Channel Compatible DigiCipher HDTV System

Submitted by
Massachusetts Institute of Technology

on behalf of
The American Television Alliance
May 14, 1992

Contents

List of Figures

List of Tables

1 Introduction

The Channel-Compatible DigiCipher (CCDC) HDTV system is an all-digital HDTV system developed by the Massachusetts Institute of Technology and General Instrument Corporation. This document is submitted for final certification of the system.

The CCDC system has a number of important features. The system is channel-compatible and will fit within the channels currently being utilized for terrestrial broadcast transmission. The system is very efficient in using the given channel bandwidth. A high resolution progressively scanned baseline video signal of 1280×720 picture elements, 59.94 fps, and 16:9 aspect ratio, can be transmitted within a single 6 MHz channel.

The CCDC system is source-adaptive. The system can recognize and adapt itself to the particular characteristics of the source format so that the highest quality video can be reconstructed. The system, for example, accommodates different frame rates. The system automatically optimizes both the encoding and decoding processes so that the highest quality video can be obtained for each source format.

The CCDC system can accommodate a variable number of compact disc quality audio channels. The system can also accommodate expanded audio formats, such as surround sound. Since the number of bits used for audio is not predetermined, any remaining bits available after the audio is encoded are then used for video.

The CCDC system is resistant to channel impairments. Very high quality video can be delivered to the home despite substantial channel degradations such as noise, multipath, and frequency distortion. Because of this notable resistance to noise and interference, high picture quality can be achieved at low transmitter power, thereby making the use of taboo channels feasible.

The CCDC system has been designed to be highly modular. Modularity, a useful feature in any engineering system, offers many advantages. For example, the video processing, audio processing and transmission systems are separated and are largely independent of one another. As a result, the video and audio processing systems can be used for applications other than HDTV. In addition, they will not require significant changes when other media are used for transmission. Modularity also leads to reduced hardware complexity which will ultimately lead to savings for consumers.

We believe that selection of the CCDC system, as represented by the hardware system that is submitted to the ATTC for testing, will result in a high performance, robust and cost effective HDTV standard for many years to come. We recognize, however, that further improvements in video coding efficiency are possible in the future and performance extensions may well be desirable. With this in mind, we have designed the data structure of the CCDC system to be flexible, so that these improvements can be incorporated in a backward compatible manner. With its flexible data structure, the CCDC system is extensible. The system will support, for example, a wide range of source adaptivity including different frame rates, resolutions, aspect ratios, progressive/interlaced scanning, and the black-and-white/color characteristics of the video.

The system is also scalable, in that performance subsets can be extracted from the CCDC HDTV signal. With proper design of the receiver, the received video may be displayed at multiple resolutions with modest amounts of signal processing. As an example, lowpass filtering and subsampling the signal is an effective method to obtain a lower spatial resolution image. In addition, row/column doubling methods can be used to increase the spatial resolution.

To achieve the features discussed above, state-of-the-art signal processing and communications technologies have been used. It is worth noting that the United States has been at the forefront in the development of these technologies. HDTV is not just a question of beautiful pictures and spectacular sound; it also affects our economy, our jobs, and our future standard of living. We hope that the adoption of American technology for the HDTV standard in the United States will open a window of opportunity through which we can regain leadership in manufacturing in the consumer electronics and semiconductor industries.

2 Overview

A block diagram of the CCDC HDTV system is shown in Figure 2.1. The system consists of three parts: video coding, audio coding and transmission. By coding, we mean to remove redundancy in the information source and to transmit only the essential information. The system architecture depends on the scope of coding operations which is set by the channel bandwidth (in this case, a single 6 MHz television broadcast channel) and by the cost of signal processing.

Signal processing was very expensive in the early days of television. This meant that the source coding was severely restricted, to an essentially uncoded format in which a video signal was scanned and the waveform directly modulated an RF carrier. When color was introduced in the mid-1950s, some source coding ideas were applied to insert chrominance into the signal at the expense of some luminance spectrum. Modern systems have taken an approach similar in spirit, and use a single channel much more efficiently than a conventional television system. This permits HDTV video signals with more than four times the spatial resolution, better motion reproduction, and high-fidelity audio to be transmitted over the same channel used for a conventional signal.

Similar remarks apply to audio. Gradual evolution shaped the FM audio systems of conventional TV, eventually adding stereo and auxiliary audio channels. Modern audio coders, using ideas from transform coding and psychoacoustics, can send several channels of compact-disc quality audio using only a small part of a broadcast channel's capacity.

A transmission or digital communications system is required to link the digital outputs of the various source coders to an analog RF channel. In a good design, the transmission system isolates the source coder from the channel and the two problems of representing the HDTV signal efficiently and transporting it from broadcaster to user become largely independent. The transmission system for a given medium (terrestrial broadcast, cable, satellite, videotape, etc.) must account for imperfections and noise in the channel and deliver a stream of bits as reliably as possible. Noise and bandwidth set theoretical limits on the capacity of a given channel, and the efficiency of a transmission system can be judged on this basis.

The data structure in the CCDC HDTV system supports a wide range of video formats. A video signal is taken to be a sequence of arrays of pixels, or frames. The system proposed as the HDTV standard uses 1280×720 pixels per frame at 59.94 fps. With the flexible data format, the system can be extended to accomodate other spatial resolutions, frame rates, and aspect ratios.

In the case of video source coding, there is redundancy along all three dimensions of the signal. Temporal redundancy is removed by a motion-compensated predictive coding scheme in which the motion between the previous and current frames is estimated, a predicted current frame is computed based on the previous frame and the extrapolated motion, and only the new information is sent to the spatial coder.

Spatial redundancy arises from objects which have extent and texture; these introduce correlations between neighboring pixels. The correlations can be removed with a linear map such as the Discrete Cosine Transform (DCT). The DCT coefficients are then weighted to reflect the relative importance of different frequency components, and coefficients which are perceptually important are quantized and entropy coded. The combination of spatial and temporal techniques yields a concise representation well suited to transmission.

In audio coding, the input signal is divided into short overlapping segments, then windowed and transformed. These transform coefficients are then divided approximately into the critical bands of the human auditory system. Psychoacoustic models are used to eliminate inaudible information in the signal spectrum. Spectral envelope information is encoded using a fixed bit allocation strategy, and this envelope is used for dynamic bit allocation for the transform coefficients.

The encoded video data, audio data and auxiliary data are multiplexed into a single bit stream for the transmission system. Block and trellis encoding are used to reintroduce some redundancy to combat channel noise, and the resulting digital symbols are transformed into analog waveforms through quadrature modulation. The system can operate at 32 QAM (preferred mode) or 16 QAM, depending on the desired noise threshold/coverage area.

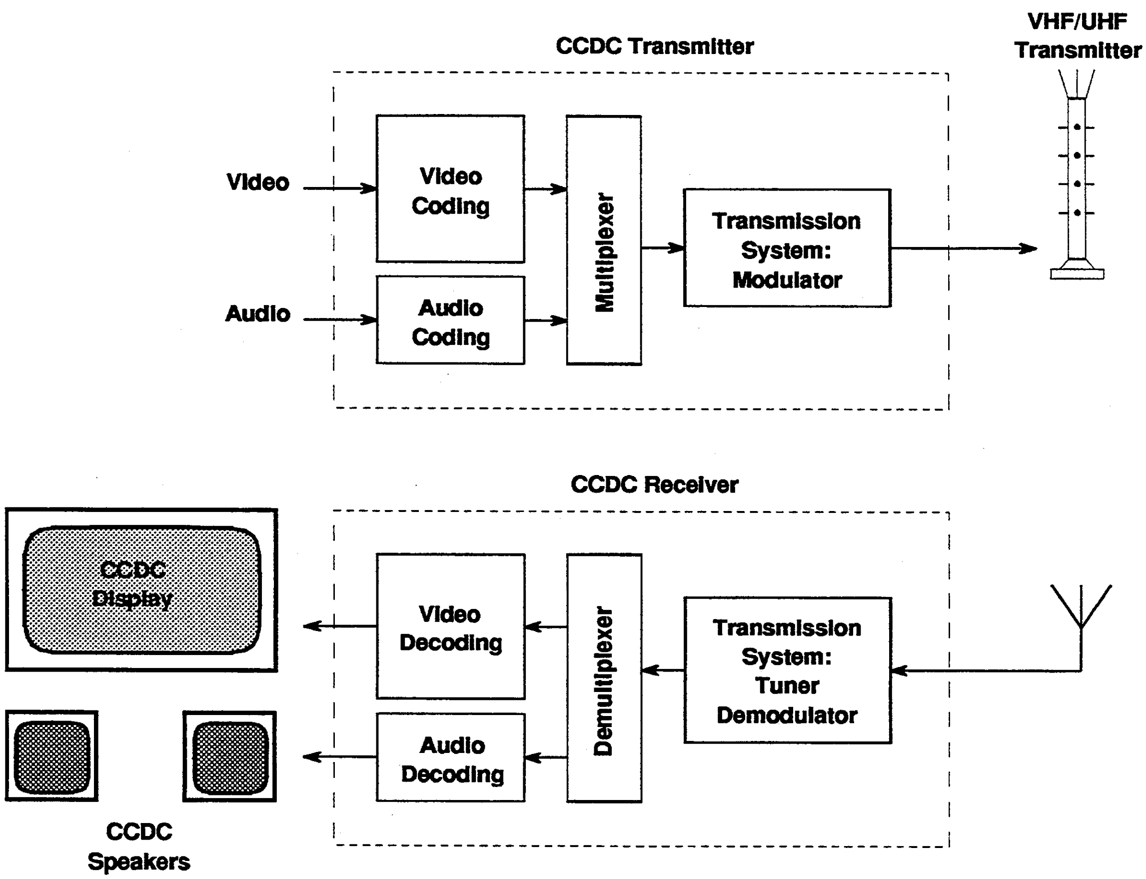

Figure 2.1: CCDC high-level block diagram.

3 Video Processing

The video processing for the CCDC HDTV system contains a number of novel and sophisticated signal processing techniques. The system is designed to achieve high video quality while keeping the cost of the receiver as low as possible. An overview of the video encoding portion of the system is shown in Figure 3.1 and consists of signal format identification, source adaptive processing, motion estimation and compensation, DCT analysis of motion compensated residuals, adaptive selection of the visually important transform coefficients, quantization of the selected coefficients, and entropy coding of these quantized coefficients. The decoding portion of the system consists of transform synthesis of the motion compensated residual.

3.1 Video Encoding

3.1.1 Video Source

A video signal consists of a sequence of frames; each has a header which contains all the information that is required or useful about the frame, such as spatial resolution and aspect ratio. The video signal to be used as a test source for the CCDC system is analog RGB with 720 active line, 59.94 fps progressive scan. Each line has 1280 pixels, so the resulting aspect ratio with square pixels is 16:9. The video signal will have originated from three

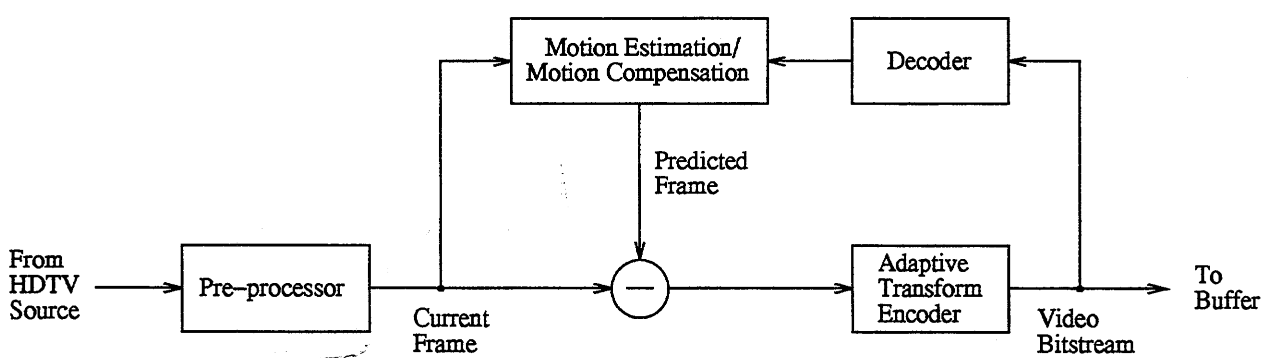

Figure 3.1: An overview of the video encoder.

possible frame rates, 59.94 fps, 30 fps, and 24 fps.

3.1.2 Extraction of Frame Header Information

The frame header includes the frame rate, spatial dimensions of the frame in pixels, aspect ratio, frame versus partial frame (such as a field), and color space (e.g., monochrome versus color). The frame header may also contain global data for the processing algorithms; for example, the entire frame may be encoded without any form of temporal processing. This will be discussed further in subsequent sections.

Since most of this information is known at the video postproduction stage, the best approach is to explicitly include it as part of the video source. The only parameter which is not fixed for the video sequences to be tested by the ATTC is the frame rate that the video signal originated from: 59.94 fps, 30 fps, or 24 fps. Since this parameter is not provided explicitly in the test signal, it became necessary to extract this information from the supplied video. When the video originates from 30 fps material, two consecutive frames are identical (after upconversion), the next two frames are identical, and so forth. When the video source used is 24 fps, three frames are identical, the next two frames are identical, and so forth. These patterns are used to determine the appropriate frame rate.

3.1.3 Source-Adaptive Encoding

A given television system must provide an interface to many kinds of imaging systems. These include other television systems, the various film standards, magnetic and optical media, and synthetic imagery. In the CCDC HDTV system, we exploit the differences that may exist in the sources to improve the performance of the video processing system. For example, consider the source material originating from film at 24 fps. Instead of converting 24 fps to 59.94 fps rate using 3:2 pulldown and then encoding 59.94 fps at the transmitter, only the original 24 fps material is encoded and transmitted. This improves the coding efficiency significantly.

The video signal supplied by the ATTC has a frame rate of 59.94 fps. This video signal originated from video source of 59.94 fps, 30 fps, or 24 fps. In each case we derive a sequence of frames at the original rate. This becomes the input to the video coder.

Depending on the source rate and characteristics, it is possible to use very different encoding strategies. The encoding method that has been used for the current frame may be included in the header. In the CCDC system, we have chosen to use the same encoding method and to change only the effective coding rate for the specific source material. We believe that this method reduces the cost of hardware significantly with little loss of performance.

3.1.4 Color Space Conversion

The color space conversion operation converts RGB components to YUV components using approximately the SMPTE 240M standard. This takes advantage of the differing signal characteristics and differing responses of the human visual system to the luminance and chrominance components. Each of the YUV components has the same full spatial resolution as the original RGB components.

In the CCDC system, the chrominance components U and V are lowpass filtered and subsampled by a factor of two along both the horizontal and vertical directions. A possible future extension is to treat the Y, U, and V components in the same manner and to exploit the difference in the relative importance among the Y, U, and V components at a later stage in the processing. This would allow full chrominance resolution as well as full luminance resolution when the bit rate is sufficient to accommodate both the luminance and chrominance components. This feature could be particularly significant for movies at 24 fps, where the effective coding rate is high. Also computer generated graphics and text containing saturated colors would benefit from the full chroma resolution.

