

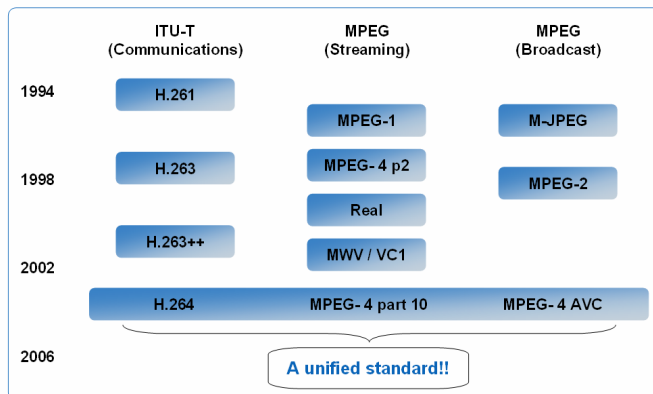
MPEG-4 AVC (H.264) and why, only now, it can save 60% of the network video bandwidth and storage requirements

There is much confusion about video compression standards in general, and the introduction of this new standard with many names compounds the confusion. “MPEG-4 AVC”, also called “H.264” and “MPEG-4 Part 10”, is rapidly being adopted across all network video industry segments because it saves money!! APPROXIMATELY 60% OF BANDWIDTH RELATED CHARGES!! Many segments have hesitated to implement high quality network video solutions. However, combined with steadily declining bandwidth charges and widespread availability of broadband connections, the introduction of MPEG-4 AVC has created a historic point of inflection towards the adoption of high quality network video throughout distributed and multinational organizations.

This paper is an executive summary of the terminology, progression, and design variables related to this dramatic new technology.

A Unified Video Compression Standard

While MPEG-4 AVC is the same as MPEG-4 Part 10 and H.264, it should not be confused with the much older internet streaming format MPEG-4 Part 2 (often referred to as simply MPEG-4). Written by the ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG), MPEG-4 AVC is now the accepted standard for communications, broadcast, and streaming applications, and is used within QuickTime, iPod, PlayStation, Nero, HD DVD, and Blu-ray Disc technologies.



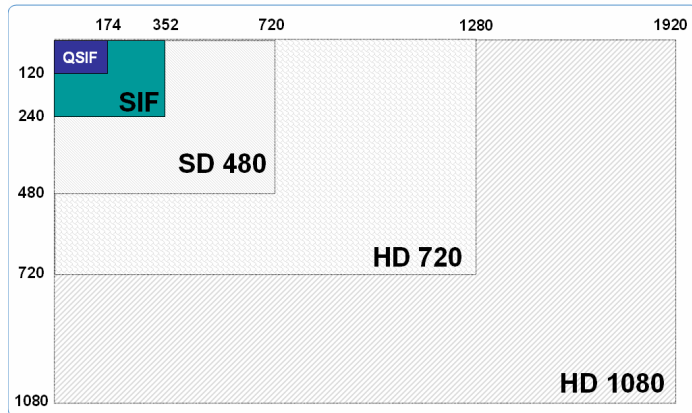
The Resolution Revolution

The broadcaster community knows few video formats. Full resolution video is called standard definition (SD), and high definition (HD) is a series of formats with progressively greater precision. Network video, on the other hand, has many resolution formats due to transmission and encoder/decoder (codec) limitations. SIF (known outside North America as CIF) is typically ¼ SD resolution, and is the most common videoconferencing resolution.

Resolution	North America (NTSC)	Europe (PAL)
QSIF	174x120	-
QCIF	-	174x144
SIF (~QVGA)	352x240	-
CIF	-	352x288
Half-Horizontal Resolution (HHR)	352x480	352x576
Half-Vertical Resolution (HVR)	720x240	720x288
SD 480i (full resolution)	720x480	720x576
HD 720p	1280x720	
HD 1080i	1920x1080	

HHR (or Half Horizontal Resolution of SD) is closely matched to the definition of less recent analog TVs and monitors. With the advent of plasma and LCD TVs, resolutions lower than SD (HHR and SIF) are no longer acceptable for full screen applications.

Once video is encoded, the bandwidths required to support advanced resolutions are usually simple multiples. For instance, SD takes 4x the bandwidth of SIF, and HD typically takes 5 to 10 times the bandwidth of SD.



The ITU (H.XXX) Standards

H.261, H.263, H.263+, H.263++ and now H.264 are progressions of the video compression standard typically used by videoconferencing devices.

