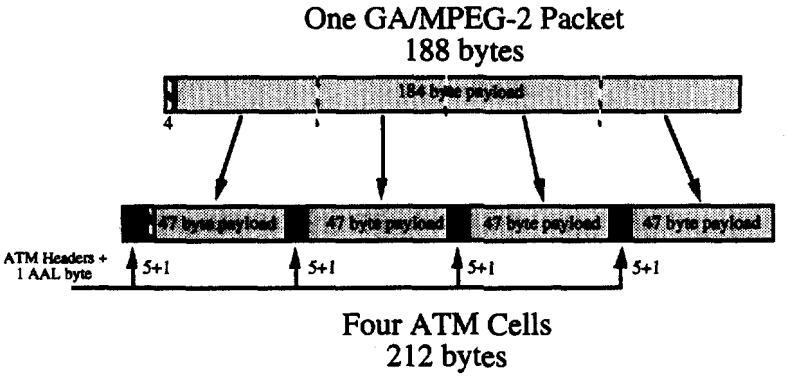

Figure 4.3. **Encapsulation of GA/MPEG-2 within ATM.** A GA/MPEG-2 packet may be encapsulated complete with its header within four ATM packets by using 1 AAL byte per ATM Header leaving 47 usable payload bytes x 4 packets for 188 bytes. Encapsulation concatenates the full functionality of the MPEG-2 Transport stream with the full functionality of the ATM transport stream.

Another ATM interoperation technique is conversion, where the GA/MPEG-2 packet header is removed and the information it contains is converted and placed in the ATM headers. As shown in Figure 4.4, the 4-byte GA/MPEG-2 packet header is stripped and converted to create ATM headers, and the remaining 184-byte payload is packed into four 46-byte ATM payloads leaving two bytes for ATM AAL.

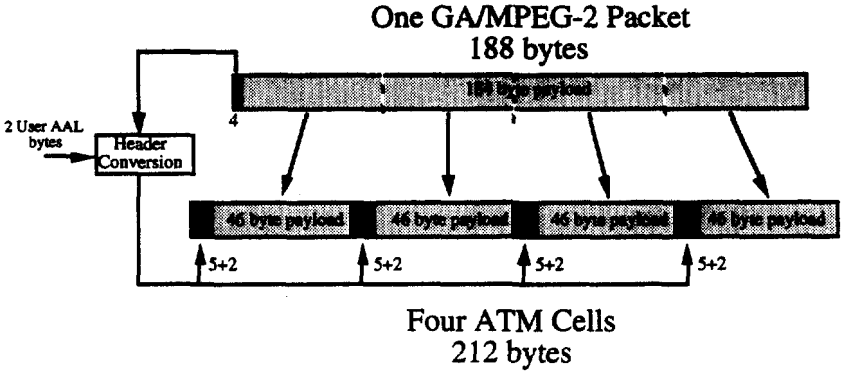

Figure 4.4. **Conversion of GA/MPEG-2 to ATM.** A GA/MPEG-2 packet stream may be converted to an ATM packet stream by removing and translating its header for use in the ATM headers. By not carrying the 4-byte GA/MPEG-2 packet header as ATM payload, one extra byte per ATM packet is freed for use as another AAL byte. Now only 46 bytes per ATM packet are needed to carry 184 GA/MPEG-2 payload bytes.

4.4 SYSTEM LEVEL MULTIPLEX -- A HEADER/DESCRIPTOR TREE

The GA/MPEG-2 transport layer provides the functionality to multiplex multiple streams of video, audio and data in order to provide multiple windows and/or programs in a single packet stream[10]. The MPEG-2 System describes how multiplexed packets within a single transport stream are to be interpreted, as illustrated in Figures 4.5 and 4.6. Each packet stream contains an

[10] In accordance with PS-WP4 recommendation # 11.

identifying number called a PID, or packet ID. The only PID that is defined by the standard is number zero (PID 0). PID 0 is the root of the tree structure, and contains a table of PID numbers that each correspond to one *program*. Each of the PID's in this *program association table* points to another packet stream that contains another table of PID's, the *program map table,* which contains one PID for each element of the program such as video, audio, and aux. data. The program map table PID's point to *elementary packet streams* each of which contain a certain type of data (video, audio, data, etc.).

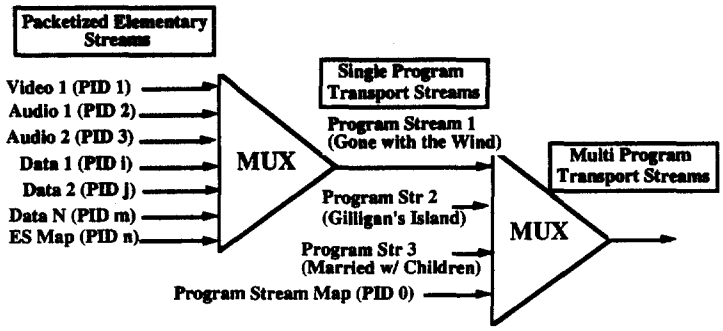

Figure 4.5. **Hierarchical Packet Multiplexing**: Packetized Elementary Streams (PES) are multiplexed to create a single program transport stream, which can be combined with additional program streams.

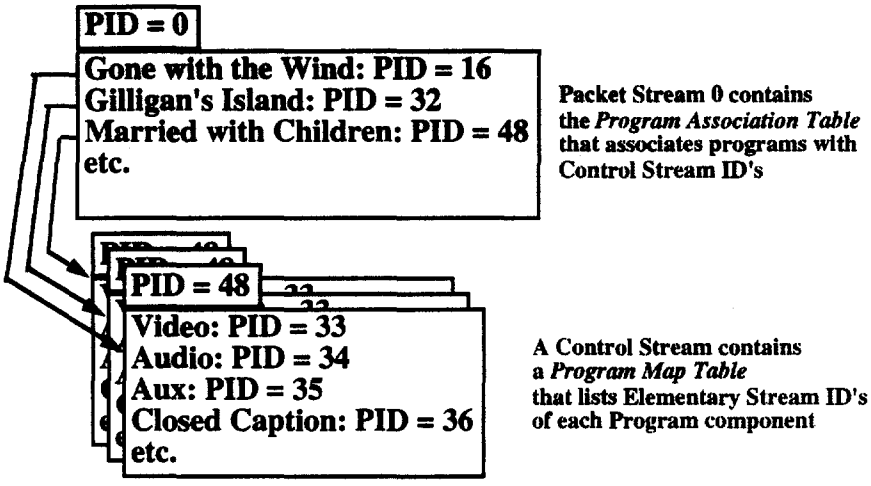

Figure 4.6. **A hierarchy of packet streams.** PID 0 is the root of the tree structure, and contains a table of PID numbers that each correspond to one *program*. Each of the PID's in this *program association table* points to another packet stream that contains another table of PID's, the *program map table,* which contains one PID for each element of the program such as video, audio, and aux. data. The program map table PID's point to *elementary packet streams* each of which contain a certain type of data (video, audio, data, etc.). Except for PID 0, all PID numbers shown are examples and are not "wired in" to the standard.

By traversing this tree a receiver may determine which packet streams it needs to decode, and which it should ignore. By using a hierarchy of tables and describing the tables in the bitstream itself, MPEG-2 creates a highly flexible and extensible packet stream architecture that contains very few hard limits.

4.5 TRANSPORT LAYER INTEROPERABILITY SUMMARY

By adopting a compatible subset of the MPEG-2 Systems Transport standard, the Grand Alliance HDTV system provides an extremely efficient and robust data transport system. The system provides flexibility and extensibility through the use of Packet ID (PID) header/descriptors. These PID headers provide a hierarchical relationship among various packet streams, with the relationship defined dynamically within the packet stream itself. This allows flexible multiplexing of any combination of video, audio and auxiliary data into a program stream, and the additional multiplexing of multiple program streams into a transport packet stream. It also allows service providers to define the structure of their own packet systems, and affiliates and headend operators to add or drop programs as their business needs require.