For the remainder of this document the following terms will be used in describing aspects of the video compression algorithm for the CCDC HDTV system to be tested.

- Pixel: A sample of video corresponding to either a luminance or chrominance component.
- Block: An 8 pixel by 8 pixel region of an image.
- Superblock: A grouping of 4 luminance blocks and the associated 2 chrominance blocks (one of each U and V) corresponding to a 16 by 16 pixel area of an image.
- Macroblock: A grouping of 20 consecutive superblocks.

3.1.5 Intraframe and Interframe Encoding Modes

Video contains information along both spatial and temporal dimensions. There is a considerable amount of redundancy among these dimensions, and the reduction of this redundancy is required for the transmission of high quality video at low bit rates. By making a prediction of the current pixel from the previous pixels and encoding only the prediction error or residual, much of the redundancy within the signal is eliminated. This is the main idea behind all differential pulse code modulation (DPCM) systems.

Motion estimation/motion compensation (ME/MC) is a form of predictive coding usually applied along the temporal dimension. The specific details of this scheme will be described in the following sections. Since ME/MC is a predictive coding scheme, there are some frames of a video sequence where a prediction can be made reasonably well, while with other frames, such as those at scene changes, a prediction is not as useful. In some cases, therefore, the system may perform worse with predictive coding than by simply coding the frame itself. Therefore, it is important to determine when these events occur and to process accordingly.

For each video frame we indicate if the entire frame should be encoded without using ME/MC. For convenience we will refer to the two encoding modes as

- Intra: frame encoded by itself, no temporal prediction, no ME/MC

- Inter: frame encoded using temporal prediction, with ME/MC

The intra mode, which does not use any prediction, is useful when there is significant scene change or new imagery so that the prediction would not be accurate. This mode is also useful for a production standard, as discussed in Section 9.5. One bit in the frame header is used to signify if the entire frame was encoded using only the intra mode.

3.1.6 Motion Estimation and Motion Compensation

A video signal is a sequence of still frames that are shown in rapid succession to give the impression of continuous motion. Even though each of the frames is distinct, the high frame rate necessary to achieve proper motion rendition usually results in a great deal of temporal redundancy among the adjacent frames. For example, frames adjacent in time often contain the same object, possibly at different spatial positions. This can be exploited in reducing the temporal redundancy. The process of estimating the motion of objects within a video sequence is known as motion estimation. The processing of images compensating for the presence of motion in the sequence is called motion compensation. The combined processes of motion estimation and motion compensation (ME/MC) produce a prediction of the current frame from the previous frame. The error in this prediction, called the motion compensated residual (MC-residual), can then be processed and transmitted. In this manner, ME/MC can be viewed as a form of predictive processing along the temporal direction.

The motion model used is locally uniform translation. Let $f(x,y)$ and $\hat{f}_p(x,y)$ denote the luminance component of the current frame and previously encoded frame, respectively, at spatial location (x,y). We assume that

$$f(x,y) = \hat{f}_p(x-d_x, y-d_y)$$

where d_x and d_y are the horizontal and vertical displacements between the previous frame and the current frame. An example of $f(x,y)$ and $\hat{f}_p(x,y)$ that satisfies the above assumption is shown in Figure 3.2. A global translational motion assumption would be very restrictive.

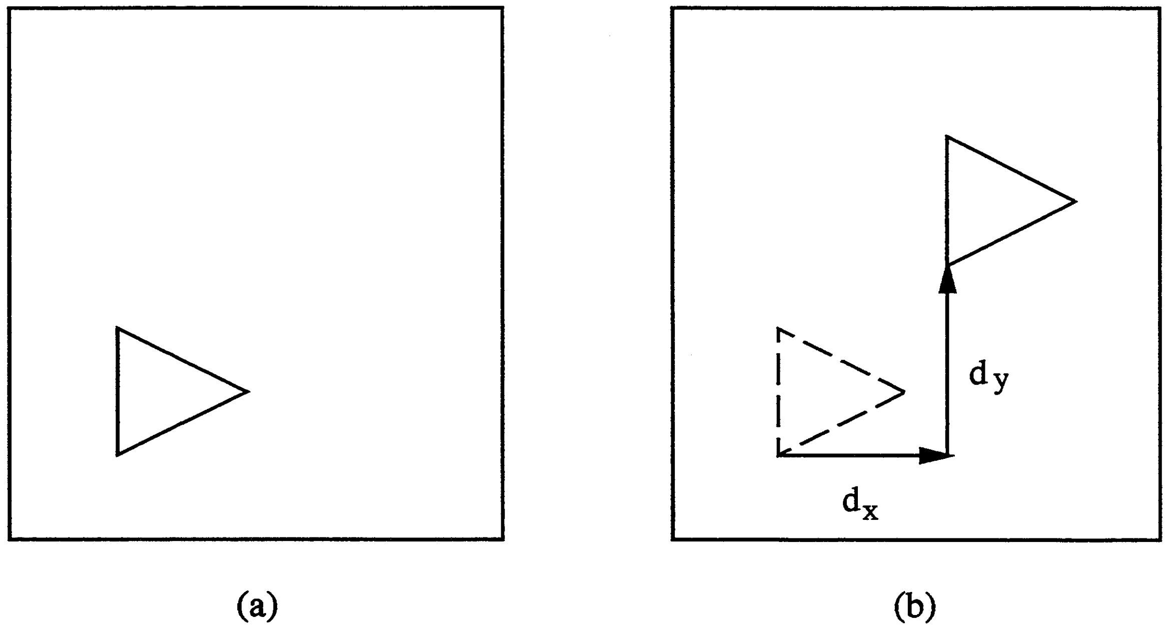

Figure 3.2: An object that is displaced by d_x and d_y.

However, by assuming the translational motion assumption only locally and estimating the displacement vector at various spatial locations, motion compensation based on the translational motion assumption has been observed to be very effective in reducing the temporal redundancy in video.

Motion estimation methods can be classified broadly into two groups: region matching methods and spatio-temporal constraint methods. Region matching methods involve considering a small region in a frame and searching for the displacement which produces the "best match" among possible regions in an adjacent frame. Spatio-temporal constraint methods use an error criterion defined in the spatio-temporal constraint equation domain.

The specific choice of a motion estimation method is standard-independent. If a better motion estimation method which gives a more accurate motion estimate is developed in the future, it can be used within the broadly defined standard. In the CCDC HDTV system, we have used a region matching method with exhaustive search. This choice was made because VLSI that estimates motion based on region matching with exhaustive search is commercially available and the method gives an accurate estimate of motion.

In the CCDC HDTV system, the displacement vector (d_x, d_y) is estimated by minimizing

$$\text{Error} = \sum_{(x,y)\in R} |f(x,y) - \hat{f}_p(x - d_x, y - d_y)|$$

where R is the search region over which the computation is performed. The search region surrounds the spatial location at which we wish to estimate the displacement vector. The block size of the region to be displaced is adaptive based on the local characteristics of the video to be encoded. This will be further discussed in Section 3.1.8.

The motion vectors are estimated only from the luminance (Y) component and the same displacement vectors are applied to both the luminance (Y) and chrominance (U and V) components. The search region for the motion vectors is +15/-16 pixels horizontally and +7/-8 pixels vertically. The larger horizontal search range corresponds to the increased likelihood of rapid motion along the horizontal direction. This search range will allow the encoder to track objects moving at up to 0.75 frame width and 0.67 frame height per second. This is especially useful for encoding video with large frame to frame displacements such as sporting events. The motion vectors have $\frac{1}{2}$ pixel accuracy.

Although a larger search region for the motion vectors was considered, it is evident, after testing on a wide range of source material, that this increase is not really necessary. In addition, camera integration with such fast pans will cause blur and reduce the need for an accurate temporal predictor.

After the motion estimation is completed, motion compensation is performed. The computed motion vectors are used to displace blocks of the previous frame in the appropriate manner to create the prediction of the current frame. The size of the blocks to be displaced is once again dependent upon the local characteristics of the video to be encoded. The motion vectors are differentially Huffman encoded for transmission.

3.1.7 Spatial Transform

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

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

The size of the DCT might be chosen as the entire frame, but much better performance can be achieved by subdividing the residual into many smaller regions each of which is individually processed. The motivation for this is very easy to see and it is one of the most important aspects of a high quality video encoder. If we compute the DCT of the entire residual frame, we treat the whole residual frame equally. For a typical image, some regions contain a large amount of detail and other regions contain very little detail. By exploiting the changing characteristics of different images and of different portions of the same image, significant improvement in performance can be realized. In order to take advantage of the varying characteristics of the residual over its spatial extent, the residual is partitioned into 8×8 blocks. The blocks are then independently transformed and adaptively processed based on their local characteristics. The partitioning of the residual into small blocks before taking the transform not only allows spatially adaptive processing, but also reduces the computational and memory requirements. Partitioning the signal into small blocks before computing the DCT is referred to as the Block DCT.

In determining the optimal method to transform/subband filter the residual, a number of schemes were examined. The Block DCT may sometimes result in artificial blocking artifacts. Schemes such as multi-resolution subband coding and overlapped transform coding could eliminate many of the blocking artifacts. On the other hand, we have observed that they create other types of artifacts. Based on current results from our studies, the overall performance of the Block DCT appears to be at least as good as other methods, while the required complexity was the least for the Block DCT. Therefore, the Block DCT was chosen for encoding the residual.

3.1.8 Spatially Adaptive Processing

As mentioned above, when using motion compensated prediction, there are times when the prediction error is large, such as during a scene change or new imagery. Coding the residual at such times may be more difficult than coding the original image. In those cases, the motion compensated prediction should be suppressed and intraframe processing, where the original image itself is coded, should be performed. Just as the ME/MC may fail over the entire frame, it may also fail over only a small region of the frame. As the ME/MC may perform better over some regions than others, the "inter/intra" decision making process should be performed in a spatially adaptive manner. The block DCT is convenient to use for spatially adaptive processing because it partitions the image into separate blocks which are processed independently. This allows inter/intra decisions on local regions of the image to be made easily.

The optimal block size over which ME/MC should be performed may also vary from one region to another. For processing some regions of the residual, a 16×16 block size may be appropriate, while for other regions, an 8×8 block size may yield higher performance. Note that if the entire residual is processed with ME/MC utilizing 8×8 blocks there would be four times as many motion vectors to be encoded relative to using 16×16 blocks. Therefore, there is a tradeoff between higher resolution in the ME/MC operation and the number of motion vectors to be encoded.

For each local region (8×8 pixel block) of the residual, a decision will be made as to whether it should be encoded using the inter-mode or intra-mode processing and if the block size used for ME/MC should be 16×16 or 8×8. Therefore, there are three possibilities (intra-mode, inter-mode with 16×16, and inter-mode with 8×8) for each block. Each superblock (16×16 pixel region) which consists of four blocks has $3^4 = 81$ different possibilities of coding. These are grouped into the following three superblock encoding modes:

1. Intra-mode encoding of all blocks
2. Intra-mode encoding and inter-mode encoding utilizing MV_{16}
3. Intra-mode encoding and inter-mode encoding utilizing both MV_{16} and MV_8

where MV_{16} corresponds to ME/MC based on 16×16 block size and MV_8 corresponds to ME/MC based on 8×8 block size. The decision criterion applied is the fewest number of bits required to achieve the same quality reconstructed video. Therefore, for each superblock of the residual each of the three schemes above is applied to see which achieves the most efficient compression. This algorithm allows adaptive inter/intra-mode processing and adaptive block size selection for ME/MC. Therefore, the video encoder has two primary modes of operation, (1) when the entire frame is encoded using intra-mode and (2) when the entire frame is encoded utilizing the super block encoding modes in a spatially adaptive manner. Note that one is a special case of the other, but a global decision requires less side information. A block diagram of the video encoder is shown in Figure 3.3.

There is an added benefit from spatially adaptive inter-mode/intra-mode encoding. Whenever a predictive encoding scheme like ME/MC is used, it is important for the receiver to have the same previous frame as the encoder. When the receiver is initially turned on, the channel is changed, or uncorrected errors occur over the channel, however, the receiver does not have the previous frame information to use in the prediction process. Two possible

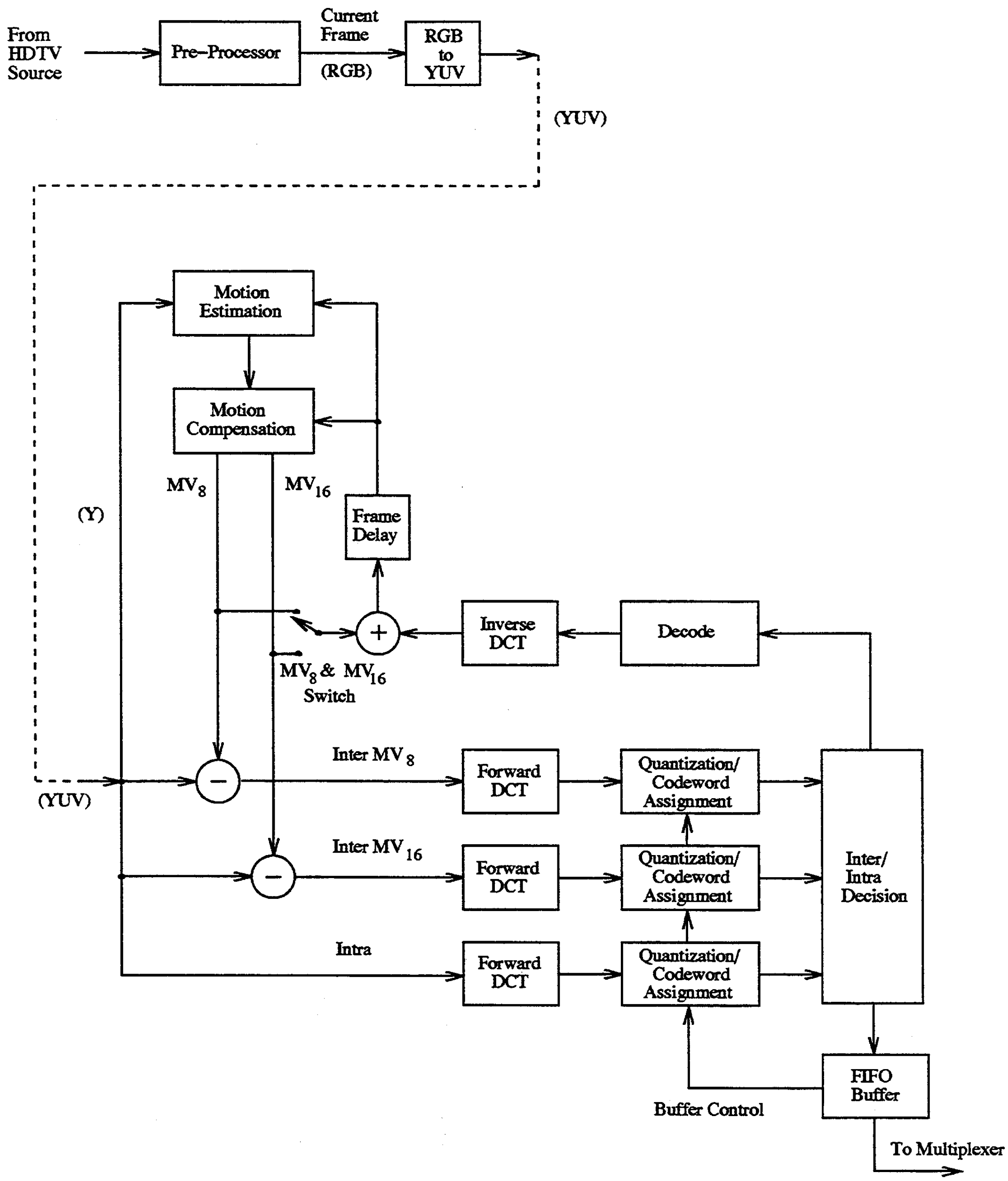

Figure 3.3: The video encoder.

solutions are having some leakage in the ME/MC, or periodically encoding a full frame of video using intra-mode. Both of these methods have drawbacks. Leakage is designed for one-way processing (forward only) and prevents simple execution of VCR commands such as reverse playback. Periodically encoding an entire frame using intra-mode may result in artifacts in the reconstructed frame because of the higher bit rate required for intra-mode encoding and the limited distribution capabilities of the buffer.

