87-268

DOCKET FILE COPY ORIGINAL

RECEIVED

MAY - 3 1994

FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

GRAND ALLIANCE HDTV SYSTEM SPECIFICATION

Version 1.0

April 14, 1994

No. of Copies rec'd 1 Copy
List A B C D E

Table of Contents

RECEIVED
MAY - 3 1994
FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF THE SECRETARY

Executive Summary

Chapter 1 System Background

Chapter 2 Video Picture Format

Chapter 3 Video Compression System

Chapter 4 Audio Compression System

Chapter 5 Transport System

Chapter 6 Transmission System

Chapter 7 Grand Alliance System Summary

Chapter 8 Prototype Hardware Implementation

Chapter 9 Projected Prototype Performance

Chapter 10 Referenced Documents

SCOPE OF DOCUMENT

This document details the System Specification of the Grand Alliance HDTV System, and is intended to form the basis for the documentation of the proposed standard. It is comprised of the documents that have evolved from the work of the Grand Alliance Specialist Groups with the guidance and cooperation of the Advisory Committee on Advanced Television Services (ACATS) Experts Groups. This system specification document also details the prototype hardware, currently under construction. This prototype implementation will be delivered to the Advanced Television Test Center (ATTC) for verification of the performance of the proposed Grand Alliance HDTV system standard.

Executive Summary

Layered System Architecture with Header/Descriptors

The Grand Alliance HDTV System is a layered digital system architecture that uses headers/descriptors to provide flexible operating characteristics. The layers of the GA HDTV System, and some of their most important capabilities, are:

- the Picture layer provides multiple picture formats and frame rates
- the Compression layer uses MPEG-2 video compression and Dolby AC-3 audio compression
- the Transport layer is a packet format based on MPEG-2 transport, that provides the flexibility to deliver a wide variety of picture, sound and data services
- the Transmission layer is a Vestigial Sideband signal that delivers a net data rate of over 19 Mbps in the 6 MHz simulcast channel

While all of the GA HDTV system's layers operate in unison as an effective simulcast system, each layer has also been designed to have outstanding interoperability. The GA HDTV system's layered digital system approach with header/descriptors will create interoperability among a wide variety of consumer electronics, telecommunications, and computing equipment.

Picture Layer

The picture layer consists of raw pixel data, organized as pixels, scan lines and frames. The GA HDTV system provides for multiple formats and frame rates, *all of which will be decoded by any GA HDTV receiver*. This approach allows program producers and application developers to make their own tradeoffs among resolution, frame rate, compression artifacts and interlace artifacts, and to choose the format that is best for their particular use. The formats are:

Spatial Format (X x Y active pixels)	Temporal rate
1280 x 720 (square pixels)	23.976/24 Hz progressive scan
	29.97/30 Hz progressive scan
	59.94/60 Hz progressive scan
1920 x 1080 (square pixels)	23.976/24 Hz progressive scan
	29.97/30 Hz progressive scan
	59.94/60 Hz interlaced scan

Compression Layer

The compression layer transforms the raw video and audio samples into a coded bit stream -- essentially a set of computer instructions and data that are executed by the receiver to recreate the pictures and sound. The compression layer of the Grand Alliance HDTV system:

- uses video compression syntax that conforms to the ISO-MPEG (International Standards Organization — Moving Picture Experts Group) MPEG-2 video data compression draft standard, at a nominal data rate of approximately 18.9 Mbps.

- uses Dolby AC-3 audio compression to provide 5.1 channel surround-sound at a nominal data rate of 384 kbps.

Transport Layer

The transport layer separately packetizes video, audio and auxiliary data and allows their mix to vary dynamically, providing the flexibility needed to innovate new services and new kinds of programming. The transport layer of the Grand Alliance HDTV system:

- uses a packet format that is a subset of the MPEG-2 transport protocol.
- provides basic service capabilities that include video, five-channel surround-sound audio and an auxiliary data capacity.
- offers great flexibility in the mix of video, audio and data services that can be provided to appropriately-featured receivers. It separately packages each type of data (e.g., video, audio, etc.) in its own set of transmission packets. Each packet has a *Packet ID* header that identifies the content of the data stream. This capability enables the creation of new services, ranging from many stereo channels of audio, to broadcast distribution of computer software, to the transmission of very high resolution still images to computers.
- allows the mix of services to be dynamically allocated. This capability will allow rapid burst-mode addressing of receivers. It will also enable broadcasters to send multiple "streams" of video, audio and data programming to their audience, all complementing or enhancing the basic program content. This capability can fundamentally change the nature of television programming, since it enables software to be broadcast to "smart receivers" that can operate in conjunction with the HDTV picture and sound. With this capability, HDTV will likely become a more interactive medium than today's television, enabling new forms of educational and entertainment programming and games.
- provides important extensibility, since a Grand Alliance HDTV receiver will disregard any packet with a PID header that it does not recognize or cannot process. This will eliminate future "backward-compatibility" problems in the installed base of receivers, removing a crucial constraint from the introduction of new services.

Transmission Layer

The transmission layer modulates a serial bit stream into a signal that can be transmitted over a 6 MHz analog channel. The transmission layer of the Grand Alliance HDTV system:

- uses a trellis-coded 8-VSB modulation technique to deliver approximately 19.3 Mbps in the 6 MHz terrestrial simulcast channel
- provides a pilot tone that facilitates rapid signal acquisition and increases pull-in range.
- provides a training signal that facilitates channel equalization for removing multipath distortion.
- provides a related 16-VSB modulation technique to deliver two 19.3 Mbps data streams in a 6 MHz cable television channel

Summary

The GA HDTV system has the flexible operating characteristics that allow it to provide broad interoperability needed to form the basis for new and innovative services and applications of HDTV in many industries.

Chapter 1

SYSTEM BACKGROUND

RECEIVED
MAY - 3 1994
FEDERAL COMMUNICATIONS COMMISSION
OFFICE OF SECRETARY

A successful HDTV standard will be used for HDTV delivery to the public by terrestrial broadcasting and a variety of alternate delivery media, and it will also form the basis for new services and new applications of HDTV across a wide range of industries. In order for the enormous potential impact of digital HDTV to be realized, a high degree of interoperability is an essential attribute of an American standard.

