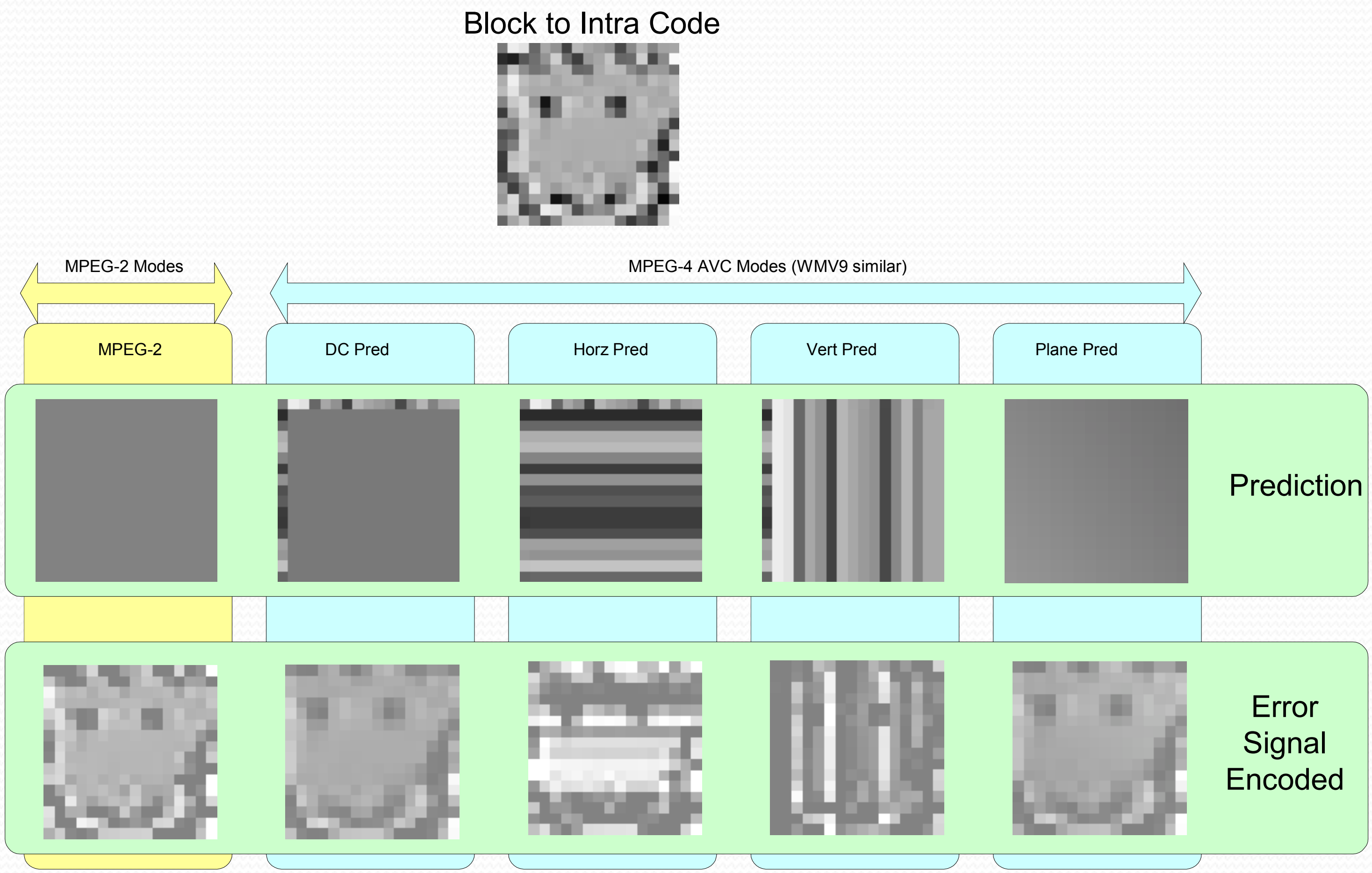
MPEG Data Hierarchy



Types of Coded Frames

- Intra-Coded Frames (I-Frames)
  - Coded similarly to a still image – no dependency on other frames
  - Generally coded every half-second
- Predictive Frames (P-Frames)
  - Uses differences between the current frame and frames in the past
- Bi-directional Predictive Frames (B-Frames)
  - Uses differences between the current frame and frames both in the past and in the future

