

Editor's Message

The paper beginning on this page is not from the recent ICCE Conference. Because it gives an excellent overview and description of the proposed digital terrestrial HDTV system for North America, it is included as the first paper in this issue. This article is adapted from an article entitled "Digital Terrestrial HDTV for North America—The Grand Alliance HDTV System" published in the EBU Technical Review, No. 260 (Summer 1994).

DIGITAL TERRESTRIAL HDTV FOR NORTH AMERICA: THE GRAND ALLIANCE HDTV SYSTEM

Robert Hopkins, Senior Member, IEEE
ATSC
Washington, DC

Abstract

The Grand Alliance HDTV System has been designed for the needs and requirements of North America. The system has a great deal of flexibility to facilitate interoperability and is heavily based on international standards. The Grand Alliance and the FCC Advisory Committee on Advanced Television Service have been working together to complete the design of the Grand Alliance HDTV System. When a technical decision is made, technical performance is the number one priority in making the decision. The prototype is under construction and testing will begin late in 1994. This paper describes the technical characteristics of the Grand Alliance HDTV System.

1. Introduction

The Advisory Committee on Advanced Television Service (Advisory Committee) was formed by the United States Federal Communications Commission (FCC) in 1987 to "advise the FCC on the facts and circumstances regarding advanced television systems" for terrestrial broadcasting. The Advisory Committee objective also stated that the Advisory Committee should recommend a technical standard in the event the FCC decides that adoption of some form of advanced broadcast television is in the public interest. The Advisory Committee is organized into three subcommittees, one for planning, one for systems analysis and testing, and one for implementation. Further information on the objectives and organization of the Advisory Committee may be found in [1, 2].

From 1987 to 1991, many technical system proposals were made to the Advisory Committee. These proposals were analyzed by technical experts. Tests were planned. Only five proposals survived the rigorous process. Then in mid 1990, the first digital high definition television (HDTV) system was proposed to the Advisory Committee. Within seven months, three other digital HDTV systems were proposed. Tests on five HDTV systems (four digital, one analog) were conducted from September 1991 through October 1992. The results and conclusions were analyzed by the Special Panel of the Advisory Committee in February 1993 and are available in [3, 4]. A summary of the conclusions may be found in [5, 6]. In short, the Special Panel found that there are major advantages in the performance of digital HDTV systems, that no further consideration should be given to analog-based systems, that all of the systems produced good HDTV pictures in a 6 MHz channel, but that none of the systems was ready to be selected as the standard without implementing improvements. The Advisory Committee adopted the Special Panel report and encouraged the proponents of the four digital systems to combine their efforts into a "Grand Alliance." The Advisory Committee also authorized its Technical Subgroup to monitor ongoing developments.

Within three months, in May 1993, the proponents of the four digital systems agreed to combine their efforts. The resulting organization was called the "Digital HDTV Grand Alliance." The members of the Grand Alliance are AT&T, David Sarnoff Research Center, General Instrument Corporation,

Massachusetts Institute of Technology, Philips Electronics North America Corporation, Thomson Consumer Electronics, and Zenith Electronics Corporation.

In June 1993, the Grand Alliance submitted a preliminary technical proposal to the Technical Subgroup (video formats of 720 active lines and 960 active lines, video compression using MPEG-2 simple profile (no B-frames) with non-MPEG-2 enhancements, and MPEG-2 Transport Stream). Some subsystems were not specified by the Grand Alliance (audio compression and modulation), but were proposed to be the winner of subsystem tests to be conducted by the Grand Alliance. The Technical Subgroup began a review of the proposal within individual Expert Groups of the Technical Subgroup. The Expert Groups agreed with some portions of the proposal, and made various suggestions on possible changes to other portions. The Grand Alliance, with assistance from the Audio Expert Group, performed tests on three different multi-channel audio compression systems in July 1993.

In a meeting of the Technical Subgroup in October 1993, the Grand Alliance reported that their experiments showed that non-compatible enhancements to MPEG-2 did not produce a sufficient gain in picture quality to offset the loss of MPEG compatibility, that higher video compression performance could be obtained using B-frames, and that the AC-3¹ audio compression system exhibited the best overall technical performance in their tests. The Grand Alliance also reported that they had decided to replace the 960 active line video format with a 1080 active line video format. They proposed system characteristics as shown in Table 1. The Technical Subgroup approved the proposed subsystems.

The Grand Alliance, with assistance from the Transmission Expert Group, performed tests on 32 QAM (quadrature amplitude modulation) and 8-VSB (vestigial sideband) subsystems in January 1994. Both subsystems were tested also for high data rate cable transmission (256 QAM and 16-VSB). In February 1994, the Grand Alliance reported that the VSB system exhibited the best overall technical performance, and proposed that the modulation subsystem be VSB. The Technical Sub-

Table 1. Grand Alliance system characteristics.

Video formats	1280 (H) x 720 (V) progressive scan at 60 Hz, 30 Hz, and 24 Hz. 1920 (H) x 1080 (V) interlaced scan at 60 Hz, progressive scan at 30 Hz and 24 Hz Vertical rates also at 59.94 Hz, 29.97 Hz, and 23.98 Hz.
Video compression	MPEG-2 (Main Profile at High Level)
Audio compression	AC-3
Transport	MPEG-2 Transport Stream

group approved the proposal.² The VSB subsystem was subsequently subjected to field tests in Charlotte, North Carolina. Measurements were made at almost 200 sites during a three month period.

This completed the selection of all subsystems of the Grand Alliance HDTV System. The Grand Alliance was authorized to construct a prototype for testing by the Advisory Committee. Laboratory tests will begin late in 1994. Field tests will begin early in 1995. It is anticipated that the Advisory Committee will recommend adoption of the Grand Alliance HDTV System to the FCC during the second quarter of 1995 as the terrestrial HDTV broadcasting standard for the United States. The Advanced Television Systems Committee (ATSC) is documenting the Grand Alliance system for the FCC. The documentation is expected to be available at the same time the Advisory Committee makes its recommendation.

2. Technical overview of the Grand Alliance HDTV System

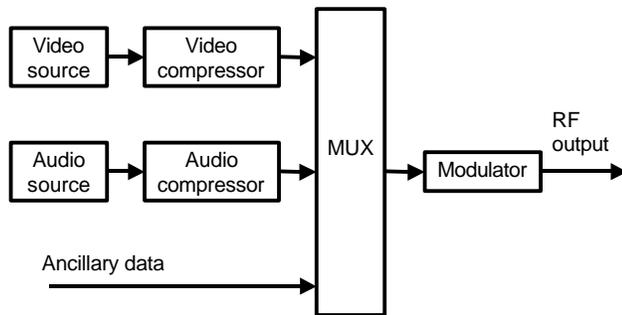
The Technical Subgroup has approved specifications of the Grand Alliance HDTV System [7]. The information contained in the technical description that follows was taken from those specifications.

A simplified diagram of the Grand Alliance HDTV System encoder is shown in Figure 1. The input video conforms to SMPTE proposed standards for the 1920x1080 system [8] or the 1280x720 system [9]. The input may contain either 1080 active lines or 720 active lines — the choice will be left to the user. In either case, the number of horizontal picture elements, 1920 or 1280, results in square pixels because the

¹ The AC-3 audio compression system was developed by Dolby Laboratories.

² The Advisory Committee also has been monitoring developments in coded orthogonal frequency-division multiplex (COFDM) technology. A number of broadcast organizations in North America have expressed interest in COFDM and are funding a program to develop and test a 6 MHz COFDM subsystem for comparison with the VSB subsystem.