The PID headers also provide extensibility, since certain PIDs can be reserved for future use, and older receivers will simply discard the packets with PIDs that they don't understand, and continue to deliver the original service. The system provides interoperability with ATM networks either through conversion or encapsulation of each of its packets into exactly 4 ATM packets, and provides functionality that addresses issues unique to program delivery, such as clock synchronization data.

5. COMPRESSION LAYER INTEROPERABILITY

The tremendous data reduction that makes HDTV-in-6-MHz possible takes place in the Compression Layer. This layer is by far the most computationally intensive layer of the entire system, and its optimization and standardization has been the result of several years of intense research and development by members of the Grand Alliance and other organizations participating in the Moving Picture Experts Group (MPEG) of the International Standards Organization. The compression layer is a standardized syntax or "language" that allows a moving image to be described with far fewer bits than are normally required. Each pixel of a color RGB image is normally described with 24 bits (8 for each color); after compression each pixel of the Grand Alliance image is described by an average of less than 0.25 bits.

5.1 COMPRESSION DESIGN ISSUES

Image coding is a statistical process that takes advantage of redundancies that occur in most image sequences. Frame-to-frame redundancies are removed using motion-compensated frame differences, and spatial redundancies are removed using the Discrete Cosine Transform (DCT). After the redundancies are removed, the statistics of the remaining differences are usually such that certain values (those near zero) are much more common than others, as shown in Figure 5.1. A set of codes with varying lengths (variable-length codes, or VLC's) is used to transmit the differences. Short codes are assigned to the most common values, the same way that the Morse telegraphy code uses the shortest code to represent the most common letters.

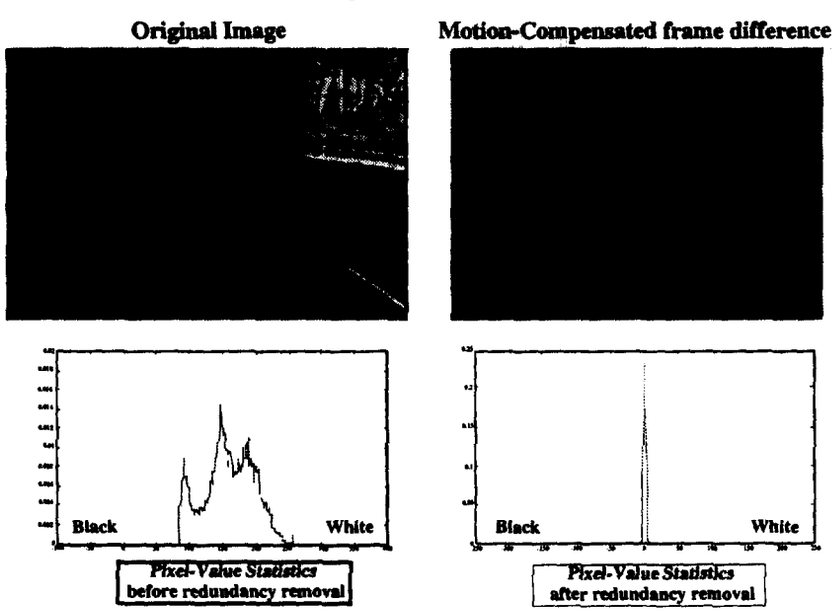

Figure 5.1. **Image statistics** before and after Motion Compensated Prediction: Motion compensated prediction removes a tremendous amount of redundancy.

Further reduction of the number of bits is achieved by *quantizing* the values to be coded with VLC's. The factor that divides the values is called the *quantization scale,* and it is used to control the number of bits generated by the encoder. Unfortunately, quantization causes an irreversible loss in image data and, when it is severe, artifacts (called *coding distortion* or *quantization noise*) may be visible. This is most likely to be the case in images that have little redundancy such as highly detailed scenes with fast motion or scene cuts, and especially in systems designed to operate at low coding rates below 0.5 bits per pixel. A example is shown in Figure 5.2.

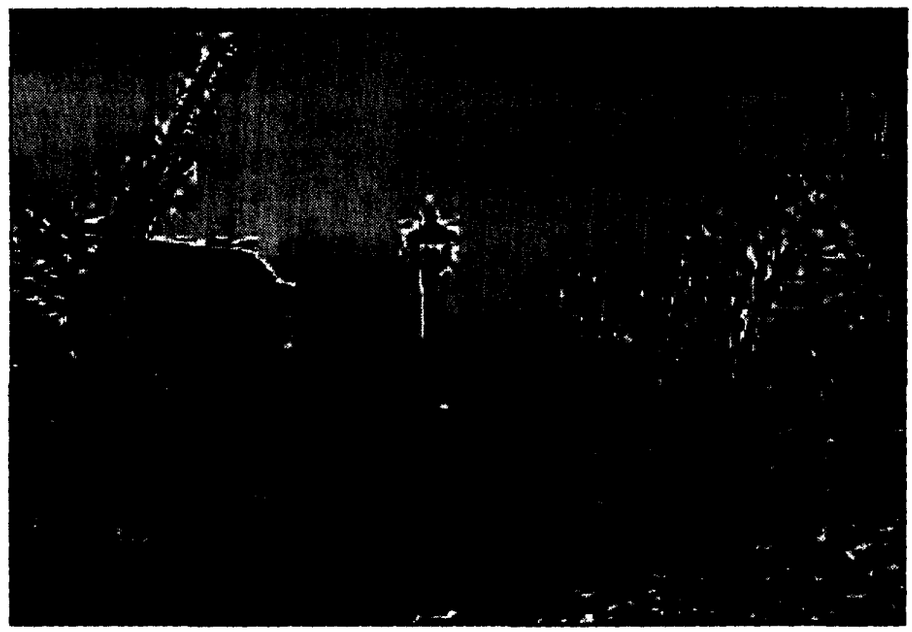

Figure 5.2. **Image Coding Artifacts**: Clearly visible above are the sorts of artifacts that arise from compressing an image into too few bits. This condition is the result of trying to send too many pixels/sec with too few bits/sec. The coding rate, bits/pixel, is a measure of the trade-off between picture fidelity and compression. (I-frame coded with a fixed quant_scale of 25.)

Another good measure of how much coding distortion can be expected is to divide the channel capacity (megabits per second) by the pixel rate (megapixels per second) to get the coding rate (bits per pixel). This is why the format chosen in the Picture Layer, which determines the pixel rate, has so much influence on the number of artifacts and the final picture quality. Stated another way, at a fixed number of bits per pixel available to produce acceptable picture quality, the system designer must make tradeoffs among horizontal, vertical and temporal resolution.

The table below shows some typical picture formats, the bits rates they might typically be coded with, and the resulting coding rate (bits/pixel).

Resolution (H x V x fps)	Data Rate (Mbits/sec)	Coding Rate (bits/pixel)	Picture Quality
1920 x 1080 x 30	45	0.48	*Distribution*
1920 x 1080 x 30	18	0.19	*Consumer HDTV*
1728 x 960 x 30	18	0.24	*Consumer HDTV*
1440 x 960 x 30	18	0.30	*Consumer HDTV*
1280 x 720 x 60	18	0.22	*Consumer HDTV*
720 x 480 x 30	9	0.56	*Studio NTSC*
720 x 480 x 30	4.5	0.28	*"NTSC" Quality*
352 x 240 x 30	1.2	0.32	*≈ VHS Quality*
176 x 120 x 30	0.3	0.32	*< VHS Quality*

5.2 THE MPEG-2 COMPRESSION FAMILY: LEVELS AND PROFILES

The Moving Picture Experts Group realized that if a syntax was created that could describe extremely large pictures and provide a rich "toolbox" of coding techniques, then a "standard" decoder capable of decoding anything specified by the syntax would be prohibitively expensive. Manufacturers might choose what portion of the syntax and what size of pictures and bits rates they would support, and no one would be sure if a particular bitstream would be decoded by a particular decoder. The resulting chaos would be an interoperability nightmare.

