3.3. VIDEO COMPRESSION APPROACH

Figure 3.3-1 shows the functional block diagram of the GA video encoder. The components are described in the following sections.

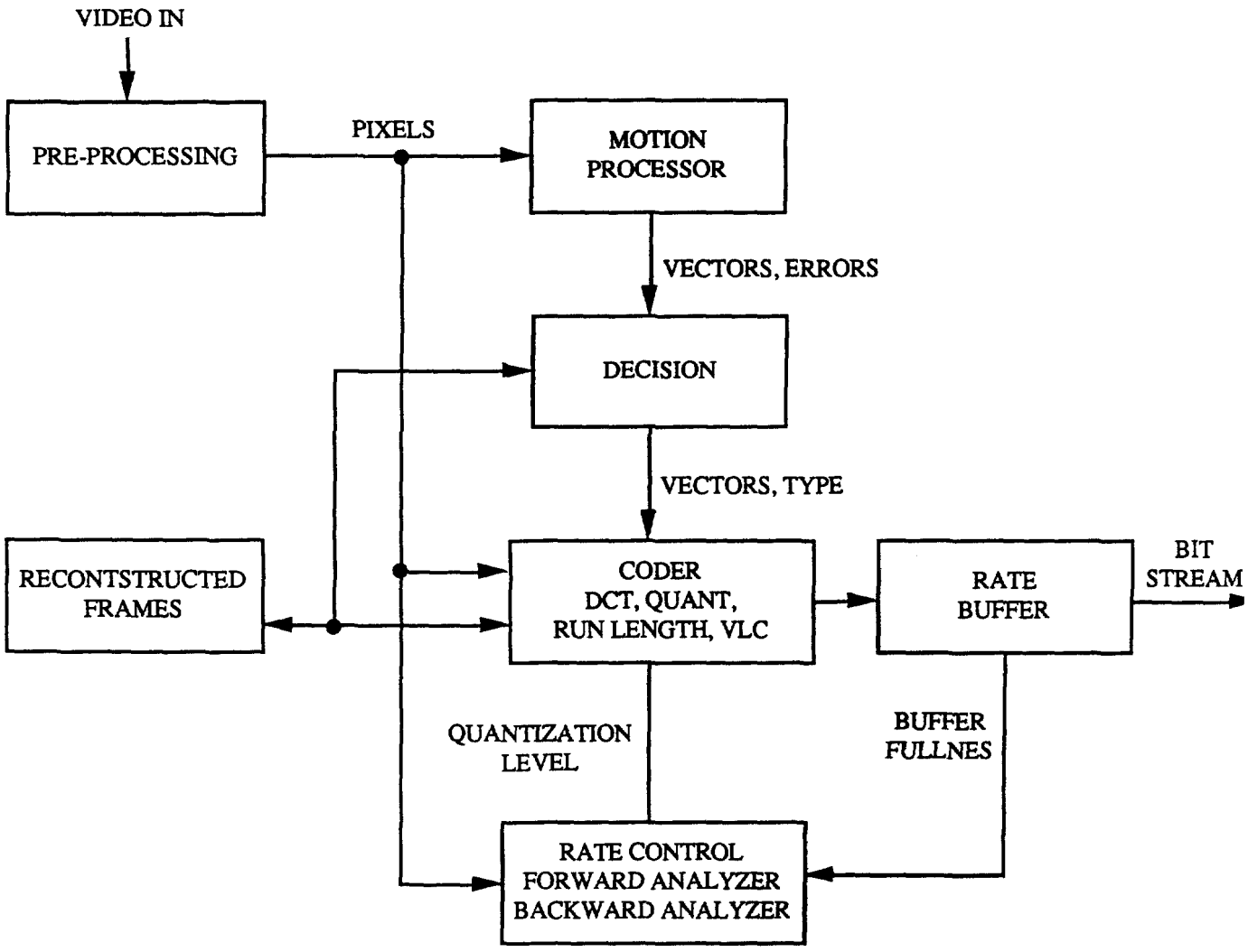

Figure 3.3-1. GA Video Encoder

3.1 Video Preprocessor

As shown in Figure 3.3-2, analog video is received in RGB format and digitized using 10-bit A/D converters. Gamma correction is then applied to each color component in order to compensate for the non-linear response of the camera. This helps to reduce the visibility of quantization noise, particularly in the dark regions of the image.

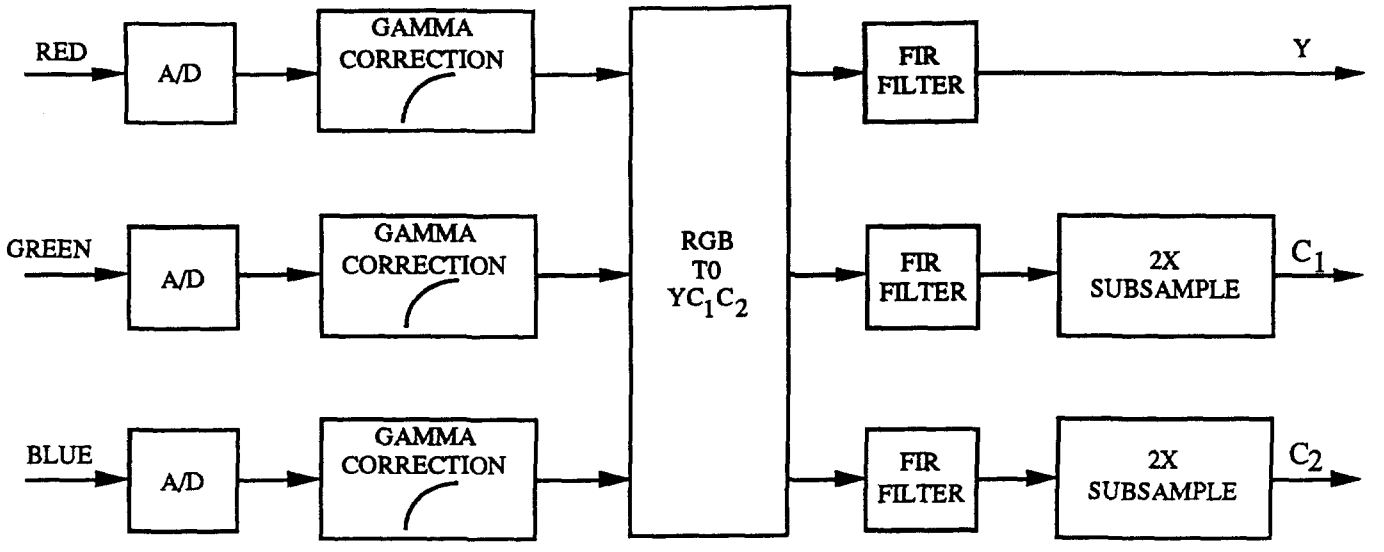

Figure 3.3-2. Video Preprocessor

The digitized and gamma-corrected RGB samples are then converted to the SMPTE 240M YC1C2 color space using a linear matrix transformation. Programmable FIR filters are used to shape the frequency response of each color component. The two chrominance components use half-band filters in order to prevent aliasing during the subsequent sub-sampling process. The sub-sampling is performed by a factor of 2, both horizontally and vertically.

3.3.2 Motion Estimation and Compensation

The GA compression algorithm utilizes multiple prediction methods to effectively provide motion compensation for progressive and interlaced pictures. The prediction methods include frame prediction, adaptive field prediction, dual prime (used for forward prediction only), and bi-directional prediction. Among these, the dual prime and the bi-directional prediction modes deserve further discussion. In the dual prime case the prediction is formed from pixels in previously displayed frames, while in the B-frames case the prediction is based on both previous frames and future frames. The dual prime and bi-directional methods are complementary since the dual prime mode is for interlaced only and is precluded when a sequence uses B-frames.

3.3.2.1 The "P-Frame" Motion Vector Prediction Mode

P-frames are frames where the prediction is in the forward direction only (i.e., predictions are formed only from pixels in previously displayed frames). These forward-predicted frames allow the exploitation of inter-frame coding techniques to improve the overall compression efficiency and picture quality.

Figure 3.3-3 illustrates a time sequence of video frames consisting of intra-coded pictures, predictive coded pictures, and bi-directionally predictive coded pictures. P-frames are predicted using the most recently encoded P-frame of I-frame in the sequence. In Figure 3-3, P-frames occur every second frame except when an I-frame is used. Each macroblock within a P-frame can be either forward-predicted or intra-frame coded. If a macroblock is forward-predicted, then either frame-based or field-based prediction may be used. The decision is made by the encoder based on the smallest prediction error using each method. Dual prime prediction may also be chosen at this stage if the prediction error using dual prime is smallest. Regardless of whether field-based,

frame-based or dual prime prediction is used, the predicted macroblock is subtracted from the original macroblock, and only the "difference" values are encoded and transmitted.