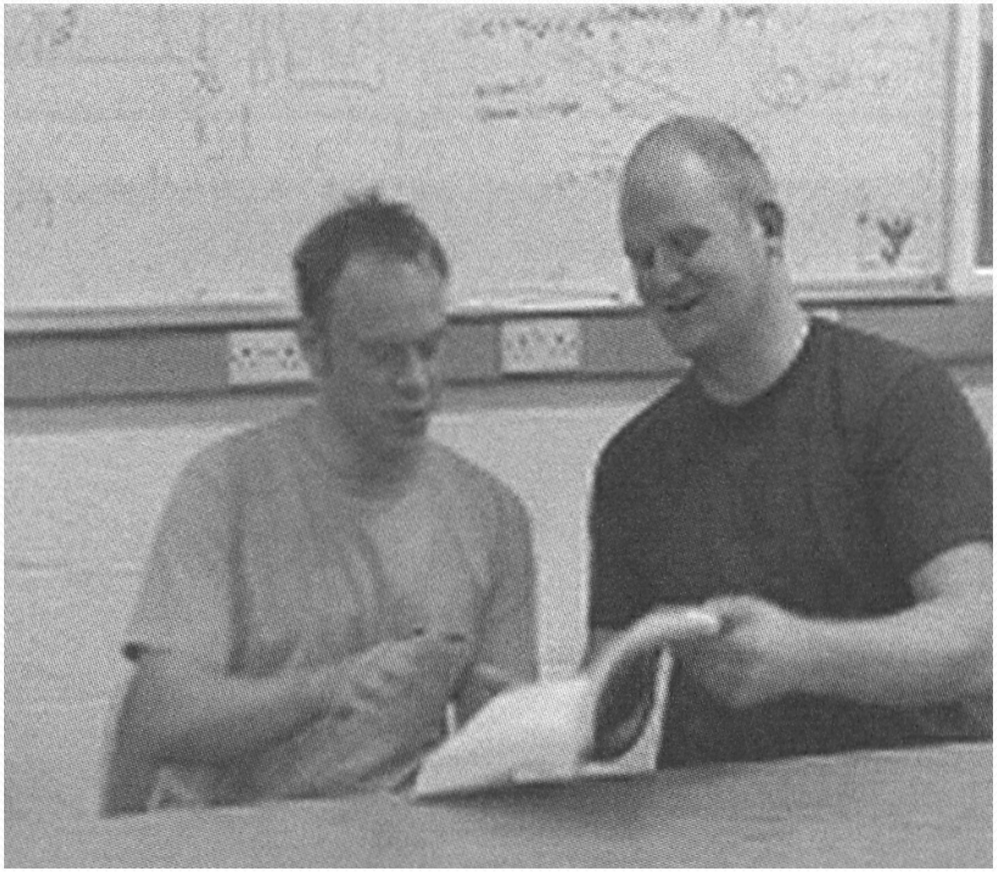
Intra Prediction Modes



- In AVC, adjacent blocks can be used to encode the current block and reduce the error signal

Motion Estimation in Detail

Motion estimation is used to identify moving image content in order to better exploit temporal redundancy