The MPEG-2 solution has been to define a standard group of subsets of the syntax. Actually, as shown in Figure 5.3, these subsets are two-dimensional: profiles and levels. A profile is a minimum set of "tools" or coding structures such as B-Frames and Spatial Scalability. Within a given profile, a given level restricts the maximum values of various parameters, such as the largest picture and the fastest bit rate that can be accommodated. To date, four profiles and four levels have been defined (although not all combinations are defined), creating a 4x4 matrix of standards within MPEG-2.

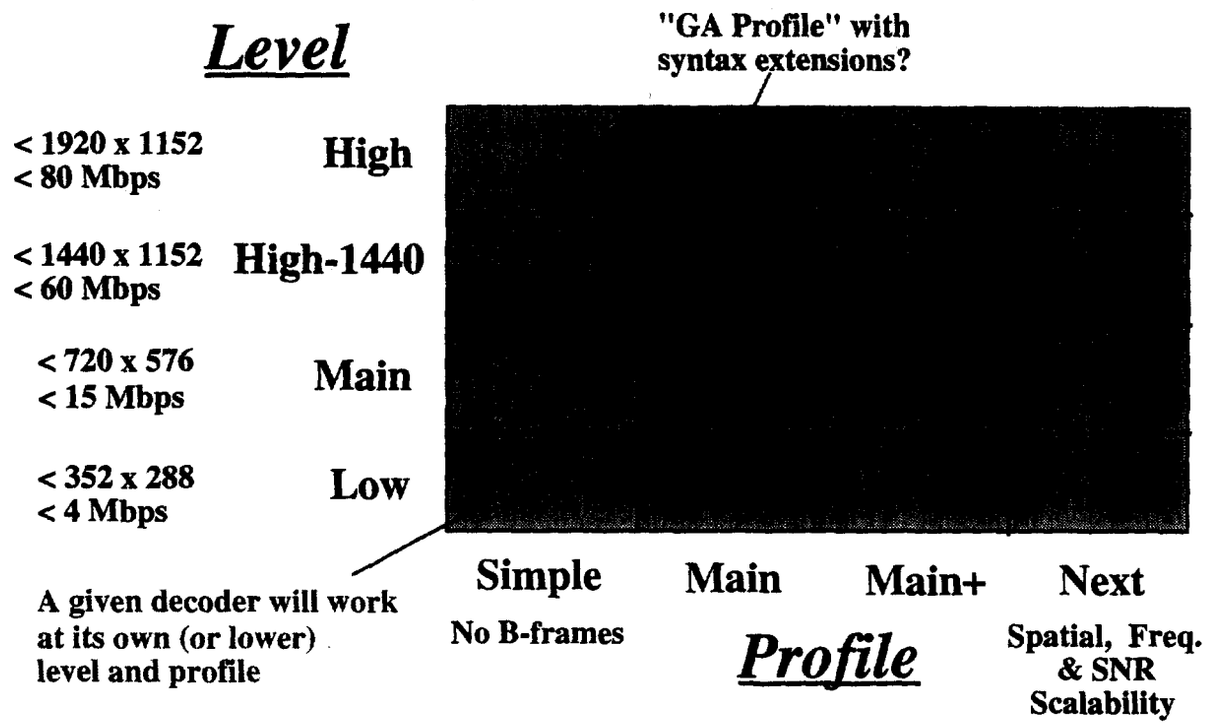

Figure 5.3. Levels and Profiles. This 4x4 matrix shows the relationship among the various Levels and Profiles defined by MPEG. A profile is a minimum set of "tools" or coding syntax elements such as B-Frames and Spatial Scalability. Within a given profile, a given level restricts the maximum values of various parameters, such as the largest picture that can be described and the fastest bit rate.

5.2 THE GA COMPRESSION SYNTAX

The GA HDTV system supports multiple formats through the use of MPEG-2's headers and descriptors, and uses syntax elements that including field/frame coding, I, P *and* B frames, and progressive refresh. The GA HDTV system uses a superset of the MPEG-2 Working Draft Syntax[11]. The GA system differs from MPEG-2 High Level at Main Profile in that it includes additional syntax for AC coefficient leaky prediction ("AC leak"). Liaison among the Grand Alliance, ACATS and MPEG is on-going to attempt to harmonize the syntaxes.

5.4 MPEG-2 INTEROPERABILITY

A Grand Alliance HDTV receiver can decode a High Level at Main Profile MPEG-2 bit stream. Whether the receiver will be designed to decode at lower levels and profiles is up to the receiver manufacturer.

---

[11] In accordance with PS-WP4 recommendation # 10.

A Main Level at Main Profile or lower MPEG-2 decoder cannot decode a GA bit stream for reasons of insufficient memory and speed, regardless of whether the particular GA bit stream contains GA extensions.

A Grand Alliance encoder can create a High Level at Main Profile MPEG-2 bit stream by "turning off" its syntax extensions.

5.5 MPEG-1 INTEROPERABILITY

MPEG-1 was the first standard to come out of the Moving Picture Experts Group, and was designed for interactive video applications (e.g. CD-ROM's) playing compressed video at about 1.5 Mbits/sec. Any MPEG-2 decode can decode an MPEG-1 bit stream (this is required for the decoder to be called "MPEG-2 compliant"). Whether Grand Alliance HDTV receivers will decode MPEG-1 bit streams is an option open to the receiver manufacturer. No currently existing MPEG-1 decoders can decode the Grand Alliance or any other MPEG-2 bit stream.

5.6 H.261 INTEROPERABILITY

H.261 is a predecessor to MPEG. It has a simpler syntax, and is intended for symmetric applications that have equal numbers of encoders and decoders, and that require low latency (i.e., a short encode/decode delay), such as videoconferencing. H.261 is almost but not strictly a subset of MPEG (it contains a filter in the coding loop that MPEG does not have). However, a programmable MPEG decoder should be able to decode H.261 fairly easily. Once again, whether a Grand Alliance HDTV receiver will decode H.261 is a manufacturer's option.

5.7. COMPUTER INTEROPERABILITY AT THE COMPRESSION LAYER

Like any other imaging system, HDTV can interoperate with computers at the picture layer of the architecture. (This topic will be discussed in the next chapter.) However, interoperability at the compression layer may be of much greater importance to computers. We anticipate that MPEG compression will become the computer industry standard for video, since it is an international standard. The GA/MPEG-2 compression has the potential to enable interoperability directly in compressed form, as shown in Figure 5.4.