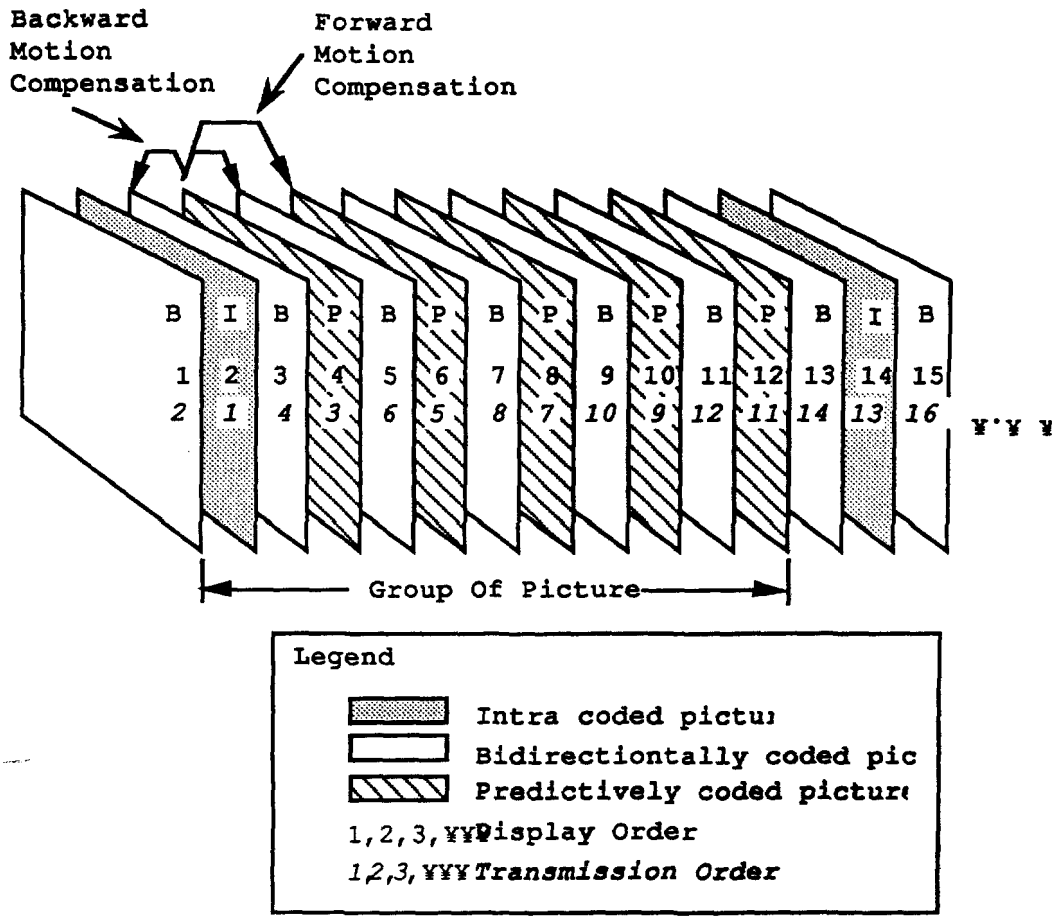

Figure 3.3-3 Example of a Coded Video Sequence Using P-Frames and B-Frames

3.3.2.1.1 Dual Prime

The dual prime prediction mode is an alternative "special" prediction mode which is based on field-based motion prediction but requires fewer transmitted motion vectors than conventional field-based prediction. This mode of prediction is available only for interlaced material and only when the encoder configuration does not use B-frames (M=1). For these reasons, this mode of prediction is particularly useful for improving encoder efficiency for low delay applications.

The basis of dual prime prediction is that field-based predictions of both fields in a macroblock (MB) are obtained by averaging two separate predictions which are predicted from the two nearest decoded fields (in time). Each of the MB fields is predicted separately, although the four vectors (one pair per field) used for prediction are all derived from a single transmitted field- based motion vector. In addition to the single field-based motion vector, a small "differential" vector (limited to vertical and horizontal component values of +1, 0 and -1, and represented by two 1-2 bit codes) is also transmitted for each MB. Together, these vectors are used to calculate the pairs of motion vectors for each MB field. the first prediction vector in the pair is simply the transmitted field-based motion vector. The second prediction vector is obtained by combining the differential vector with a scaled version of the first vector. Once both predictions are obtained, a single prediction for

each macroblock field is calculated simply by averaging each pel from the two original predictions. The final averaged prediction is then subtracted from the macroblock field being encoded.

Figure 3.3-4 illustrates the relationship between the transmitted vectors (one field-based vector and one differential vector) and the prediction vectors for each of the fields in a macroblock. The two separate sets of vectors shown in Figure 3-4 correspond to the predictions for the two fields which make up the macroblock. The transmitted field-based motion vector and the transmitted differential vector (identical for each set) are represented by solid lines. The differential motion vector is the smaller vertical vector. The scaled vectors are represented by dotted lines. The second (calculated) field-based motion vectors are represented by the dashed lines.

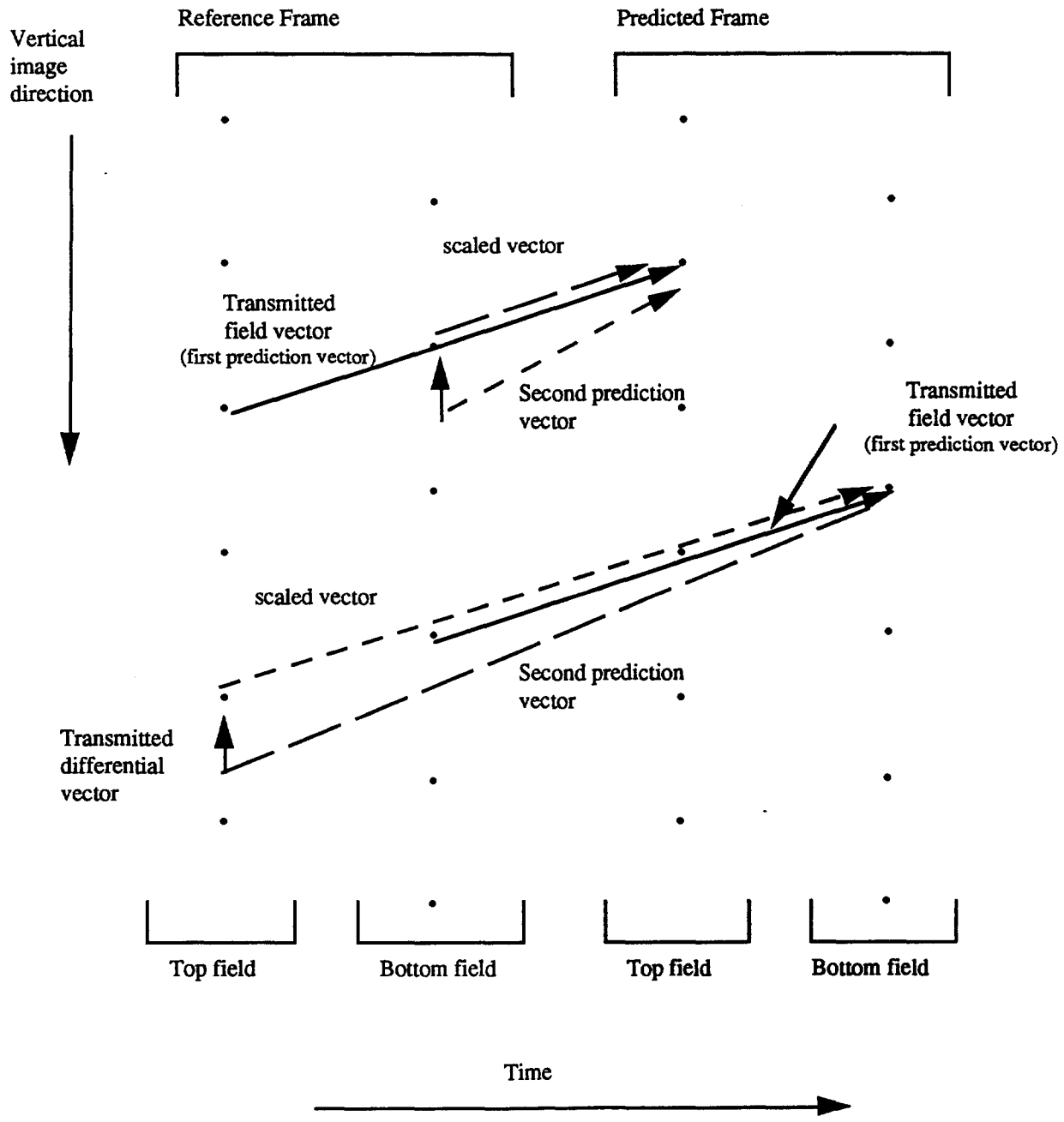

Figure 3.3-4. Dual Prime Prediction

3.3.2.2 The "B-Frame" Motion Vector Prediction Mode

The B-frame is a picture type within the coded video sequence which has some special prediction modes available to the encoder. It is useful for increasing the compression efficiency and perceived picture quality when encoding latency is not an important factor. It is also an attractive tool since it works equally well with both interlaced and progressive material. However, it has impact on the receiver cost since it requires additional memory. B-frames are included in the GA system because the increase in compression efficiency is noticeable especially with progressive scanning where techniques such as dual prime are not available. The improvement in performance was found to be sufficient to justify the cost of the additional memory required by the ATV receiver.