1.1.1 Introduction

The formation of the Grand Alliance moves the HDTV standardization process from a competitive phase into a new collaborative phase. The previous competitive phase stimulated the rapid development of technology, as well as broad industry debate over various attributes in which the proposed systems differed. In the new collaborative phase, the Grand Alliance and the Advisory Committee must rapidly establish a national consensus and move forward into laboratory and field performance verification, and the documentation of a standard for FCC approval. Failure to achieve these objectives will harm every industry that is potentially advantaged by the deployment of HDTV systems and technology.

In some cases where the Grand Alliance confronted the dilemma of conflicting goals, it was technically feasible and reasonably cost-effective to allow a multiplicity of modes within the system. In these cases, the presence of modes favored by a particular industry do not preclude the inclusion of modes that are more favorably viewed by a different industry. The result is a system that is maximally useful to all industries, rather than one that burdens one industry at the expense of another.

In other cases where the Grand Alliance confronted the dilemma of conflicting goals, a choice had to be made among the conflicting alternatives. In these cases, there are good technical reasons for the choice that was made, and careful consideration has been given to interface requirements that bridge the differences between the approaches of different industries and make the system interoperable.

1.1.2 Scope of HDTV Standards and Their Impact

HDTV technology will not only revolutionize the television industry, but it will have a profound impact across a wide range of industries. The proliferation of high speed digital communications needed to deliver HDTV to the home will make the associated industries a key part of our National Information Infrastructure. Further, the core capability of cost-effectively delivering full-motion high resolution pictures and high-quality surround-sound will become a

cornerstone of image communications that may be adopted by a variety of industries and applications.

However, it is important to understand the scope of the GA HDTV System, which is a proposed transmission standard that is under consideration by the FCC through its Advisory Committee on Advanced Television Service (ACATS). The Grand Alliance HDTV system (and presumably the FCC standard) is NOT a production standard, or a display standard, or a consumer equipment interface standard. In fact, a digital compression and transmission system effectively DECOUPLES these various elements of the end-to-end imaging chain.

A digital HDTV transmission standard must specify sufficient information that allows any manufacturer to build a receiver that can receive a radio frequency (RF) transmission, and from it, produce pictures and sound. For a digital system, this means specifying the meaning of a bit stream and the signal format for transmitting those bits in a 6 MHz television channel. Thus, the picture formats, the compression syntax, the packet format and the RF signal format are examples of items that require standardization. While these standardized elements establish a minimum functionality for HDTV encoders and receivers, they do not specify encoder or receiver implementation. In the opinion of the GA, this traditional approach to transmission standards is highly desirable, because it enables natural competitive and economic forces to influence the implementation of encoders and receivers, which can evolve over time as underlying technology as cost changes take place.

1.1.3 Grand Alliance Goals

Since we are considering a proposed HDTV transmission standard that contains many complex tradeoffs, it is important to state the design goals that the GA established for itself. These goals essentially summarize the selection criteria that were established by ACATS during the competitive phase of the standards process. The GA design goals are to provide:

- High quality HDTV pictures and sound
- An effective simulcast system that:
  - Provides wide HDTV service area for terrestrial broadcast
  - Avoids unacceptable interference to existing NTSC service during its lifespan
- A cost effective solution for consumers, producers and all users at the time of introduction and over the life of the HDTV standard
- Interoperability with other media (e.g., cable, satellite, computers, etc.) and applications
- The potential for a worldwide standard

In this simple list of design goals, there are many conflicting desires. For example, there is a three-way tradeoff among HDTV picture quality, HDTV service area and interference to existing NTSC service. In order to have outstanding HDTV picture quality, a high data rate is desired.

But achieving a high data rate over a large HDTV service area requires high power transmission that results in too much interference to existing NTSC service. Reducing the HDTV service area is unacceptable, because there would be insufficient audience to make the service economically viable. Fortunately, with some balancing of objectives and associated engineering tradeoffs, the combination of video compression technology and digital transmission technology is sufficient to enable the creation of an HDTV simulcast service.

As another example of conflicting goals, even when considered solely as an entertainment medium, the HDTV transmission standard must have a high degree of interoperability with its common production sources, in order to ensure economical operation from production through delivery to the home. Even in this limited context, interoperability of one kind may be in conflict with interoperability of another kind. The production sources for HDTV transmission will include: the HDTV production standard, film, computer graphics, and NTSC video. The conflicting interoperability desires are that while the HDTV production standard will likely be interlaced with square pixels at 59.94 Hz, the desire for NTSC interoperability argues for preserving non-square pixels and the 59.94 Hz temporal rate, while the desire for film interoperability argues for a 24 Hz temporal rate, while the desire for computer interoperability argues for square pixels, progressive scan and possibly a frame rate higher than 70 Hz. Clearly, these conflicting interoperability desires are mutually exclusive, yet the interoperability of an HDTV transmission standard with all of these sources is important. A solution can be found that takes into account existing and widely accepted interoperability practices, such as the conversion of 24 Hz film to 59.94 Hz through a combination of a 1000/1001 slowdown to 23.976 Hz and a 3:2 frame repeat sequence to speed up the temporal rate to 59.94 Hz. In addition, the conversion between square pixels and non-square pixels must take place at some interface boundary; the question is simply where. Understanding the technical difficulty and cost of one conversion compared to another is an important element of the complex tradeoffs that must be made in the design of an HDTV system.

To further complicate our decision space, we observe that HDTV is on the vanguard of an explosion in digital technology. HDTV products will exist in a world where digital delivery of television by cable, satellite and pre-recorded media are proliferating; where the availability of low cost computational power and digital data networks are enabling the creation of a National Information Infrastructure; and where the development of new interactive services and applications will drive the creation of information appliances and products that may be hybrids of today's products, or entirely new products based on technologies that rapidly cross the computing, communications and consumer electronics industries.

1.2. System Overview

The Grand Alliance HDTV system employs two fundamental system principles that make it a highly interoperable system. First, it is designed with a layered digital system architecture, that results in the ability to interface with other systems at any layer. This means that many different applications can make use of various layers of the HDTV architecture. In addition, the GA has designed each individual layer of the system to be highly interoperable with other systems at corresponding layers. Secondly, the Grand Alliance HDTV system takes full advantage of the potential flexibility offered by a digital approach by using header/descriptor design principles that allow maximum flexibility to be achieved. As a truly flexible digital system, the Grand Alliance HDTV system provides the basis for an HDTV standard that will have a revolutionary impact on many industries and applications that are well beyond the scope of television today.

1.2.1 Layered Digital System Architecture

The GA HDTV system is a layered digital system that consists of four primary layers: the Picture layer, the Compression (video and audio) layer, the Transport (packetization) layer and the Transmission layer. These layers are conceptually illustrated in Figure 2.1.