Figure 1. Grand Alliance encoder.



aspect ratio is 16:9. With 1080 active lines, the vertical rate can be 60 (or 59.94) fields per second with interlaced scan. With 720 active lines, the vertical rate can be 60 (or 59.94) frames per second with progressive scan. If the video input is from scanned film, the encoder will detect the frame rate (30, 29.97, 24, or 23.98 Hz) and convert the 60 Hz video to progressive scan video at the film frame rate³. Although the Grand Alliance prototype will not be designed to directly accept inputs at the 30 or 24 Hz frame rate, this would be possible in Grand Alliance encoders in the future. Anticipating this possibility, SMPTE plans to document the 1080 and 720 proposed standards also at picture rates of 30 and 24 Hz.

Video compression is accomplished in accordance with the MPEG-2 Video standard [10] at the Main Profile/High Level. The video encoder output is packetized in variable-length packets of data called Packetized Elementary Stream (PES) packets. The video compression is explained in Section 3 of this paper.

Audio compression is accomplished using the AC-3 system [11, 12]. A standard for AC-3 is being documented currently by ATSC [13]. The audio encoder output also is packetized in PES packets. The audio compression is explained in Section 4 of this paper.

The video and audio PES packets, along with any ancillary data (which could be in the form of PES packets), are presented to the multiplexer. The output of the multiplexer is a stream of fixed-length 188-byte MPEG-2 Transport Stream packets. Both the PES packets and the Transport packets are formed in accordance with the MPEG-2 Systems standard [14]. The multiplex and transport are explained in Section 5 of this paper.

³ Throughout the remainder of this article, vertical rates of 60, 30, or 24 will be used. It should be understood that in each case, the vertical rate also can be 59.94, 29.97, or 23.98 (1000/1001 times 60, 30, and 24). The capability to use either set of numbers allows eventual phase-out of the NTSC-based vertical rates.

Table 2. Video specifications.

Video parameter	Format 1	Format 2
Active pixels	1280 (H) x 720 (V)	1920 (H) x 1080 (V)
Total samples	1600 (H) x 787.5 (V)	2200 (H) x 1125 (V)
Frame rate	60 Hz progressive / 30 Hz progressive / 24 Hz progressive	60 Hz interlaced / 30 Hz progressive / 24 Hz progressive
Chrominance sampling	4:2:0	
Aspect ratio	16:9	
Data rate	Selected fixed rate (10 - 45 Mbits/s) / Variable	
Colorimetry	SMPTE 240M	
Picture coding types	Intra coded (I) / Predictive coded (P) / Bidirectionally predictive coded (B)	
Video refresh	I-picture / Progressive	
Picture structure	Frame	Frame / Field (60 Hz only)
Coefficient scan pattern	Zigzag	Zigzag / Alternate zigzag
DCT modes	Frame	Frame / Field (60 Hz only)
Motion compensation modes	Frame	Frame / Field (60 Hz only) / Dual prime (60 Hz only)
P-frame motion vector range	Horizontal: Unlimited by syntax Vertical: -128, +127.5	
B-frame motion vector range (forward and backward)	Horizontal: Unlimited by syntax Vertical: -128, +127.5	
Motion vector precision	1/2 pixel	
DC coefficient precision	8 bits / 9 bits / 10 bits	
Rate control	Modified TM5 with forward analyzer	
Film mode processing	Automated 3:2 pulldown detection and coding	
Maximum VBV buffer size	8 Mbits	
Intra/Inter quantization	Downloadable matrices (scene dependent)	
VLC coding	Separate intra and inter run-length / Amplitude codebooks	
Error concealment	Motion compensated frame holding (slice level)	

The MPEG-2 Transport Stream packets are presented to the modulator where the data are encoded for the channel and a modulated carrier is generated. The channel coding and modulation are explained in Section 6 of this paper.

A summary of the specifications of the Grand Alliance HDTV System is given in the tables. Table 2 lists video specifications, Table 3 lists audio specifications, Table 4 lists transport specifications, and Table 5 lists transmission specifications.

Table 3. Audio specifications.

Audio parameter	
Number of channels	5.1
Audio bandwidth	10 - 20 kHz
Sampling frequency	48 kHz
Dynamic range	100 dB
Compressed data rate	384 kbits/s

Table 4. Transport specifications.

Transport parameter	
Multiplex technique	MPEG-2 Systems Layer
Packet size	188 bytes
Packet header	4 bytes including sync
Number of services	
Conditional access	Payload scrambled on service basis
Error handling	4-bit continuity counter
Prioritization	1 bit/packet
System multiplex	Multiple program capability described in PSI stream

3. Video compression

The bit rate required for an RGB HDTV studio signal with 1080 active lines, 1920 samples per active line, 8 bits per sample, and 30 pictures per second is $3 \times 1080 \times 1920 \times 8 \times 30 \cong 1.5$ Gbits/s with no bit rate reduction. To broadcast such a signal in a 6 MHz channel, with a service area comparable to the NTSC service area, requires the data rate to be compressed to something less than 20 Mbits/s, a factor of 75. Techniques that can be used to accomplish this compression are source-adaptive processing, reduction of temporal redundancy, reduction of spatial redundancy, exploitation of the human visual system, and increased coding efficiency.

Table 5. Transmission specifications.

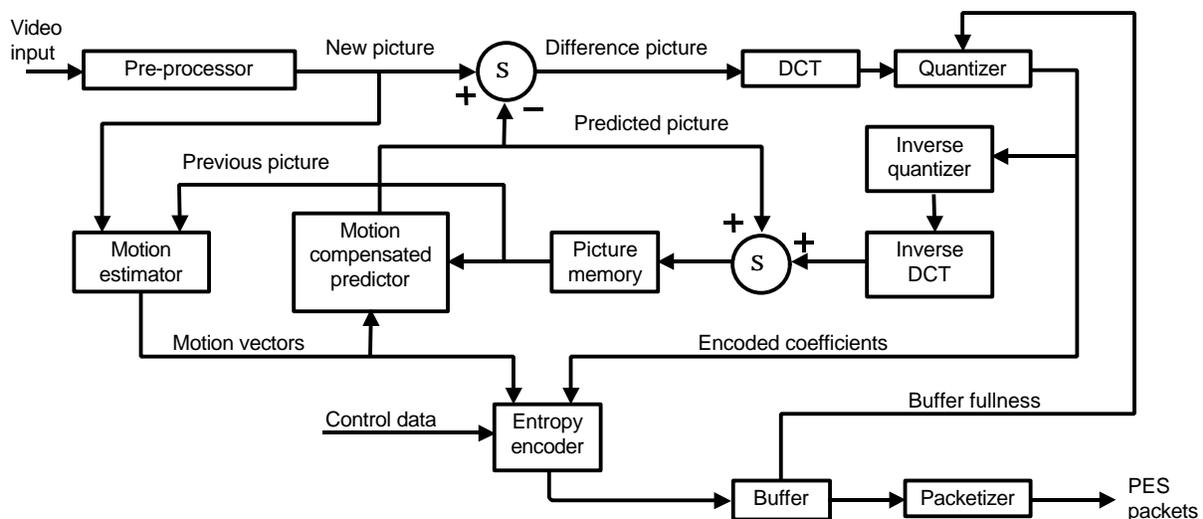
Transmission parameter	Terrestrial mode	High data rate cable mode
Channel bandwidth	6 MHz	6 MHz
Excess bandwidth	11.5%	11.5%
Symbol rate	10.76 Msymbols/s	10.76 Msymbols/s
Bits per symbol	3	4
Trellis FEC	2/3 rate	None
Reed-Solomon FEC	(208,188) T = 10	(208,188) T = 10
Segment length	836 symbols	836 symbols
Segment sync	4 symbols per segment	4 symbols per segment
Frame sync	1 per 313 segments	1 per 313 segments
Payload data rate	19.3 Mbits/s	38.6 Mbits/s
NTSC co-channel rejection	NTSC rejection filter in receiver	N/A
Pilot power contribution	0.3 dB	0.3 dB
C/N threshold	14.9 dB	28.3 dB

3.1. Video encoder

Source-adaptive processing is applied to the RGB components which, to a human observer, are highly correlated with each other. The RGB signal is changed to luminance and chrominance components to take advantage of this correlation. Furthermore, the human visual system is more sensitive to high frequencies in the luminance component than to high frequencies in the chrominance components. To take advantage of these characteristics, the chrominance components are low-pass filtered, and sub-sampled by a factor of two both horizontally and vertically.