The basis of the B-frame prediction is that a video frame is correlated both with frames which occur in the past and the frames in the future. Consequently, if a future frame is available to the decoder, a superior prediction can be formed, thus saving bits and improving performance. Some of the consequences of using future frames in the prediction are: the B-frame cannot be used for predicting future frames, the transmission order of frames is different than the displayed order of frames, and the encoder and decoder must reorder the video frames, thereby increasing the total latency. In the example illustrated in Figure 3-3, there is one B-frame between each pair of I/P frames. Each frame is labeled with both its display order and transmission order. Although a B-frame is displayed between its adjacent I/P frames, it is transmitted out of sequence. This is so that the video decoder has the adjacent frames decoded and available for prediction. For a given macroblock within the B-frame, the encoder has four options for its prediction. They are: forward prediction, backward prediction, bi-directional prediction, and intra-frame coding. When bi-directional prediction is used, the forward and backward predictors are averaged and then subtracted from the target macroblock to form the prediction error. The prediction error is then transformed, quantized and transmitted in the usual manner.

3.3.3 Refreshing Options

A motion compensated prediction loop is not practical without some form of refreshing. This refreshing may be constant or variable and periodic. If a sequence of input images were perfectly predictable, the decoder would not receive any coefficient data and therefore could not reconstruct the picture after initialization. Decoder initialization occurs after the channel is changed, or the signal is lost and then recovered.

Consider the coding of a still image. The motion vectors and the prediction error would be zero. If the decoder were started with a sequence of zero prediction frames, then a blank or zero reconstructed frame would result. If the viewer changed to a channel transmitting a coded still image, the still image would never appear. A portion of the original picture must be spatially or temporally mixed with the prediction in the encoder to allow the decoder to synchronize to the encoder after decoder initialization.

The GA system satisfies the need for refreshing with two different mechanisms. I-frame refreshing uses periodic intra-coded frames. The advantages are that it provides clean insertion points in the compressed bit stream and nominal picture quality after acquisition (typically .5 seconds). The disadvantages are that it requires a larger bit buffer, it increases latency, and it complicates rate control.

Progressive refreshing uses periodic intra-coded (16x16) macroblocks. The advantages over I-frame refreshing are that it reduces the required buffer size, it simplifies rate control, and it reduces latency. One disadvantage is that motion vectors must be restricted in order to guarantee complete picture buildup after channel acquisition or channel errors.

3.3.4 Adaptive Field/Frame Processing

There are two options when processing interlaced video signals. The first option is to separate each frame into its two field components and then process the two fields independently (Figure 3.3-5). The second option is to process the two fields as a single frame by interleaving the lines of field 1 and field 2 (Figure 3.3-6).

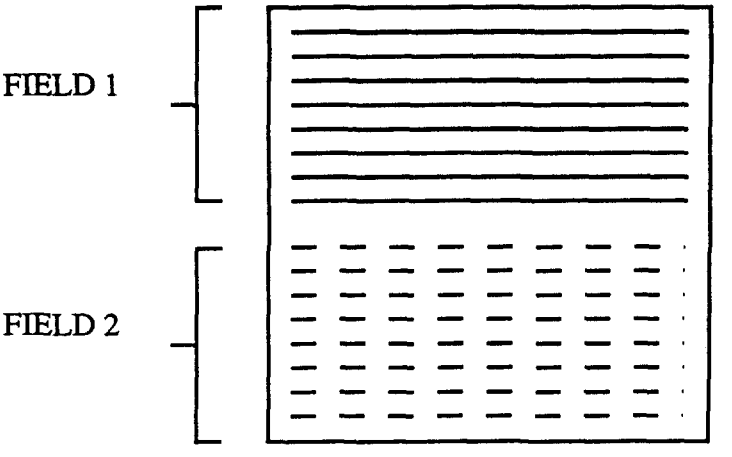

Figure 3.3-5. Field Processing

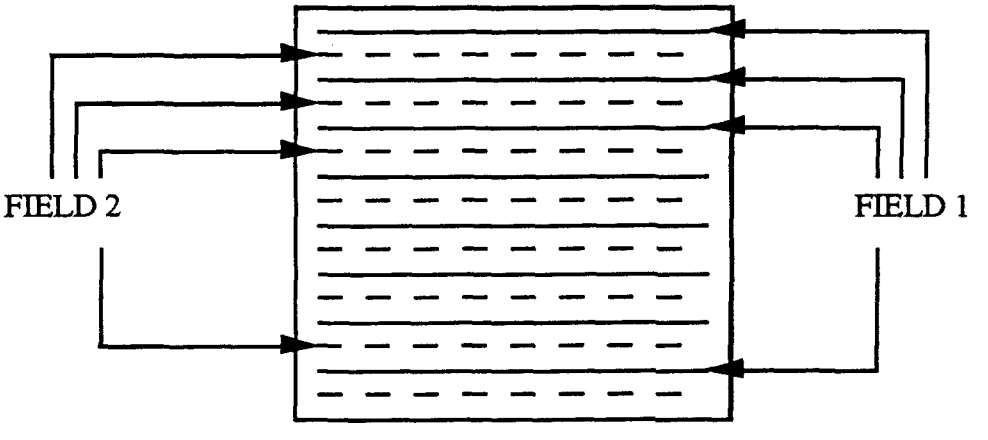

Figure 3.3-6. Frame Processing

Frame processing works better than field processing when there is little or no motion. Since each frame has twice as many lines or samples for a given picture height, there will be more correlation between samples and hence compressibility will be increased. Therefore, to achieve the same accuracy, field processing will require a higher bit rate, or alternatively, for equal bit rates, frame processing will achieve greater accuracy.

Similar advantages over field processing will be realized if horizontally moving features have little horizontal detail or if vertically moving features have little vertical detail. In other regions, where there is little detail of any sort, frame processing may still work better than field processing, no matter how rapidly changes occur.

Field processing generally works better than frame processing in detailed moving areas. In such cases, the interleaving of the first and second fields would introduce spurious high vertical frequencies if frame processing were used. This would reduce the correlation between lines and therefore the effectiveness of the compression process.

The GA system combines the advantages of both frame processing and field processing by adaptively selecting one of the two modes on a block-by-block basis. The mode which is selected is the one which maximizes correlation between adjacent lines within the block. Since the decision is made on a local basis, the system can adjust to scenes containing both moving and non-moving features, and therefore accurately reproduce vertical details in non-moving areas, and deliver good motion rendition in others.

The adaptive field/frame feature is only enabled when processing interlaced source material. Non-interlaced sources are always processed using frame mode.

3.3.5 Forward Estimation for Rate Prediction

Motion compensation, adaptive quantization and variable length coding produce highly variable amounts of compressed video data as a function of time. For example, the compressed bit-rate after a scene change can be several times greater than the bit-rate of the channel. Therefore, compressed video data buffering in the encoder and decoder is required for efficient channel utilization.

Buffer size is constrained by the maximum tolerable delay through the system and by cost. The fullness of the buffer is controlled by adjusting the amount of distortion or quantization error in each image. In the encoder, the buffer will fill more quickly if the distortion is low. A feedback control system is required to regulate the distortion level which controls the buffer fullness to prevent overflow. The design of this control system is complicated by:

- Delay between a change in the distortion level and the subsequent change in buffer fullness.

- Perceptual constraints on the instantaneous and averag distortion levels.

  Difficulties in modeling the bit-rate as a function of distortion level.

The visibility of distortion due to an increase in scene complexity is minimized by smoothly increasing the distortion level. This is facilitated by accurately modeling the rate versus distortion level since an accurate model allows the desired buffer fullness to be achieved for each frame. The model allows an estimate of complexity to be translated into a bit-rate for a given frame.

In the GA rate control system, the complexity estimate is the variance of the ideal (unquantized) displaced frame difference (DFD). Since the whole-pel motion vectors are available one frame ahead of the loop, the ideal displaced frame difference is calculated using whole-pel accuracy (i.e., no sub-pel estimation). The pixel values of the ideal DFD are squared and summed over the picture.

3.3.6 Adaptive Intra and Non-Intra Quantization Matrices

The MPEG-2 syntax allows the quantization matrices to be specified for every picture for improved coding efficiency. A certain probability distribution is associated with the VLC codes for quantized coefficients in the MPEG standard. Although one cannot change the VLC distribution to match the

actual distribution of the data, the quantization matrices can be adjusted to help match the distribution of the data to the distribution of the VLC. Over the course of encoding the frame data, the variance of each spatial frequency band is calculated for both Intra and Non-Intra data. Applying upper and lower bounds per band to ensure reasonable operation in all cases, the desired quantization matrix can be derived.