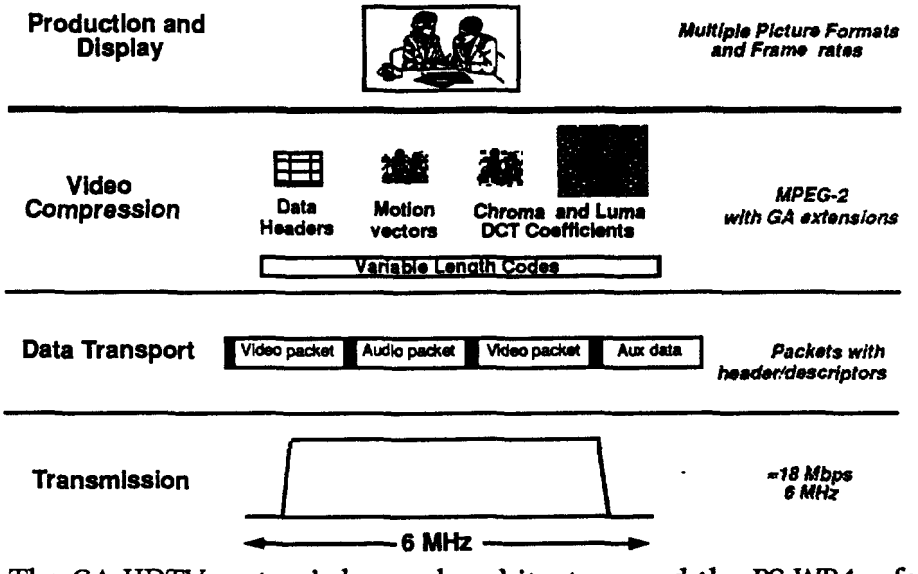


Figure 1.2.1 - The GA HDTV system's layered architecture and the PS-WP4 reference model.

While Picture layer Production and Display standards are not within the scope of a transmission standard, header/descriptors in the compression layer of the GA system support the transmission of multiple picture formats and frame rates that are related to both high-definition video and film production, and that are also appropriate for computer-based applications. At the Video Compression Layer, the GA HDTV system is based on MPEG-2 compression. At the

transport layer, the GA HDTV system is also based on MPEG-2, which uses a flexible ATM-like packet protocol with headers/descriptors. At the Transmission Layer, the VSB modulation system will provide a net user data rate of over 18 Mbps delivered in the 6 MHz terrestrial simulcast channel.

The GA HDTV system has been designed to provide interoperability at every layer of its architecture, and the PS-WP4 reference model provides an excellent framework to explain its interoperability features. Figure 1.2.2 shows the protocol stack that consists of the layers of the GA HDTV system (shown on the left) expanded in a form that corresponds to the reference model established by PS-WP4 (shown in the middle) as a basis for evaluating interoperability. Also shown is the ISO Open Systems Interconnect (OSI) reference model for data communications (on the right).

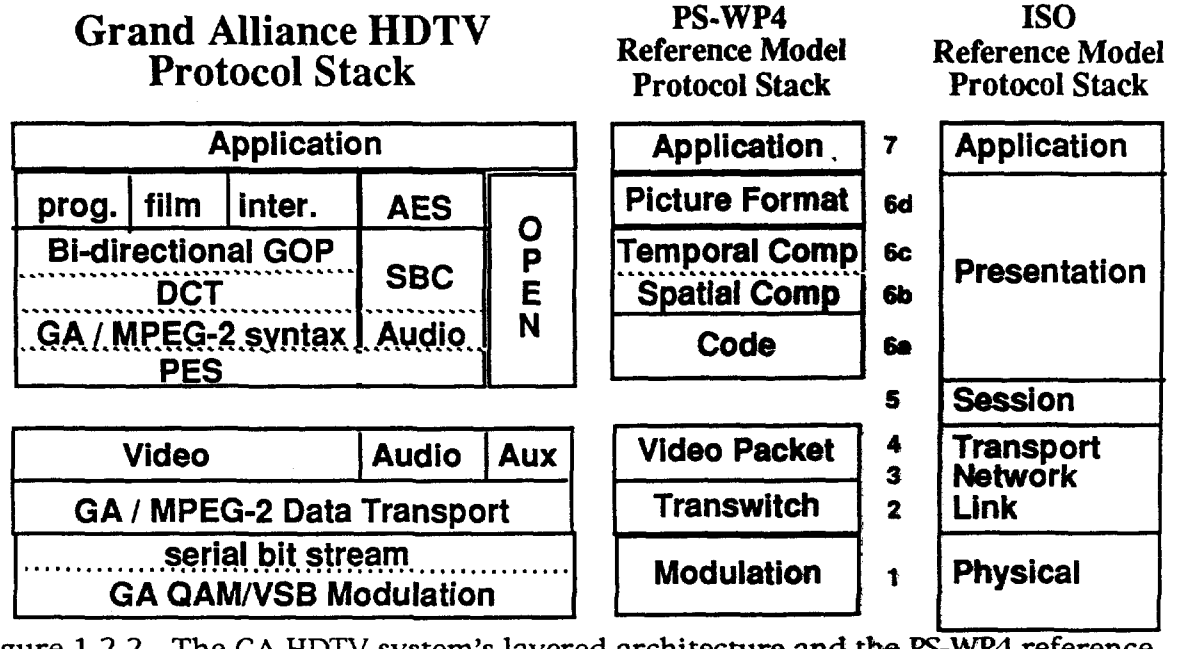


Figure 1.2.2 - The GA HDTV system's layered architecture and the PS-WP4 reference model

At the picture layer, progressive scan formats are provided at both video and film frame rates, in addition to an interlaced format. The compression layer conceptually consists of temporal compression, spatial compression and code (syntax) sublayers. For temporal compression, the MPEG Group of Pictures (GOP) approach is used to allow bi-directional motion-compensation across a block of frames. Spatial compression is achieved through the Discrete Cosine Transform (DCT). For the compressed code representation, the GA compression syntax is based on MPEG-2, but has additional syntax required to support AC coefficient leaky prediction. The Program Elementary Stream (PES) layer is the fundamental data stream that comprises a compressed video bit stream. At the video packet/transwitch layer, the GA system uses the MPEG-2 packet format, which provides the capability to synchronize the delivery of video,

audio and auxiliary data packets in an ATM-like multiplexing environment. Finally, at the transmission layer, the serial bit stream is converted to a 6 MHz signal by Vestigal Sideband Modulation. Subsequent chapters of this document will describe each layer of the architecture.