Figure 2 is a diagram showing the essential elements in video compression. Temporal redundancy is reduced using the following process. In the motion es-

Figure 2. Video encoder.

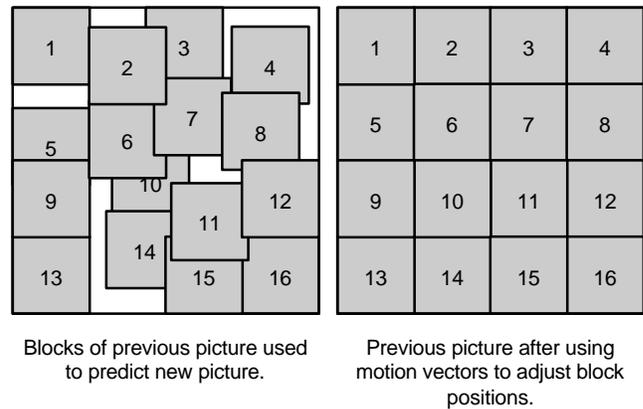


imator, an input video frame, called a new picture, is compared with a previously transmitted picture held in the picture memory. Macroblocks (an area 16 picture elements wide and 16 picture elements high) of the previous picture are examined to determine if a close match can be found in the new picture. When a close match is found, a motion vector is produced describing the direction and distance the macroblock moved. A predicted picture is generated by the combination of all the close matches as shown in Figure 3. Finally, the new picture is compared with the predicted picture on a picture element by picture element basis to produce a difference picture.

The process of reducing spatial redundancy is begun by performing a discrete cosine transform (DCT) on the difference picture using 8x8 blocks. The first value in the DCT matrix (top left corner) represents the DC value of the 64 picture elements of the 8x8 block. The other 63 values in the matrix represent the AC values of the DCT with higher horizontal and vertical frequencies as one moves to the bottom right corner of the matrix. If there is little detail in the picture, these higher frequency values become very small. The DCT values are presented to a quantizer which, in an irreversible manner, can “round-off” the values. Quantization noise arises because of rounding-off the coefficients. It is important that the round-off be done in a manner that maintains the highest possible picture quality. When quantizing the coefficients, the perceptual importance of the various coefficients can be exploited by allocating the bits to the perceptually more important areas. The quantizer coarseness is adaptive, and is coarsest (fewest bits) when the quantization errors are expected to be least noticeable. The DCT coefficients are transmitted in a zigzag order as shown in Figure 4. When the picture is interlaced, the DCT coefficients are read in an alternate zigzag fashion. After rounding, the higher frequency coefficients often have zero-value. This leads to frequent occurrence of several sequential zero-value coefficients.

The quantizer output is presented to an entropy encoder which increases the coding efficiency by assigning shorter codes to more frequently occurring sequences. An example of entropy encoding is the Morse Code. The frequently occurring letter “e” is given the shortest one symbol code while the infrequently occurring letter “q” is given a longer four symbol code. Another example is run-length coding where several sequential same-value coefficients can be represented with fewer bits by encoding the value of the coefficients and the number of times the coefficient is repeated rather than encod-

Figure 3. Predicted picture.

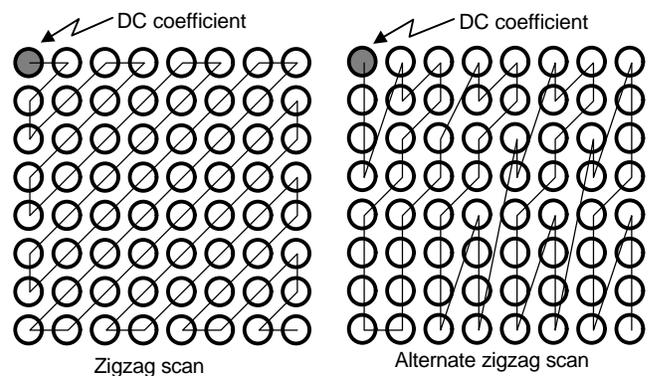


ing the value of each and every repeated coefficient. This is especially useful when the higher frequency DCT coefficients have zero-value. Run-length coding is used in the Grand Alliance system. Huffman coding, also used in the Grand Alliance system, is one of the most common entropy encoding schemes.

The entropy encoder bit stream is placed in a buffer at a variable input rate, but taken from the buffer at a constant output rate. This is done to match the capacity of the transmission channel and to protect the decoder rate buffer from overflow or underflow. If the encoder buffer approaches maximum fullness, the quantizer is signaled to decrease the precision of coefficients to reduce the instantaneous bit rate. If the encoder buffer approaches minimum fullness, the quantizer is allowed to increase the precision of coefficients. The output of the buffer is packetized as a stream of PES packets.

Because the transmitted picture is required also at the encoder for the motion compensated prediction loop,

Figure 4. Scanning of DCT coefficients.



the quantizer output is presented to the inverse quantizer, then to the inverse DCT, summed with the predicted picture, and then placed in the picture memory.

In the description thus far, it has been assumed that the picture used to predict the new picture was, in fact, the previous picture from the video source. An advantage may be gained, in some cases, by predicting the new picture from a future picture, or from both a past and a future picture. For example, after a video switch, a future frame is a better predictor of the current frame than is a past frame. In the MPEG standard, three types of frames are defined. An I-frame is a picture which is transmitted as a new picture, not as a difference picture. A P-frame is a picture which is predicted from a previous P- or I-frame. A B-frame is a picture which is predicted from both a past P- or I-frame and a future P- or I-frame. This is illustrated in Figure 5. Inclusion of B-frames requires an additional frame of storage in the decoder. Before the information describing the B-frame can be transmitted, the information for both anchor frames must be transmitted and stored. As a result, the transmission order is different from the display order.

Because the two fields of an interlaced picture represent two different points in time, they can vary significantly when there is a lot of motion. In such a case, it may be preferable to make the motion compensated prediction based on fields rather than frames. This choice is facilitated by allowing both prediction modes. Another prediction mode, dual prime, is supported also. Dual prime is available only for interlaced video material and only when B-frames are not in use.

It allows motion vectors determined in one field to be used in the other field.

With a motion compensated prediction loop, refreshing the received image is necessary whenever the receiver is first turned on or tuned to another channel, after a loss of signal, or when major transmission errors occur. In each case, the picture in the receiver memory will be different from the picture in the memory at the encoder. Because the transmitter cannot know when the pictures are different, it is necessary to transmit periodically the new picture, rather than the difference picture. Otherwise, errors will propagate in the receiver. Two refresh methods are allowed, I-frame refresh and progressive refresh. With I-frame refresh, an entire frame is transmitted at a periodic rate. This is accomplished by transmitting the DCT coefficients of the new picture in place of the DCT coefficients of the difference picture. With progressive refresh, the DCT coefficients of a group of blocks (macroblock) of the new picture are transmitted at a periodic rate in place of the DCT coefficients of the same group of blocks of the difference picture.