Spatially adaptive inter-mode/intra-mode processing, on the other hand, allows us to solve this problem in a very simple and elegant manner. For every frame that is transmitted, successive columns of the video are transmitted using intra-mode. In this manner during a predetermined length of time the entire video frame is encoded using intra-mode and the increase in bit rate due to this intra-mode encoding is uniformly distributed over time. As an example, if a column of four superblocks in width (64 pixels wide) is encoded using intra-mode for each frame, and this column is swept across the video data, every block of video would be encoded using the intra mode once every 20 frames. Similarly, the increase in required bit rate for intra-mode encoding would be uniformly spread over 20 frames as opposed to a single frame receiving the full hit in performance. Consequently, the video would be refreshed three times per second and the channel acquisition time would be 0.33 seconds, excluding receiver synchronization to be discussed in Section 7.

3.1.9 Coefficient Quantization and Entropy Coding

The key step taken toward reducing the required bit rate is the quantization and encoding of the DCT coefficients. This process is illustrated in Figure 3.4. One wants to code each coefficient with as few bits as possible. The quality of the HDTV video we perceive is directly related to how the human visual system responds to the video. By taking the human visual system into consideration the performance of a system can be greatly enhanced. For example, some of the DCT coefficients are visually more important than the others. The human visual system can perceive detail more easily in the luminance than in the chrominance components. Also errors in coding the high frequency coefficients are not as visible as errors in coding the low frequency coefficients.

To exploit our understanding of the human visual system, an adaptive quantizer is used for encoding the DCT coefficients. There are 6 times 32 different sets of 8×8 weighting factors that may be applied to each 8×8 block of DCT coefficients. Specifically, we have different sets depending on the interframe or intraframe processing mode. We also have different sets depending on the Y, U, or V component. For each of these six different modes, we have a different set for each of the 32 different quantization levels. We, therefore, have 192 (6×32) different sets of 8×8 weighting factors. Clearly, the quantizer is adaptive in that the choice of weighting factors depends upon the interframe or intraframe mode, the luminance/chrominance (Y, U or V) component of the coefficients, and measures of the local scene complexity and buffer fullness. The weighting factors effectively make the individual DCT coefficient quantizer more coarse (increasing the stepsize) or fine (decreasing the stepsize) in order to take advantage of the differing sensitivity of the human visual system. For example, each set of 8×8 weighting factors is designed to vary with frequency. This is because the human visual system has reduced sensitivity to high-frequency quantization noise in comparison with low-frequency quantization noise. Therefore, the set of 8×8 weighting factors is selected to produce coarser quantization of the high frequency coefficients as compared to the low frequency coefficients. The human visual system appears to be less susceptible to quantization noise in the chrominance than in the luminance components, and the chrominance is therefore quantized more coarsely.

The DCT coefficients are also weighted depending upon the buffer fullness and the local scene complexity. The motivation for weighting based on the buffer fullness will be discussed in the next section. The idea behind weighting based on the local scene complexity is that the human eye is less sensitive to artifacts in detailed regions as compared to artifacts in less detailed regions. This phenomenon is often referred to as spatial masking. A local scene complexity indicator is used to determine the complexity of a superblock, and based

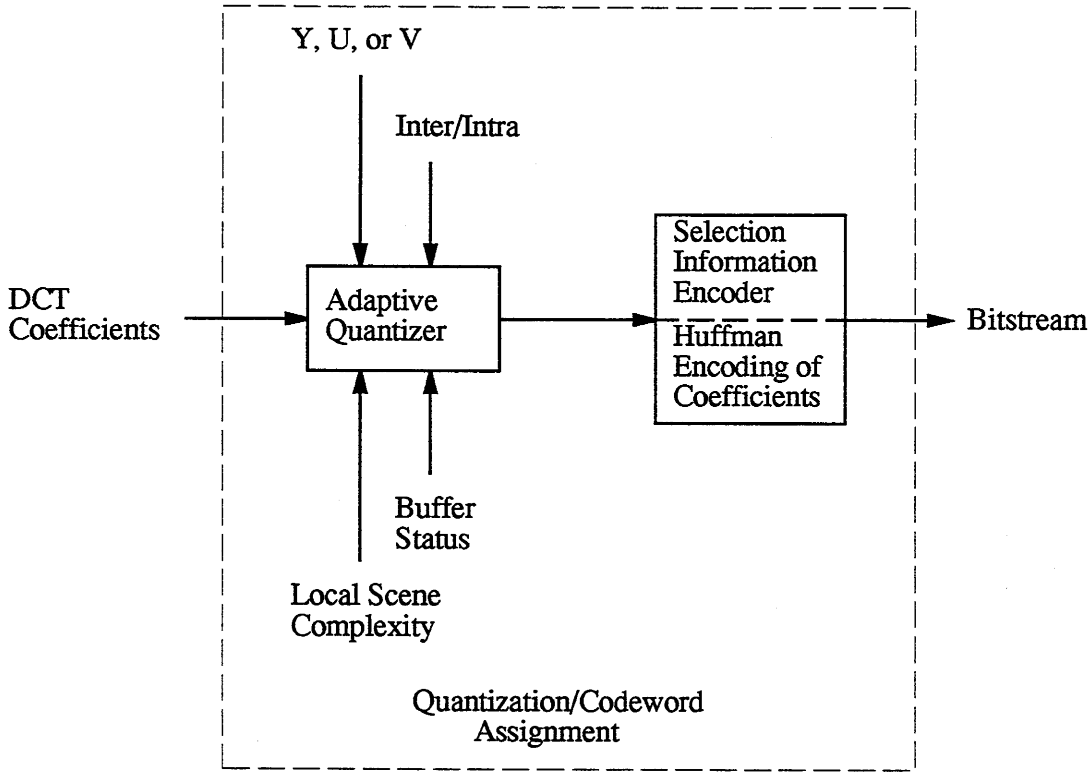

Figure 3.4: Quantization and encoding of the DCT coefficients.

on the determined complexity, all the DCT coefficients are weighted for that superblock in an appropriate manner. If the superblock is determined to be very detailed, the DCT coefficients are quantized more coarsely.

The DCT coefficients are first weighted by the appropriate weighting factors and then passed through a uniform quantizer. The quantization results in most of the coefficients being quantized to zero. Since a large percentage of the coefficients are quantized to zero, it is more efficient to tell the receiver which coefficients are nonzero and transmit them, instead of transmitting a large number of zero coefficients. The coefficients which are nonzero will hereto be referred to as the selected coefficients, and the information describing which coefficients have been selected will be called the selection information.

There are a number of different methods to encode the selection information. One widely used method is runlength encoding. In this method, all the coefficients are ordered in some manner, and the runs between one selected coefficient and the next selected coefficient are encoded. For image coding applications the coefficients are usually scanned in a zigzag pattern as shown in Figure 3.5 for highest efficiency. This usually results in long runlengths which, in general, result in higher performance. One difference between typical intra-frame image coding applications and the HDTV system is that the HDTV system encodes a residual. By taking the transform of an image, most of the selected coefficients tend to be in the low-frequency region of the spectrum. The runlengths are therefore usually rather long and normal runlength encoding works very well. On the other hand, a residual typically has its energy more uniformly distributed in the frequency domain. The selected coefficients are then more sparsely spread, resulting in many more short and inefficient runlengths to be encoded.

To efficiently represent the selection information, a number of schemes have been examined. In the CCDC system, a Vector Coding (VC) scheme is utilized. This scheme outperforms conventional runlength encoding methods.

In the VC method, the 8×8 block of DCT coefficients is first divided into 4 regions,

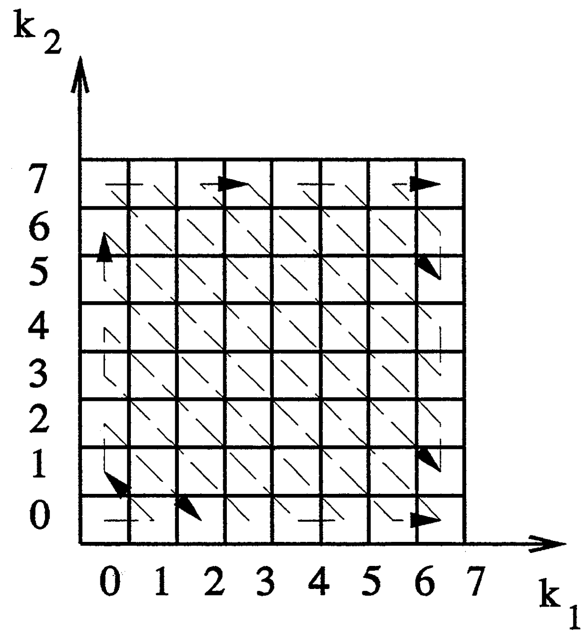

Figure 3.5: Zigzag scanning of the DCT coefficients.

each containing 16 transform coefficients. As shown in Figure 3.6, the division is such that the first region contains the low frequency coefficients which are most likely to be selected, and the last region contains the high frequency coefficients which are least likely to be selected. Since there are 16 coefficients in each region, we have 2^{16} (65536) possible selection patterns. As we represent each possible selection pattern distinctly, the encoding of the selection information is lossless. This is one reason why this method is called vector coding, as opposed to vector quantization.

To reduce the average bit rate required to describe the selection patterns, Huffman encoding is used. For lossless coding, it is necessary to provide codewords for all possible selection patterns within each of the four regions. However, since the probability of many of the selection patterns is very low, a special fixed-length codeword limits the maximum length of the codeword describing any selection pattern. This special codeword consists of a short escape sequence followed by a 16 bit word identifying the subset of coefficients selected for transmission. In addition, the codewords for the first three regions indicate if all

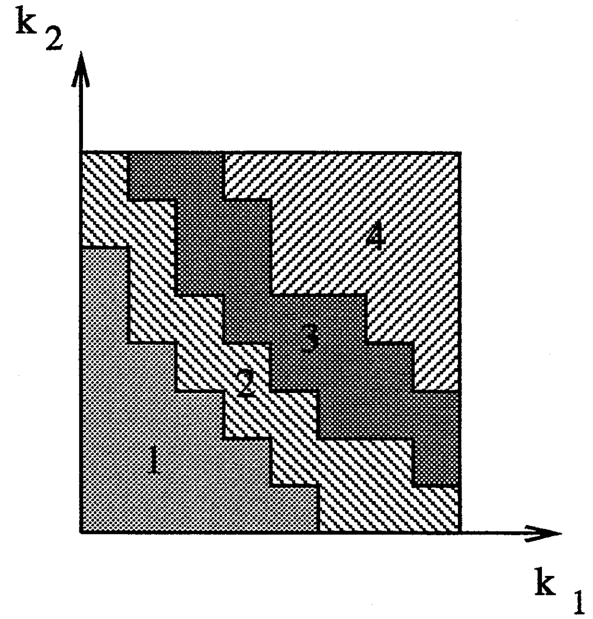

Figure 3.6: The four regions for the vector coding scheme.

of the coefficients in the following regions are zero. If this is the case, then the codewords for those regions are not transmitted.

To efficiently encode the selection information, we are also examining other approaches. Since the encoded methods we consider are lossless, the primary comparison criteria are the required bit rate and hardware complexity. The CCDC system has been designed with enough flexibility to incorporate some of these possible future improvements.

The amplitudes of the selected coefficients are also statistically encoded. The codebook is adaptive based upon the particular coefficient to be encoded as well as its inter/intra and luminance/chrominance characteristics. Specifically, each set of 8×8 DCT coefficients is arranged into four groups for the codebook design. The DC coefficient has one codebook for itself, while the remaining 63 coefficients are organized into 3 other codebooks. There are also separate codebooks for inter and intra, and Y and UV. This gives a total of 16 (2 (inter/inta) $\times$ 2 (Y/UV) $\times$ 4 (separated DCT coefficients)) codebooks.

3.1.10 Buffer Control

The adaptive allocation of bits to DCT coefficients results in a locally variable bit rate system, which must be coupled to a constant bit rate channel. Our solution is a buffer between the video encoder and the channel, and a buffer control mechanism that is coupled to the quantization of the DCT coefficients. As the buffer begins to fill, the DCT weighting factors are increased, thereby quantizing all the DCT coefficients more coarsely and reducing the bit rate. This also prevents the possibility of buffer overflow. As the buffer begins to empty, the DCT weighting factors are decreased, thereby quantizing all the DCT coefficients more finely, increasing the bit rate and filling up the buffer. The buffer control mechanism is designed to ensure stability and to ensure that the buffer can never overflow or underflow.

3.2 Video Decoding

The video decoding process at the receiver is the inverse of the encoding process described in the previous section. This is shown in Figure 3.7. Using the selection information the selected DCT coefficients are identified, reconstructed, and appropriately inverse weighted. An inverse block DCT operation is performed to obtain the residual signal. The residual signal is combined in a spatially adaptive manner with the previously reconstructed frame to reconstruct the current frame. Every step of the video processing algorithm has been designed so that the decoder would be kept as simple and inexpensive as possible. For example, the computationally expensive motion estimation, which is part of the encoder, is absent in the decoder.

Through the error control built into the system, which will be discussed in section 6.3, the receiver will be able to detect and correct transmission errors that may result from unfavorable channel conditions. When the channel is very poor, uncorrected errors may reach the video decoder. Under these circumstances, the CCDC system uses a number of error concealment schemes.