These were initially optimized for low bandwidth and unstable bandwidth conditions. Combined with their related communications standards (H.320 and H.323), the ITU standards accommodate ubiquity through concessions in resolution, frame rate, audio/video synchronization, etc. Implementations of this technology have typically been optimized for SIF resolution and a low frame rate (15 frames per second instead of 30).

The Streaming Standards

The most widely adopted video compression formats for streaming are targeted to the bandwidth-constrained user and require a very easy-to-deliver or widely adopted soft player. RealPlayer initially had a strong position in this market, but they have been replaced by the current market leaders Microsoft (with the Windows Media formats) and QuickTime. Other standards such as DIVX (based on MPEG), MPEG-1 and MPEG-4 Part 2 have held position as well. MPEG-4 Part 2 is known mostly for its ability to incorporate object and other data. Typically, bitrates higher than 1150 kbit/s and resolutions higher than 352 x 288 are not used, and latency is not a design consideration (since the video is not used for communication). These standards have also been incorporated into low-cost streaming encoders (security, etc.), compression board sets, and IP cameras.

Standards used in Broadcast and Network Video

Because broadcasters require full resolution, the streaming and ITU standards were never attractive. MPEG-2 (and previously, to a lesser extent Motion JPEG) provided broadcasters with full resolution performance. MPEG-2 has also been extended to HD for broadcast purposes, and adopted as the on-disc compression standard for DVD, as well as the standard for digital cable, making MPEG-2 the most prolific video compression format. For non-broadcast network video applications, MPEG-2 has excellent picture quality, but historically has been seen as requiring too much bandwidth for WAN applications. Outside of broadcast, currently only the demanding users in distance education, high-end corporate collaboration suites, government/military, and medical markets transport MPEG-2 across the WAN.

Why choose MPEG-4 AVC (H.264)

- MPEG-4 AVC (H.264) is the new standard for *all* broadcast, streaming, and communication challenges, so your new systems will be compatible.
- MPEG-4 AVC (H.264) is a highly efficient encoding standard that can dramatically lower bandwidth and storage requirements. This is extremely important for any WAN application, reducing the recurring costs (bandwidth charges) by 60%!!
- MPEG-4 AVC (H.264) is resolution independent, and will be the basis of most HD implementations.

To understand the complexity of the new compression standard, compared to MPEG-2, MPEG-4 AVC (H.264) requires 8 times the processing power for encoding and 3 times the processing power for decoding. So developing encoder and decoder technology is not trivial, especially if performance is a concern. Most

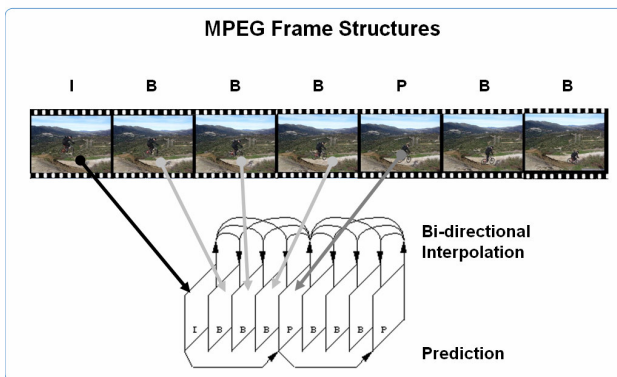
of today's implementations are based on high power FPGA (field programmable gate array) technology, and have been brought to market using Windows XP-based systems. Only a few vendors have non-Windows based systems, typically built on a Carrier Grade Linux (CGL) infrastructure. Remember, just because the operating system is not visible to the user does not mean that it is carrier grade.

It is important as well to consider the audio codec. While most vendors use the AAC audio codec along with the H.264 video codec, some use alternate audio codecs. This is important to note when trying to achieve compatibilities for current installation as well as forward looking application (storage).

Also, most VTC (videoteleconference) products now support H.264 encoding up to a certain bandwidth, based on the available processing power within their system. But past this bandwidth (768 kbps in some systems), they then fail-over to H.263 mode. (Being "H.264 compliant" does not mean Plug and Play communications with *any* H.264 system.) So both the audio codec and the communications protocol must be taken into consideration.

Interactive High Performance MPEG-4 AVC Products

Today's communications suites require encoder/decoder systems designed for very high multi-stream performance supporting multiple bi-directional video channels in a single chassis system. Critical to performance of the suite is latency (the time to encode and decode). Standard video conferencing devices typically provide ~300 millisecond latency end-to-end without network considerations. Broadcast streaming devices typically sustain over 800 ms. Communications suites require end to end latency below 150 ms and 200ms respectively.