Transmitting the quantizer matrices costs bits in the compressed data stream; if sent every picture in the 60 Hz progressive mode, the matrices consume 0.1% of the channel bandwidth. This modest amount of overhead is reduced by updating the quantization matrix at the start of each GOP, at least every 400 ms (for start-up), or when the difference between the desired quantizer matrix and the prevailing quantizer matrix becomes significant.

Sufficient compression cannot be achieved unless a large fraction of the DCT coefficients are dropped and therefore not selected for transmission. The coefficients which are not selected are assumed to have zero value in the decoder. The GA system encodes the selections and runs of zeros following a zigzag pattern through the array of frequency ordered coefficients. The DC coefficients are coded differentially to take advantage of high spatial correlation.

3.3.7 Perceptual Weighting and Coefficient Selection by Perceptual Sensitivity

The human visual system is not uniformly sensitive to coefficient quantization error. Perceptual weighting of each source of coefficient quantization error is used to increase quantization error in order to lower the bit-rate. The amount of visible distortion resulting from quantization error for a given coefficient depends on the coefficient number or frequency, the local brightness in the original image and the duration or the temporal characteristic of the error. DC coefficient error results in mean value distortion for the corresponding block of pixels which can expose block boundaries. This is more visible than higher frequency coefficient error which appears as noise or texture.

Displays and human visual systems exhibit non-uniform sensitivity to detail as a function of local average brightness. Loss of detail in dark areas of the picture is not as visible as it is in brighter areas. Another opportunity for bit savings is presented in textured areas of the picture where high frequency coefficient error is much less visible than in relatively flat areas. Brightness and texture weighting require analysis of the original image since these areas may be well-predicted in the DFD. Finally, distortion is easily masked by limiting its duration to one or two frames. This effect is most profitably used after scene changes where the first frame or two can be greatly distorted without perceptible distortion at normal speed.

Traditionally, a given coefficient is transmitted whenever its quantized level is non-zero. By selecting some non-zero coefficients for elimination, a fine level of quantization can be achieved on the remaining coefficients, potentially improving the overall picture. Perceptual selection is such a method.

Perceptual Selection is used in the GA system using the properties of the human visual system to code pictures using fewer bits within a perceptually consistent level of quality. For each transform block, the perceptual selection method determines the acceptable amount of distortion per each frequency band; if the magnitude of the coefficient in the loop is smaller than the acceptable distortion level, then the coefficient is set to zero, regardless of the quantization step size level.

3.3.8 Adapting M and N

The motion estimator range of the GA prototype system is ±32 V x ±128 H between two successive frames. The use of B-frames reduces the effective motion tracking range by a factor of M where M-1 is the number of B-frames between a given pair of I- or P-frames.

Scenes with rapid large area motion may exceed the range of the motion estimator if M is too high. This condition is detected in the forward analyzer which then signals the frame reorder section and motion estimator to use a lower M value for subsequent frames. The increased bit- rate caused by the temporary motion estimator overrun is absorbed by the rate buffer to maintain picture quality.

3.3.9 Film Mode

The GA system is able to detect source material originating from 24 fps film that has been converted to 60 fps using the 3:2 pulldown process. As shown in Figure 3.3-7, some fields (or some frames) are known to be identical due to the 3:2 pull down method. The GA system detects this redundancy through utilization of a film detection method that looks for the 3:2 pull down pattern. In the future, when the film mode is detected, the GA system removes the redundancy and converts it back to the original 24 fps source format prior to video encoding. The removal of redundancy ensures that the same information is not transmitted repeatedly, thus significantly improving the efficiency of the video compression system.

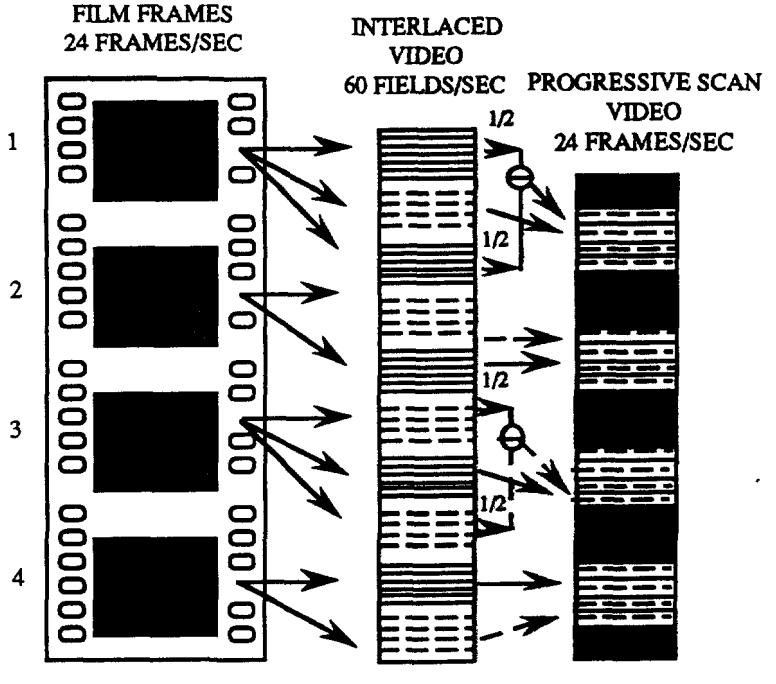

Figure 3.3-7. Conversion of Film to Interlaced Video and Restoration of 24 fps Progressive Scan

The GA system is also able to detect 30 fps material that has been converted to a 60 fpsprogressive scan format by simple frame repetition. When this conversion is detected, theencoder will discard the redundant frame and indicate to the decoder that each frame is to bedisplayed twice.

3.4. INTER-OPERABILITY

The GA compression algorithm is based on the emerging MPEG-2 compression standard. Consequently, the HDTV decoders will have resident decoder functionality to implement a device with a high degree of inter-operability. This inter-operability is a manufacturers option, and is not mandated.

3.4.1 Inter-operability with MPEG

The opportunities for inter-working are with MPEG-2 at High and High 1440 Level, CATV, DBS, and Teleco Standard Definition MPEG-2 and MPEG-1 services, and inter-working with MPEG-1 services such as those expected to be implemented for computer multimedia terminals. In each case, the HDTV decoder has the required decode engine and frame memory to decode all of these services. The implementation of image scaling for display, or multi-scan displays is left to the manufacturer.

The inter-operability choices with MPEG-2 are shown in Figure 3.4-1. Each line illustrates a potential inter-working scenario. The lines originate at the encoding side, with the arrow pointing to the decoder. A solid line indicates that inter-operability is achieved, while a dashed line indicates that the system is not inter-operable unless the manufacturer adds functionality to the decoding device. The lines labeled A and E show a US HDTV decoder inter-working with a MPEG-2 bit stream. Line A is dashed, while line B is solid, since the GA decoder is a super set of an MPEG-2 decoder. In case additional syntax elements are added to the GA system in the future to improve the performance, then line B becomes dashed and an MPEG-2 decoder will not decode the GA bit streams unless the manufacturer adds the additional functionality to the decoding device. Line C is also dashed, since there is not yet a requirement that a GA television be able to decode and display a compressed digital standard definition bit stream. This level of inter-operability does not require any change to the decoding engine, since MPEG-2 is a subset of the GA algorithm. It will require that the display interpolation functionality be resident in the GA decoder. There is no line from the GA system to the MPEG-2 main because MPEG-2 main level decoder will not be able to decode the GA bit stream due to memory and speed constraints.

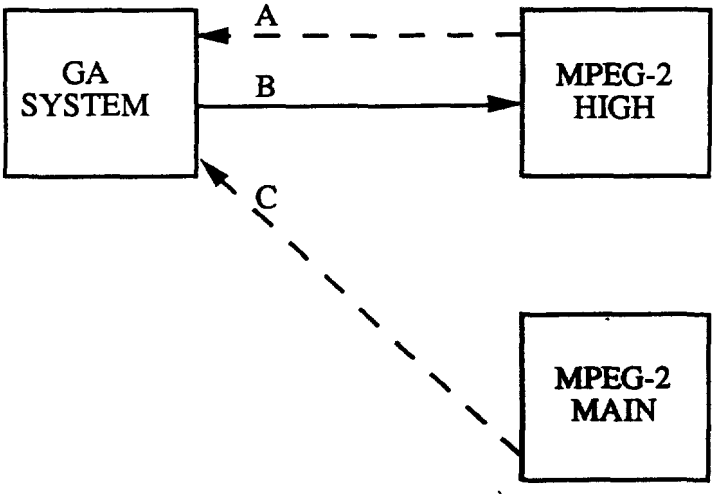

Figure 3.4-1. Inter-operability between GA and MPEG-2

3.4.2 Inter-operability with Computers

The GA system will inter-operate with computer displays in a similar manner, as shown in Figure 3.4-2. Here, the bit stream inter-operates at either the Transport Layer or the Compression Layer. The compression layer will syntactically inter-operate, however, the computer will require some display processing to convert the video frame-rate assuming a computer display refresh rate of * 66 Hz as is commonly the case. The windowing system will also manage the format conversion between GA video formats and the display format native to the computer. The format conversions required will be: frame-rate conversion, picture aspect ratio (16:9 versus 4:3), interlaced to progressive (when required), and pel aspect ratio (when required).