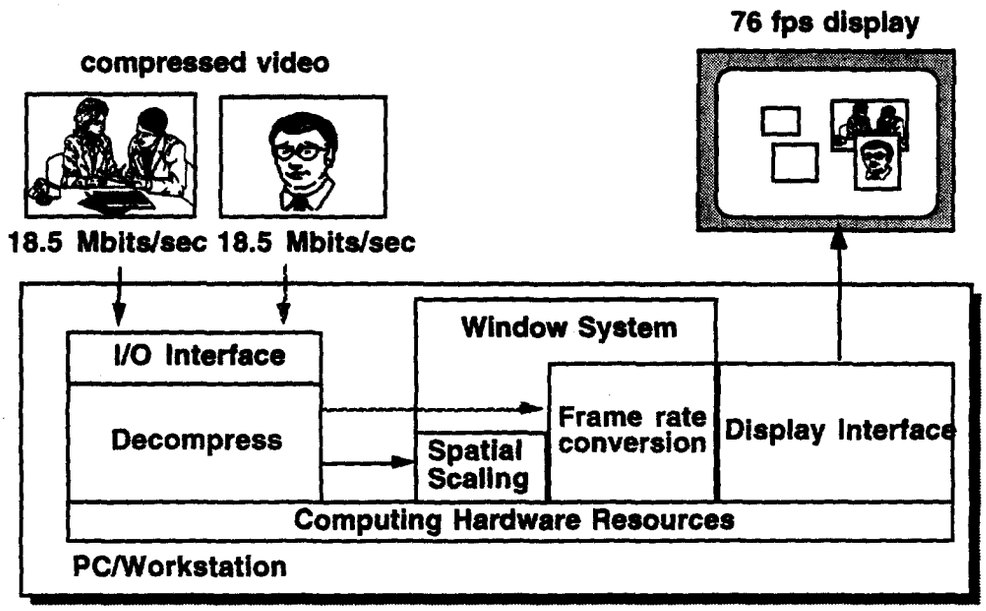

Figure 5.4 - Compression-level computer interoperability

This approach facilitates video and audio data interchange at rates of about 18 Mbps, instead of the approximately 500 Mbps that are required for full picture-layer interchange. Such lower data rate interfaces obviously result in significantly lower cost. Perhaps more importantly, such data rates with pictures in compressed form will form the basis for I/O of video from storage media and networks in multimedia applications. Computer software (and possibly some accelerator hardware) will perform decompression, with spatial scaling and frame rate conversion performed by the windowing system as part of its role of managing the computer's display.

5.8. AUDIO INTEROPERABILITY

The Grand Alliance is currently evaluating three proposed audio systems through subjective quality testing and plans to choose one of these in the near future. The three systems under test are:

- AC (MIT)
- AC-3 (Dolby)
- MUSICAM Surround (Philips)

All of the systems encode 5 channels of audio (left, right, center, and two channels of surround) using 320 to 500 kbits/sec. Each system has distinctly different interoperability advantages. MIT's AC system is the most flexible, encoding the five channels completely independently. Dolby's AC-3 system is already used for distribution in the film industry. MUSICAM Surround is a five-channel version of Musicam that is compatible with its two-channel predecessor that is part of the MPEG-1 standard and is also used in the Digital Compact Cassette

(DCC). Other influences include the direction or directions the U.S. Cable TV industry will move for audio for digital television over cable.

Audio interoperability is independent of video interoperability, due to the ability to perform the computation with programmable devices, and because of the packetized transport layer. Audio is relatively less difficult for HDTV because of the overwhelming preponderance of video data: audio represents abbot 2% of the HDTV data payload. For comparison, consider that audio for MPEG-1 at 1.5 Mbps represents about 17% of the data.

5.9. COMPRESSION LAYER INTEROPERABILITY SUMMARY

The Grand Alliance Compression Layer uses a superset of the MPEG-2 Working Draft Syntax which provides support for multiple formats at the Picture Layer through the use of MPEG-2's headers and descriptors. The GA HDTV system uses syntax elements that include field/frame coding, I, P, *and* B frames, and progressive refresh.

The GA system differs from MPEG-2 High Level at Main Profile in that it includes additional syntax for AC coefficient leaky prediction ("AC leak"). Liaison among the Grand Alliance, ACATS, and MPEG is on-going to harmonize the syntaxes.

The GA System interoperates with other compression syntaxes in various ways. Whether a GA receiver will be designed to decode other compression syntaxes will be determined by each receiver manufacturer individually, based on cost and market influences.

Grand Alliance

6. PICTURE LAYER INTEROPERABILITY

Perhaps the greatest debate over HDTV systems is about the use of a progressive scan or an interlaced scan picture format. There are many very compelling reasons to use each, and there seems to be no possible compromise. Fortunately, with digital technology and particularly through the use of MPEG's headers and descriptors, it is a simple matter to describe the particular picture format in the bit stream, and this is exactly what is done in the Grand Alliance System. The Grand Alliance HDTV system can encode either progressive or interlaced images, depending on the source. This allows application-by-application optimization of the picture format, all within a single transmission format.

6.1 A PLETHORA OF PICTURE FORMATS

There are a large number of existing and emerging formats in today's growing world of imaging, and choosing a single format to interoperate with all of them is impossible. The three formats that an HDTV system must clearly work well with in order to be successful are HDTV Production, Film and NTSC (including CCIR-601 component video).

NTSC is today's analog broadcast television system with 480 lines per frame, interlaced and no particular horizontal pixel structure (due to its analog nature). Each 4x3 aspect-ratio frame is made up of a pair of fields that are interlaced, and the field rate is 59.94 fields per second (1000/1001 x 60). CCIR-601 refers to a component digital version of the NTSC format with 720 x 480 pixels per interlaced frame. Note that CCIR-601 does not have square pixels.

Film provides very high analog spatial resolution, but relatively poor motion rendition because of its 24 Hz frame rate. Film is the oldest image format, and the tremendous number of new and old movies provide an extremely important source of programming material for both current television and HDTV. Film-originated productions currently make up about 80% of U.S. prime-time television. In addition, film is the primary format for international image exchange. There are several aspect ratios used in film production: 4x3, 5x3 (used in Europe), 16x9, 5.6x3, and 7x3 (Cinemascope). Film aspect ratio interoperates with TV using the technique of pan-and-scan, where a 4x3 window is panned to follow the area of interest. Since the film frame rate of 24 Hz is so much slower than the 59.94 Hz rate of NTSC, it maintains its progressive scan nature even during NTSC transmission, since a full 480 lines per film frame can be reconstructed. The frame rate of film is made to interoperate with the field rate of NTSC and CCIR-601 by showing odd film frames for three fields and even film frames for two fields (and slowing the film rate by 1000/1001 to 23.98). This technique is known as 3-2 pulldown.

Grand Alliance

HDTV Production standards are evolving at the same time as the transmission standard. The standard under development by the Advanced Television Standards Committee (ATSC) is a 16x9 aspect ratio, 1080 line interlaced system, at first with 59.94 frames per second and later allowing 60 frames per second (as the need for NTSC support decays). The ATSC has also stated that a future step to a 1080 line progressive scan format is planned (as camera technology allows). These two transitions parallel the planned migration path for Grand Alliance HDTV (see Chapter 8). Although the pixel formats for HDTV production and the Grand Alliance transmission formats may not be the same (e.g. 1920x1080 vs. 1280x720), the first round of HDTV systems testing verified that transcoding among picture formats can be achieved with extremely high quality.

There are also many other standard picture formats in use, and it is also important to interoperate with them as well. Here is a partial list of some current picture formats.

Major Picture Format Influences

Type	Format	Scanning	Frame/Field Rate	Square Pixels
NTSC	(analog) x 480	int	59.94	n.a.
CCIR-601 (U.S.)	720 x 480	int	59.94	No
Film	(analog)	pro	24	n.a.
U.S. HDTV Production	1920 x 1080	int (pro later)	59.94 (60 later)	Yes
SMPTE 240 M	1920 x 1035	int	60	No
Cable & Satellite Digital SDTV (CCIR-601)	720 x 480	int	59.94	No
Computer: VGA, Sun, HP, etc.	*wide variety*	pro	*wide variety*	Yes
PAL & SECAM	(analog) x 576	int	50	n.a.
Euro CCIR-601	720 x 576	int	50	No

6.2 FORMAT CONVERSION -- SPATIAL INTERPOLATION AND DEINTERLACING

With so many formats, it would seem that the world of imaging would be incapable of interoperating, but this is not so. A picture of one format may be converted to another simply by interpolating neighboring pixels either horizontally, vertically, or both, as shown in Figure 6.1. This is done, for example, when a computer window containing an image is re-sized by the user,

or when a television picture is resized with a special effects unit. Many approaches have been developed to do this *spatial interpolation*, and the tradeoff between conversion fidelity and interpolation complexity is well understood. The sophistication required of the interpolation algorithm depends on the application.