MPEG Basics


To help you get a clear picture of the performances described below, here is a brief introduction to some of the technical aspects:

- I-Frames (Intra Pictures) are stand-alone compressed frames – they are like stand-alone JPEG images
- P-Frames are predicted from the closest I or P frames available
- B-Frames are a second level of prediction (actually polled from adjacent images on either side)


A group of pictures (GOP) is the number for frames from one I-Frame to another. So in one GOP, there is a full frame, and a number of frames predicting the motion associated with that full frame. Higher levels of compression (less bandwidth for a given clip) are achieved through larger GOP sizes and deeper frame prediction (i.e., using an IBP compression structure instead of an IP structure). However, deeper frame compressions require more time to encode, resulting in higher latencies. Note that all of these parameters are typically established during system setup and not changeable during sessions.

ALL BANDWIDTH ESTIMATES NEED TO BE PLACED IN PERSPECTIVE OF THE CONTENT

We typically speak of:



– “Talking Heads” - stable video shots of low contrast and low motion with very few if any scene changes. Talking heads typically require lower bandwidth.



– “Action Video” – think of a blockbuster movie; dramatic transitions and motion, high contrast. Action video tests the metal of an encoder – but is this really your application?

Performance-Tuning MPEG Systems

Low bandwidth / high latency - If you have a tight bandwidth budget, you may wish to select a deep compression (IBBP) with a large GOP size (~40) – although latency will suffer to some extent.

High bandwidth / low latency - If you require 100% accuracy and immediacy, you should select low compression (I frame only, or IP) and a very small GOP size (~4).

MPEG-2 Performance

With MPEG-2, the minimum bandwidth for full resolution video is typically just over 2 Mbps. In use, a minimum for a “Talking Head” might be about 2.2 Mbps and for “Action Video” about 5 Mbps. These numbers are dramatically affected by the content, and by the users’ tolerance for MPEG artifacts, as the limits of the settings are pushed by the video content – Can you tolerate “blockiness” under extreme circumstances?

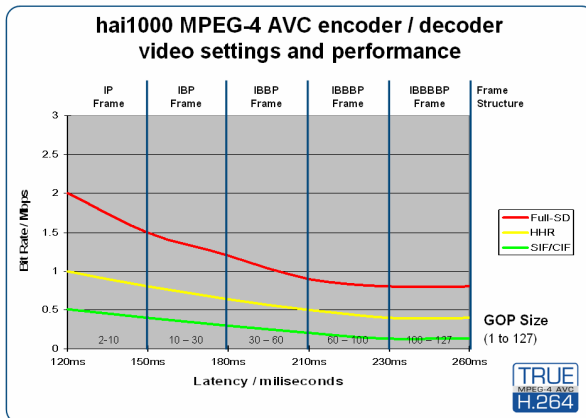
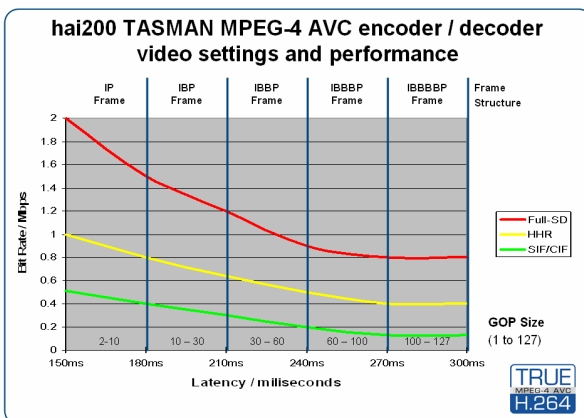
Keep in mind that all bandwidths noted within this discussion are video bandwidths, referred to as the VIDEO BIT RATE. In addition to the VIDEO BIT RATE, you must add the AUDIO BIT RATE and a provision for the overhead related to IP (about 25-30%). Audio bit rates with MPEG-4 AVC are about half that required for MPEG-2 (or your standard MP3 song), so 128 kbps is typically fine for MPEG-2, and 64 kbps for MPEG-4 AVC. Approximate TOTAL BANDWIDTH REQUIRED = (VIDEO BIT RATE + AUDIO BIT RATE) x 1.25.