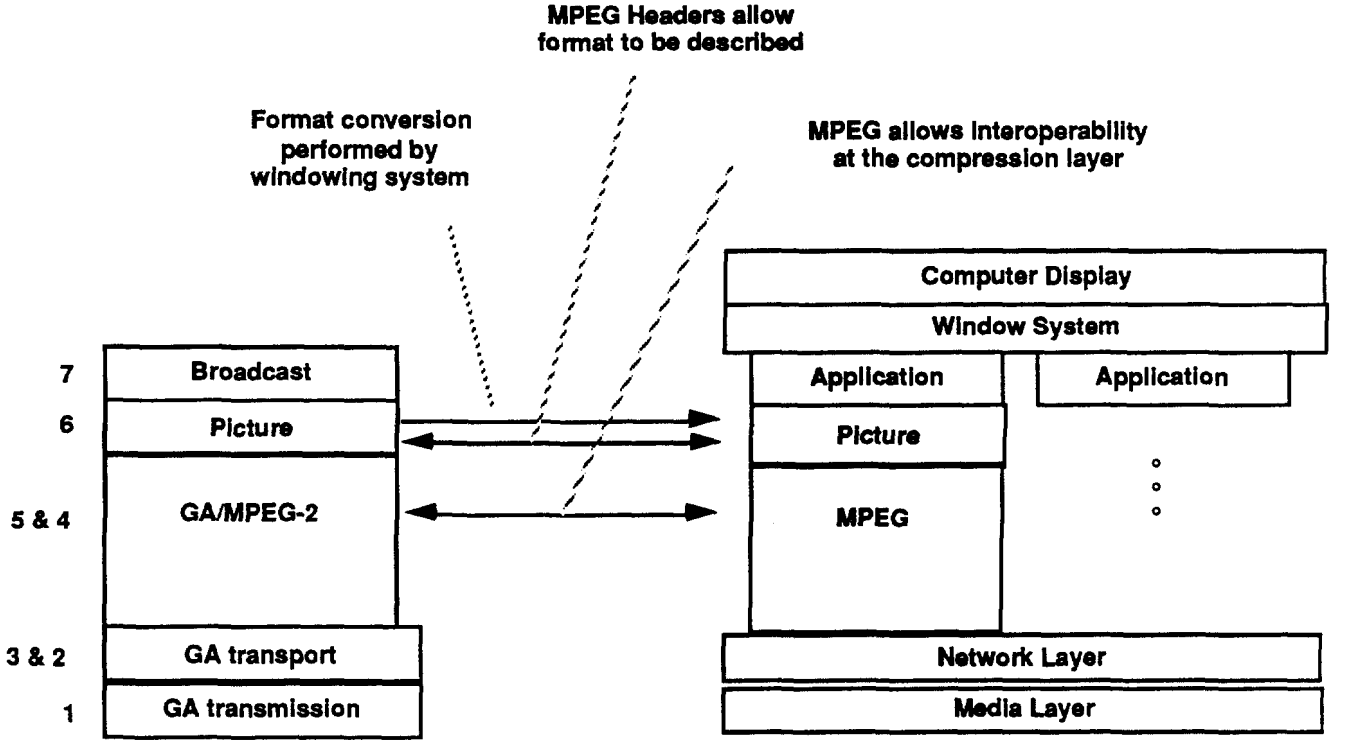

Figure 3.4-2. GA Inter-operability with Computers

3.3.5. PROTOTYPE IMPLEMENTATION ISSUES

A prototype video coder-decoder (codec) that demonstrates the salient features of the GA video compression specification (as described in this document) will be built. This section describes some of the features. The coding parameters and algorithms of the compression system are being fine tuned through simulation of the system in software. All those elements in the compression specification that add to the quality of the compressed/uncompressed video will be implemented. It must be pointed out that the GA video coding specification, being a bit stream/decoder standard, will allow further improvements to the encoder in the future without any incompatibility with video decoders that already exist.

The video codec will be able to compress multiple video formats depending upon the source (both interlaced and progressive). The compression technique is essentially independent of the video format, although the coding parameters change. The codec will operate at compressed bit- rates of around 17 Mbps.

3.5.1 Features

3.5.1.1 Frame Types, M and N, Refresh

Each video frame is coded in one of three modes: Intra-coding (MPEG I-pictures), Predictive coding (MPEG P-pictures) or Bi-directional predictive coding (MPEG B-pictures). The period of I-pictures (Group-of-pictures size, N) and the distance in number of frames between two anchor (I or P) pictures (known as M), can be programmed.

In the above mentioned mode of operation, I-pictures provide refresh. The codec will allow an alternate way of refreshing, which is achieved through progressive refresh. In this mode, parts of P-pictures (I-slices) are refreshed progressively. Periodic I-frames will be present in progressive refresh to facilitate editing.

3.5.1.2 Motion Estimation

Motion of up to one-half pixel accuracy will be estimated. Integer pixel motion will be estimated on the original video frames, while the sub-pixel refinement will be done on reconstructed pictures. Motion vector search range of up to ± 32 pixels vertically and ± 128 pixels horizontally will be used for P-pictures. For B-pictures, the motion search range will be ± 32 pixels vertically and ± 64 pixels horizontally.

The specification of the GA compression group includes the dual prime prediction technique for interlaced video (it does not add additional cost to the decoder). This technique is useful for low-delayapplications. In the interest of time, this technique will not be implemented.

3.5.1.3 Field Structure Pictures

The specification of the GA compression group also includes the option of using field-structure pictures when processing interlaced video. As with dual prime, this technique will not be implemented in the GA prototype system.

3.5.1.4 Movie Material

When the video codec is presented with material that has been "three-two pulled-down" (originally from a 24 fps movie source), the encoder detects this condition and utilizes the inherent

redundancy in the input video to better compress it. The decoder decodes the pictures and displays them as three-two pulled down material.

3.5.1.5 Adaptive Field/Frame Motion Estimation and DCT

Interlaced pictures can be predicted and coded adaptively as fields or frames on a local (macroblock) basis. The adaptive field/frame motion prediction decision will be based on mean absolute error criterion, whereas, the adaptive field/frame DCT will be based on quantized errors or mean absolute error criteria.

3.5.1.6 Rate Control

In order to provide good overall subjective picture quality, a sophisticated adaptive quantization and rate control algorithm will be used. This algorithm will be implemented on a DSP-based subsystem. In addition, a forward analyzer (to assist the rate controller) will be implemented.

3.5.1.7 DCT Coefficient Coding

Compressed video (motion vectors, DCT coefficients, modes, and other headers) will be coded using MPEG-2 VLC tables and syntax. In particular, DCT coefficients will be coded using two VLC tables (one each for inter and intra) in one of two scan orders (zigzag and alternate scans) as prescribed by MPEG-2.

Chapter 4

AUDIO COMPRESSION SYSTEM

4.1. Introduction

The AC-3 audio coding technology is a sophisticated and flexible system for the digital representation of high quality (including multi-channel) sound. A wide range of bit rates and audio coding modes are supported. Features have been incorporated which allow the coded AC-3 data stream to service nearly all listeners, not just those with a particularly ideal listening situation. A section of the data stream referred to as *bit stream information* (bsi) has a number of fixed and optional fields that describe the coded bit stream in some detail. Some of these fields, concerned with audio levels and types of audio production facilities, allow coded bit streams from different sources to be decoded with improved uniformity of level and intended frequency response.

4.2. Technical Details

4.2.1. Overview

An AC-3 serial coded audio bit stream is made up of synchronization frames containing 6 coded Audio Blocks (AB), each of which represent 256 audio samples. A Synchronization Information (SI) header at the beginning of each frame contains information needed to acquire and maintain synchronization. A Bit Stream Information (BSI) header may follow SI, and contains parameters describing the mode of the coded audio service(s). The Audio Blocks may be followed by an Auxiliary Data (Aux Data) field.

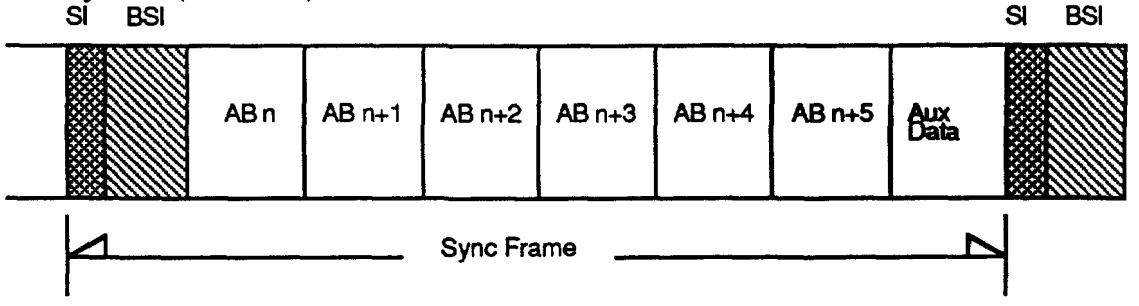

Figure 1 AC-3 Data Stream

4.2.2. Filterbank

AC-3 is a transform coder (see Figure 2) using the Time Domain Aliasing Cancellation (TDAC) transform. The filterbank is critically sampled, and of low computational complexity. Each input time point is represented in two transforms by means of the 50% overlap/add windows. The Fielder window is used which offers an optimum trade-off of close in frequency selectivity and far away rejection. The block size is 512 points which is the optimal transform length for most program material, being a good trade-off between frequency resolution (which determines coding gain and thus minimum bit rate), and complexity (primarily the buffer lengths and memory requirement in the decoder). The frequency resolution of the AC-3 filter bank is 93 Hz, uniform across the spectrum.