Chapter 2

VIDEO PICTURE FORMAT

2.1. Introduction

This document provides more detailed specifications of the scanning formats proposed on June 30, 1993 by the Grand Alliance, and includes modifications made since June 30 to the initial proposals. The document reflects extensive and valuable interactions between the Grand Alliance Specialist Group on Formats and the ACATS Technical Subgroup's Expert Group on Scanning Formats and Compression.

The parameters needed for specification of the HDTV systems are in two categories: those related to the interface to the encoder (on the transmission side), and those related to the encoding within the encoder in preparing the compressed data for transmission.

For the interface specification, the total line and pixel counts and the pixel clock rate are important. We have chosen to restrict the interface variations for the prototype to two formats, with 720 active lines and 1080 active lines. This limitation does not preclude other natural extensions for frame rates or for progressive scanning when such extensions become useful and equipment for new format variations is available.

For specifying the transmission format, which deals with the compressed representation of the active picture information, the total line and pixel counts are not relevant, but the frame rate variations can be distinguished and coded according to their frame rates are significant.

Parameters to be specified

The parameters that need to be specified for the interface and transmission formats include the following:

A. Number of frames per second.

B. Number of fields per frame (i.e., whether interlaced or progressive scanning is specified).

C. Number of active lines per frame.

D. Total number of lines per frame.

E. Number of active video samples per line.

F. Total number of video clock cycles per line.

G. Sampling frequency for video samples (derivative from the above specifications).

H. Allowable combinations of parameters.

Note that the HDTV system proposed by the Grand Alliance intrinsically includes more than one format, so that for some parameters there is more than one allowable value. Also note that 59.94 Hz and 60.0Hz frame/field rates are considered implementation variations, not defining separate formats.

2.2. Format Specifications

The Grand Alliance proposal includes two main format variations, with different numbers of lines per frame. These formats have 720 active lines and 1080 active lines per frame.

For interfacing to the encoder, the 720-line format uses 1280 active samples per line, while the 1080-line format uses 1920 active samples per line. These choices yield square pixels for all formats. Also for interfacing to the encoder, the 720-line format is progressively scanned, with 60 Hz (nominal) frame rate, while the 1080-line format is interlaced 2:1 with a 60 Hz nominal field rate. The prototype encoder input accepts 787.5 total lines per frame and 1125 total lines per frame for the two interface formats.

For compression and transmission, the frame rate for the 720-line format and for the 1080-line format can be 60.0 Hz, 24 Hz, or 30 Hz. The same formats are also supported with the NTSC-friendly numbers, i.e., 59.94 Hz, 23.98 Hz, and 29.97 Hz.

The 60.0 Hz and 59.94 Hz variations for the 1080-line format are encoded as interlaced scanned images for the initial HDTV systems. Other formats are encoded as progressively scanned.

2.2.1. Number of frames per second

The frame rates for the Grand Alliance HDTV system will include both 59.94 Hz and 60.0 Hz, since the 59.94 Hz rate may simplify interworking with NTSC material during the simulcast period, and because the 60.0 Hz rate may have interoperability benefits in the future.

The proposal includes 24 Hz and 30 Hz (including the corresponding adjustments for the NTSC-friendly frequencies) as film modes. This means that the encoder will be designed to encode the reduced pixel rate from image sequences that originated at 24 Hz or 30 Hz. It also means that for both the 720-line and 1080-line formats, the encoder will be designed to identify and take advantage of the lesser pixel rate in the film modes if the film mode material is presented to the encoder at 59.94 Hz or 60.0 Hz rate, using a pull-down technique.

Initial prototypes of the HDTV equipment will not contain interfaces for direct connection to 24 Hz or 30 Hz material, although the encoder architecture can support such interfaces when they become useful.

2.2.2. Number of fields per frame

The 1080-line format with 59.94 Hz or 60.0 Hz rate will be interlaced 2:1, yielding a frame rate of 29.97 Hz or 30.0 Hz. All the other format variations will use progressive scanning.

2.2.3 Number of active lines per frame

The number of active lines per frame will be either 720 active lines or 1080 active lines.

2.2.4. Total number of lines per frame

The total line count in the prototype for the 720-line format will be 787.5 lines per frame.

The total line count in the prototype for the 1080-line format will be 1125 lines per frame.

2.2.5 Number of active video samples per line

The number of active samples per line for the 720-line format will be 1280 samples of pixels.

For the 1080-line format, the progressive-scan variations with 24 Hz or 30 Hz frame rates will include 1920 active video samples per line. For the 59.94 Hz and 60.0 Hz interlaced variations of the 1080-line format, the number of active samples per line will be 1920 for the input interface.

The prototype encoder will be designed to support both 1440 and 1920 active samples per line for encoding and transmission, for the interlaced format only. The encoded data stream will include a signal that indicates the number of samples per line.

2.2.6 Total number of video samples per line

The total number of video samples per line for the 720-line format used in the prototype will be 1600, for all variations.

For the 1800-line format in the prototype, each line will represent 2200 cycles of the pixel rate clock, or 2200 sample times.

The Grand Alliance prototype also incorporates a variation of the 1080-line format with 1440 active samples per line.

2.2.7 Sampling frequency for video samples

For the 720-line format, the pixel rate clock used in the prototype will be 75,600,000 Hz for the 60.0 Hz frame rate, and approximately 75,524,476 Hz for the 59.94 Hz frame rate (which is 1000/1001 times the 60.0 Hz rate).

The clock rates used in the prototype for the 1080-line format will be 74,250,000 Hz for the 60.0 Hz interlaced variation, and approximately 74,175,824 Hz for the 59.94 Hz interlaced variation, at the input interface to the encoder.

Note that the ratio of clock rates for the specified 60 Hz inputs is 56:55 (75.6 MHz: 74.25 MHz).

2.2.8 Allowable combinations of parameters

Internally, the encoder for the 720-line format uses three frame rates, to allow the encoder to take advantage of film sequence redundancies. For the input to the encoder, a single format at either 59.94 Hz or 60.0 Hz will be interfaced to the encoder for the initial system.