Another important characteristic of a receiver is the resolution of the displayed video.

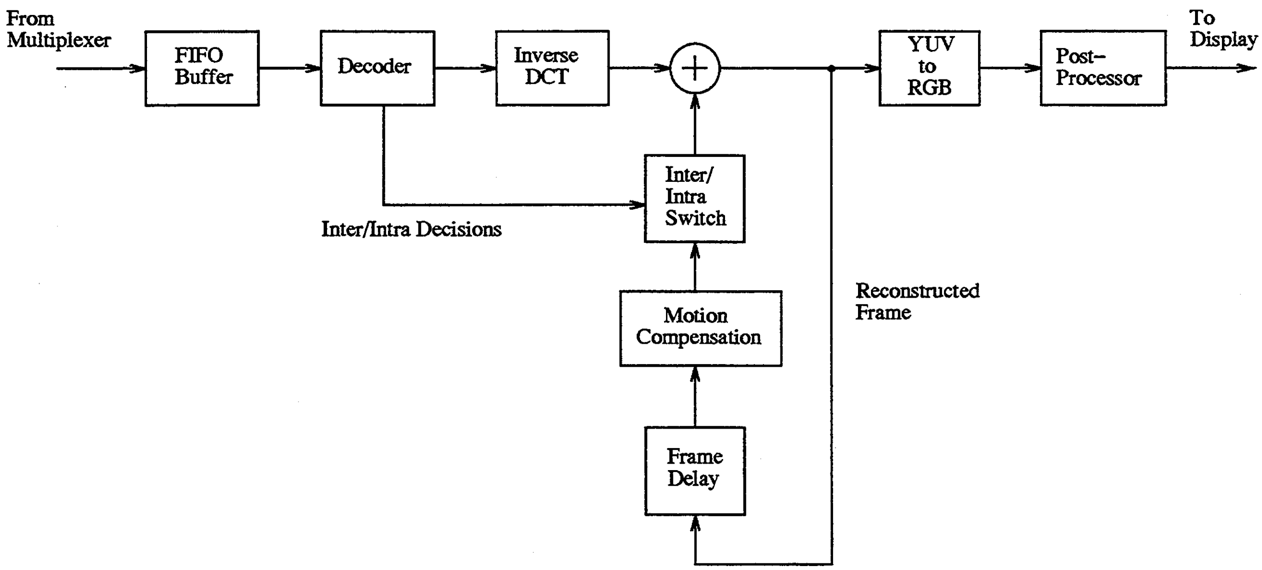

Figure 3.7: Decoder for the video compression.

With relatively simple processing the resolution can be either increased or decreased. The receiver is also free to upsample temporally, using either sample-and-hold or motion compensated interpolation. Increased resolution may be useful for large screen viewing. Lower resolution may be useful for smaller, hand-held receivers. As the display will probably be the largest cost element of any receiver, different display sizes will introduce different price/complexity/performance classes for receivers.

3.3 Video Specifications

The specifications for the video processing of the CCDC HDTV system are shown in Table 3.1.

Video Signal	Scanning	Progressive
	Aspect Ratio	16:9
	Frame Rate	59.94 Hz
	Bandwidth	
	Luminance	34.0 MHz
	Chrominance	17.0 MHz
	Active Pixels	
	Luminance	1280×720
	Chrominance	640×360
	Sampling Frequency	75.5 MHz
	Colorimetry	SMPTE 240M (Approximate)
Video Compression	Source Adaptivity	59.94 fps, 30 fps, 24 fps
	Algorithm	MC-DCT
	Motion Compensation	Causal
	Motion Vectors	
	Search Range	+15/-16 (H) +7/-8 (V)
	Tracking Range	.75 frame width/sec .67 frame height/sec
	Accuracy	$\frac{1}{2}$ pixel
	Encoding	Differential and Huffman Coding
	Motion Block Size	
	Luminance	16×16 and 8×8
	Chrominance	8×8
	Spatial Transform	DCT
	Block Size	8×8
	Encoding	Perceptual Processing Adaptive Vector Coding Uniform Quantization Huffman Coding
	Superblock Processing	Adaptive Inter/Intra

Table 3.1: Specifications for the video processing portion of the CCDC HDTV system.

3.4 Video Processing Summary

The video compression portion of the CCDC HDTV system contains a number of specially designed features that will ensure the production of very high quality video at the home receiver. These features include:

- Progressive scanning of the video
- Source adaptive processing
- Motion estimation/motion compensation with adaptive block size and half pixel resolution
- DCT based transform encoding
- Spatially adaptive inter/intra-mode encoding
- Compression designed to be transparent to the human visual system
- Detection of scene changes and appropriate processing of them
- Novel encoding of the selection information
- Fast channel acquisition
- Resistance to channel errors
- Simple scalability and extensibility
- Low complexity and low cost receiver
- Simple VCR functionality
- Future extensibility including general source adaptivity and varying luminance/chrominance resolution

4 Audio Processing

The CCDC HDTV system uses the MIT Audio Coder (MIT-AC) system for audio compression. MIT-AC can maintain, at a significantly reduced bit rate, a sound quality comparable to that of the Compact Disc. The bit rate reduction allows the CCDC system to accommodate a number of CD-quality audio channels. Since the number of bits used for audio is flexible, the bits that remain after the audio is encoded are used for video.

In the CCDC system, four channels of digital audio are provided. Even though the four channels are likely to be used as two stereo pairs, the coding algorithm treats each channel independently. This minimizes the possibility of stereo artifacts and allows the four channels to be used separately.

The input to and output from MIT-AC is digital audio sampled at 48 kHz in the AES/EBU standard format, which ensures a clean digital interface to both the studio and the user. The basic method used is an adaptive transform coder. Extensive literature exists on the general concepts behind adaptive transform coding of audio. The encoding part consists of transform analysis, spectral envelope encoding, and transform coefficient encoding. Audio compression is achieved by exploiting properties of the human auditory system. Decoding consists of error concealment and transform synthesis.

4.1 Audio Encoding

A block diagram of the audio encoder is shown in Figure 4.1. The digital audio signal is converted from time domain to frequency domain by a critically sampled single-sideband filterbank which is implemented with a fast transform algorithm. In this method, the data is first segmented by overlapping windows. The choice of window involves a time-frequency tradeoff—a longer window gives better frequency resolution, but poorer time resolution. In MIT-AC, a raised-cosine window with a duration of approximately 20 ms is used. For each window, the temporal samples are transformed into a frame of spectral coefficients. Since each window is overlapped by 50% with the adjacent windows, the number of coefficients must be undersampled by a factor of two to maintain critical sampling. Although aliasing is introduced by undersampling, the original signal can still be reconstructed exactly at the decoder with an overlap-add process. This representation is also called "time domain aliasing cancellation" (TDAC) in the literature.

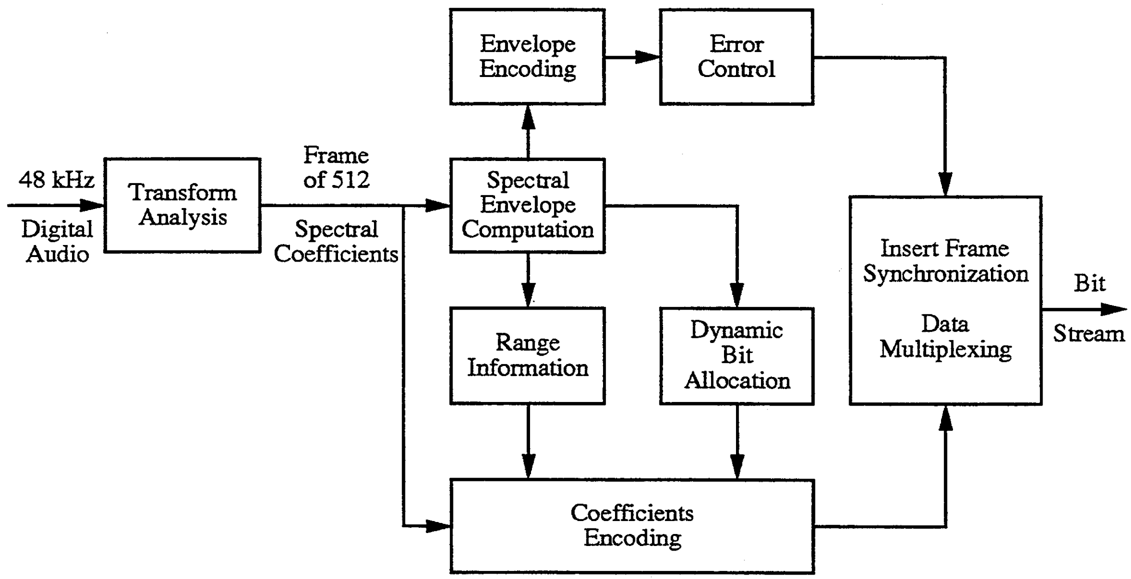

Figure 4.1: Audio encoder block diagram.

In MIT-AC, the same transform is used for every windowed data segment. This avoids the synchronization issue that occurs in "evenly-stacked" schemes where different transforms are used for alternate frames.

The transform coefficients that result from each frame are grouped into subbands of varying lengths. These lengths are chosen to match the critical bands of the human auditory system. Within each subband, a reference coefficient is selected and encoded accurately with a fixed bit allocation scheme. The encoded bits form the spectral envelope information for the frame. Since this information is used to dynamically allocate bits for encoding transform coefficients, it is protected with error control coding (in addition to the error correction applied to all data, audio and video, as discussed in Section 6.3). Approximately 20% of the total number of bits available for audio is used to encode the spectral envelope information.

After encoding the spectral envelope, the remaining bits are dynamically allocated to the rest of the coefficients. The spectral envelope provides a fair approximation to the energy within each subband and can therefore serve, along with an interband masking model, as an indicator of the relative perceptual importance of each subband within the frame. This is exploited in the dynamic bit allocation step.

The coefficients are then quantized within the range specified by the spectral envelope using the allocated bits for the subband. Approximately 80% of the total number of bits available for audio is used to encode these coefficients.

4.2 Audio Decoding

A block diagram of the audio decoder is shown in Figure 4.2. The audio decoding process at the receiver is the inverse of the encoding process described in the previous section. The spectral envelope information is first decoded. Because of the additional error protection, the spectral envelope information is very robust. This ensures that the dynamic bit allocation, which is based on the spectral envelope information, is performed accurately. In addition, the spectral envelope provides the range information for the coefficients within each subband. It is therefore required for decoding the coefficient amplitudes.

Through the error correction built into the system, which will be discussed in Section 6.3, the receiver will be able to detect and correct transmission errors that may result from unfavorable channel conditions. When the channel is very poor, there may be errors detected but not corrected. In these circumstances, error concealment is necessary. In MIT-AC, error concealment is performed by repeating previous subbands of spectral coefficients. This method is satisfactory for maintaining undegraded audio up to a bit error rate of about 10^{-4} and maintaining usable audio up to a bit error rate of about 10^{-2}.

In the present implementation, the capacity used for one channel of audio is 125 kb/s. At this rate, audio transparency is maintained for nearly all of the difficult audio material tested.

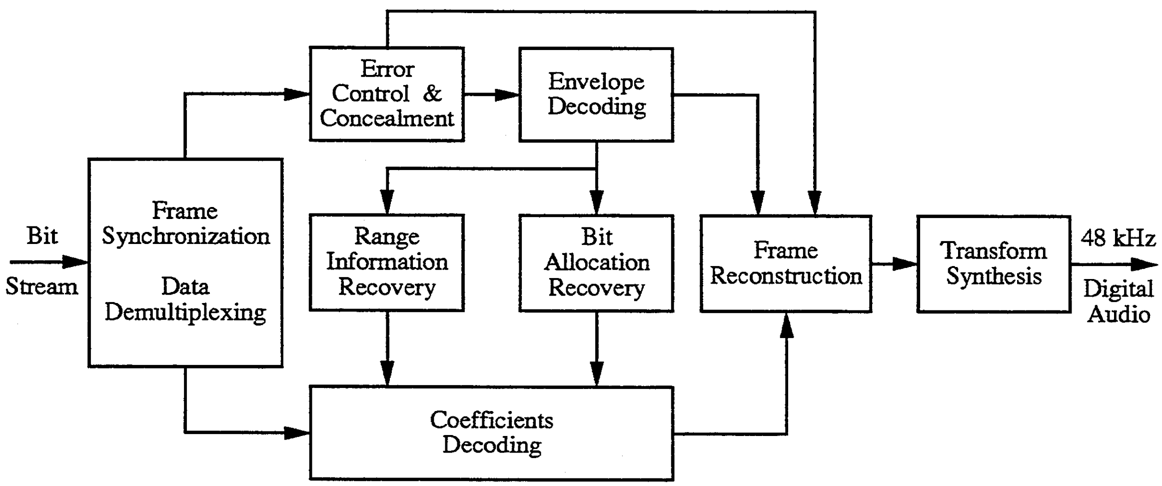

Figure 4.2: Audio decoder block diagram.

5 Auxiliary Data and Data Multiplexing

The CCDC HDTV system is designed to adaptively allocate data capacity to support auxiliary services, which may include closed captioning for the hearing impaired, teletext, program guides, conditional access information (authorization and decryption keys), and other services that may be developed in the future. A total of 0.25 Mb/s is allocated for these services. Additionally, a portion of 755 Kb/s allocated for audio can be used for these services. Any unused data capacity will be automatically allocated back to the video processor to optimize the video encoding performance.

The video, audio, auxiliary data, and control data are multiplexed together into a single bit stream for transmission. A high-level view of the data format is shown in Figure 5.1. The diagram shows the data transmitted in one CCDC HDTV frame time, which will be called a frame of data. This frame of data should not be confused with the data necessary to represent a frame of video. The bits to be transmitted are organized in this manner solely to simplify the timing. One frame of data can represent, for example, more or less than one frame of video depending on such factors as source frame rate, difficulty of scene and buffer fullness.

There are 525 data lines within each frame of data. Each line time in this frame of data is equivalent to one NTSC line time or three HDTV line times. Each line contains control bits, auxiliary data, audio and video, and check bits. The first data line of each frame also contains a 24 bit sync word. The numbers in the figure correspond to 8-bit bytes for 16-QAM and 32-QAM modes. The first three bytes of control bits are used to send an additional 24 bit sync to help the acquisition of deinterleaver synchronization.