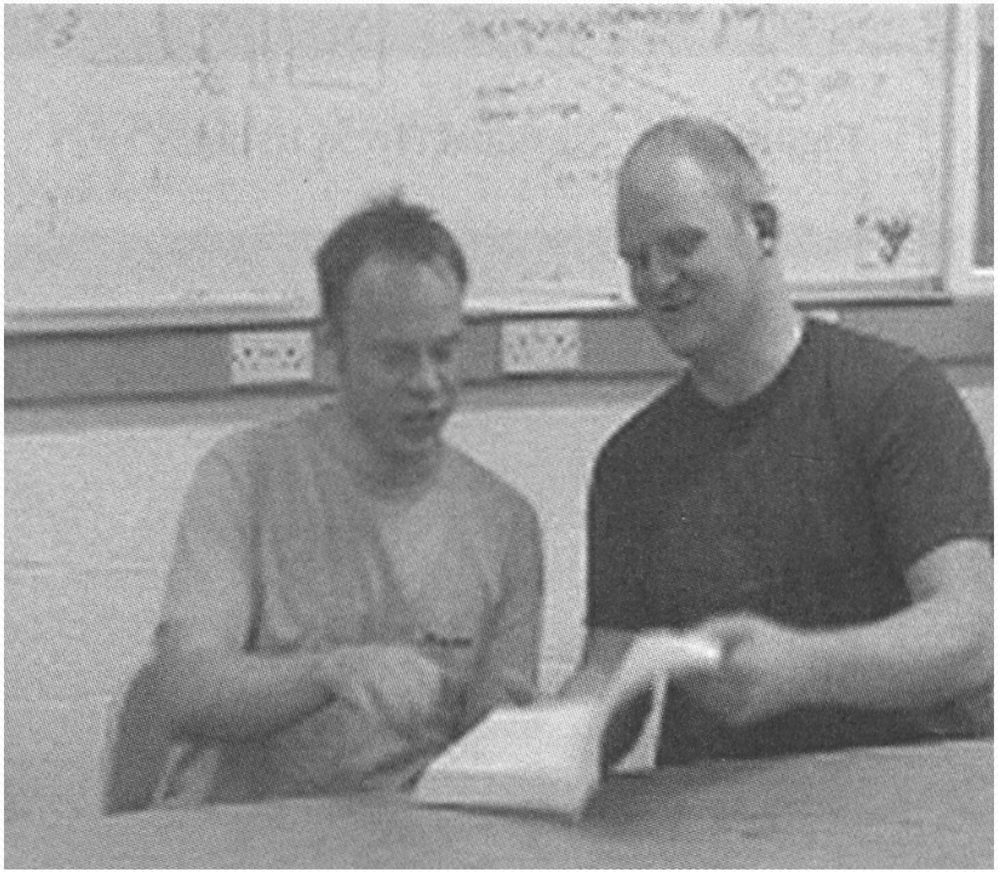


Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Frame N

Motion Estimation

Motion searching during encoding identifies the “best fit” between the current picture macroblock and the reference picture(s)



Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Frame N + 1

Motion Estimation

Only differences between pictures are encoded

Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

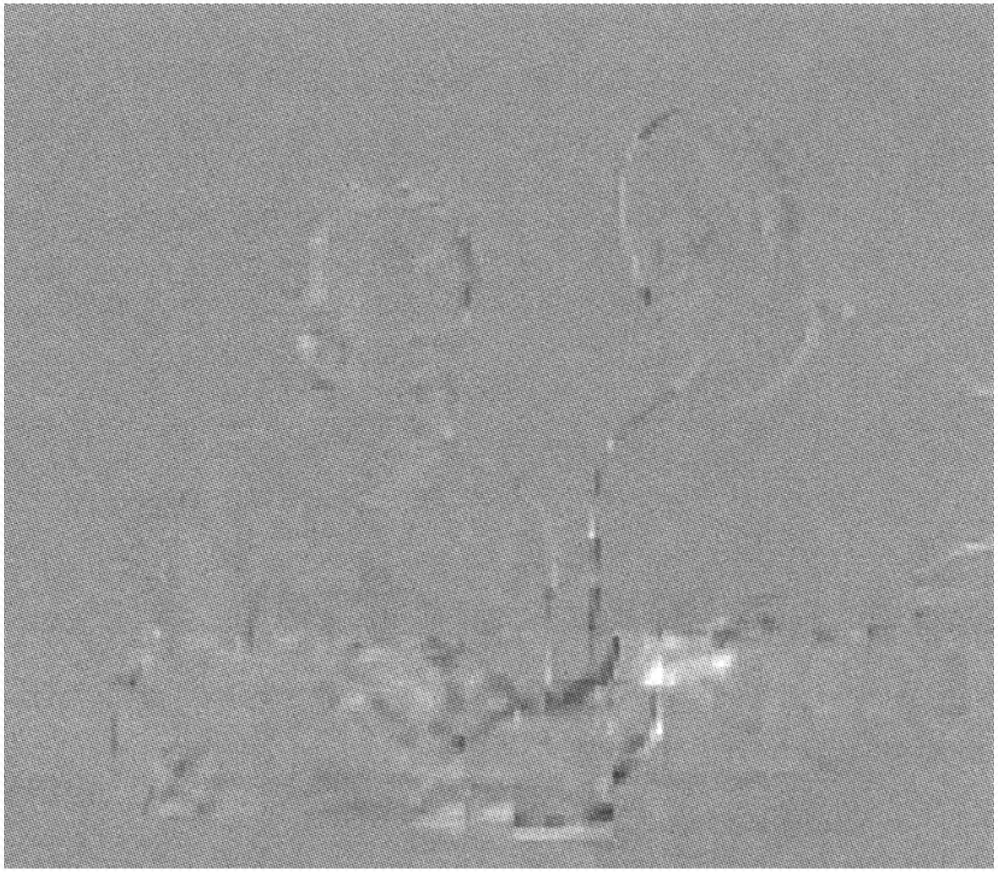
Delta No ME

Motion Estimation

Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Delta 16x16 ME

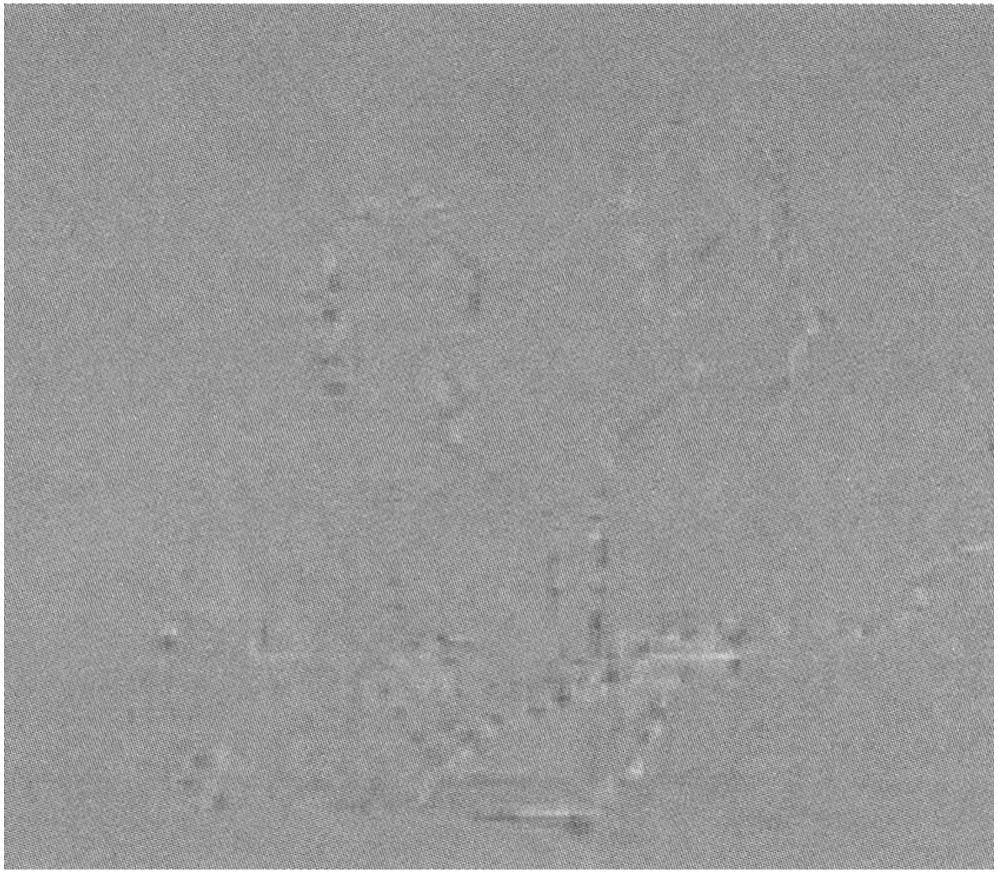
Motion Estimation



Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Delta 8x8 ME