3.2. Video decoder

The video decoder is shown in Figure 6. Following de-packetizing of the PES packets, the encoded coefficients and motion vectors are held in a buffer until they are needed to decode the next picture. The entropy decoder performs the inverse function of the entropy encoder. The encoded coefficients, after in-

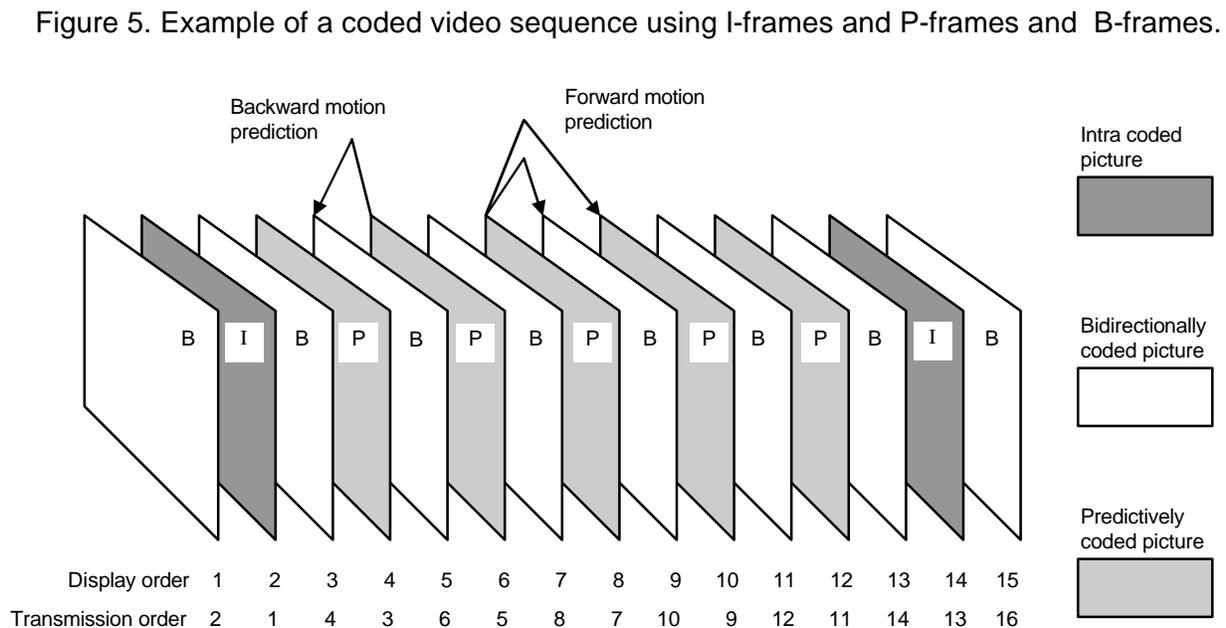
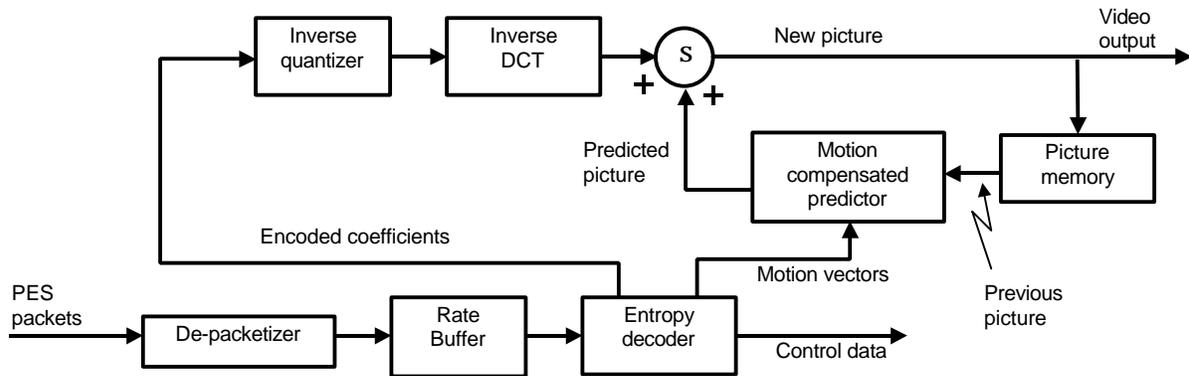


Figure 6. Video decoder.



verse quantization and inverse DCT, are added to the predicted picture to produce the new picture. The predicted picture was obtained by using the received motion vectors to move portions of the previously transmitted picture.

4. Audio compression

The Grand Alliance audio system uses AC-3 technology. The main audio service can range from a simple monophonic service, through stereo, up to a six channel surround sound service (left, center, right, left surround, right surround, and subwoofer). The sixth channel conveys only low frequency (subwoofer) information and is often referred to as 0.1 channel for a total of 5.1 channels. Several services, in addition to the main audio service, can be provided. Examples are services for the hearing or visually impaired, dynamic range control, and multiple languages.

When the audio service is a multi-channel service and mono or stereo outputs are required in the receiver, the downmix is done in the decoder. The downmix may be done in the frequency domain, reducing the complexity of mono and stereo receivers. The program originator can indicate in the bit stream which downmix coefficients are appropriate for a given program.

The audio sampling rate is 48 kHz. With six channels and 18 bits per sample, the total bit rate before compression is $48000 \times 6 \times 18 \cong 5$ Mbits/s. The compressed data rate is 384 kbits/s for the 5.1 channel service representing a compression factor of 13.

4.1. AC-3 encoder

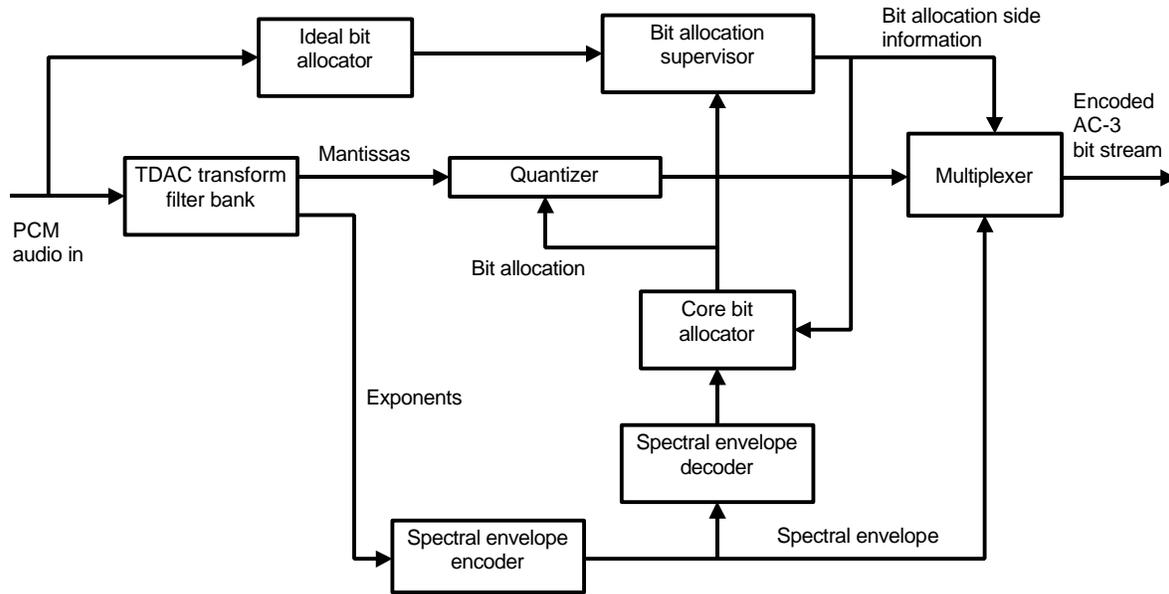
Due to the frequency masking properties of human hearing, a frequency domain representation of audio is used in the bit rate compression. As shown in the diagram of the AC-3 encoder in Figure 7, the audio input

channels are transformed from the time domain to the frequency domain using the Time Domain Aliasing Cancellation (TDAC) transform. The block size is 512 points. Each input time-point is represented in two transforms. The 512-point transform is done every 256 points providing a time resolution of 5.3 ms at a 48 kHz sampling rate. The frequency resolution is 93 Hz and is uniform across the spectrum. During transients, the encoder switches to a 256-point transform giving a time resolution of 2.7 ms.