As the video data is entropy coded, the number of bits per frame is variable. The video data for a given frame may begin anywhere within the frame of data. Likewise, the video data for a given frame may end anywhere within the same data frame or another data frame. The transmitted video is arranged in macroblocks, where a macroblock is 20 consecutive superblocks. The arrangement of the video in macroblocks simplifies the communication

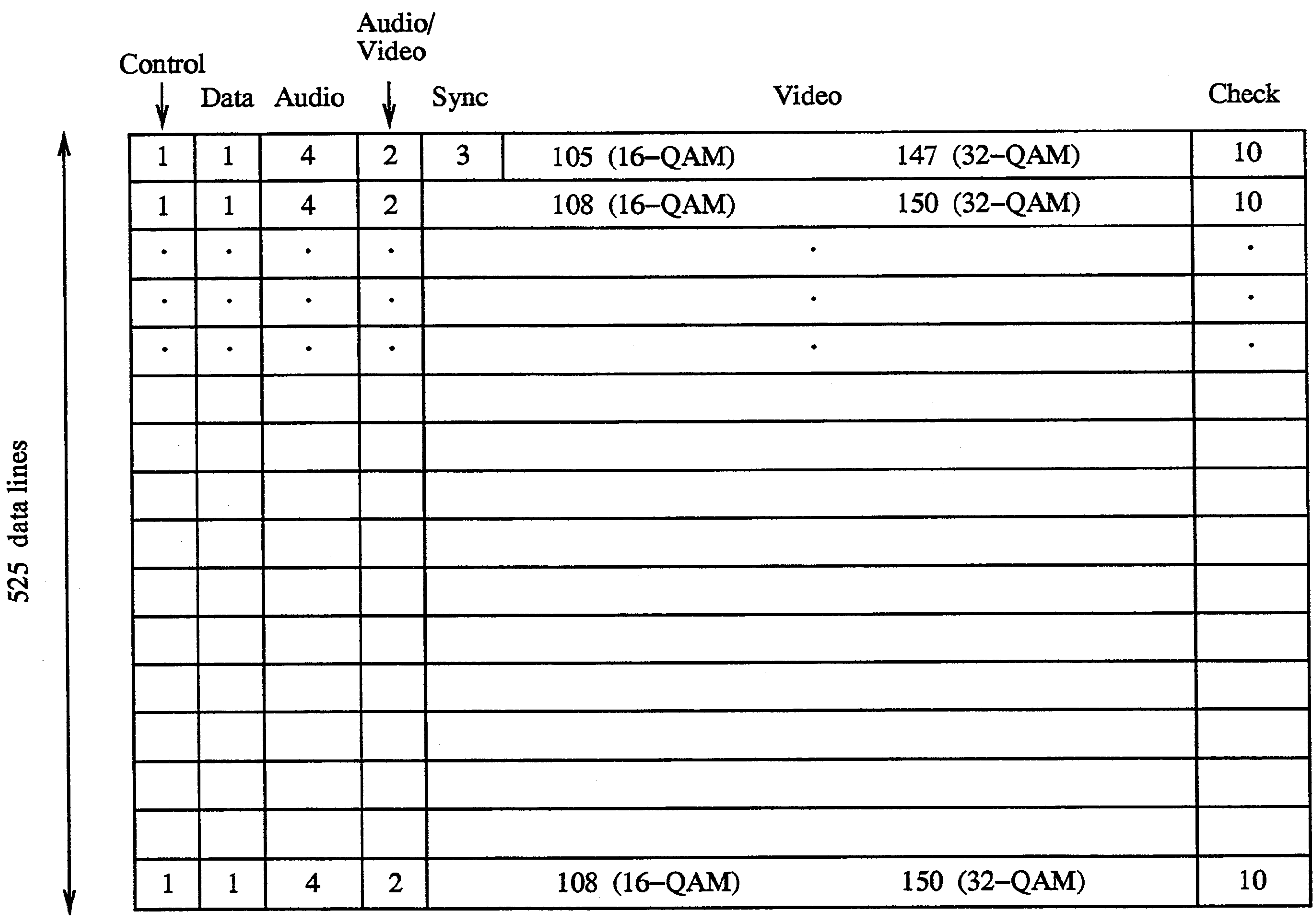

Figure 5.1: A high-level view of the data format. Numbers given are in 8-bit bytes and correspond to 16-QAM and 32-QAM modes.

between the transmitter and receiver, and it allows easy synchronization during the initial acquisition of the video data, or resynchronization after an uncorrected channel error. Synchronization of the video data is very important. As video data is entropy coded, and therefore of variable length, one uncorrected bit error may destroy a large amount of video data. The macroblock is the smallest unit of independent data. Each line of video data contains a pointer to the next macroblock, so the largest amount of variable length data that can be lost by a bit error is limited to one macroblock.

6 Transmission System

The CCDC HDTV transmission system is designed to deliver a high rate digital stream with high reliability, even under severe channel conditions. This result is achieved by using the most modern modulation and coding techniques available, several levels of interleaving and fast adaptive equalization.

The transmission system can operate in one of two modes, coded 32-QAM or 16-QAM. The CCDC system uses a single, selectable, transmission mode. The selection is based on system threshold and coverage considerations. The 32-QAM mode is the preferred mode of operation and will be used in most cases. There are many factors that influenced the choice of a uniform single carrier modulation scheme. One important factor is modularity. In the CCDC system, source and channel coding are clearly separated and largely independent of one another. This has the advantage that the video and audio processing described in sections 3 and 4 will not require substantial changes when using other media for transmission. Another advantage stems from the reduced hardware complexity as compared to systems combining source and channel coding.

One common objection to single rate systems is that the system performance does not degrade gracefully. We have determined through simulations, however, that bi-rate systems must compromise performance in the primary area of coverage in order to provide a marginal extension of the fringe area. Furthermore, in those outer areas, the video signal received is no longer of HDTV quality, and is likely to be of less than NTSC quality, while the NTSC simulast service is available. The simulation results pertaining to the coverage issues are detailed in section 8.

Another consideration in the design of a transmission system relates to NTSC cochannel interference. We have chosen not to apply any form of precoding or formatting to the signal at the transmitter end. Such methods imply either an increase of the carrier-to-noise (C/N) threshold or a reduction of the data rate. Furthermore, such inefficiency continues after the HDTV service becomes widely available and the NTSC service discontinues. It is our belief that the same goal of increased NTSC cochannel interference immunity can be achieved in a standard independent manner by a clever design of the receiver system. Although these schemes are not implemented in the CCDC system, they will be proposed for consideration at a later stage.

To summarize, the main features of the transmission system are:

- Efficient use of the bandwidth: 32-QAM/16-QAM modulation
- Low noise threshold: concatenated trellis/RS coding
- Fast and reliable acquisition in the presence of noise and interference
- Robustness against burst noise and interference: multilayer interleaving
- Hierarchical error detection, correction and concealment.
- Powerful adaptive equalizer to combat multipath and other channel distortions.

6.1 Transmission System Overview

A block diagram of the transmission system is shown in Fig. 6.1. The multiplexed data, which comprises digital video information, four digital audio channels, and auxiliary and control data is first converted into two coded digital waveforms by block and trellis encoding and digital filtering. The digital streams are then transformed into analog waveforms, quadrature modulated at an IF frequency of 44 MHz and further filtered with a SAW filter to remove remaining spurious signals before being translated to the assigned frequency in the VHF or UHF band. The receiver uses a double conversion tuner to recover the analog waveform at a center frequency of 43.5 MHz. The waveform is then filtered with a SAW filter at IF, demodulated, converted to a digital stream and decoded by trellis and Reed-Solomon decoders.

The next four sections describe the functional blocks that have similar counterparts in both the transmitter and the receiver and are better understood in pairs. These are the modulation/demodulation, error control, interleaving and spectral shaping. Finally, sections 6.6 and 6.7 describe the features that are specific to the transmitter and receiver.

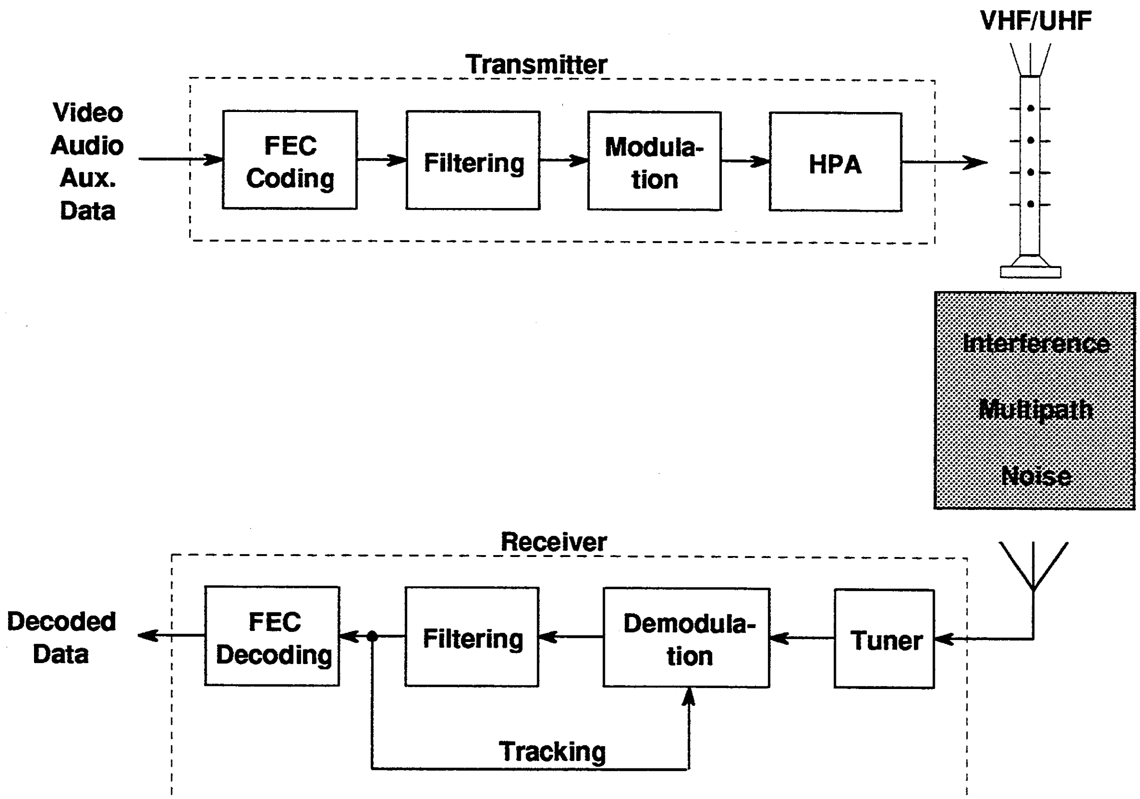

Figure 6.1: CCDC HDTV transmission block diagram.

6.2 Modulation

The CCDC HDTV system uses quadrature amplitude modulation (QAM) to achieve high spectral and power efficiency. The transmission system can operate in one of two modes, 32-QAM or 16-QAM, at a signaling rate of 5.287 MHz. The primary modulation scheme is 32-QAM and it conveys 5 bits per signaling interval. The 16-QAM mode is provided for those cases where a lower noise threshold is desired. Automatic mode detection is performed by the receiver. The specific mapping between bits and symbols is determined by the trellis code described in subsection 6.3.1. The signal constellations for 32-QAM and 16-QAM modulations are shown in Fig. 6.2. The peak to average ratio for the 32-QAM cross constellation is slightly lower than that of the 16-QAM constellation (by approximately 0.1 dB).

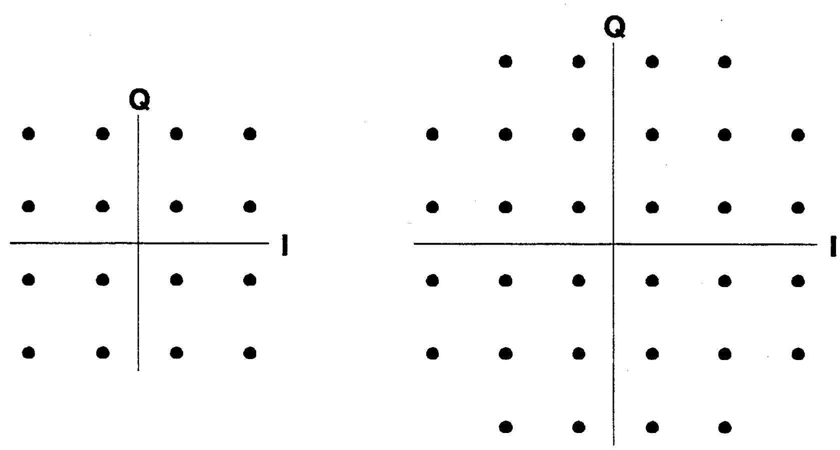

Figure 6.2: Signal constellations for 16-QAM and 32-QAM.

6.3 Error Control

The CCDC HDTV system uses concatenated trellis coding and Reed-Solomon coding to reduce the effect of channel errors and achieve the desired error performance. The trellis inner code interfaces with the modulator/demodulator and is configured to correct most of the channel errors. The Reed-Solomon outer code then reduces the probability of error to the desired level. The choice of a concatenated code is motivated by the desire to achieve a given probability of error at a specified signal-to-noise ratio (SNR) with an overall implementation complexity that is lower than that of a single code.

6.3.1 Trellis Coding

The gap between uncoded QAM and the Shannon channel capacity is approximately 9 dB for a probability of symbol error of 10^{-6}. Trellis coding is a powerful technique for reducing this gap without bandwidth expansion. This is accomplished by coding onto an expanded signal constellation so that the "free distance" (minimum Euclidean distance) between coded signal sequences is maximized. It consists of a convolutional encoder followed by a special signal mapping. The trellis decoder computes the most likely signal sequence using the Viterbi algorithm with soft-decision decoding. Nominal coding gains as high as 6 dB can be achieved using optimized trellis codes. The choice of the current trellis code was based on hardware availability and may not represent the best cost-performance tradeoff for volume production.

6.3.2 Reed-Solomon Coding

Further error control is achieved by block coding. Trellis coded modulation (TCM) in itself is very effective for combatting random and short impulsive noise but tends to generate burst errors in cases of strong interference and burst noise in general. Reed-Solomon coding is ideal for correcting this type of error.

For 32-QAM TCM, each group of 158 information bytes are encoded using a t=5 error correcting Reed-Solomon code. For 16-QAM TCM, groups of 116 information bytes are encoded with the same 5 byte error correcting Reed-Solomon code used for the 32-QAM TCM.

Fig. 6.3 shows the measured performance of the CCDC HDTV transmission system.