Frame Structure	I Frame	IP Frame	IBP Frame	IBBP Frame
Bit Rates (400 kbps to 15 Mbps)	6 – 15 Mbps	2 – 10 Mbps	1 – 6 Mbps	0.4 – 5 Mbps
Picture Size Full-SD: 720x480 HHR (½ SD): 352x480 SIF (¼ SD): 352x240	Full SD	Full SD (> 3 Mbps) HHR (> 2 Mbps)	Full SD (> 3 Mbps) HHR (> 2 Mbps) SIF (> 800 kbps)	Full SD (> 2 Mbps) HHR (> 1.4 Mbps) SIF (> 400 kbps)
GOP Length (1 to 127)	1	2 – 30	30 – 60	60 – 127
End-to-End Latency (ms)	90 ms	170 ms	200 ms	230 ms

Frame Structure	I Frame	IP Frame	IBP Frame	IBBP Frame
Bit Rates (800 kbps to 15 Mbps)	4 – 8 Mbps	2 – 8 Mbps	1 – 6 Mbps	800 kbps – 4 Mbps
Picture Size Full-SD: 720x480 HHR (½ SD): 352x480 SIF (¼ SD): 352x240	Full-SD	Full-SD (> 3 Mbps) HHR (> 2 Mbps)	Full-SD or HHR (> 2 Mbps) SIF (> 1 Mbps)	Full-SD (> 3 Mbps) SIF (> 800 kbps)
GOP Length (4 to 19)	N/A	4 – 19	4 – 19	4 – 19
End-to-End Latency (ms)	200 – 330	300 – 640	370 – 690	400 – 700

MPEG-4 AVC (H.264) Performance

With MPEG-4 AVC, the minimum bandwidth for full resolution video is typically just over 800 kbps. In use, a minimum for a “Talking Head” might be about 800 kbps and for “Action Video” about 1.8 Mbps. This makes “Talking Head” transmission ideal for T-1 connections, and “Action Video” ideal for a 2xT-1 connection (also called a “bonded T-1”).



MPEG-4 AVC is based on a much more efficient encoding scheme than MPEG-2. It also has a smaller block size and intelligent algorithms that result in much softer degradation of the images as limitations are met. So, while the limits of MPEG-2 were best avoided, since large, very visible blocks would occur, with MPEG-4 AVC, as you approach the limits, the video will degrade much more gracefully.

MPEG-2 vs. MPEG-4 AVC Performance

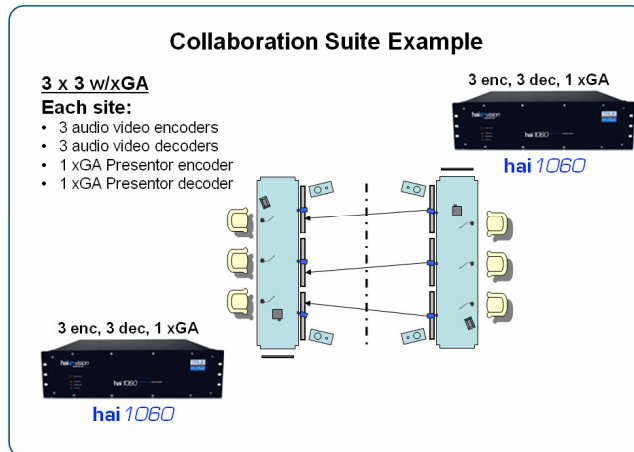
What has been considered acceptable video quality for MPEG-2 at 6 Mbps can now be experienced using MPEG-4 AVC at about 2 Mbps for “Action Video”.

What has been considered acceptable video quality for MPEG-2 at 2.5 Mbps can now be experienced using MPEG-4 AVC at about 900 kbps for “Talking Heads”.

The overall savings on bandwidth due to MPEG-4 AVC implementation is about 60%!!

That means a 60% savings on the recurring monthly bill for any video activity over the WAN!!

MPEG-4 AVC vs. MPEG-2		
	MPEG-2	MPEG-4 AVC (H.264)
Talking Head (Steady content, low contrast)	2.2 – 4.0 Mbps	0.7 – 1.4 Mbps
Action Video (Dramatic changes, high contrast)	4.0 – 7.0 Mbps	1.6 – 3.0 Mbps
End-to-End Latency (ms)	90 ms (lowest hai500) 170 ms (average)	120-150 ms (hai1000)



Practical Application

A collaboration suite with 3 cameras and 3 video displays may have required approximately 12 Mbps full duplex bandwidth with MPEG-2 (estimate 3 Mbps video bit rate plus 256 kbps audio and 25% IP overhead per channel = 6.6 Mbps per channel times 3 full duplex channels).

With MPEG-4 AVC, the same suite can be accommodated with approximately 5 Mbps (estimate 1.2 Mbps video bit rate plus 128 kbps audio and 25% IP overhead per channel = 1.66 Mbps per channel times 3 full duplex channels).

Keep in mind that the bandwidth is full duplex by nature, so you need only account for the largest of either the inbound or outbound bandwidth.