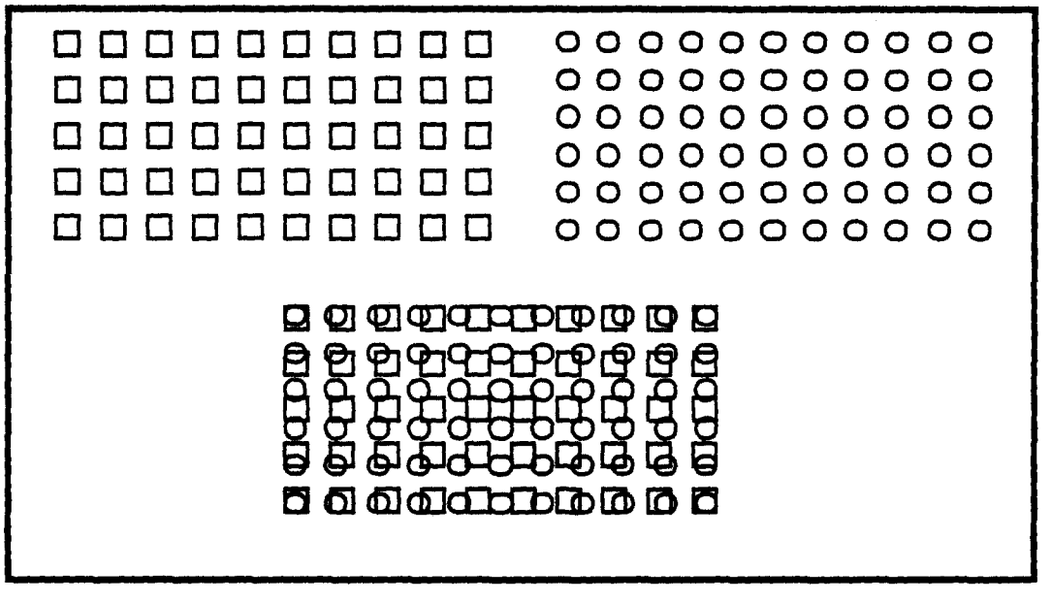

Figure 6.1. **Spatial Interpolation.** A picture may be converted from one scanning format to another by interpolation. A new pixel is created by calculating its position "in the cracks" relative to the original pixels, and interpolating a new value accordingly.

The concept of interpolation can be extended to the problem of conversion from interlaced scan to progressive scan images. When an interlaced image is converted to progressive scan, spatial and/or temporal interpolation is used to fill in the missing lines, as shown in Figure 6.2. A large number of approaches have been developed to solve this problem (driven by the desire to display progressive scan from NTSC in high-end NTSC receivers). As with spatial interpolation, there is a trade-off between deinterlacer complexity and picture quality, and different approaches are appropriate for different applications.

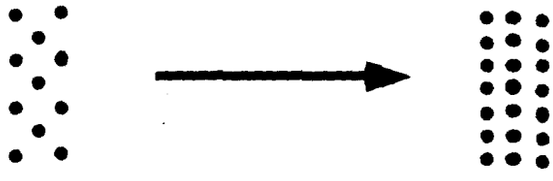

Figure 6.2. **Deinterlacing.** The problem of deinterlacing is very similar to that of spatial interpolation: missing samples (lines) must be filled in through the use of spatial and/or temporal interpolation. Many algorithms of varying degrees of complexity and accuracy currently exist.

It is essential to note that for more than a decade, the broadcast and post-production industries have been using mature solutions to the problem of interpolating and manipulating interlaced images with non-square pixels. Special effects equipment suppliers have developed signal processing algorithms and hardware that achieve performance nearly equivalent to optical-quality image manipulations.

6.3 GA PICTURE FORMATS

The long-term goal of the Grand Alliance is to deliver a progressively scanned 16x9 aspect ratio picture with about 1000 lines at all frames rates (24, 30 and 60). However, sending 1920 x 1080 x 60 pixels per second in an 18 Mbit/sec channel (0.096 bits/pixel) is well beyond the state of the art of practical video compression.

At the introduction of the GA system, two 16x9 aspect ratio picture formats will be supported at a variety of frame rates, as listed in Figure 6.3. The 720 line format is progressively scanned and has square pixels and is provided at three different frame rates[12]. The ≈1000 line format is progressively scanned at two frame rates, but interlaced at the highest temporal rate. *(The exact line number, 960 or 1080, and pixel format is not yet determined; see further discussion of the issues below.)*

Spatial	Temporal	
≈ X x 1000	23.97 / 24	progressive
	29.97 / 30	progressive
	59.94 / 60	interlaced
1280 x 720	23.97 / 24	progressive
	29.97 / 30	progressive
	59.94 / 60	progressive

Figure 6.3 **Grand Alliance Picture Formats and Frame Rates.** Two picture formats are supported each at three frame rates (and NTSC-friendly rates as well). All formats are progressively scanned except for the highest frame rate in the ≈1000-line format.

Each Grand Alliance receiver *must* decode and display a picture for each of these formats. Note that because of the decoupled nature of the transmission format and displays, the particular display a receiver manufacturer chooses to use (interlaced or progressive) may not correspond to the format in the transmitted bit stream.

---

12 In accordance with PS-WP4 recommendation # 9.

Thus even if the Grand Alliance bit stream delivered only progressive scan formats, this would not guarantee that receivers would use progressive scan in their displays. Indeed, because of the increased cost of progressive scan displays, some manufacturers might choose to use an interlaced display for all received formats. Consider the possible receiver architecture in Figure 6.4.