The output of the TDAC transform is a set of frequency coefficients for each channel. Each transform coefficient is encoded into an exponent and a mantissa. The exponent provides a wide dynamic range. The mantissa is encoded with limited precision, resulting in quantizing noise. The exponents of each channel are encoded into a representation of the overall signal spectrum, referred to as the spectral envelope. The time and frequency resolution of each spectral envelope is signal dependent. The frequency resolution varies from 93 Hz to 750 Hz, depending on the signal. The time resolution varies from 5.3 ms to 32 ms. The algorithm that determines the time and frequency resolution of the spectral envelope is in the AC-3 encoder only, and thus may be improved in the future without affecting decoders in the field. The AC-3 encoder decodes the spectral envelope to make use of the identical information that will be available in the receiver. The decoded version is used as a reference in quantizing the transform coefficients and in determining the bit allocation.

Allocation of bits to the various frequency components of the audio signals is a critical part of the encoder design. AC-3 makes use of hybrid forward/backward adaptive bit allocation. With forward bit allocation, the encoder calculates the bit allocation and explicitly encodes the bit allocation into the bit stream. This method allows for the most accurate

Figure 7. AC-3 audio encoder.



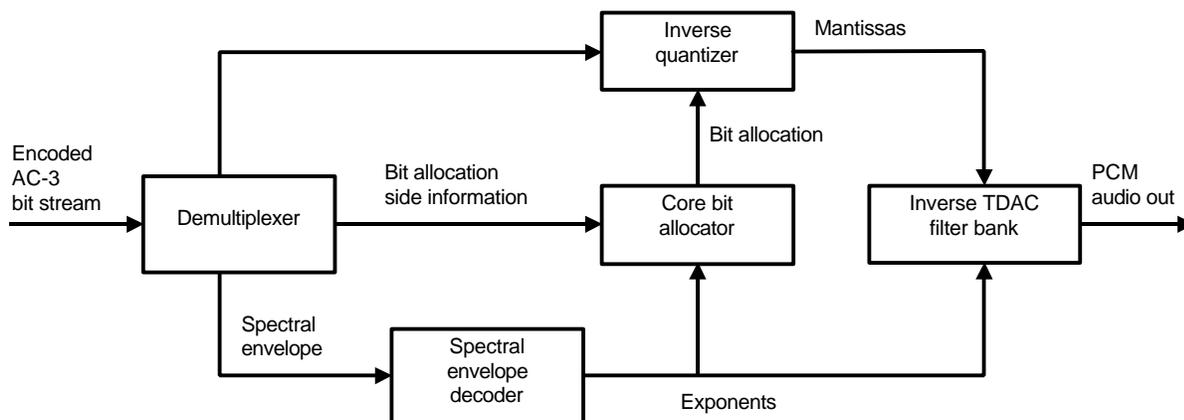
bit allocation because the encoder has full knowledge of the input signal. Also, the psychoacoustic model is resident only in the encoder and may be improved without affecting decoders in the field. With backward bit allocation, the bit allocation is calculated from the encoded data without explicit information from the encoder. This method is more efficient because all of the bits are available for encoding audio. Disadvantages of backward bit allocation are that the bit allocation must be computed from information in the bit stream which is not fully accurate, and that the psychoacoustic model cannot be updated because it is included in the decoder. The AC-3 encoder, with a hybrid forward/backward adaptive bit allocation, has a relatively simple backward adaptive core bit allocation routine which runs in both the encoder and the decoder. The decoder psychoacoustic model can be

adjusted by sending some parameters of the model forward in the bit stream. The encoder may compare the results of the bit allocation based on the core routine to an ideal allocation. If a better match can be made, the encoder can cause the core bit allocation in both the encoder and decoder to change. When it is not possible to approach the ideal allocation by varying parameters, the encoder can send bit allocation information directly. The multiple channels are allocated bits from a common bit pool.

4.2. AC-3 decoder

The AC-3 decoder, shown in Figure 8, performs the inverse functions of the encoder. The input serial data are demultiplexed producing the quantized mantissas, spectral envelope, and bit allocation side information.

Figure 8. AC-3 audio decoder.



The spectral envelopes are decoded and the bit allocation is computed. After inverse quantization of the mantissas, they are combined with the exponents to form the frequency coefficients. The frequency coefficients are inverse transformed to reproduce the original PCM audio signals.

5. Transport

The Grand Alliance HDTV System uses a constrained subset of the MPEG-2 Transport Stream syntax. MPEG-2 defines two alternative approaches, Program Streams and Transport Streams. Program Streams are designed for use in relatively error-free environments. Transport Streams are designed for use in environments where errors are likely, such as transmission in noisy media. Because the Grand Alliance system is designed for terrestrial broadcasting, an environment where errors are likely, Transport Streams are the proper choice. Both approaches, however, are described here to illustrate the differences.

Both Program Streams and Transport Streams provide syntax to synchronize the decoding of the video and audio information while ensuring that data buffers in the decoders do not overflow or underflow. Both streams include time stamp information required for synchronizing the video and audio. Both stream definitions are packet-oriented multiplexes. Program Streams use variable-length packets. Transport Streams use fixed-length 188 byte packets.

Another type of packet is the Packetized Elementary Stream (PES) packet. After compression, video data and audio data are packaged into separate PES

packets. PES packets may be fixed-length or variable-length. PES packets contain the complete information required to reconstruct the video or the audio. A program consists of elementary streams with a common timebase, for example, video PES packets, audio PES packets, and possibly ancillary data PES packets, along with a control data stream.

The Program Stream results from combining one or more streams of PES packets, with a common time base, into a single stream. The Transport Stream results from combining one or more programs (each program consisting of one or more streams of PES packets with a common time base), with one or more independent time bases, into a single stream.

The three different types of packets being discussed here, Program Stream packets, Transport Stream packets, and PES packets, are illustrated Figure 9. A system level multiplex of two different programs is illustrated in Figure 10.

Each Transport packet begins with a four byte header. The contents of the packet and the nature of the data are identified by the packet header. The remaining 184 bytes are the payload. Individual PES packets, including the PES headers, are transmitted as the payload. The beginning of each PES packet is aligned with the beginning of the payload of a Transport packet — stuffing bytes are used to fill partially-full Transport packets. This means that every Transport packet contains only one type of data — video, audio, or ancillary. The four byte Transport header also provides the functions of packet synchronization, error handling, and conditional access.

Figure 9. MPEG-2 packets.