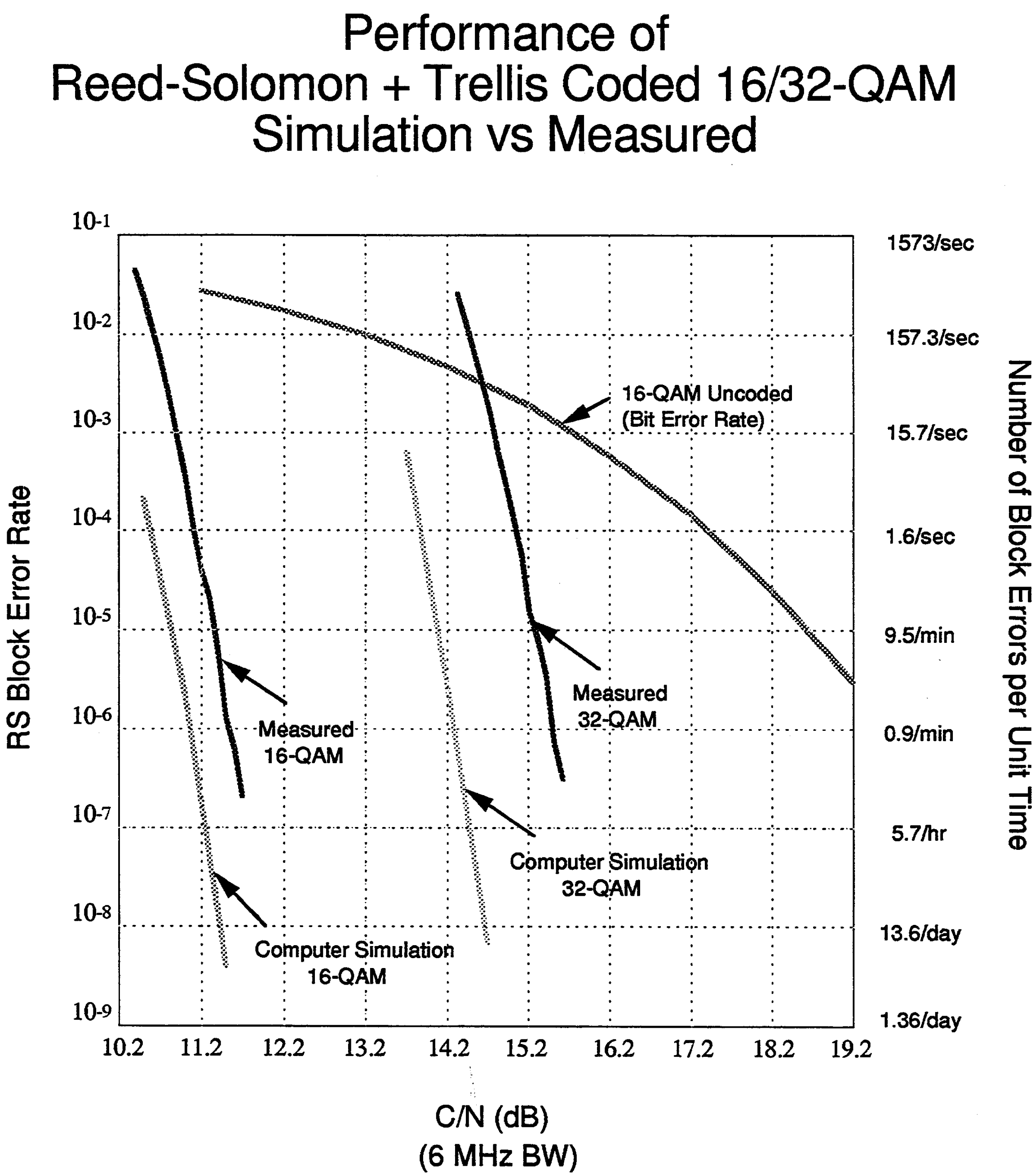

Figure 6.3: Simulated and measured performance of the CCDC HDTV system.

6.4 Interleaving

Two levels of interleaving are provided in the CCDC HDTV system. The first level (interleaver #2) is used to protect against short impulsive noise. The inner interleaver provides effective protection against 3 μs bursts. The second level of interleaving (interleaver #1) is used to gain protection against burst errors generated by the trellis decoder. The interleavers and deinterleavers are shown in Fig. 6.5 and 6.6 respectively.

6.5 Spectral Shaping

Pulse shaping of the QAM signals is performed both at the transmitter and the receiver in order to ensure zero intersymbol interference (ISI) for the ideal (flat) channel frequency response while limiting the length of the impulse response. The cascade of the transmitter and receiver pulse shaping filters results in a close approximation to a raised-cosine filter with 10% excess bandwidth. The choice of this roll-off factor represents a good compromise between bandwidth efficiency and peak power. The spectral shaping is divided equally between transmitter and receiver, each implementing a square-root raised-cosine filter. Correct shaping at band edges is attained by filtering each of the I and Q channels with a 64 tap FIR filter using 2 times oversampling. After D/A conversion, the two waveforms are combined and passed through a SAW filter. The function of the SAW filter is to remove spurious signals and strongly attenuate the sidelobes that may have resulted from the digital filtering. Fig. 6.4 shows the IF spectrum of the signal at the output of the SAW filter.

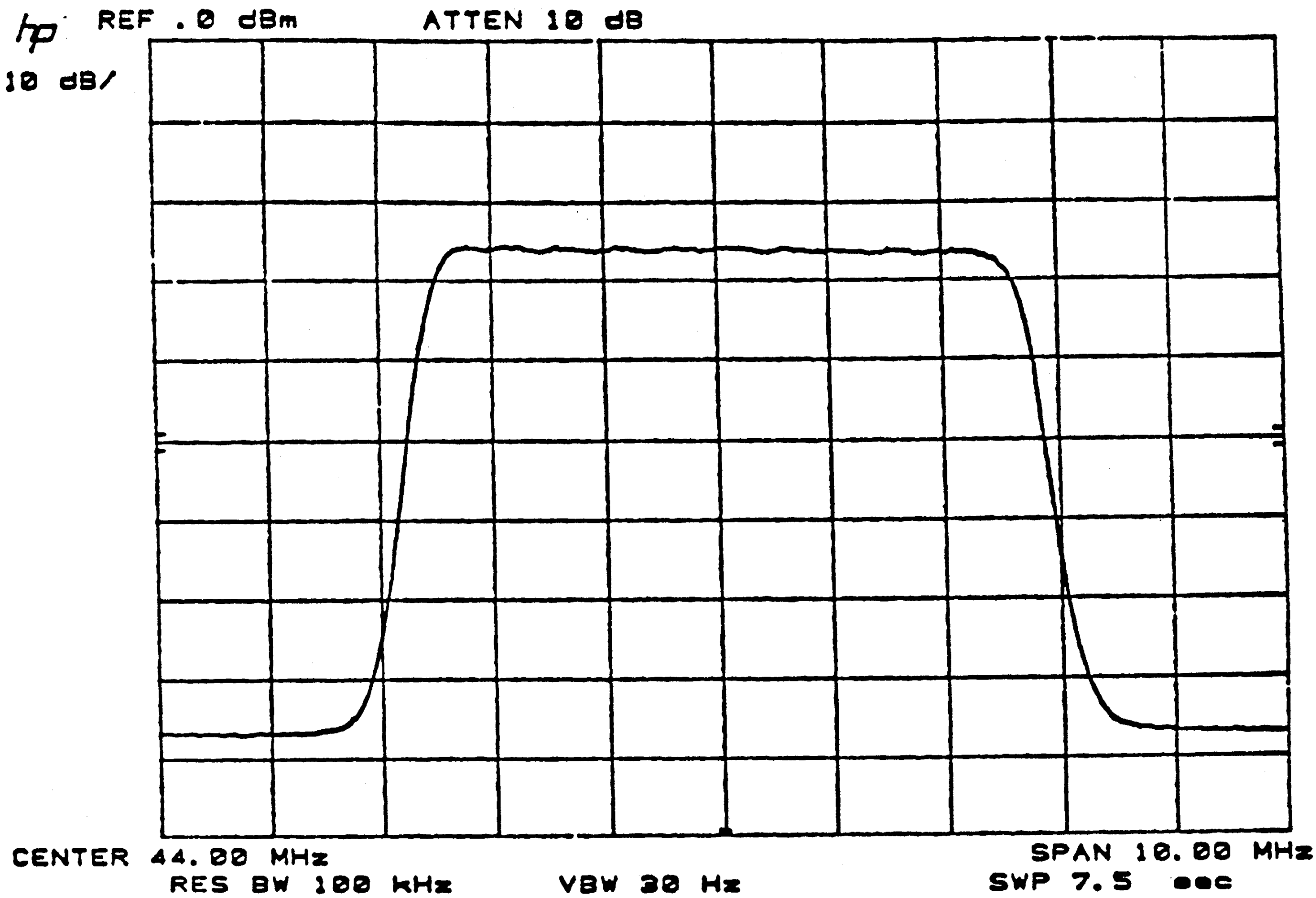

Figure 6.4: IF Output Spectrum.

6.6 Transmitter

Fig. 6.5 shows the functional block diagram of the CCDC HDTV transmitter. The input to the transmitter is a multiplexed stream comprising video, audio and ancillary data. The data is formatted in blocks and encoded by the RS encoder, as described in section 6.3.2. Next, an interleaver is used to reduce the effect of long error bursts that may be generated in the decoder. After proper formatting, the data is trellis encoded and interleaved to mitigate the effect of short error bursts, as described in sections 6.3.1 and 6.4. The modulator performs the spectral shaping, D/A conversion and frequency translation required for over-the-air transmission.

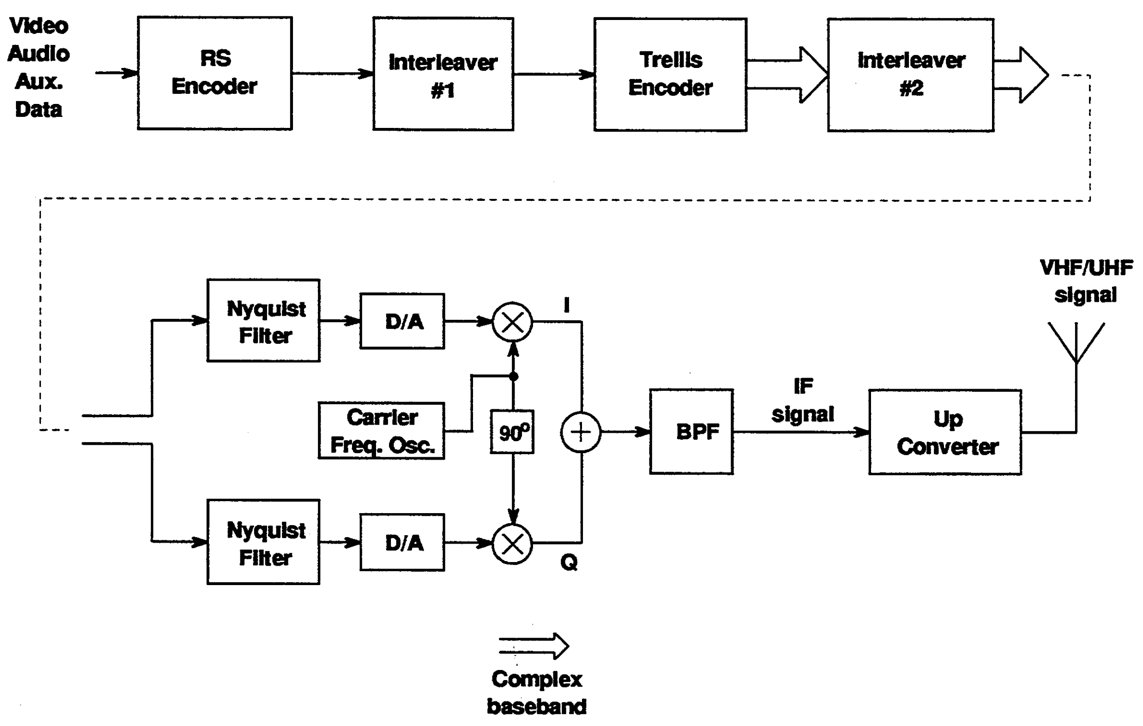

Figure 6.5: Transmitter functional block diagram.

6.7 Receiver

Fig. 6.6 shows the functional block diagram of the CCDC HDTV receiver. The blocks pertaining to modulation, error control and spectral shaping have corresponding functions to those of the transmitter and have already been explained. The blocks specific to the receiver and which have not been introduced before include the tuner and equalizer circuits. These are detailed in the following sections.

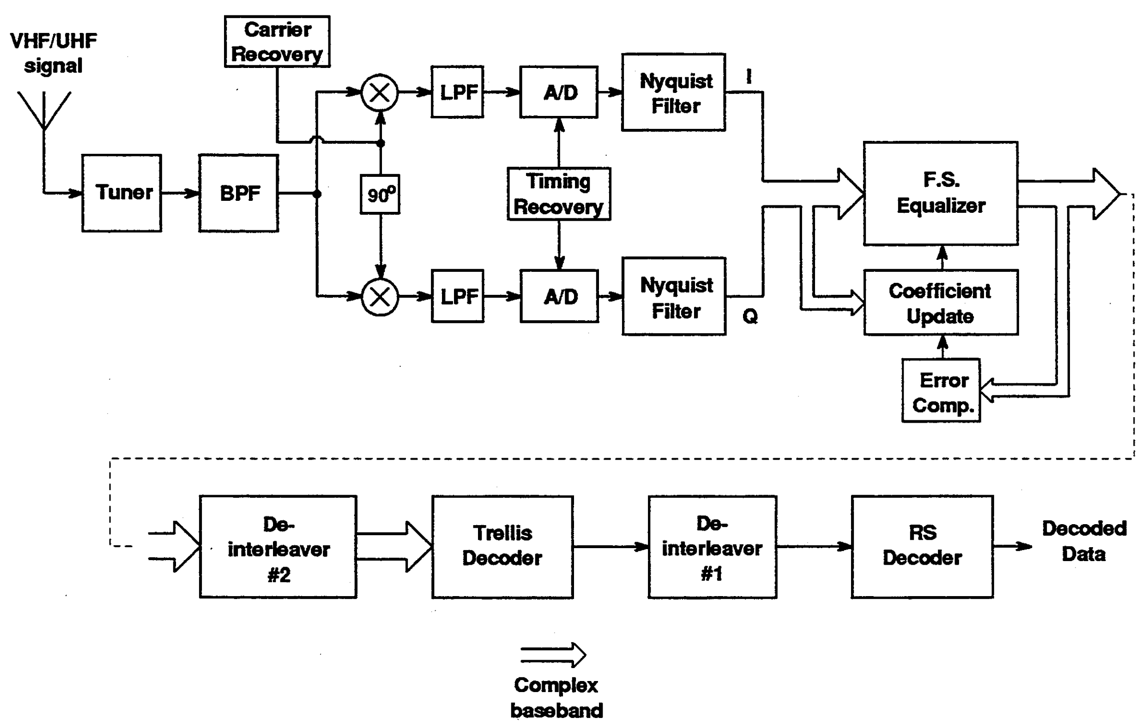

Figure 6.6: Receiver functional block diagram.

6.7.1 Tuner

The tuner used by the CCDC HDTV system is of the double conversion type. It has improved characteristics compared to a conventional single-conversion tuner in the area of spurious response rejection capabilities.