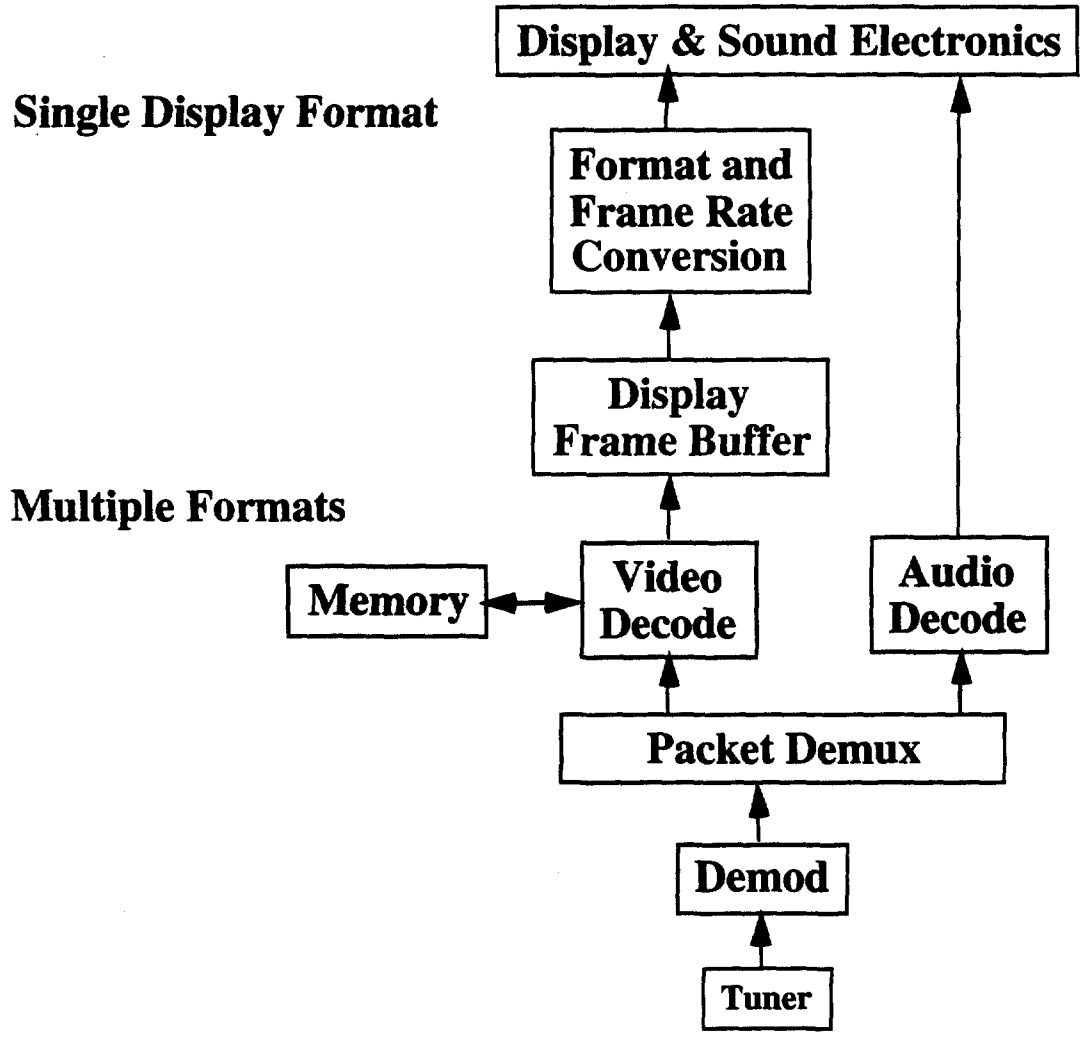

Figure 6.4. **A Single-Display Format GA Receiver Architecture.** Even though the Grand Alliance system supports multiple formats in its bit stream, for cost reasons some receivers may be built with a single-format display, and convert the various formats to match that display. This conversion is facilitated by the existence of a frame store required for decoding.

The receiver demodulates the carrier and separates the video and audio packet streams. Then the video stream is decoded and each picture, whatever its format, is placed in a frame buffer. Then this picture format is converted, both spatially (X and Y) and temporally (frame rate and interlace/progressive) for display.

If the receiver's display is 720 line progressive, Figure 6.5 lists the conversion operations required for each of the GA transmission formats:

Transmitted Format			Conversions to 720 Line Progressive Display
Spatial	**Temporal**		
≈ X x 1000	23.97 / 24	progressive	3:2 frame repeat and 3:2 interpolation
	29.97 / 30	progressive	2:1 frame repeat and 3:2 interpolation
	59.94 / 60	interlaced	2:3 interpolation/de-interlace
1280 x 720	23.97 / 24	progressive	3:2 frame repeat
	29.97 / 30	progressive	2:1 frame repeat
	59.94 / 60	progressive	none

Figure 6.5 **Receiver Format Conversions for 720/60 Progressive Display.** Low frame rates are increased to 60 frames per second simply by repeating each frame two or three times. Spatial formats are converted by interpolation, and deinterlacing is performed on the one interlaced format.

Low frame rates are increased to 60 frames per second simply by repeating each frame two or three times. Spatial formats are converted by interpolation, and deinterlacing is performed on the one interlaced format. If the receiver manufacturer chooses to use a fixed-format display at ≈1000 lines interlaced, Figure 6.6 lists the conversion operations:

Transmitted Format			Conversions to 1000 Line Interlaced Display
Spatial	**Temporal**		
≈ X x 1000	23.97 / 24	progressive	3:2 frame repeat and interlace
	29.97 / 30	progressive	2:1 frame repeat and interlace
	59.94 / 60	interlaced	none
1280 x 720	23.97 / 24	progressive	3:2 frame repeat and 3:2 interpolation
	29.97 / 30	progressive	2:1 frame repeat and 3:2 interpolation
	59.94 / 60	progressive	3:2 interpolation

Figure 6.6. **Receiver Format Conversions for ≈1000/60 Interlaced Display.** Low frame rates are increased to 60 frames per second simply by repeating each frame two or three times. Spatial formats are converted by interpolation, and interlacing is performed on the progressive formats.

6.4 THE ≈1000-LINE FORMAT ISSUE

The Grand Alliance and the Advisory Committee on Advanced Television Systems (ACATS) are currently discussing which spatial format to use for the ≈1000-line format. Three proposals are under consideration: 1728 x 960, 1440 x 1080, and 1920 x 1080. Each has its advantages, as conceptually illustrated in Figure 6.7.

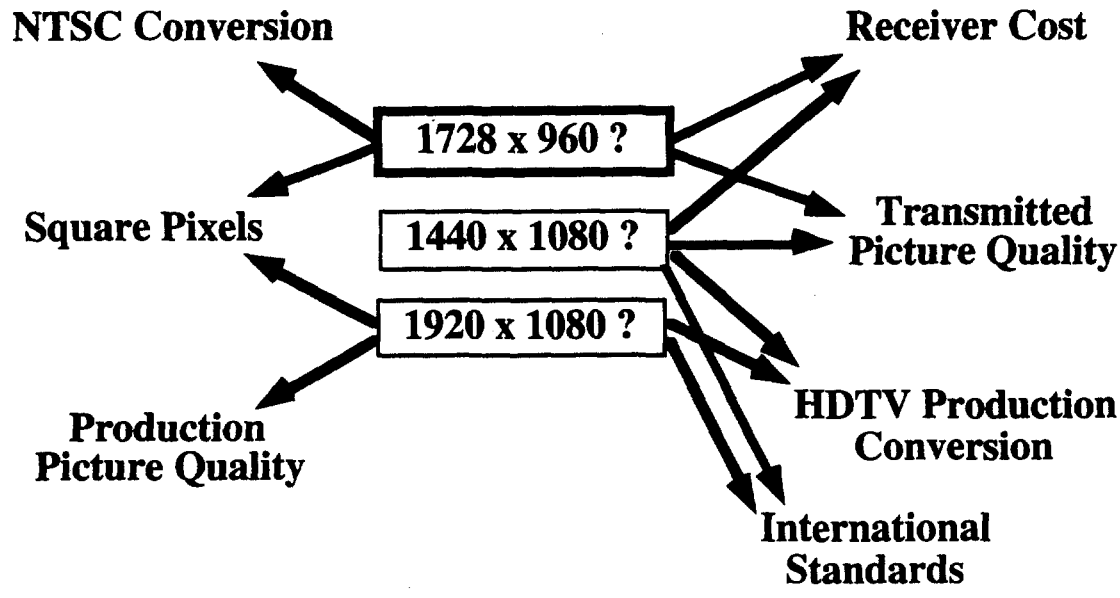

Figure 6.7. ≈1000 line formats under consideration by GA and ACATS. Each format has several important advantages. Two of three have square pixels. One (1728 x 960) has a 2:1 relationship to NTSC. Formats with lower total pixel counts will produce fewer compression artifacts (they operate at a higher bits/pixel coding rate). 1920 x 1080 will interoperate directly with the emerging international HDTV production standard.