Motion Estimation



Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Delta 4x4 ME

Tradeoffs

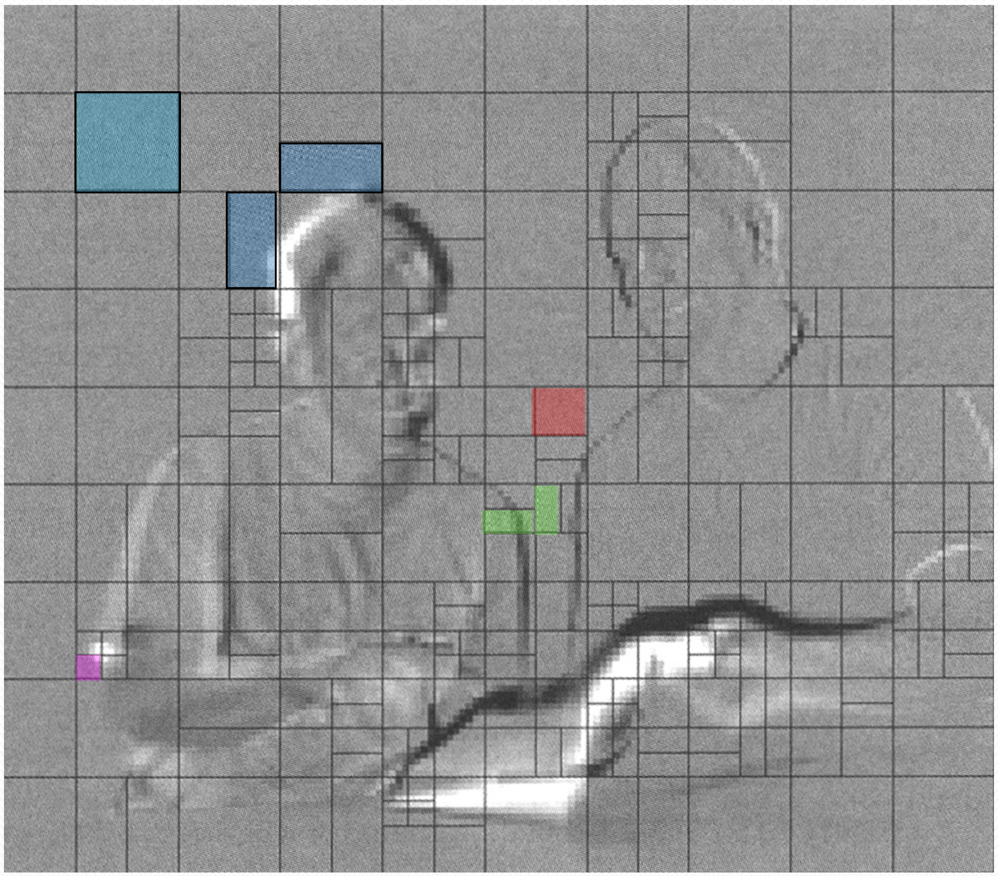
- Real-life objects are not rectangular
- Using smaller ME block sizes improves the fit of the prediction, but also increases the number of vectors and information to be sent
  - Going for smaller block sizes is not always necessarily good.
- A smart encoder can choose the block size depending on the characteristics of the region