The 512-point transform is done every 256 points, providing a time resolution of 5.3 ms (at 48 kHz sampling rate). This time resolution is sometimes insufficient depending on the temporal characteristics of the input signal. During transients a finer time resolution is needed in order to prevent pre-noise artifacts at low bit-rates. The encoder controls the transform length, and during transients will switch to a 256-point transform for a time resolution of 2.7 ms. The algorithm which determines the optimum transform length resides only in the encoder, and may be improved

or updated without affecting the decoders already in the field. The transition between blocks is seamless, and the aliasing cancellation of the transform is maintained along with critical sampling. During block switching, there is no increase in filterbank computation rate.

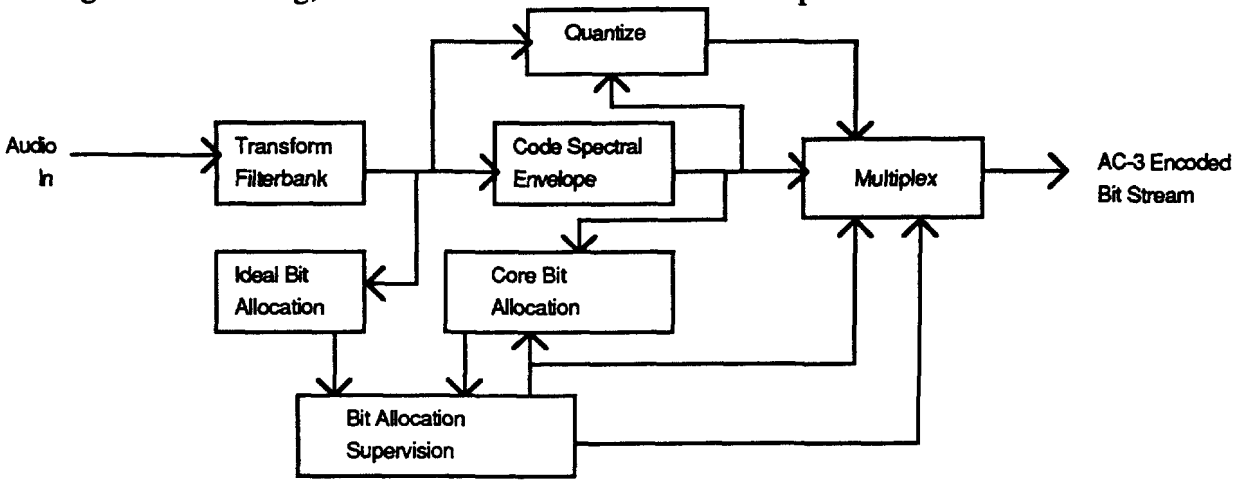

Figure 2 AC-3 Encoder

4.2.3. Spectral Envelope

The output of the 512-point TDAC transform gives 256 real-valued frequency coefficients. The logarithm of the absolute values of these coefficients forms the spectral envelope. The spectral envelope is coded with time and frequency resolution which is signal dependent. The frequency resolution of the spectral envelope can vary from the fundamental filterbank resolution of 93 Hz, up to 750 Hz, depending on the signal. The time resolution of the spectral evelope can vary from 5.3 msec to 32 msec, and is signal dependent. The algorithm which determines the time and frequency resolution of the spectral envelope is in the encoder, and may be updated or improved without affecting the decoders already in the field. The encoder decodes the encoded spectral envelope, so as to make use of the identical information that is available to the decoder. The decoded representation is used both as a reference for the quantization of the transform coefficients, and to drive the bit allocation routine.

4.2.4. Bit Allocation

The decoded spectral envelope has sufficient precision to be used for a psychoacoustic based bit allocation procedure. The encoder performs an iterative bit allocation routine altering the noise to mask ratio of all channels until the number of available bits is used up. Key hint information is coded and transmitted to the decoder which performs the core bit allocation routine only once and arrives at the identical bit allocation which was obtained by the encoder. While not as ideal as a bit allocation routine that uses the full knowledge available at the encoder, this approach saves the overhead required to explicitly transmit the bit allocation to the decoder, and allows more frequent updating of the allocation to be performed. The bits are allocated to each channel out of a common bit pool. Each individual channel may be allocated a differing number of bits.

The bit allocation may be modified by one of two methods. First, parameters of the psycoacoustic model may be altered by side information which is transmitted from the encoder to the decoder. These psycoacoustic model parameters are used in the core bit allocation routine which is computed in both the encoder and decoder. Second, the encoder may optionally send additional side information in order to explicitly alter some or all of the allocation values derived in the core bit allocation routine.

An AC-3 encoder may compute an ideal bit allocation based on full knowledge of the input signal, and compare the results of this allocation to that of the core routine. If the encoder determines that

improved subjective results can be obtained by altering the parameters of the psycoacoustic model used by the core routine, it can do so and send the appropriate side information to the decoder. If the results of the core routine are still sub-optimal, the encoder may send additional side information to explicitly modify the output of the core bit allocation routine.
Since the bit allocation is dependent on the spectral envelope, the bit allocation has time and frequency resolution equal to that of the spectral envelope. The bit allocation may change as often as every 5.3 msec, and the frequency resolution may be as fine as 93 Hz.
The core bit allocation routine used by AC-3 only requires the use of a rudimentary ALU with a pair of 16 bit accumulators. Only simple instructions such as ADD, SUBTRACT, COMPARE, AND, OR, and conditional branch are required. The psychoacoustic model used by the AC-3 encoder may be changed at any time without requiring a change in the installed base of decoders. The performance of AC-3 may thus improve over time as more knowledge is gained about the human auditory system.

4.2.5. Channel Coupling

At high frequencies the human ear determines directionality based on the signal envelope rather than on the particular signal waveform. Additionally, the human auditory system cannot independently detect directionality for multiple signals which are very close together in frequency. During periods of high bit demand, AC-3 takes advantage of these phenomena by selectively "coupling" (a form of combination) channels together at high frequencies. The method used maintains, within a critical band (a narrow region of frequency), the original powers of both the individual signals and the individual speaker outputs in order to avoid audible artifacts.
Coupling derives coding gains by combining number of individual channel transform coefficients (at a given frequency) into a common coefficient (at that frequency). The combinations only occur at high frequencies, above the "coupling frequency". The set of combined coefficients form the "coupling channel". Only the spectral envelope of the coupling channel is transmitted (instead of all of the individual spectral envelopes), along with the quantized values of the coupled coefficients (instead of all of the individual coefficients). The coupled coefficients are formed into bands, with widths similar to critical band widths of the human auditory system. For each band in each coupled channel a "coupling coordinate" is also transmitted. The coupling coordinate is used by the decoder to reproduce each coupled band out of each coupled channel loudspeaker at the original signal power level.
The channels which are coupled, the threshold frequency above which coupling takes place, and the widths of the coupled bands is determined by the encoder, and may dependent on the audio program content. The encoder algorithms used to determine these items may be updated at any time without requiring changes to decoders already in the field.

4.2.6. Synchronization and Acquisition

The AC-3 bit stream syntax has been designed to allow rapid synchronization and acquisition. The sync word is 16 bits long and has good autocorrelation properties. The probability of randomly finding a sync within the encoded data in a 32 msec sync frame is less than 3% (assuming byte level registration of the transport system). The first sync found will be correct 97% of the time, and it will be found within 32 msec. The sync frame may be verified by use of an embedded CRC check. If the CRC checks, one can be assured that true sync has been achieved, and decoding of audio may begin.

4.3. Features

4.3.1. Rates and Modes

The AC-3 algorithm supports industry standard sample rates of 48 kHz, 44.1 kHz, and 32 kHz. The Grand Alliance system may not support the use of more than one sample rate. Nominal bit rates are indicated by a bit field which indicates (via table lookup dependent on sample rate) the

number of bytes between sync codes. Nineteen nominal bit rates of 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320, 384, 448, 512, 576, and 640 kb/s are defined in the AC-3 algorithm, with an additional 13 rate codes reserved for future definition. Again, the Grand Alliance system may not support the use of all available bit rates. The main audio service mode can range from simple monophonic (1/0) through stereo (2/0) and up to all combinations of multiple channels (2/1, 2/2, 3/0, 3/1, 3/2). Any of the modes may optionally use a low frequency effects channel (0.1 channel).