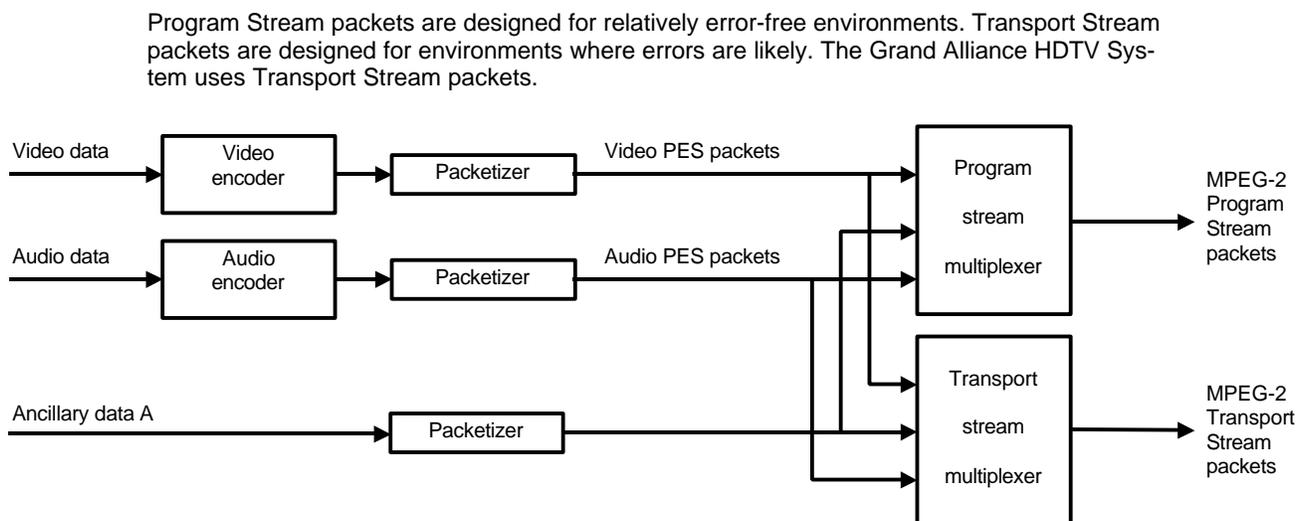
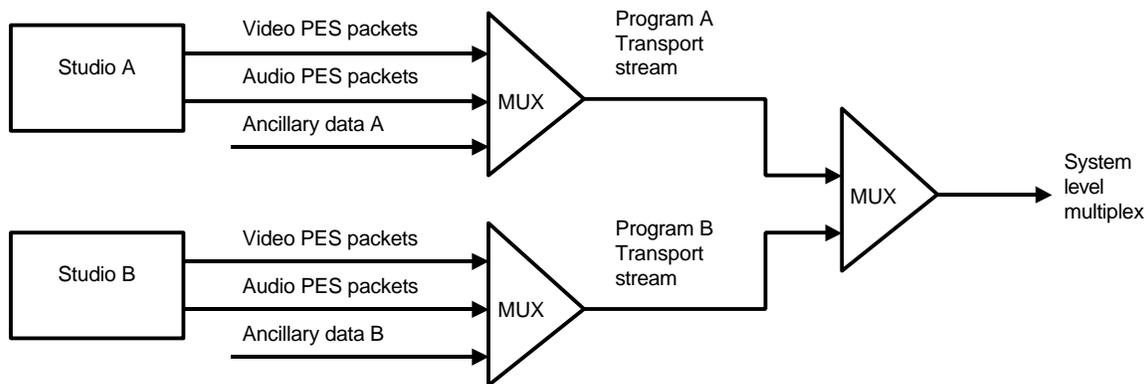


Figure 10. System level multiplex.



For conditional access, audio, video, and ancillary data can be scrambled independently. Information in the Transport header of the individual packets indicates whether the payload in that packet is scrambled. The Transport header is always transmitted in the clear. In the Grand Alliance system, scrambling is implemented only within Transport packets, not within PES packets.

Sometimes additional header information is required. This is provided by the adaptation header, a variable-length field placed in the payload of the Transport packet. Its presence is flagged in the Transport header. Functions of this layer include synchronization (audio and video program timing), support for random entry into the compressed bit stream (tuning to a new channel), and support for local program insertion (inserting local programming into a network program).

The Transport stream provides easy interoperability with Asynchronous Transfer Mode (ATM) transmission. ATM cells consist of a five byte header and a 48 byte payload. The ATM header is used primarily for networking purposes. There are various ways the Transport packets can be mapped into ATM cells. The Transport packet size was selected to ease this transfer. Note that one Transport packet (188 bytes including header) can fit into four ATM cells ($4 \times 48 = 192$ byte payload).

6. Modulation

The VSB transmission system provides two modes, one for terrestrial broadcasting (8-VSB) and one for high data rate cable transmission (16-VSB). Both modes make use of Reed-Solomon coding, segment sync, a pilot, and a training signal. The terrestrial mode adds trellis coding. The symbol rate for both modes is 10.76 Msymbols/s. The terrestrial mode uses 3 bits/symbol. Because the cable environment is less

severe, a higher data rate is transmitted by using 4 bits/symbol and no trellis overhead. The C/N threshold for the terrestrial mode is 14.9 dB. The C/N threshold for the high data rate cable mode is 28.3 dB. The terrestrial mode has a payload data rate of 19.3 Mbits/s. The high data rate cable mode has a payload data rate of 38.6 Mbits/s.

The Reed-Solomon code is a (208,188) $t=10$ code (the data block size is 188 bytes with 20 parity bytes added for error correction) and can correct up to 10 byte errors per block. A 2/3 rate trellis code is used in the terrestrial mode (one input bit is encoded into two output bits while the other input bit is not encoded).

Data are transmitted according to the data frame shown in Figure 11. The data frame begins with a first data field sync segment followed by 312 data segments, then a second data field sync segment followed by another 312 data segments. Each segment consists of 4 symbols of segment sync followed by 832 symbols of data. The symbols during segment sync and data field sync carry only 1 bit/symbol in order to make packet and clock recovery rugged.

In the terrestrial mode, one segment corresponds to one MPEG-2 Transport packet, as follows. The number of bits of data plus FEC per segment is 2,496 (832 symbols times 3 bits/symbol). The MPEG-2 Transport packet contains 188 bytes. Because Reed-Solomon encoding adds 20 bytes for every 188 payload bytes, the total becomes 208 bytes. Because trellis coding adds one bit for every two input bits, this number must be increased by the ratio 3/2, making the total 312 bytes, or 2,496 bits. Thus, one segment is 2,496 bits and one MPEG-2 Transport packet requires 2,496 bits in transmission.

The symbols modulate a single carrier using suppressed-carrier modulation. Before transmission, most of the lower sideband is removed. The resulting spec-

trum is flat except for the band edges. A small pilot, used in the receiver to achieve carrier lock, is added 310 kHz above the lower band edge.