The 1728 x 960 format has exactly twice the line count of NTSC, while 1440 x 1080 has exactly twice the horizontal pixel count of CCIR-601, which would ease spatial interpolation for conversion to NTSC. Formats with lower total pixel counts (1728 x 960 and 1440 x 1080) will produce fewer compression artifacts due to operating at a higher coding rate (bits/pixel). Formats with square pixels (1920 x 1080 and 1728 x 960) ease computation during geometric manipulation and provide better interoperability with computers. The 1920 x 1080 format is the emerging U.S. HDTV production standard. Thus, it is not immediately clear which ≈1000-line transmission format is appropriate; each has advantages and drawbacks. *As of this writing, the Grand Alliance is continuing its dialog with ACATS to select the ≈1000 line format.*

6.5 FILM INTEROPERABILITY

The Grand Alliance HDTV system has been designed to transmit film frames directly, encoded as progressively scanned frames at 24 frames per second using either ≈1000 or 720 lines. Film is relatively easy to compress because of its low frame rate.

Broadcasters and post-production houses will handle film exactly the same way as they do today: by transferring to 59.94/60 Hz tape using 3-2 pulldown and performing all edit and graphics operations at 59.94/60 Hz (not 24 Hz). The Grand Alliance HDTV encoder will automatically detect the presence of repeated frames and remove this redundancy, transmitting the images in 24 Hz progressive scan format. Conversion to a higher display rate at the receiver is facilitated by the

existence of a display frame store in the receiver. This frame buffer is read out multiple times to create 72 frames per second ("triple shutter") or 60 fields per second using 3-2 pulldown.

6.6 NTSC INTEROPERABILITY

The FCC mandate that the HDTV system operate in a simulcast mode with NTSC makes interoperation with NTSC of great economic importance.

Initially the frame rate for all formats will be 59.94 frames (fields) per second. This will simplify transcoding between HDTV and NTSC simulcast. Additionally, this will greatly simplify production for broadcast, avoiding mixed frame rate production and the timing difficulties which would result (e.g. calculating commercial insertion points in both drop-frame and non-drop-frame timecode for the same production).

In all cases, the formats in use or under consideration by the Grand Alliance have simple interoperability with NTSC:

- The 1280 x 720 format has a 3:2 vertical relationship to NTSC. Receivers can use a 3H (three times NTSC) horizontal deflection system for the lowest cost progressive scan receivers.
- The 1728 x 960 format has a 2:1 vertical relation to NTSC. Receivers can use a 2H (twice NTSC) horizontal deflection system for the lowest cost receivers.
- The 1440 x 1080 format has a 9:4 vertical relation to NTSC and a 2:1 horizontal relation to CCIR-601, but NTSC transcoding is slightly more difficult. Receivers can use a ≈2H horizontal deflection system for the low cost receivers.
- The 1920 x 1080 format has a 9:4 vertical relation to NTSC and an 8:3 horizontal relation to CCIR-601, but NTSC transcoding is slightly more difficult. Receivers can use a ≈2H horizontal deflection system for the low cost receivers.

Note that receivers can elect to use whatever display is appropriate for their target market. Future top-of-the-line progressive scan receivers displaying ≈1000 lines can use a 4H horizontal deflection system.

6.7 HDTV PRODUCTION INTEROPERABILITY

Perhaps it is stating the obvious to say that an HDTV transmission system must interoperate well with the HDTV production standard. The U.S. HDTV production standard under development by the Advanced Television Standards Committee (ATSC) is 1920 x 1080 59.94/60 Hz interlaced, with a future step to progressive scan. While conversion from the HDTV production format to the various Grand Alliance transmission formats may have important economic

implications, the first series of HDTV system tests showed that these transcoding operations could be performed with exceptionally high quality.

In all cases, the formats in use or under consideration by the Grand Alliance have simple interoperability with this production standard:

- The 1280 x 720 format has a 2:3 horizontal relation and a 2:3 vertical relation to the production format.
- The 1728 x 960 has a ≈8:9 horizontal relation and a 8:9 vertical relation to the production format.
- The 1440 x 1080 format has a 3:4 horizontal relation and a 1:1 vertical relation to the production format.
- The 1920 x 1080 format is identical to the production format.

6.8 30 FPS FILM INTEROPERABILITY

Film shot at 30 fps is sometimes used for film-for-TV production, especially for TV commercials. The increased frame rate yields reduced motion artifacts and avoids the use of 3-2 pulldown. It is a format that is particularly appealing for use in multimedia: it has a 25% faster frame rate than film, and yields a high-resolution, progressively scanned image.

The Grand Alliance HDTV System encodes and transmits 30 fps film frames directly at 30 frames per second as progressive frames using 720 or ≈1000 lines per frame and square pixels. Conversion to flicker-free display is done at receivers by repeating each frame twice for 60 fps display. The frame store required for doing this already exists due to the need for frame buffering for decompression.

The Grand Alliance HDTV system has been designed to transmit film frames directly, encoded as progressively scanned frames at 30 frames per second using either ≈1000 or 720 lines and square pixels. Film is relatively easy to compress because of its low frame rate.

6.9 COMPUTER INTEROPERABILITY

In cases where computers must import material produced by other industries/applications, conversion to square pixels and progressive scanning involves straightforward video signal processing as discussed in Section 6.2. Furthermore, in modern computer architectures, all applications update the screen through the control of a windowing system, which is a highly complex piece of software with commensurate hardware computing resources, as shown in Figure 6.8. In the windowing system environment, window resizing is a desirable operation. For images and video, it will become standard practice to scale the entire picture (rather than to simply crop it,

as is common in the case of text), by interpolation processing. This may be performed in dedicated hardware, or in software on sufficiently powerful general purpose processors. Thus, in computers with such windowing system capabilities, changes in pixel aspect ratio and deinterlacing can be readily supported. This approach is further necessitated by the inevitable requirement to deal with archived standard definition source material in NTSC, D1 and D2 formats, which are all based on rectangular pixels and interlaced scanning.

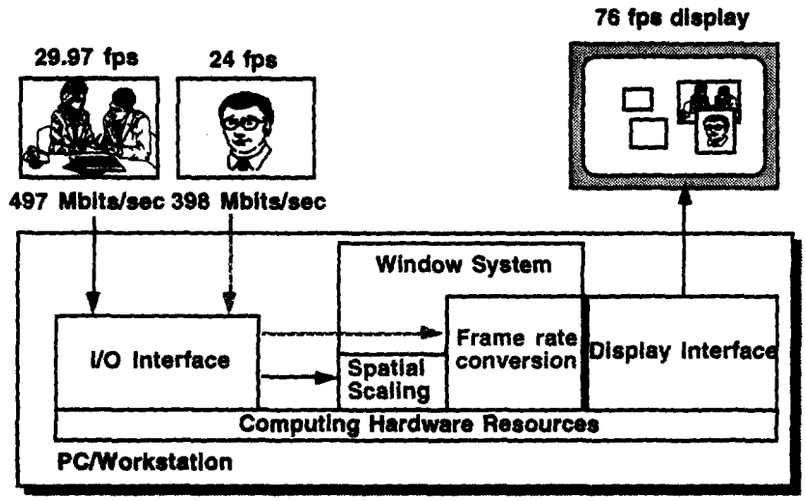

Figure 6.8 - Picture-level computer interoperability

The GA HDTV system's interoperability with computers is summarized in Figure 6.9. At the picture layer, the five progressive scan and square pixel format/frame rate combinations directly provide these essential attributes for computer applications. Likewise, due to MPEG header/descriptors, computer generated material can be directly delivered by the GA HDTV system. For computers importing material in other formats, format conversion processing will be performed under the control of advanced computer windowing systems that will perform spatial format conversions and interlaced-to-progressive scan conversions. Most importantly, as described in the previous chapter, the GA/MPEG-2 compression ensures that at the compression layer, compressed data streams can be interchanged between HDTV and computer applications and devices.