Fig. 6.7 shows a block diagram of the tuner. The VHF/UHF signal is filtered through switchable bandpass filters and passed through a wideband amplifier before being mixed with a local oscillator (LO) reference signal. The first local oscillator, whose frequency ranges from 1257 MHz to 2087 MHz, is used to translate the selected channel to a first IF frequency of 1200 MHz. The first IF frequency has been chosen high enough for spurious free reception of all VHF and UHF frequencies, yet low enough for low cost implementation of the tuner. After bandpass filtering at 1200 MHz and amplification, the signal is translated to a second IF frequency of 43.5 MHz using a second LO at 1156.5 MHz. The selection of the 43.5 MHz IF as compared to 44 MHz is due to the availability of off-the-shelf SAW filters. The operating range of the VHF/UHF input to the tuner is -70 to -5 dBm.

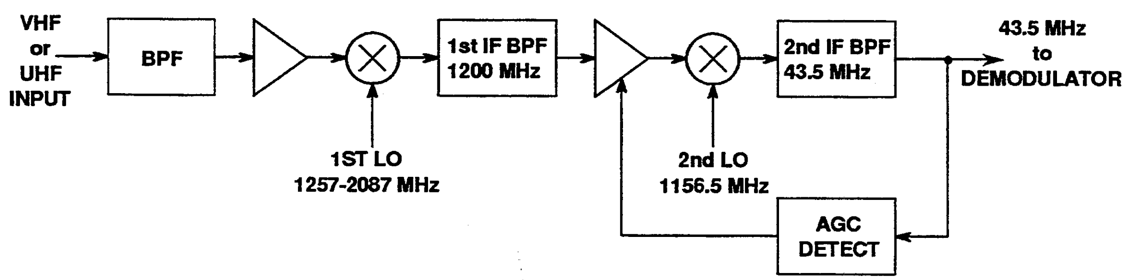

Figure 6.7: CCDC HDTV Tuner block diagram.

6.7.2 Equalization

Adaptive equalization is used in the CCDC HDTV system to correct for linear channel distortions such as multipath, non-ideal frequency response and group delay. Fig. 6.8 shows a block diagram of the adaptive equalizer. Each filter consists of 256 taps fractionally spaced, yielding an effective range of -2 μs to +24 μs. The equalizer can handle single or multiple echoes within that range, with levels as high as -6 dB for close-in echoes (-2 to +4 μs) and -12 dB for long echoes (+4 to +24 μs).

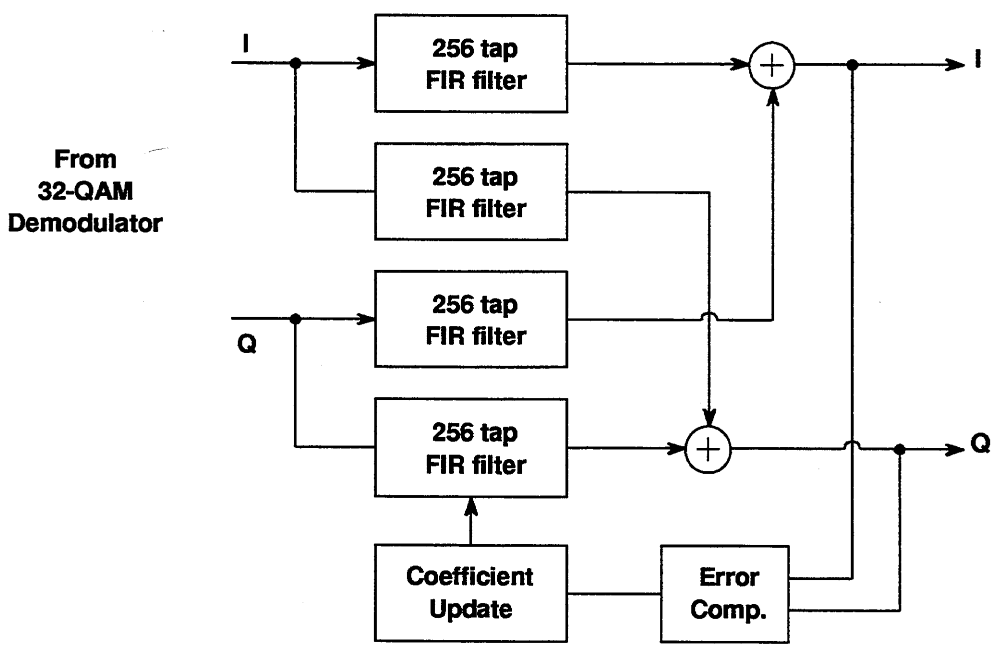

Figure 6.8: CCDC HDTV Adaptive Equalizer.

7 Synchronization

The CCDC HDTV system has been designed to provide fast, reliable acquisition in the presence of noise, multipath, and interference. Through the use of advanced digital processing, the acquisition and synchronization time has been minimized.

7.1 Clock Synchronization

Figure 7.1 shows the block diagram of the clock/sync generator used in the encoder and the decoder. A voltage controlled crystal oscillator (VCXO) generates a master clock at 528.6713 MHz. All the clocks used for the digital video processing and the 32-QAM modem are derived from the master clock. The encoder synchronizes the master clock to the 47.2 kHz horizontal sync provided externally. The decoder synchronizes the master clock to the 32-QAM symbol rate. The phase-locked loops have been designed to minimize the clock jitter (a few nsec).

The encoder also makes use of vertical sync provided externally to synchronize the frame vertical sync. The decoder acquires the internal vertical sync by using the 24-bit sync generated by the encoder.

7.2 Acquisition

When the HDTV receiver is tuned to a new channel, the decoder goes through a number of synchronization processes. The CCDC HDTV system takes approximately 0.4 seconds for the demodulator and overall synchronization and 0.33 seconds for complete refreshing of the decoded video. The total acquisition time of the CCDC HDTV system, therefore, is 0.73 seconds. However, subscribers will be able to start to see portions of the new channel and hear the sounds in 0.4 seconds. Table 7.1 lists the breakdown of the synchronization process.

32-QAM Demodulator and Overall Synchronization	
AGC	0.05 sec
Bit Sync	0.1 sec
Adaptive Equalizer	0.1 sec
Carrier Sync	0.1 sec
Interleaver	0.02 sec
24-bit Sync	0.03 sec
Total	0.40 sec
Digital Video Decoder	
DPCM Refreshing	0.33 sec
Total Acquisition Time	
Total Acquisition Time	**0.73 sec**

Table 7.1: Signal acquisition time.

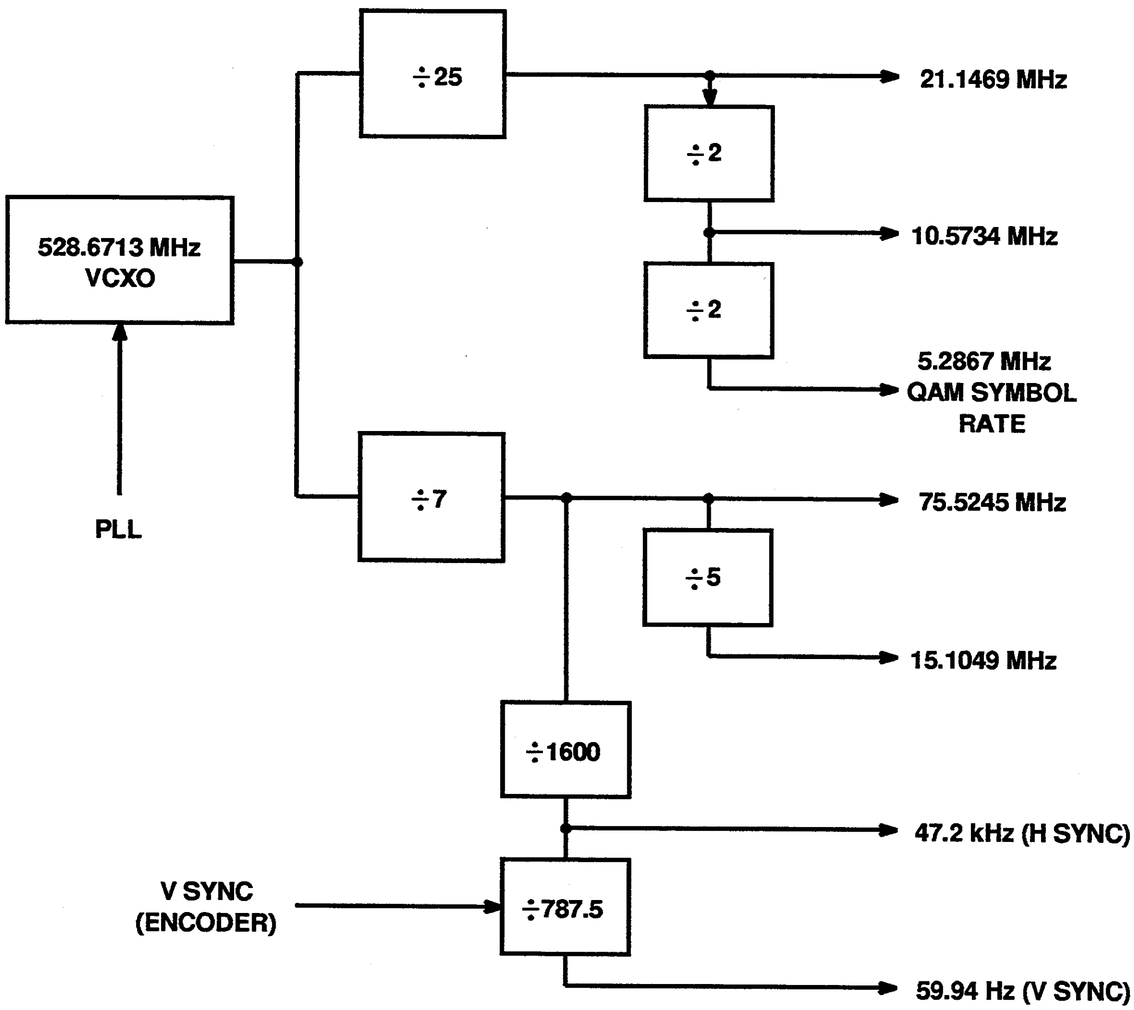

Figure 7.1: Clock/Sync Generator.

8 Coverage Area Analysis

The coverage area analysis provided in this section is based on a set of assumptions as well as the measured performance of the CCDC HDTV transmission system. This analysis shows that the CCDC system provides the same service area as current NTSC systems, but with much less power and closer cochannel spacing.

8.1 Coverage Area Criterion

The following assumptions are based on work to date by the ATS Advisory Committee:

1. The ATV service area should be comparable to the current NTSC service area.
2. The interference to an existing NTSC service by an ATV signal should be no worse than the current interference into an NTSC signal by another NTSC signal.
3. In order to meet simulcast channel allocation goals, cochannel transmitters may have to be separated by as little as 100 miles.

In light of the above assumptions, the chosen coverage area criterion is the service area which results when two UHF NTSC transmitters of maximum allowable power (37 dBk) are separated by the minimum allowable cochannel distance (155 miles).

8.2 Coverage Area Assumptions

8.2.1 NTSC Assumptions

Table 8.1 shows the assumptions used in the coverage calculations involving NTSC transmission/reception. The transmit antenna height of 1200 feet is in agreement with PS/WP-3. The Effective Radiated Power (ERP), receiver antenna gain, antenna front-to-back ratio, downlead loss, and noise figure are the FCC planning factors assumed for NTSC receivers for UHF Grade B service.

The NTSC Grade B service boundary is the locus of points for which the carrier-to-noise ratio (C/N) is 28.5 dB (peak sync to RMS noise in a 6 MHz bandwidth). This level of service is based on an F(50,90) field, i.e., a field strength available at at least 50% of locations at least 90% of the time, for a receiving antenna height of 30 feet above ground level. F(50,90) field strength data are derived from F(50,50) data provided by the FCC, by assuming that the field strength is log-normal distributed.

Cochannel NTSC interference is measured in terms of a carrier-to-interference (C/I) ratio. For the case of NTSC carriers employing nominal frequency offset, a 28 dB Desired/Undesired (D/U) ratio defines the acceptable limit of interference.

The nature of digital ATV interference into NTSC is markedly different from that of NTSC interference into NTSC. The impairment of a digital ATV interferer in an NTSC picture is subjectively similar to that caused by white noise. Thus cochannel ATV interference into NTSC will merely appear to add to the thermal noise level at the reception site, yielding an effective C/N value lower than that due to thermal noise alone. In order to be consistent with current coverage analysis practice, in which plots are made for fixed D/U ratios, a fixed C/I value is assumed that, for practical purposes, results in nearly constant effective C/N criteria for the interference boundary. This is the reason for the value of 30 dB C/I shown in Table 8.1.

The field strength ratio of F(50,50) to F(50,10), corresponding to a given D/U (C/I) ratio, is assumed throughout the calculations.

NTSC ASSUMPTIONS	
Tx Antenna Height Above Avg. Terrain (ft)	1200
Transmitter ERP (dBk)	37
Receiver Antenna Gain (dB)	13
Downlead Loss (dB)	5
Dipole factor (dB/meter) (75 ohms, 609 MHz)	-22
Thermal Noise at room temp (dBuv) (75 ohms, 6 MHz)	2.6
Receiver Noise Figure (dB)	10
90% Time Avail. Factor at 56 mi (dB)	8.7
C/N for Grade B Contour (dB) (6 MHz BW)	28.5
Antenna Front-To-Back ratio (dB)	6.0
NTSC into NTSC Interference D/U (dB)	28
ATV into NTSC C/I Threshold (dB) (peak sync-to-avg)	30

Table 8.1: NTSC Assumptions

8.2.2 ATV Assumptions

Table 8.2 shows the assumptions used in the coverage calculations involving ATV transmission/reception. The transmit antenna height of 1200 feet is the same used in NTSC calculations. The ERP of 23.9 dBk is derived from the coverage calculations under the aforementioned criterion. The receiver antenna gain, antenna front-to-back ratio, downlead loss, and noise figure are believed realistic for ATV receiving systems.

As in the case for NTSC Grade B service, the perimeter for ATV service is defined using F(50,90) field strength data. Based on measurements of the CCDC system, 11.7 dB defines the 16-QAM ATV noise threshold point, while 15.7 dB defines the threshold for 32-QAM. These values are measured as average signal power to average noise power in a 6 MHz bandwidth.

Based on measurements, the ATV-to-ATV cochannel C/I threshold is 11.2 dB for the 16-QAM mode, and 15.2 dB for the 32-QAM mode. This measurement is based on average desired ATV power to average undesired ATV power in a 6 MHz bandwidth. Note that the ATV-to-ATV C/I performance is 0.5 dB better than the C/N performance. The difference is due to the non-Gaussian nature of the QAM signal.

Based on measurements, the CCDC system can reject NTSC interference for C/I values of 0 dB or more in the 16-QAM mode, and 6 dB or more in the 32-QAM mode. This C/I measurement corresponds to average ATV power to peak sync NTSC interference power.

The thresholds assumed for ATV calculations correspond to a Reed-Solomon block error rate of 10^{-6}, which corresponds to about one block error per minute (with a 15 kHz block rate).

The field strength ratio of F(50,50) to F(50,10), corresponding to a given D/U (C/I) ratio, is assumed throughout the calculations.