4.3.2. Error Handling

The AC-3 syntax includes the use of 16 bit CRC checks for error detection. Some implementations of receiver architecture allow error flags to be delivered to the audio decoder to identify portions of the bit stream which may contain errors. Concatenation of the use of error flags and CRC checking is a useful technique to reduce the frequency of reproduced errors. The audio decoder must handle an occasional errored packet or dummy packet without significant audio disturbance. In the event of an error, previous audio data may be repeated. The window used for the FFT 50% overlap/add process is a very good window to crossfade the repeated time segment into the output time waveform. The implementation of error concealment does require some additional memory in the decoder, since old data must be kept available until a new data is verified as error free. Since the input data rate to the decoder is much lower than the output data rate, it is most efficient to keep a section of the previous input data.

4.3.3. Program Related Information

Several types of static program related information are contained in the bit stream information section of the audio frame. Program related information which is dynamic (such as dynamic range control) is coded into each individual audio block.

4.3.3.1. Center Mix Level

When a program with three front channels is mixed down to one or two channels for mono or stereo reproduction, artistic considerations sometimes require a different mixture of the Center channel signal with respect to the Left and Right signals. In order that the original multi-channel program mix not be compromised for the sake of mono or stereo reproduction, the AC-3 syntax allows different center channel downmix coefficients to be indicated to the decoder.

4.3.3.2. Surround Mix Level

When multi-channel sound which contains surround information is mixed down to stereo or mono, the optimum mixing level of the surround content which is mixed into the output channels may vary depending on the program type. The AC-3 syntax allows differing levels for the surround mixing coefficients to be indicated to the decoder.

4.3.3.3. Dialogue Level

In order to assure that different AC-3 encoded programs reproduce a uniform level of dialogue, there is an indication of dialogue level encoded into the data stream. Audio decoders make use of this information in order to reproduce all encoded programs with uniform loudness, consistent with the requirements of ATSC document T3-186.

4.3.3.4. Dynamic Range Control

The coded audio blocks contain a dynamic range control word which is used to modify the output level of the decoder. This word is encoded by the broadcaster or production studio in order to intentionally restrict the dynamic range of the reproduced sound to that which is suitable for the broad audience. The AC-3 decoder is required to, by default, reproduce audio with the intended (by those artistically responsible for the program) restricted dynamic range. The decoder may permit the listener the option to restrict the use of the dynamic range control word during reproduction, which will have the effect of reproducing more (or all) of the original program dynamic range. The coded AC-3 bit stream can thus serve a wider range of audience, including

those who require restricted dynamic range reproduction as well as those who wish to experience the full dynamic range of the original audio production.
The amplitude resolution of the dynamic range gain control word is less than 0.25 dB. This resolution, coupled with the smoothing effect of the gentle overlapping blocks of the TDAC alternating transforms, insures that there will be no audible modulation products generated by changes in the control word. During transient events, the control word may be updated as often as every 2.6 msec. The encoder may use any amount of time look ahead practical (including off-line disc based processing) to generate the most intelligent gain control words. Any algorithm may be used to generate these words (even direct manual control) and all decoders in the field will respond correctly.

4.3.3.5. Production Mix Room Level, Type

Optional fields allow the indication of the type of mixing room used to prepare the program (large cinema or small studio, resulting in slightly differing equalization) as well as the absolute sound pressure level (SPL) of the mixing session. The SPL code may be used along with the dialogue level indication to allow the reproduction of the program at the identical level used during the mixing session. This information may also be used to adjust the reproduced frequency response to compensate for acoustic reproduction at a different level SPL level so as to maintain the same perceived frequency response (true loudness compensation based on differential equal loudness hearing contours).

4.3.3.6. Time Synchronization

The AC-3 data stream may be tagged with an hour, minute, second, frame, and fractional frame time code which is useful for synchronization with SMPTE time code. This information is useful to keep audio and video bit streams in sync in the production environment, and may be useful in the home environment as systems become more complex and interoperable.

4.3.3.7. Additional Information

Additional information may be appended to bit stream information and will be ignored by decoders not intended to recognize it.

4.3.4. Associated Services

The AC-3 syntax and the Grand Alliance transport system allow AC-3 bit streams to convey associated services. Each associated service may be one of the following types: visually impaired, hearing impaired, commentary, dialogue, or emergency flash message. Each associated service may have an optional indication of language.

4.3.5. Auxiliary Data

If the audio data rate is restricted to less than the indicated rate, the excess capacity will result in unused bits at the end of the data frame. This auxiliary data will be ignored by the audio decoder, and may be used for any purpose.

4.4. Compatibility

4.4.1. Compatibility with Mono and Stereo Reproduction

The multi-channel AC-3 bit-stream can serve listeners requiring mono and stereo outputs. When fewer than the full number of output channels is required, the particular down mix is done in the decoder and can be optimum for each listening situation. The program originator can indicate which downmix coefficients are appropriate for a given program. The down mix may be done in the frequency domain, so that only the desired number of output channels needs to be transformed back into the frequency domain and buffered for output. This saves on complexity for AC-3 decoders which only have mono or stereo outputs.

4.4.2. Compatibility with Dolby Surround Encoded Programs

When operated in the 2/0 mode, AC-3 will convey 2 channel Dolby Surround matrix encoded soundtracks without degrading the performance subsequently obtained when further decoding the 2 signals into 4 signals with a Dolby Pro Logic surround decoder. The AC-3 coder makes use of internal re-matrixing in order to avoid the unmasking of coding artifacts which may occur when matrix decoding. Re-matrixing places the dominant coding error in the same location as the dominant signal energy, and avoids the case of a loudspeaker reproducing only coding artifacts which may not be masked by signals from other loudspeakers. This feature is especially important when operating at low data rates, such as 2/0 mode at 192 kb/s.

4.4.3. Compatibility with Dolby Surround Decoders

The AC-3 decoder can provide a 2 channel stereo output from a multi-channel bit stream which is compatible with Dolby Surround matrix decoding. This is a useful feature due to the large number of surround decoders already in use in the field which only have 2 channel inputs.

4.5. ATSC Document T3-186

Portions of the ATSC document T3-186 are concerned with audio coding. AC-3 has been designed with the ATSC concerns in mind, and satisfies all of the requirements of T3-186. The audio related sections of T3-186 are referenced below, and the methods by which AC-3 satisfies the requirements are described.

4.5.1. T3-186 Section 1.5.4, Multi-Channel Audio

This section states the requirement for 5 channel coding, with the low frequency enhancement channel optional. Monophonic and two channel stereophonic modes are also required, along with the ability to decode the full 5 channel service into stereo or mono. AC-3 offers all of the necessary coding modes, which are further elucidated below. Lower cost mono or stereo AC-3 decoders may be built which only partially decode a full 5.1 channel data stream, mix the channels down in the frequency domain, and perform the inverse filter bank only on the needed output channels (see section 7 on complexity).

4.5.2. T3-186 Section 1.5.5, Multiple Languages

This section states the requirement to handle multiple languages. Multiple AC-3 data streams may be provided, each conveying a main service in a different language. The AC-3 bit stream information syntax allows a main audio service to be tagged with an internationally recognized language code. Associated services such as commentary and dialogue may be conveyed by AC-3 data streams and each of these may also be tagged with a language code.

4.5.3. T3-186 Section 1.5.6, Audio Services to the Visually and Hearing Impaired

The section states the requirement to allow associated services such as VI or HI to be provided along with a main service. The AC-3 syntax and Grand Alliance transport system allows single channel associated services to be delivered. These services can be tagged as to type: visually impaired, hearing impaired, or commentary.

4.5.4. T3-186 Section 1.5.7, Uniform Loudness

This section deals with uniform loudness. The AC-3 data stream contains a field for a reference level indication, and the decoder must use this information to adjust the reproduce level. This allows the inter-cutting of bit streams from different sources with different loudness calibrations and headroom without undesirable level variations occurring in the level of reproduced dialogue.

4.5.5. T3-186 Section 1.5.8, Dynamic Range Control

This section considers dynamic range control. AC-3 contains an integral dynamic range control system. The broadcast encoder, or another piece of equipment, may insert codes into the data stream which the decoder uses to dynamically alter the reproduced level. This allows the broadcaster to intentionally compress the dynamic range, but allows the listener the option to reproduce either the compressed or the original (or something in-between) dynamic range.

4.5.6. T3-186 Section 1.5.11, Error Correction and Concealment for Audio Services

The section directs the audio system to take effective methods to minimize audible disturbances caused by uncorrectable errors. The AC-3 data stream contains error detecting codes which will allow the AC-3 decoder to determine if the audio data is valid. If the audio data is not valid error concealment is performed, in which a previous block of data is decoded instead of the current errored block. The repeated block is innocuously spliced in with the gentle FFT window function and this process is typically inaudible. In the case of a television set, the transport layer may deliver error flags to indicate that certain packets of data may be in error. In the event entire AC-3 bit stream is removed from the ATV set and delivered to another piece of equipment (interoperability), the error flag may no longer be available. In this case, the final AC-3 decoder can still perform error concealment triggered by the imbedded error detecting codes.