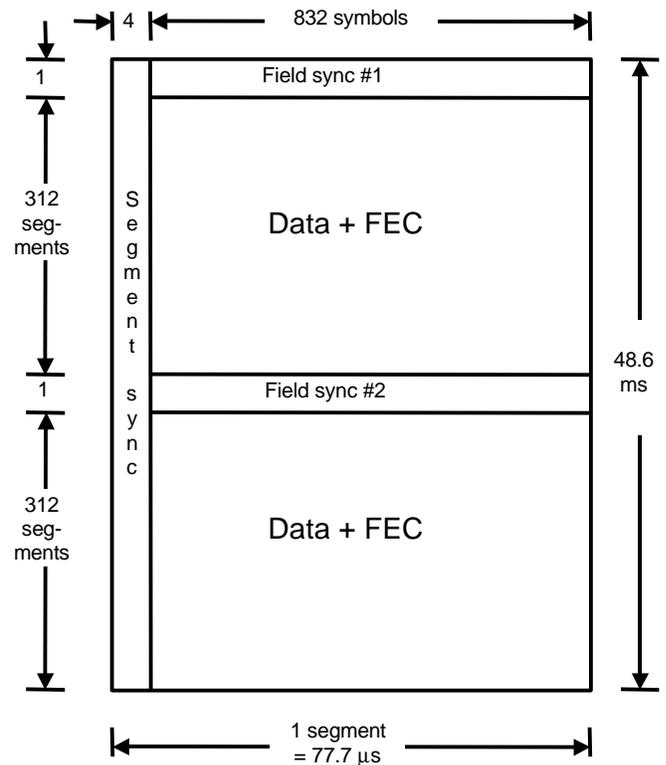
6.1. VSB transmitter

A diagram of the VSB transmitter is shown in Figure 12. The data randomizer performs an exclusive OR on the incoming data with a 16-bit maximum length pseudo-random sequence (PRS) that is locked to the data frame. The data are randomized to ensure that random data are transmitted, even when the data are constant. Segment sync, data field sync, and Reed-Solomon parity bytes are not randomized. After randomizing, the signal is encoded using a (208,188) $t=10$ Reed-Solomon code. The interleaver, an 87 data segment (intersegment) diagonal byte interleaver, spreads data from one Reed-Solomon block over a longer time to give protection against burst errors.

The terrestrial transmission mode uses a $2/3$ rate trellis code. The signaling waveform is a 3-bit 1-dimensional constellation. To help protect the trellis decoder against short burst interference, such as impulse noise or NTSC co-channel interference, 12-symbol code interleaving is employed in the transmitter. Twelve identical trellis encoders operate on interleaved data symbols. In the high data rate cable mode, there is only a simple mapper that converts data to multi-level symbols, as opposed to the trellis encoder/mapper used in the terrestrial mode.

Segment sync and field sync symbols are not Reed-Solomon encoded, trellis encoded, or interleaved. Field sync can serve five purposes. It can provide a means to determine the beginning of each data field. It can be used as a training reference signal in the receiver. It can be used in the receiver to determine whether the NTSC rejection filter should be used. It

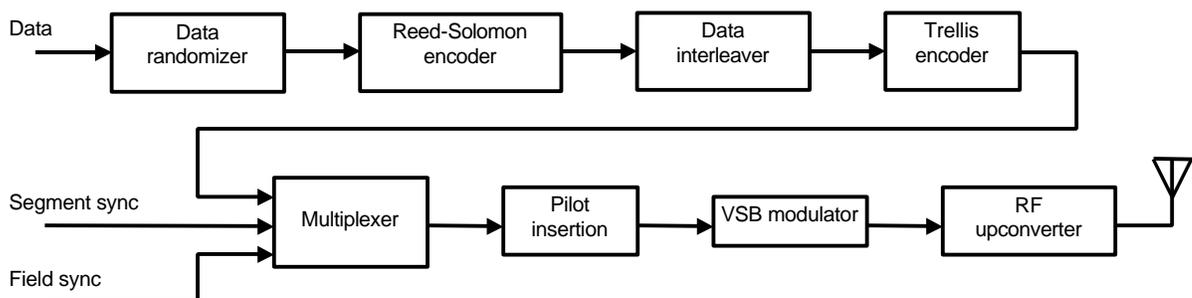
Figure 11. VSB data frame.



can be used for system diagnostics. Finally, it can be used as a reset by the receiver phase tracker.

A small pilot, at the suppressed-carrier frequency, is added to the suppressed-carrier RF signal to allow robust carrier recovery in the receiver during extreme conditions. At the output of the multiplexer, the data signal takes the relative values of ± 1 , ± 3 , ± 5 , and ± 7 . To add the pilot, the relative value of 1.25 is added to every data and sync value. This has the effect of adding a small in-phase pilot to the baseband data signal in a digital manner providing a highly stable and accurate pilot.

Figure 12. VSB transmitter.



The baseband data signal is filtered with a complex filter to produce in-phase and quadrature components for orthogonal modulation. These two signals are converted to analog signals and then used to quadrature-modulate the IF carrier creating a vestigial sideband IF signal by sideband cancellation.

The frequency of the RF upconverter oscillator in advanced television (ATV) terrestrial broadcasts will typically be the same as the nominal NTSC carrier frequency and not an offset NTSC carrier frequency. ATV co-channel interference into NTSC is noise-like and does not change with offset. Even the pilot interference into NTSC is not significantly reduced with offset because it is so small and falls far down the Nyquist slope of NTSC receivers.

With ATV co-channel interference into ATV, carrier offset can prevent misconvergence of the receiver's adaptive equalizer. If the data field sync of the interfering signal occurs during the data field sync of the desired signal, the adaptive equalizer could misinterpret the interference as a ghost. A carrier offset equal to half the data segment frequency will cause the interference to have no effect in the adaptive equalizer.

6.2. VSB receiver

A diagram of the Grand Alliance prototype VSB receiver is shown in Figure 13. After the signal has traversed the tuner, IF, and synchronous detector stages, and the clocks and syncs have been recovered, the data will be switched into an NTSC rejection filter if NTSC co-channel interference is detected. The NTSC comb filter is designed with seven nulls in the 6 MHz channel. The NTSC picture carrier falls near the second null; the NTSC color subcarrier falls at the sixth null; and the NTSC sound carrier falls near the seventh null. The filter is a 12-symbol feedforward subtractive comb filter.

Although the comb filter reduces NTSC co-channel interference, the data are also affected. Also, white noise performance is degraded by 3 dB. Therefore, if little or no NTSC interference is present, the comb filter is automatically switched out of the signal path. The NTSC comb filter is not required in the high data rate cable mode because co-channel interference is not present on cable.

The equalizer/ghost canceler delivered for the Grand Alliance test in January 1994 used a Least-Mean-Square (LMS) algorithm adapting on the data field sync. By adapting on a known training signal, the circuit converges even in extreme conditions. After reaching convergence on the data field sync, the circuit is switched to equalize on the random data for high speed tracking of moving ghosts like airplane flutter. A diagram of the equalizer is shown in Figure 14. The equalizer filter consists of two parts, a 78-tap feedforward transversal filter followed by a 177-tap decision-feedback section. Following the equalizer, the data symbols are used to detect and remove phase noise.

Because 12-symbol code interleaving is used in the trellis encoder, the receiver uses 12 trellis decoders in parallel. The trellis decoder has two modes depending on whether the NTSC rejection filter is in use. When NTSC co-channel interference is detected, the NTSC rejection filter is switched into the signal path and a trellis decoder optimized for use in tandem with the comb filter is used. When NTSC interference is not detected, the NTSC rejection filter is switched out of the signal path and an optimal trellis decoder is used. In the high data rate cable mode, the trellis decoder is replaced by a slicer that translates the multi-level symbols into data.

The de-interleaver performs the inverse function of the transmitter interleaver. The (208,188) $t=10$ Reed-

Figure 13. Grand Alliance VSB receiver.