Motion Estimation

MPEG-4 AVC provides more ME block sizes

MPEG-2

16x16



Source: H.264 and MPEG-4 Video Compression, Iain E. G. Richardson, Wiley

Appropriate block size choices

AVC

16x16
16x8, 8x16
8x8
8x4, 4x8
4x4

Entropy Encoder

- Removes statistical redundancy from bit stream
- MPEG-2: Variable Length Coding
- MPEG-4 AVC: CABAC (**C**ontext-based **A**daptive **B**inary **A**rithmetic **C**oding)
- Context-based coding –> Optimizes encoded bit string = reduces bits used

Applications → Video Profiles/Levels

Application	*MPEG-2 Video (H.262)*	*MPEG-4 AVC (H.264)*
Mobile Devices **Video Conferencing** **Internet Streaming**	**-**	**Baseline**
Broadcast Quality (TV, Cinema, IPTV)	**Main Profile** **SD: Main Level** **HD: High Level**	**Main Profile** **SD: Level 3** **HD: Level 4**
		High Profile **SD: Level 3** **HD: Level 4**

What about audio?

Technology Name (the standard)	**Vendor Brand Name**
AAC-LC	(*AAC*)
HE-AAC v1	aacPlus v1
HE-AAC v2	aacPlus v2
AC-3	Dolby® Digital
E-AC-3	Dolby® Digital Plus

Dictionary

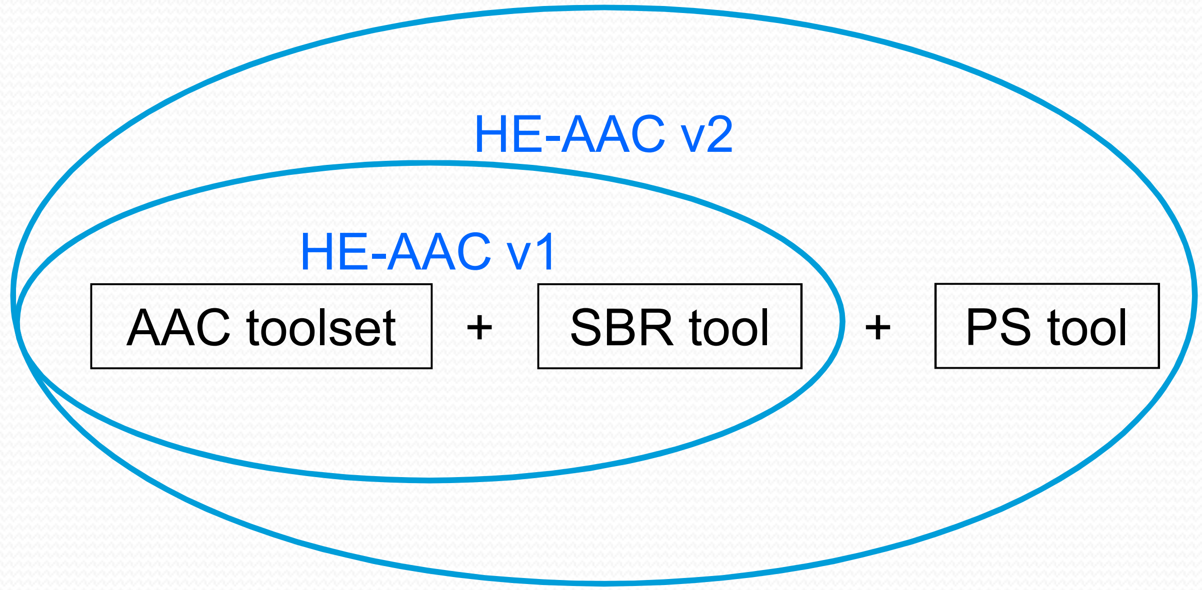
AAC	Advanced Audio Compression
HE	High Efficiency
LC	Low Complexity
AC	Audio Compression
E-AC	Enhanced Audio Compression

The AAC compression algorithm family

- AAC
  - Core audio compression algorithms common to MPEG-2 and MPEG-4 AVC standards
- HE-AAC v1