The 1080-line format will have an input interface with 1920 active samples per line. As previously indicated, the Grand Alliance may reduce the number of samples per line for compression, for the interlaced format only, if necessary to preserve image quality. In that case, the number of samples encoded would be 1440 samples per line, for the interlaced variation at 59.94 or 60.0 Hz field rates. The progressive scan 24 Hz and 30 Hz variations with 1080 active lines will use 1920 active samples per line.

2.3 Summary Table for formats

The table attached shows the allowable variations. The table indicates entries for 59.94 Hz and 60.0 Hz field/frame rates, although these should not be considered distinct formats, merely frame rate implementation options.

GRAND ALLIANCE VIDEO SCANNING FORMATS
Encoder Format for Transmission

Active Lines	Active Pixels/Lines	Rate (Hz)	Pro/Int
720	1280	59.94/60.0	1:1
	1280	23.98/24.0	1:1
	1280	29.97/30.0	1:1
1080	1920/1440	59.94/60.0	2:1
	1920	23.98/24.0	1:1
	1920	29.97/30.0	1:1

Chapter 3

VIDEO COMPRESSION SYSTEM

3.1 OVERVIEW

The Grand Alliance (GA) HDTV system is an all-digital system designed to transmit high quality video and audio over a single 6 MHz terrestrial channel. A major challenge in the system design is video compression. This report summarizes the current status of the video compression subsystem of the GA HDTV system. The Grand Alliance compression specialist group is in the process of further evaluating the video compression subsystem and additional improvements maybe added in the future.

Modern digital transmission technologies deliver approximately 17 Mbps to about 20 Mbps to encode video data within a single 6 MHz terrestrial channel. This means that encoding the HDTV video source whose resolution is typically six times that of the NTSC resolution requires a bit rate reduction by a factor of 50 or higher. To achieve this bit rate reduction, the GA HDTV system is designed to be very efficient in utilizing available channel capacity by exploiting modern, sophisticated video compression technology. The methods utilized, for example, include source adaptive processing, motion estimation/compensation, transform representation, and statistical coding. The GA system has been designed to incorporate most key video compression technology utilized in previously tested systems. The GA system has also been designed to perform well with progressive as well as interlaced pictures. Due to efficient channel utilization, the GA video compression system is an excellent choice, not only for terrestrial broadcasting, but for cable and satellite broadcasting.

In designing the GA video compression system, we have emphasized compatibility with MPEG (Moving Picture Experts Group) syntax. For example, video compression techniques which are not compatible with MPEG syntax have not been considered unless significant benefit was clearly demonstrated. There are several reasons behind the emphasis on MPEG compatibility. The MPEG standard was developed utilizing modern sophisticated video compression methods and is very efficient in video compression. In addition, the MPEG standard was established by cooperative efforts of a number of organizations throughout the world. By placing special emphasis on compatibility with MPEG syntax, we have designed a video compression system which can serve as the HDTV standard not only for North America, but also for world-wide use.

We believe that the video compression subsystem of the GA HDTV system can deliver excellent HDTV quality video within a single 6 MHz terrestrial broadcast channel in North America. Furthermore, we believe that the system is an excellent choice for world-wide HDTV delivery for cable and satellite channels.

3.2 BASICS OF VIDEO COMPRESSION

A major objective of video compression is to represent a video source with as few bits as possible while preserving the level of quality required for the given application. Video compression has two major applications. One is efficient utilization of channel bandwidth for video transmission systems. The other is reduction of storage requirements. Both of these applications apply to the HDTV system.

Video compression for HDTV application is shown in Figure 3.2-1. The video source is encoded by the video encoder. The output of the video encoder is a string of bits that represents the video source. The channel coder transforms the string of bits to a form suitable for transmission over a channel through some form of modulation. The modulated signal is then transmitted over a communication channel. The communication channel typically introduces some noise, and provision for error correction is made in the channel coder to compensate for this channel noise. At the receiver, the received signal is demodulated and transformed back into a string of bits by a channel decoder. The video decoder reconstructs the video from the string of bits for human viewing. This report focuses on video compression, namely video encoding and decoding.