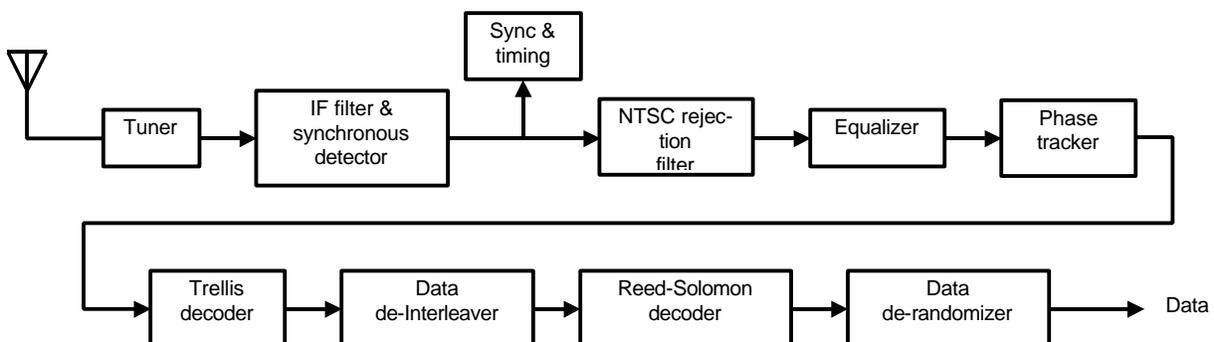
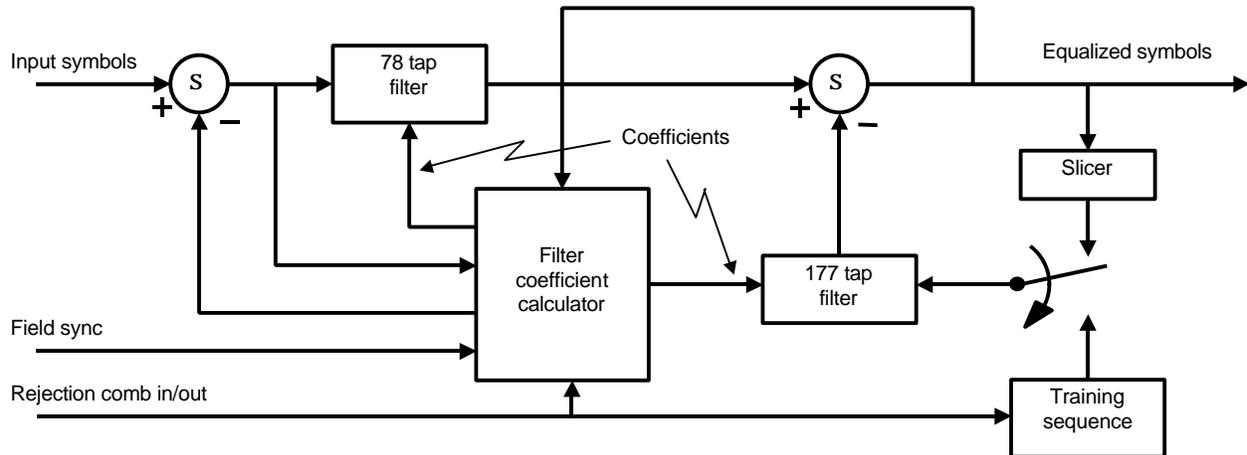


Figure 14. Grand Alliance VSB equalizer.



Solomon decoder uses the 20 parity bytes to perform the byte-error correction on a segment-by-segment basis. The de-randomizer accepts the error-corrected data bytes from the Reed-Solomon decoder, and applies to the data the same PRS code that was used at the transmitter.

7. Conclusions

The Grand Alliance HDTV System is the product of many people's efforts over many years. The visible effort began when the Advisory Committee on Advanced Television Service was formed in 1987. Not so visible at the outset was the effort of many engineers from several different organizations designing proposed systems. Those efforts really began to show in 1991 when testing of the proposed systems began.

The testing showed strong points and weak points in the original designs. One extremely strong point was the digital design that had been adopted in four of the five HDTV systems tested. After the testing and analyses were complete, the Grand Alliance was formed by the proponents of the digital systems. The Grand Alliance, working with the Advisory Committee, has designed a system that will satisfy the needs of North America. Subsystems have been selected based on technical excellence. The system has a great deal of flexibility to facilitate interoperability and is heavily based on international standards.

Acknowledgment

The author wishes to thank several persons who have reviewed this paper for accuracy. They are: Stan Baron of NBC, David Bryan of Philips Laboratories, Lynn Claudy of NAB, Carl Eilers of Zenith Electron-

ics Corporation, James Gaspar of Panasonic Advanced Television Laboratory, John Henderson of Hitachi America, Robert Keeler of AT&T Bell Laboratories, Bernard Lechner, James McKinney of ATSC, Woo Paik and Robert Rast of General Instrument Corporation, Terrence Smith and Joel Zdepski of David Sarnoff Research Center, and Craig Todd of Dolby Laboratories.

References

- [1] Hopkins, R. and Davies, K.P., "Development of HDTV Emission Systems in North America," *IEEE Transactions on Broadcasting*, Vol. 35, No. 3, September 1989.
- [2] Hopkins, R. and Davies, K.P., "HDTV emission systems approach in North America," *ITU Telecommunication Journal*, Vol. 57, May 1990, pp. 330-336.
- [3] "ATV System Recommendation," *IEEE Transactions on Broadcasting*, Vol. 39, No. 1, March 1993, pp. 2-245.
- [4] "ATV System Recommendation," 1993 NAB HDTV World Conference Proceedings, pp. 237-493.
- [5] Hopkins, R., "Progress on HDTV broadcasting standards in the United States," *Image Communication*, Vol. 5, Nos. 5-6, December 1993, pp. 355-378.
- [6] Hopkins, R., "Choosing an American Digital HDTV Terrestrial Broadcasting System," *Proceedings of the IEEE*, Vol. 82, No. 4, April 1994, pp. 554-563.
- [7] "Grand Alliance HDTV System Specification," Version 1.0, April 14, 1994. Available from International Transcription Services, 2100 M Street NW, Suite 140, Washington, DC 20037, telephone 202-857-3800, fax 202-857-3805. Available also on the Internet via anonymous ftp to "ga-doc.sarnoff.com".

- [8] SMPTE S17.394, "1920x1080 Scanning and Interface," proposed SMPTE standard for television.
- [9] SMPTE S17.392, "1280x720 Scanning and Interface," proposed SMPTE standard for television.
- [10] ISO/IEC DIS 13818-2, "MPEG-2 Video," draft international standard.
- [11] Davis, M., "The AC-3 Multichannel Coder," AES 95th Convention, Preprint 3774, October 1993.
- [12] Todd, C. C., Davidson, G. A., Davis, M. F., Felder, L. D., Link, B. D., and Vernon, S., "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," AES 96th Convention, Preprint 3796, February 1994.
- [13] ATSC T3/S7-016, "Digital Audio Compression (AC-3)," draft ATSC standard.
- [14] ISO/IEC DIS 13818-1, "MPEG-2 Systems," draft international standard.

Biography

Robert Hopkins received the B.S. degree in electrical engineering from Purdue University, West Lafayette, Indiana, and the M.S. and Ph.D. degrees from Rutgers University, New Brunswick, New Jersey. He is also a graduate of the Harvard Business School Program for Management Development.

Since 1985 he has been the Executive Director of the United States Advanced Television Systems Committee (ATSC), a standards organization sponsored by more than 50 companies involved in HDTV. He is responsible for both the technical and administrative guidance of the ATSC. He was employed by RCA from 1964 to 1985 at the David Sarnoff Research Center, the Broadcast Systems Division, and as managing director of RCA Jersey Limited, Channel Islands, Great Britain.

Dr. Hopkins is a Fellow of SMPTE and a Senior Member of IEEE. He serves as the United States representative on HDTV to the ITU-R.