  = AAC + SBR tool
- HE-AAC v2

  = AAC + SBR + PS tools
  - Decoded as mono by HE-AAC v1 decoder



Dictionary

AAC	Advanced Audio Compression
HE	High Efficiency
LC	Low Complexity
SBR	Spectral Bandwidth Replication
PS	Parametric Stereo

Quick reference guide to audio codecs

Audio Codec	Application	Bitrate	Market acceptance	Comment
MPEG Layer II (stereo)	SD	32-384 kbps	High	192 kbps* Outperforms mp3 >256 kbps
MPEG Layer III (mp3, stereo)	music	32-320 kbps	High	128 kbps* Outperforms Layer II <192kbps
MPEG AAC-LC (stereo)	SD or music	8-529 kbps	High	96 kbps* Outperforms others at any bitrate, but requires more processing
MPEG AAC-LC (5.1)	SD or HD	8-529 kbps	High	256 kbps* Outperforms others at any bitrate, but requires more processing
MPEG HE-AAC (stereo) (5.1)	PiP/handheld SD or HD	8-256 kbps	Medium	v1: 48 kbps* v2: 32 kbps, 48 kbps* 160 kbps*
DOLBY DIGITAL (AC-3, stereo)	SD or HD	32-640 kbps (448 kbps limit in many apps)	High	192 kbps* best for special effects
DOLBY DIGITAL (AC-3, 5.1)	SD or HD	32-640 kbps (448 kbps limit)	High	384-448 kbps* best for special effects