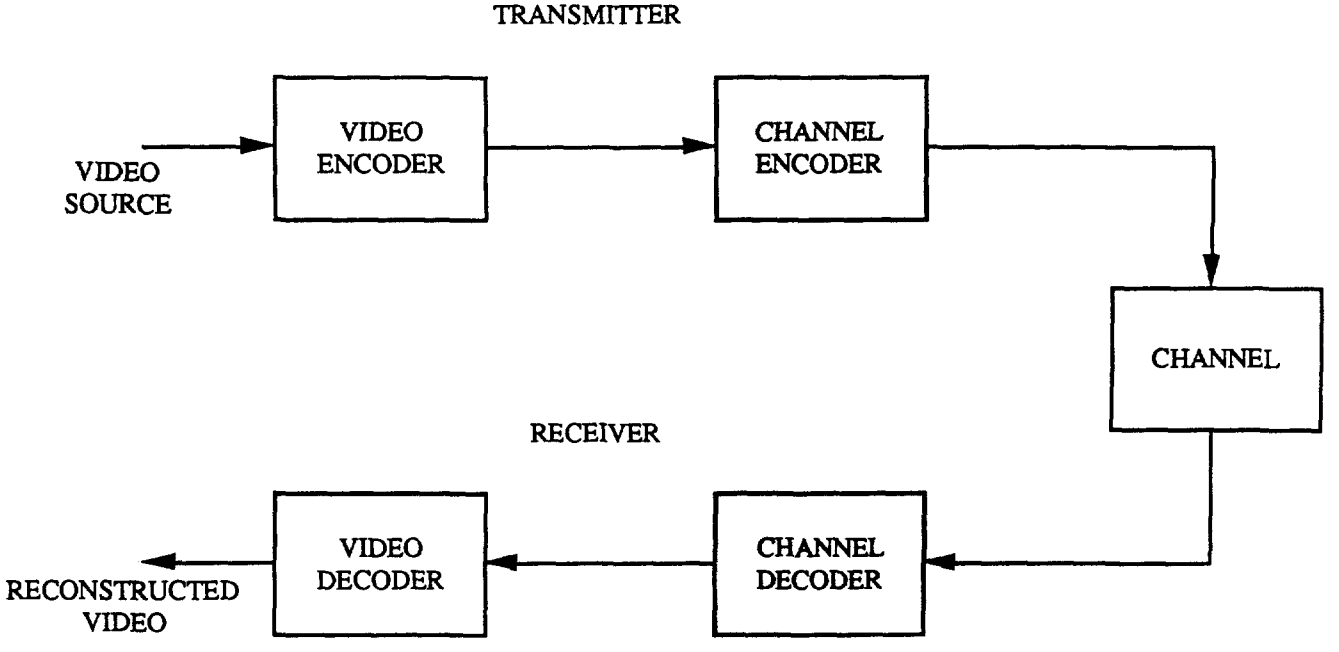


Figure 3.2-1. Video Compression for HDTV Application

The video compression system consists of three basic operations, as shown in Figure 3.2-2. In the first stage, the video signal is expressed in a more efficient representation which facilitates the process of compression. In essence, the representation determines what specifically is coded. The representation may contain more pieces of information to describe the signal than the signal itself, but most of the important information will be concentrated in only a small fraction of this description. In an efficient representation, only this small fraction of the data needs to be transmitted for an appropriate reconstruction of the video signal. The second operation, quantization, performs the discretization of the representation information. To transmit video over a digital channel, the representation information is quantized to a finite number of levels. The third operation is assignment of codewords, which are the strings of bits used to represent the quantization levels.

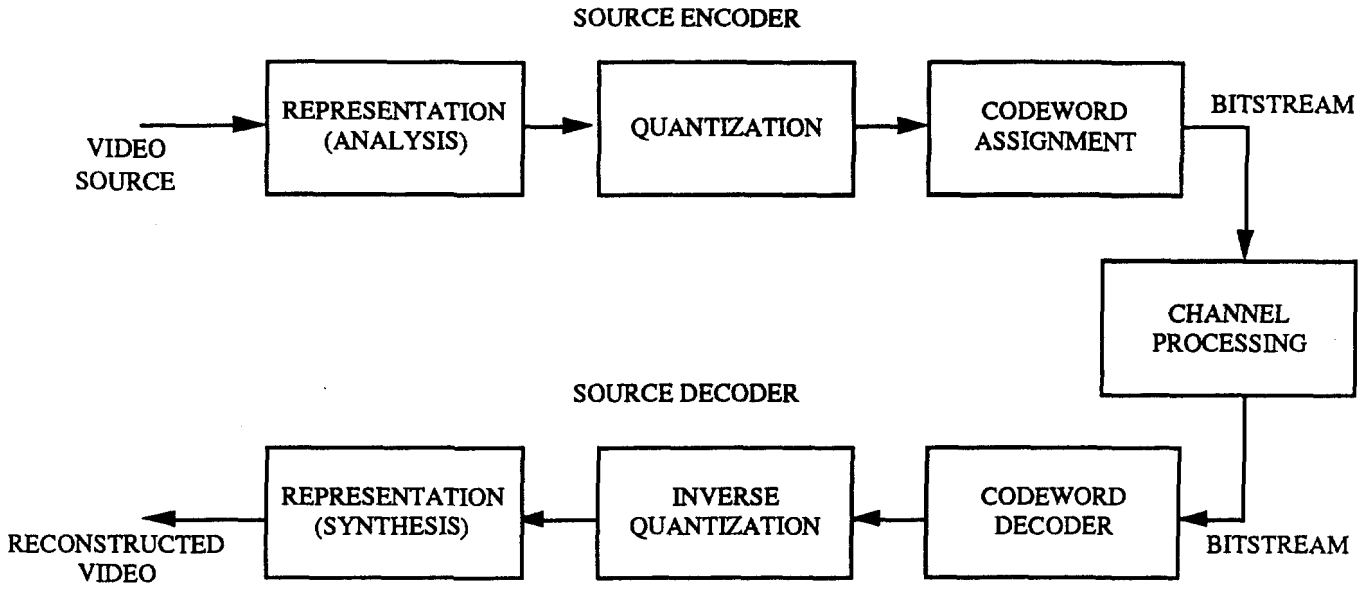


Figure 3.2-2. Overview of Video Compression System

Each of the three operations attempts to exploit the redundancy present in the video source and the limitations of the human visual system. By removing the redundancy present in the video source, the same or related information does not have to be transmitted repeatedly. By exploiting the limitations of the human visual system, the information in the video source that is not utilized by the human visual system does not need to be transmitted. In the following sections, we discuss the methods used in the GA video compression system.

3.2.1 Representation

To remove the redundancy in the signal and to eliminate the information irrelevant to the human viewer, the GA system utilizes many techniques discussed in this section.

3.2.1.1 Source-Adaptive Processing

A given television system must provide an interface to many kinds of video source formats. These include video from a video camera, the various film standards, magnetic and optical media, and synthetic imagery. In the GA system, we exploit the differences that may exist among the various sources to improve the performance of the video compression system.

3.2.1.2 Color-Space Processing

The input video source to the GA video compression system is in the form of R (red), G (green), and B (blue) components. The RGB components are highly correlated with each other. Furthermore, the human visual system responds differently to the luminance and chrominance components. To reduce the correlation and exploit this difference in the human visual system, the RGB components are converted to the YC1C2 color space through a linear transformation. Y corresponds to the luminance (intensity or black-and-white picture) while C1 and C2 correspond to the chrominance.

In the YC1C2 color space, most of the high frequency components are concentrated in the Y component. Furthermore, the human visual system is less sensitive to high frequency components of the chrominance components than of the luminance component. To exploit these characteristics,

in the GA video compression system the chrominance components are low-pass filtered and sub-sampled by a factor of two along both the horizontal and vertical dimensions, producing chrominance components that are one-fourth the spatial resolution of the luminance component.

3.2.1.3 Motion Estimation/Compensation

A video sequence is a series of still images shown in rapid succession to give the impression of continuous motion. Even though each of the frames is distinct, the high frame rate necessary to achieve proper motion rendition usually results in much temporal redundancy among the adjacent frames. Motion compensation attempts to remove this temporal redundancy.