4.5.7. T3-186 Section 3.1.1, Independent Coding Modes

AC-3 can support all of the following bit-rates for an independently coded channel.
Independent 1/0: 32 kb/s, 40 kb/s, 48 kb/s, 56 kb/s, 64 kb/s, 80 kb/s, 96 kb/s, 112 kb/s, 128 kb/s, 160 kb/s, 192 kb/s, 224 kb/s, 256 kb/s, 320 kb/s, 384 kb/s.
(Note: for reasons of simplicity, the Grand Alliance system may not require decoders to recognize all modes, and thus a subset of these may be selected. A reasonable subset would those suggested by the ATSC: 64kb/s, 96 kb/s, and 128 kb/s although some consideration will be given to an additional lower rate.)

4.5.8. T3-186 Section 3.1.2, Composite Coding Modes

AC-3 can support the following composite coding modes at bit-rates exceeding the recommended minimum indicated below. Any of the modes may optionally support the low frequency effects channel.

Composite 3/2: ≥ 256 kb/s
Composite 3/1: ≥ 256 kb/s
Composite 3/0: ≥ 192 kb/s
Composite 2/2: ≥ 256 kb/s
Composite 2/1: ≥ 192 kb/s
Composite 2/0: ≥ 112 kb/s

4.5.9. T3-186 Section 3.1.2, Associated Audio Data

The reference level field in the AC-3 data stream (discussed under 1.5.7 above) satisfies the 3.1.3.1 headroom requirement. The dynamic range control codes (discussed under 1.5.8 above) satisfies the 3.1.3.2 dynamic range requirement. The AC-3 data stream has optional fields available for additional information satisfying the 3.1.3.3 requirement for descriptor information, and the 3.1.3.4 requirement for user bits.

4.6. AC-3 Decoder Architecture

An AC-3 decoder may be efficiently implemented by five functional blocks, as shown in Figure 4. The first block contains an 8-bit serial-parallel converter and input buffer RAM. The coded AC-3 data stream enters the decoder serially and is converted to bytes and stored in RAM.

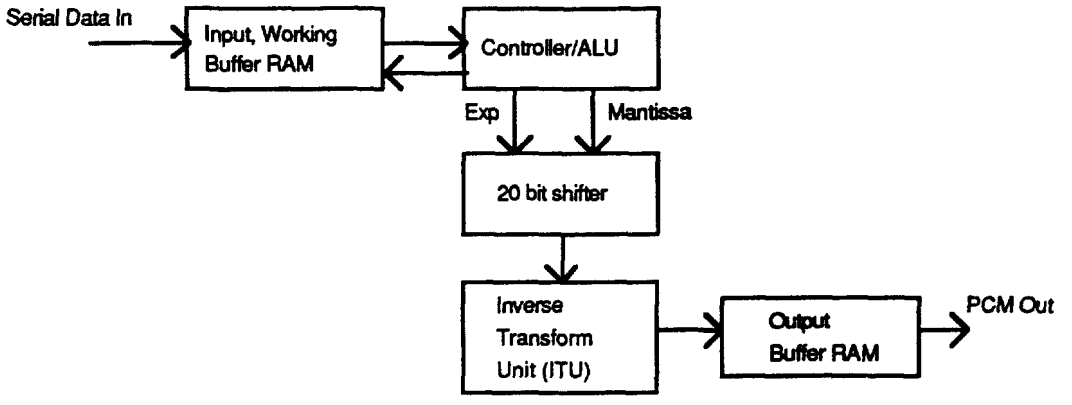

Figure 4 AC-3 Decoder

The second block contains a simple Controller/ALU incorporating two 16-bit accumulators, but no multiplier. The ALU requires the following instructions: add, subtract, compare, and, or, xor, conditional branch, and shift. The third block contains a semi-logarithmic shifter capable of implementing a 0 to 20-bit right shift on mantissas prior to inverse transformation. The semi-log design saves routing area compared to a straightforward 2^5-bit log-shifter. The Controller/ALU coordinates the input buffer and the 20-bit shifter. These three blocks perform the following functions:

1. Accept bit stream input interrupts from the serial-to-parallel converter and load the RAM.
2. Determine frame boundaries (frame synchronization).
3. Unpack control data.
4. Decode the spectral envelope.
5. Perform the core bit allocation routine.
6. Unpack mantissas, right shift up to 20 bits, and deliver data to the ITU.

The fourth circuit block, the Inverse Transform Unit (ITU), performs the frequency-to-time domain conversion (reconstruction synthesis filterbank). The ITU incorporates its own controller circuitry, and may operate somewhat autonomously from the previous circuit elements. Audio blocks representing 256 audio samples are inverse transformed using an efficient fast TDAC transform technique which minimizes computation and RAM requirements. The ITU requires the following instructions: multiply, multiply-add, and multiply-subtract.

Chapter 5
TRANSPORT SYSTEM

The Grand Alliance Transport System

Table of Contents

5.1. Introduction
- 5.1.1. Program vs. Transport stream multiplexing
- 5.1.2. Advantages of the fixed length packetization approach
- 5.1.3. Overview of the transport subsystem
- 5.1.4. General bit stream interoperability issues

5.2. The packetization approach and functionality
- 5.2.1. The "link" layer
  - 5.2.1.1. Packet synchronization
  - 5.2.1.2. Packet Identification
  - 5.2.1.3. Error handling
  - 5.2.1.4. Conditional Access
- 5.2.2. The Adaptation layer
  - 5.2.2.1. Synchronization and timing
  - 5.2.2.2. Random entry into the compressed bit stream
  - 5.2.2.3. Local program insertion

55.3. Higher Level Multiplexing functionality
- 5.3.1. Single Program Transport Multiplex
- 5.3.2. System Multiplex

5.4. Features and Services Supported by the GA ATV System
- 5.4.1. Features Supported within the Grand Alliance Video Syntax
- 5.4.2. Features Supported as Multiplexed Services within the Grand Alliance Transport System
- 5.4.3. Support of Closed Captioning
- 5.4.4. Features not Anticipated to be Transmitted by the Grand Alliance System to the Consumer Receiver

5.5. The Transport format and protocol
- 5.5.1. Link level headers
- 5.5.2. Adaptation level headers
  - 5.5.2.1. The PCR and OPCR fields
  - 5.5.2.2. The transport_private_data and adaptation_field_extension fields
  - 5.5.2.3. The splice_countdown field
- 5.5.3. The program_association_table
- 5.5.4. The program_map_table
  - 5.4.1. The overall TS_program_map_segment header format
  - 5.4.2. Header format for the individual program_definition_slices
    - 5.4.2.1. The format for the elementary stream description

5.6. The PES packet format
- 5.6.1. PES header Flags
- 5.6.2. The PES header

5.7. Conditional Access
- 5.7.1. General Description
- 5.7.2. Example of Conditional Access Implementation

8. Local Program Insertion
- 5.8.1. Systems level view
- 5.8.2. Basics of elementary bit stream insertion
- 5.8.3. Restrictions
- 5.8.4. Imperfect program insertion

9. Compatibility with other Transport Systems
- 5.9.1. Interoperability with MPEG-2
- 5.9.2. Interoperability with ATM
  - 5.9.2.1. ATM Cell and Transport Packet Structures
  - 5.9.2.2. Null AAL Byte ATM Cell Formation
  - 5.9.2.3. Single AAL Byte ATM Cell Formation
  - 5.9.2.4. Dual AAL Byte ATM Cell Formation

Appendix 1 Grand Alliance Submissions to MPEG systems Committee

The Grand Alliance Transport System

5.1. Introduction

This document provides a description of the functionality and format of the Grand Alliance transport system. While tutorial in nature, the document provides sufficient technical detail to serve as the operational specification of the transport layer. Issues related to terrestrial broadcast and cable delivery of the ATV service, in the context of ACATS discussions, are addressed in this document. The authors have attempted to make this a stand-alone document, though the reader would gain additional insight from associated ISO-MPEG documents on this and related topics.

In developing the specification of the Grand Alliance transport layer, we have drawn upon the collective experience of the member companies in developing individual systems, as well as the excellent body of work created by the ISO-MPEG standards process. While any system design requires intelligent tradeoffs to be made, selection of a format based on fixed-length packets has maintained a number of simultaneous goals.

5.1.1. Program vs. Transport stream multiplexing

In general there are two approaches for multiplexing elementary bit streams from multiple applications on to a single channel. One approach is based on the use of fixed length packets and the other on variable length packetization. Both approaches have been used in the MPEG-2 standard. As illustrated in Fig. 5.1.1, the video and audio elementary streams in both cases go through an initial stage of PES packetization (discussed in greater depth later), which results in variable length PES packets. The process of generating the transmitted bit streams for the two approaches is shown to involve a difference in processing only at the final multiplexing stage.