**claimed rate for CD quality for average-complexity content*

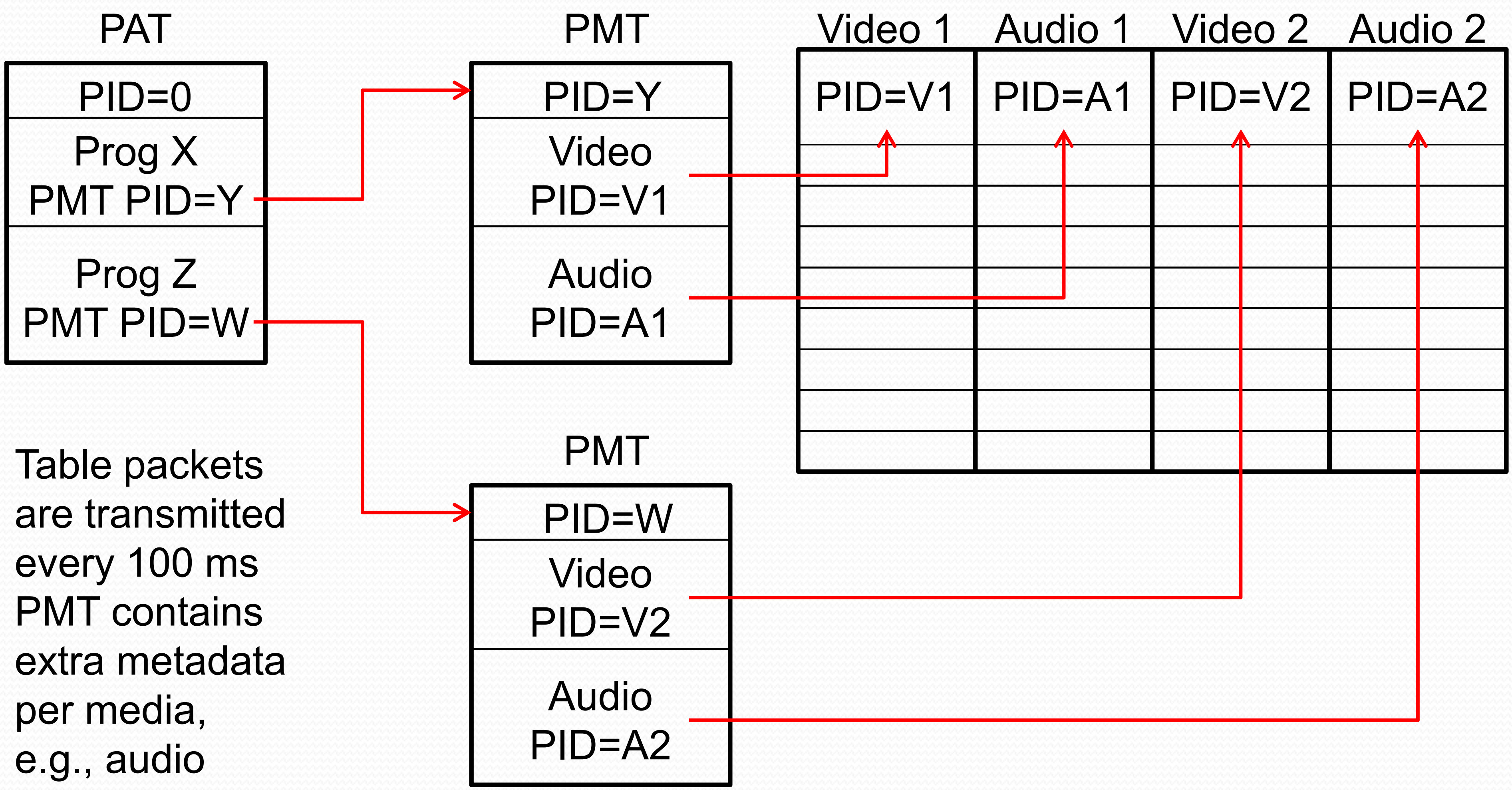
Multiplexing Audio and Video

- Video and Audio encoders generate **elementary streams**
  - Elementary streams are enough to reconstruct the media, but have no synchronization information
  - MP3 players play audio elementary streams
- The elementary streams are broken into packets called PES ("Packetized Elementary Streams")
  - PES packets contain timestamps for A/V sync
  - Audio PES packets are typically fixed size
  - Video PES packets are typically done at frame boundaries

Multiplexing: Transport Stream

- For transmission, PES packets are further broken down into **Transport Packets**
- Transport Packets from audio and video are then interleaved (multiplexed)
- Transport packet basics:
  - Fixed size: 188 bytes (4 bytes header, 184 bytes payload)
  - Header always starts with byte **`0x47`**
  - Header includes a 12-bit Packet ID (PID), which identifies a data flow within the transport
    - Audio and Video have different PIDs
    - PIDs 0 to 15 are reserved for administrative information
- A **T**ransport **S**tream may contain:
  - A **S**ingle audio/video **P**rogram: **SPTS**
  - **M**ultiple audio/video **P**rograms: **MPTS**

MPEG Tables: PAT and PMT



- Table packets are transmitted every 100 ms
- PMT contains extra metadata per media, e.g., audio language code

Other Tables

- All transport streams comply with the basic MPEG table set (PAT/PMT)
  - MPEG also specifies tables to deal with scrambling
- There are multiple standards for what "other" tables can be present in the transport
- These tables provide the following:
  - Channel names
  - Program guides
  - Date/Time
  - Tuning information (frequencies, symbol rates, etc)
  - Channel numbering information

Table Systems

- Table Systems used in practice:
  - **DVB:** International Standard
    - Used primarily in Europe, also by some US satellite providers
    - Variations: DVB-S, DVB-C, DVB-T (satellite, cable, terrestrial)
    - DVB reserves PIDs 16 to 31 for additional tables
  - **ATSC:** Used for terrestrial (and some cable) transmissions in the US
  - **ANSI/SCTE-57:** This is the old GI/Motorola proprietary tables, used in many US satellite and cable systems. ATSC is derived from this system.
  - **ARIB:** Japanese system, adopted in some other countries.
- There are a number of companies who supply table generators

The Service Description Table

- The SDT (Service Description Table) is a DVB table used to give a name to the program
- Virtually all decoders support this table
- Runs on fixed PID **`0x11 (17)`**
- Transmitted once every 2 seconds
- Supported by ImmediaTV encoders