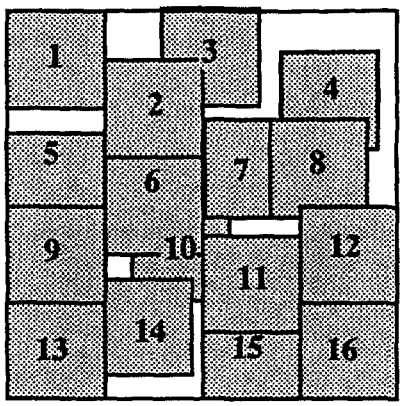
Much of the variation in intensity from one frame to the next is due to object motion. In motion-compensated coding, the current frame is predicted from the previously encoded frame by estimating the motion between the two frames and compensating for the motion. The difference between the current frame and the prediction of the current frame is called the motion- compensated residual and this residual is encoded. For a typical video sequence, the energy in the residual is much less than that in the original video due to the removal of the temporal redundancy. Encoding the residual rather than the video itself ensures that the temporally redundant information is not encoded repeatedly. In motion estimation, the same imagery is assumed to appear in consecutive video frames, although possibly at different locations. The motion may be global or local within the frame. To optimize the performance, an estimate of the motion is computed for each local region within a frame. The most common model for the local motion is simple translational motion. This model is highly restrictive, and cannot represent the large number of possible motions, such as rotations, scale changes, and other complex motions. Nevertheless, by assuming these motions only locally and by identifying and processing those regions separately where the model fails, excellent performance can be achieved.

One approach for motion estimation, which is very popular and is utilized in the GA video compression system, is based on block matching. In this approach, the current frame is partitioned into rectangular regions or blocks, and a search is performed for the displacement which produces the "best match" among possible blocks in an adjacent frame. Block matching is popular, because it achieves high performance and exhibits a simple, periodic structure which allows straightforward VLSI implementation. The process of motion-compensated prediction used in the GA system is illustrated in Figure 2-3. The frame to be encoded is partitioned into blocks and each block is predicted by displacing a block from the reference frame. The difference between the current frame and its prediction is the motion-compensated residual.

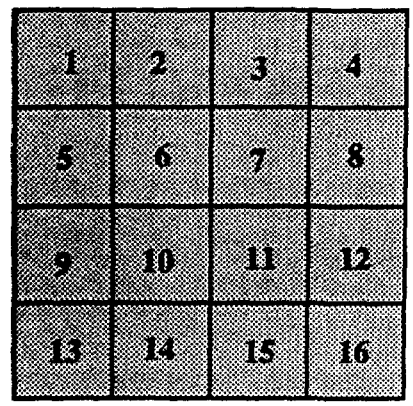
3.2.1.4 Intra-frame and Inter-frame Encoding

Motion-compensated coding is an example of inter-frame encoding. In some cases, it is useful to encode the video itself (intra-frame coding) without motion compensation. The GA video compression system utilizes both intra-frame coding and inter-frame coding.

Motion compensation is a form of predictive coding applied along the temporal dimension. For typical video frames, prediction by motion compensation is quite good. For some frames such as those at scene changes, however, a prediction is not as good. Even within a given frame, some regions may be predicted well with motion compensation while other regions may not be predicted very well using motion compensation based on translational motion. In some cases, therefore, the system may perform worse with predictive coding than by simply coding the frame itself. The GA video compression system utilizes inter-frame encoding when the motion- compensated prediction is quite good, while utilizing intra-frame encoding otherwise.



BLOCKS OF PREVIOUS FRAME USED TO PREDICT NEXT FRAME.



PREVIOUS FRAME AFTER USING MOTION VECTORS TO ADJUST BLOCK POSITIONS.

Figure 3.2-3. Motion-Compensated Prediction

The intra-frame coding mode is also utilized in the GA system for receiver initializations and channel acquisition (when the receiver is turned on or the channel is changed). It is also used when non-correctable channel errors occur. With the motion-compensated prediction, an initial frame must be available at the decoder to start the prediction loop. Therefore, a mechanism must be built into the system so that if the decoder loses synchronization for any reason, it can rapidly reacquire tracking. One approach used is periodic intra-frame encoding (refresh). In this approach, the entire frame is encoded using the intra-frame mode (I frames). Alternatively, a part of the frame is encoded using the intra-frame mode, and by encoding different parts of the frame in different frames, the entire frame is refreshed (progressive refresh). This refresh approach produces a periodic re-initialization of the temporal prediction which allows the decoder to re-acquire tracking. The GA system utilizes both progressive refresh and complete-frame refresh through the use of I-frames.

3.2.1.5 Discrete Cosine Transform

Motion compensation reduces the temporal redundancy of the video signal, but there still remains spatial redundancy in the motion-compensated residual (MC-residual). This is especially true if no MC processing is performed and the original frame is to be encoded. For simplicity, the term residual will be applied to whatever frame of data is to be spatially processed, irrespective of whether MC has been applied or not. To reduce the spatial redundancy of the residual, the Discrete Cosine Transform is used.

The Discrete Cosine Transform (DCT) compacts most of the energy of the residual into only a small fraction of the transform coefficients. The coding and transmission of only these energetic coefficients can result in the reconstruction of high quality video. The DCT was chosen in the GA system because it has good energy compaction properties, results in real coefficients, and there exist numerous fast computational algorithms for its implementation. The DCT of a two-dimensional signal may be computed by applying the one-dimensional DCT separably first to the rows and then to the columns of the signal.

The size of the DCT might be chosen as the entire frame, but much better performance can be achieved by subdividing the residual into many smaller regions each of which is individually processed. The motivation for this is very easy to see and it is one of the most important aspects of

a high quality video encoder. If we compute the DCT of the entire residual frame, we treat the whole residual frame equally. For a typical image, some regions contain a large amount of detail and other regions contain very little detail. By exploiting the changing characteristics of different images and of different portions of the same image, significant improvements in performance can be realized. In order to take advantage of the varying characteristics of the residual over its spatial extent, the residual is partitioned into 8x8 blocks. The blocks are then independently transformed and adaptively processed based on their local characteristics. The partitioning of the residual into small blocks before taking the transform not only allows spatially adaptive processing, but also reduces the computational and memory requirements. Partitioning the signal into small blocks before computing the DCT is referred to as the Block DCT.

The DCT tends to reduce the spatial correlation of the residual. This makes the DCT domain an appropriate representation since the DCT coefficients tend to have less redundancy. In addition, the DCT coefficients contain information about the spatial frequency content of the residual. By exploiting the spatial frequency properties of the human visual system, the DCT coefficients can be encoded to match the human visual system so that only the perceptually important DCT coefficients are encoded and transmitted.

3.2.2 Quantization

Through the processing discussed up to this point, an elegant representation in the form of the motion field, spatial frequency coefficients, and luminance/chrominance components has been created; however, no compression has been achieved. In fact, an expansion of data has resulted, since there are currently more pieces of information used to describe the video than before. However, most of the perceptually important information has been compressed into only a fraction of these "pieces of information", and this data can be selected and encoded for transmission.