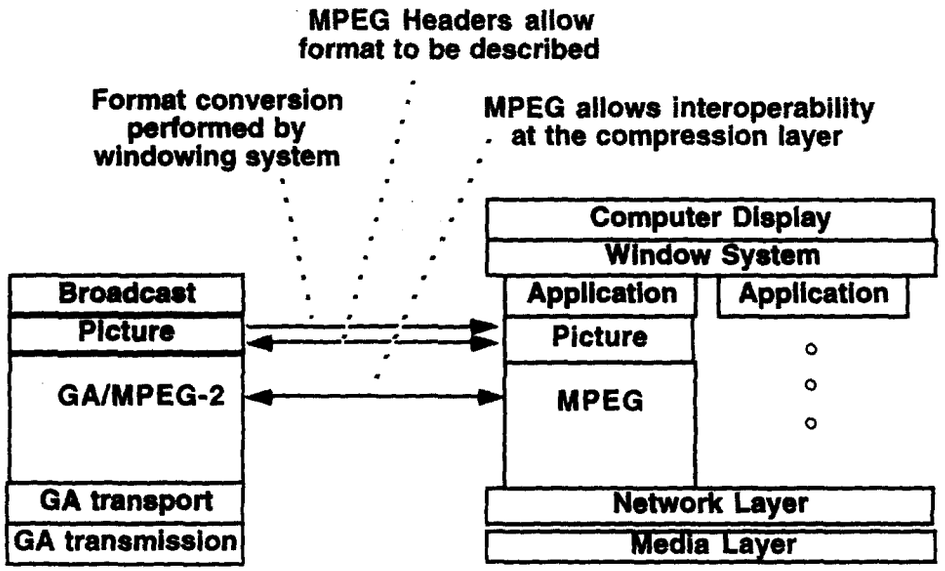

Figure 6.9 - Computer Interoperability Architectural Summary

6.10 PICTURE LAYER INTEROPERABILITY SUMMARY

The Grand Alliance provides syntax support for two picture formats, one with 720 lines with square pixels, and the other, yet to be determined, with about 1000 lines and possibly square pixels[13]. The multiple formats are described in the video stream itself using the standard MPEG-2 Sequence Header descriptor. Each format may operate at 24, 30, or 60 frames per second, or NTSC-friendly rates that are 1000/1001 slower (23.98, 29.97, 59.94). All of the format and frame rate combinations provide progressive scan except one, the ≈1000-line format which is interlaced when run at a temporal rate of 59.94/60 Hz.

Receiver display formats may not correspond to the transmission format. Rather, receiver manufacturers may choose a single display mode (e.g. 1280 x 720 at 60 progressive) and perform spatial and temporal conversion on the particular transmission format. This conversion is made less costly by already having frame buffers in order to do the image decompression.

[13]Square pixels in the ~1000-line format reduces the coding rate (bits/pixel) and thus the picture quality, but rectangular pixels causes interoperability problems with other systems. This issue will be decided soon.

7. OTHER INTEROPERABILITY CONSIDERATIONS

Creating a mass consumer market for HDTV will require low cost receivers and VCRs, both initially and in the future, as the market matures and consumer expectations rise. Also, interactive applications that will build on basic HDTV capabilities must be accommodated. The Grand Alliance system provides for these considerations in the following ways.

7.1 LOW-COST HDTV RECEIVERS

It is widely accepted that low cost HDTV receivers are essential to establish the high volume consumer market for HDTV[14]. During the competitive phase of the ACATS process, projected receiver cost data was collected by SS-WP3 from a wide variety of industry sources. This projected cost data was used to estimate receiver parts cost for each of the Grand Alliance predecessor systems. To the surprise of many, the systems were extremely close in parts cost, at a level of about $1000 for a 34" direct-view CRT receiver and about $1,500 for a CRT-based projection set. Of greater surprise to those not familiar with the cost structure of the television receiver industry, the largest cost factor is the display, sound and associated driver electronics. SS-WP3 data shows that for the $1000 parts cost direct-view CRT receiver, about $750 is display cost. Likewise, the $1500 parts-cost projection set has about a $1225 display cost. In both cases, all of the tuner, demodulator, packet decoder and decompression electronics cost about $125.

This data shows conclusively that the dominant factor in reducing the cost of HDTV receivers is to lower the cost of the display. The Grand Alliance approach of multiple format transmission does not require any particular display approach in the receiver. This means that the lowest cost HDTV receivers can use a ≈1000 line interlaced display with the associated 2H (twice the scan rate of NTSC) horizontal deflection. Other receivers can use a 720 line progressive display, with the associated 3H deflection. Still other receivers can use a ≈1000 line progressive display, with a 4H deflection. Any GA HDTV receiver must decode all of the transmitted picture formats, and perform a conversion between each transmitted format and its particular display. While this may sound complex, it involves relatively simple spatial interpolation and/or frame repetition.

The freedom of receiver manufacturers to choose an appropriate display to meet each quality/capability vs. cost segment of the market is essential to establish an HDTV industry in the near term. Equally important is the freedom to migrate that industry to higher quality displays over the lifetime of the HDTV standard, as consumer expectations and product demands grow to reflect the introduction of new services and applications.

---

[14] In accordance with PS-WP4 recommendation # 6.

7.2 LOW-COST HDTV VCRS

VCRs are an integral part of today's consumer television environment, and they will continue to enjoy this place in HDTV. Low cost digital VCRs for HDTV are another essential element of creating an HDTV market[15]. The dominant factor in VCR cost is the data rate that must be recorded. Digital VCRs for NTSC recording are already being planned. Current indications are that an NTSC DVCR will receive the analog signal, convert it to digital, compress it using inter-field spatial compression and will record a data rate of about 25 Mbps. In the case of digital HDTV transmissions, the data is already highly compressed to a data rate of about 18 Mbps. Therefore, the first-generation DVCRs will have adequate data rate capacity to directly record a compressed HDTV bit stream. This means that HD-VCRs will most likely be able to immediately make use of an established consumer technology base for the production of low cost digital VCRs.

Some of the most complex and important VCR considerations relate to "trick modes" such as fast forward and reverse. In these modes, only a small subset of the recorded data may be recovered, due to the different physical helix geometry traversed by a play head while the tape is moving at a speed other than 1x. The GA/MPEG-2 packet format has important advantages for such modes, since it offers fixed-length cells of constant size and format. As an encoder implementation option, certain packets can be identified as containing spatially coded I-frame data and/or intra-frame coded blocks. This allows special provisions to be made in the development of recording formats that use such data to advantageously implement trick modes. Of particular consequence is that spatially coded I-frame data is regularly available, and constitutes an entire picture. This means that trick mode picture quality may exhibit motion artifacts, but it will be temporally coherent. (The alternative of spatially coded data that is distributed over many frames will inherently exhibit significant picture breakup on scenes with high motion content.) This attribute of MPEG compression was specifically developed to allow trick modes for disk recording media, and the same advantages apply to VCRs as well.

7.3 LOW LATENCY ENCODING

For interactive applications, we anticipate that MPEG and/or H.261 compression coupled with ATM packet transmission will provide the basis. While throughput delay from MPEG's bidirectional motion estimation may appear to be a concern, in such applications, MPEG does not require the use of this capability. MPEG coding for applications requiring low throughput delay may in fact use forward predictive coding modes only, which reduces the minimum latency to 1 frame (33 msec). A similar approach will likely be used for low cost encoders in consumer

---

[15] In accordance with PS-WP4 recommendation # 7.