The goal of video compression in the HDTV application is to maximize the video quality at a given bit rate. This requires a wise distribution of the limited number of available bits. By exploiting the statistical redundancy and perceptual irrelevancy within the new representation, an appropriate bit allocation can yield significant improvements in performance. Quantization is performed to discretize the values, and through quantization and codeword assignment, the actual bit rate compression is achieved. The quantization process can be made the only loss step in the compression algorithm. This is very important, as it simplifies the design process and facilitates fine tuning of the system. Quantization may be applied to elements individually (scalar quantization) or to a group of elements simultaneously (vector quantization).
In scalar quantization (SQ), each element may be quantized with a uniform (linear) or nonuniform (nonlinear) quantizer. The quantizer may also include a dead zone (enlarged interval around zero) to quantize or core to zero small, noise-like perturbations of the element value. The close relationship between quantization and codeword assignment suggests that separate optimization of each may not necessarily yield the optimum performance. On the other hand, joint optimization of quantization and codeword assignment is a highly nonlinear and complex process. Experiments have shown that a linear quantizer with an appropriate step-size individually chosen for each element to be quantized, followed by proper entropy coding, may yield close to optimum performance. This is the approach taken in the GA system.

When quantizing transform coefficients, the differing perceptual importance of the various coefficients can be exploited by "allocating the bits" to shape the quantization noise into the perceptually less important areas. This can be accomplished by varying the relative step-sizes of the quantizers for the different coefficients. The perceptually important coefficients may be quantized with a finer step-size than the others. For example, low spatial frequency coefficients

may be quantized finely, while the less important high frequency coefficients may be quantized more coarsely. A simple method to achieve different step-sizes is to normalize or weight each coefficient based on its visual importance. All of the normalized coefficients may then be quantized in the same manner, such as rounding to the nearest integer (uniform quantization). Normalization or weighting effectively scales the quantizer from one coefficient to another. The GA video compression system utilizes perceptual weighting, where the different DCT coefficients are weighted according to a perceptual criterion prior to uniform quantization. The perceptual weighting is determined by quantization matrices and the GA video compression system uses adaptive quantization matrices.

In video compression, most of the transform coefficients are quantized to zero. There may be a few non-zero low-frequency coefficients and a sparse scattering of non-zero high-frequency coefficients, but the great majority of coefficients will have been quantized to zero. To exploit this phenomenon in the GA system, the two-dimensional array of transform coefficients is reformatted and prioritized into a one-dimensional sequence through a zigzag scanning. This results in most of the important non-zero coefficients (in terms of energy and visual perception) being grouped together early in the sequence. They will be followed by long runs of coefficients that are quantized to zero. These zero-valued coefficients can be efficiently represented through run length encoding. In run length encoding, the number (run) of consecutive zero coefficients before a non-zero coefficient is encoded, followed by the non-zero coefficient value. The run length and the coefficient value can be entropy coded, either separately or jointly. The scanning separates most of the zero and the non-zero coefficients into groups, thereby enhancing the efficiency of the run length encoding process. Also, a special End Of Block (EOB) marker is used to signify when all of the remaining coefficients in the sequence are equal to zero. This approach is extremely efficient, yielding a significant degree of compression.

3.2.3 Codeword Assignment

Quantization creates an efficient discrete representation for the data to be transmitted. Codeword assignment takes the quantized values and produces a digital bit stream for transmission. The quantized values can be simply represented using uniform or fixed length codewords. Using this approach, every quantized value will be represented with the same number of bits. Greater efficiency, in terms of bit rate, can be achieved by employing entropy coding. Entropy coding attempts to exploit the statistical properties of the signal to be encoded. A signal, whether it is a pixel value or a transform coefficient, has a certain amount of information, or entropy, based on the probability of the different possible values or events occurring. For example, an event that occurs infrequently conveys much more new information than one that occurs often. By realizing that some events occur more frequently than others, the average bit rate may be reduced.

There are two important issues that arise when considering the application of entropy coding. First, entropy coding involves increased complexity and memory requirements over fixed length codes. Second, entropy coding coupled with the non-stationarity of the video signal results in a time-varying bit rate. (Other aspects of the source coding may also raise this issue). A buffer control mechanism is necessary when the variable bit rate source coder is to be coupled with a constant bit rate channel.

In the GA system, an entropy coder is used to reduce the statistical redundancy inherent in the elements encoded for transmission. The primary redundancy is the nonuniform probability distribution over the possible range of each element. The more the probability distribution deviates from a uniform distribution, the greater improvement can be achieved via entropy coding. Other sources of statistical redundancy which may exist include the statisticaldependence among the encoded elements.

Huffman coding which is utilized in the GA system is one of the most common entropy coding schemes. In Huffman coding, a code book is generated which minimizes the entropy subject to the codeword constraints of integer lengths and unique decodability. Events which are more likely to occur will be assigned shorter length codewords while those which are less likely to occur will be assigned longer length codewords. Huffman coding results in a variable bit rate per event; more importantly, the average bit rate is reduced. The training or generation of the code book is achieved by using a representative set of data to estimate the probability of each event. Optimal performance can be achieved by designing an individual code book for each element to be encoded. However, this results in a large number of code books. Close to optimal performance is achieved by using a new code books where elements with similar statistics are grouped and encoded together. Similarly, the size of each code book can be reduced by grouping together very unlikely events into a single entry within the code book. When any event belonging to this group occurs, the codeword for this group is transmitted followed by an exact description of the event.

Whenever entropy coding is employed, the bit rate produced by the encoder is variable and is a function of the video statistics. If the application requires a constant bit rate output, a buffer is necessary to couple the two. The buffering must be carefully designed. Random spikes in the bit rate can overflow the buffer while dips in the bit rate can produce an underflow. What is needed is some form of buffer control that would allow efficient allocation of bits to encode the video while ensuring that no overflow or underflow occurs.

The buffer control typically involves a feedback mechanism to the compression algorithm whereby the amplitude resolution (quantization) and/or spatial, temporal and color resolution may be varied in accordance with the instantaneous bit rate requirements. The goal is to keep the average bit rate constant and equal to the available channel rate. If the bit rate decreases significantly, a finer quantization can be performed to increase it. The buffer control mechanism is an essential part of any high-performance constant-output bit rate HDTV system. The GA system utilizes a highly sophisticated buffer control mechanism to deliver the highest video quality within a reasonable buffer size.

In this section, some of the basic video compression elements utilized on the GA video compression system have been discussed. Further details can be found in references 1 and 2.