ATV ASSUMPTIONS	
Tx Antenna Height Above Avg. Terrain (ft)	1200
Transmitter ERP (dBk)	32-QAM 23.9
	16-QAM 20.4
Receiver Antenna Gain (dB)	10
Downlead Loss (dB)	4
Dipole factor (dB/meter) (75 ohms, 609 MHz)	-22
Thermal Noise at room temp (dBμv) (75 ohms, 6 MHz)	2.6
Receiver Noise Figure (dB)	10
Antenna Front-To-Back ratio (dB)	15
C/N Threshold in 6 MHz BW (dB) (avg-to-avg)	32-QAM 15.7
	16-QAM 11.7
ATV into ATV C/I Thresh in 6 MHz BW (dB) (avg-to-avg)	32-QAM 15.2
	16-QAM 11.2
NTSC into ATV C/I Threshold (dB) (avg-to-peak sync)	32-QAM 6
	16-QAM 0

Table 8.2: ATV Assumptions

8.3 Coverage Area Analysis Results

As a reference point, Figure 8.1 shows two cochannel UHF NTSC stations and the resulting service area of 8,433 square miles, the area within the Grade B service contour not affected by the cochannel interference. To create the same service areas with cochannel NTSC and ATV stations, the NTSC power was held fixed, and the ATV power and cochannel spacing were allowed to vary until both the ATV and NTSC service areas were the same and equal to the reference area. The result is shown in Figures 8.2 and 8.3 for 32-QAM and 16-QAM modes, respectively. Notice that in 32-QAM mode the ATV power is about 13 dB less than the NTSC power and the cochannel spacing is about 40 miles less than the minimum allowed NTSC cochannel spacing of 155 miles.

As another example, two ATV stations operating in 32-QAM mode and having the same 23.9 dBk power (found in the previous case) were brought together until the service areas were equal to the reference area. As shown in Figure 8.4, the cochannel spacing was very close to the 100 mile separation simulcast channel allocation goal. Furthermore, Figure 8.5 shows that in 16-QAM mode the cochannel spacing between two ATV stations using 20.4 dBk power is only 89 miles.

The examples presented here illustrate the coverage resulting from the CCDC HDTV system's low noise threshold and excellent interference rejection. The vast majority of broadcasters will benefit most from using the 32-QAM mode, which provides near 16-QAM coverage at a higher bit rate for excellent picture quality. For those situations where cochannel requirements restrict transmitted power, a broadcaster can use 16-QAM to extend his coverage area, with a modest impact on picture quality.

8.4 Extended Coverage Area

As an alternative to using 16-QAM mode to extend the coverage area, a cellular scheme can be used as a supplement to the current broadcast approach. Extended coverage is achieved by distributing low power transmitters about the fringe of the ATV service area. Each of the transmitters broadcasts the same encoded HDTV signal. Figure 8.6 shows an example of such a scheme, where each low power transmitter provides an extended service radius of 5 miles. Interference from the primary ATV signal and/or the low power signals is handled by the CCDC's superior adaptive equalizer. This situation is equivalent to multiple multipath conditions, which the CCDC's equalizer can handle very well.

Signal distribution to the cell transmitters can be handled by fiber optic links or via microwave links. In the latter case, the down converter accuracy is not critical, for the adaptive equalizer can handle more than 5 Hz frequency offset.

Since the coverage area of each low power station is relatively small, the transmitter antenna does not need to be very high. Low antenna height coupled with low power makes the cell stations relatively inexpensive.

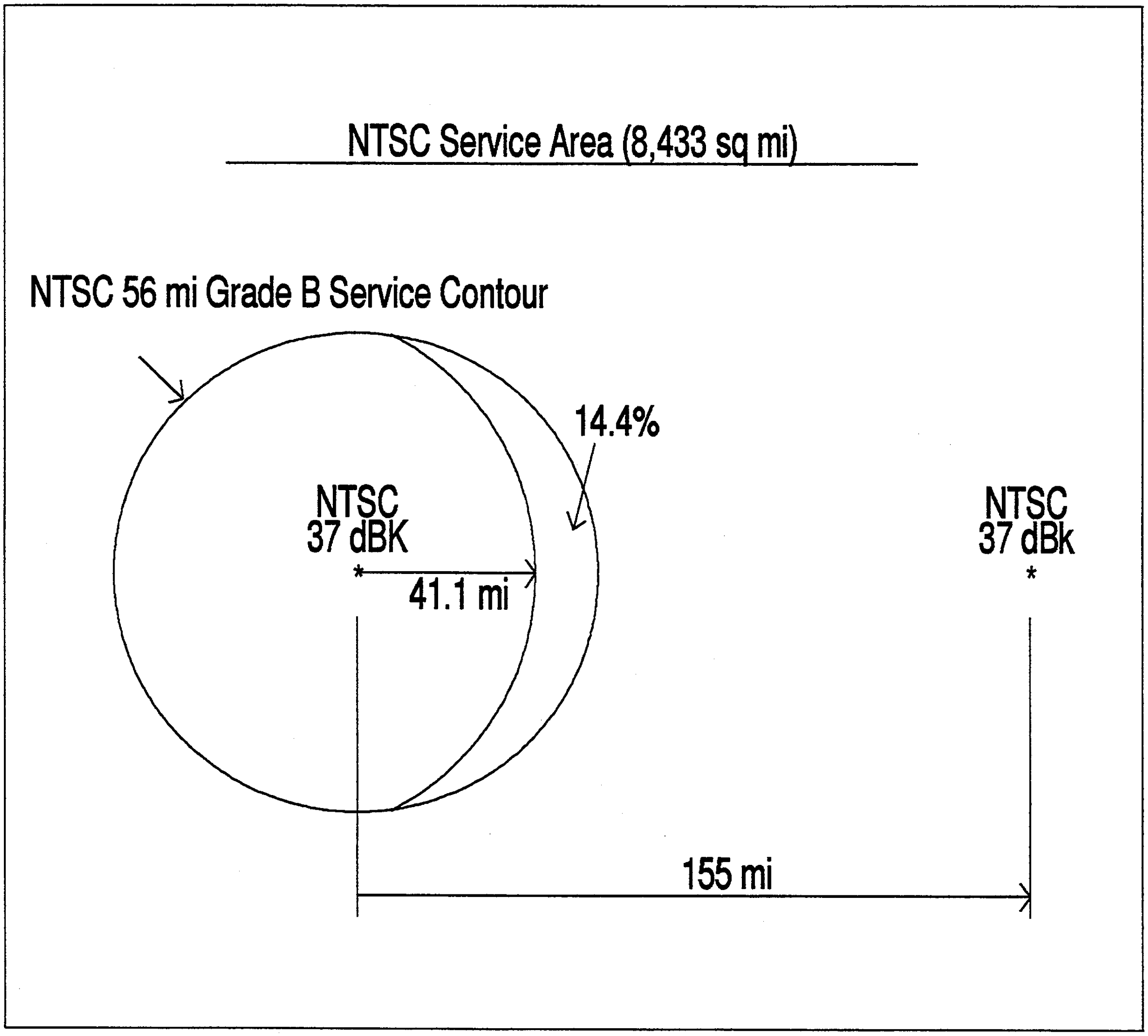

Figure 8.1: NTSC Service Area.

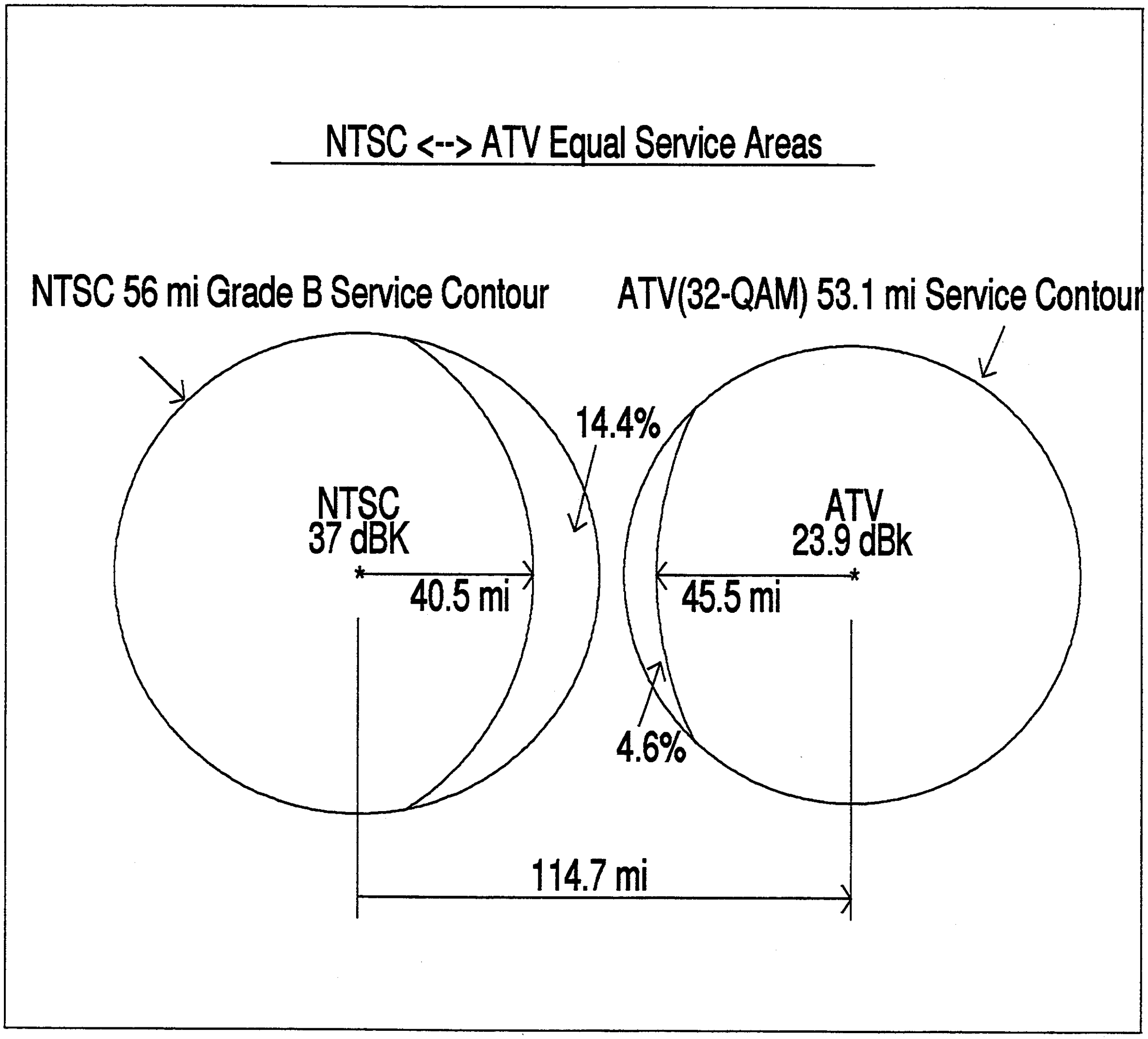

Figure 8.2: NTSC-ATV (32-QAM) Equal Service Areas.

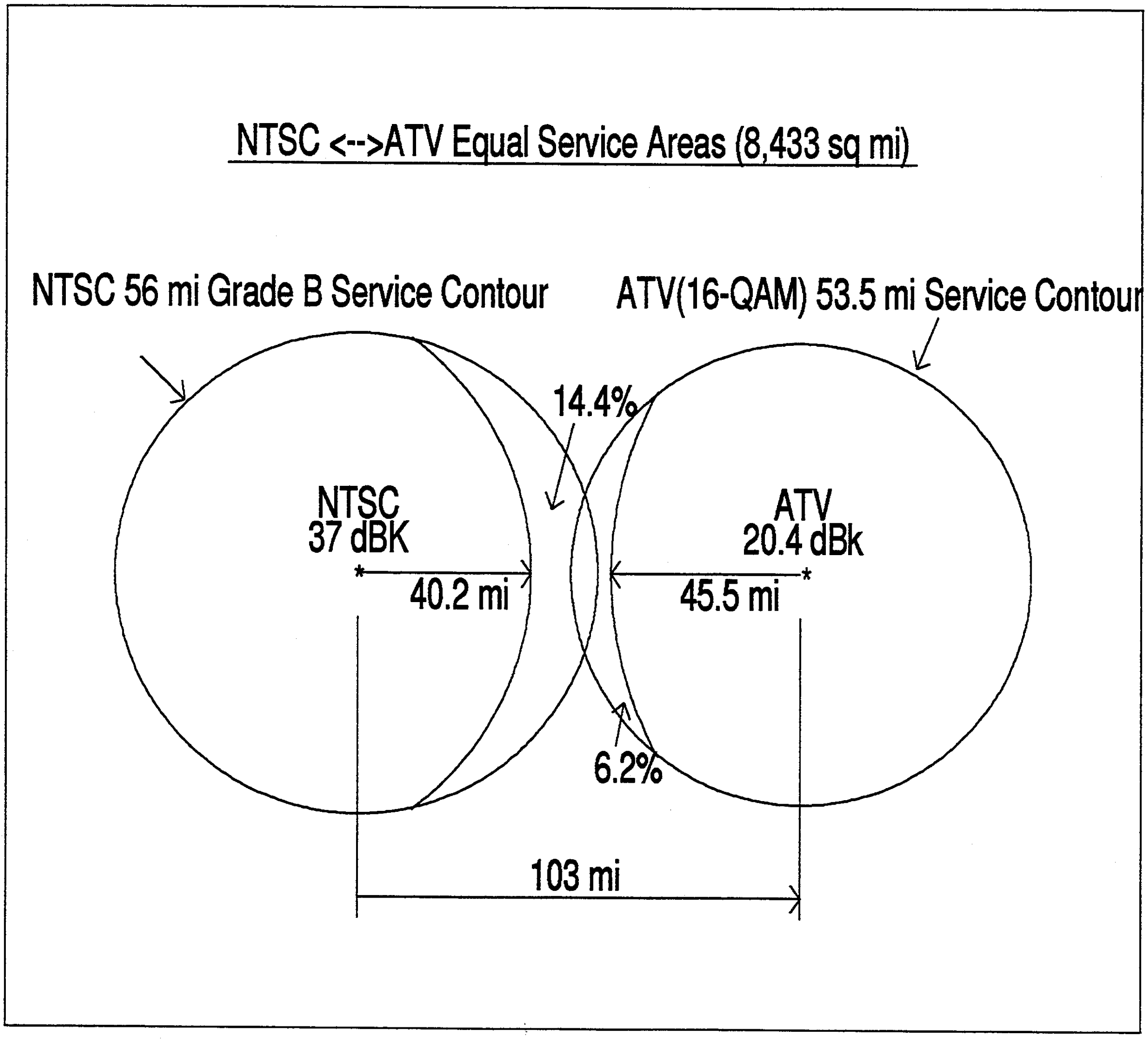

Figure 8.3: NTSC-ATV (16-QAM) Equal Service Areas.

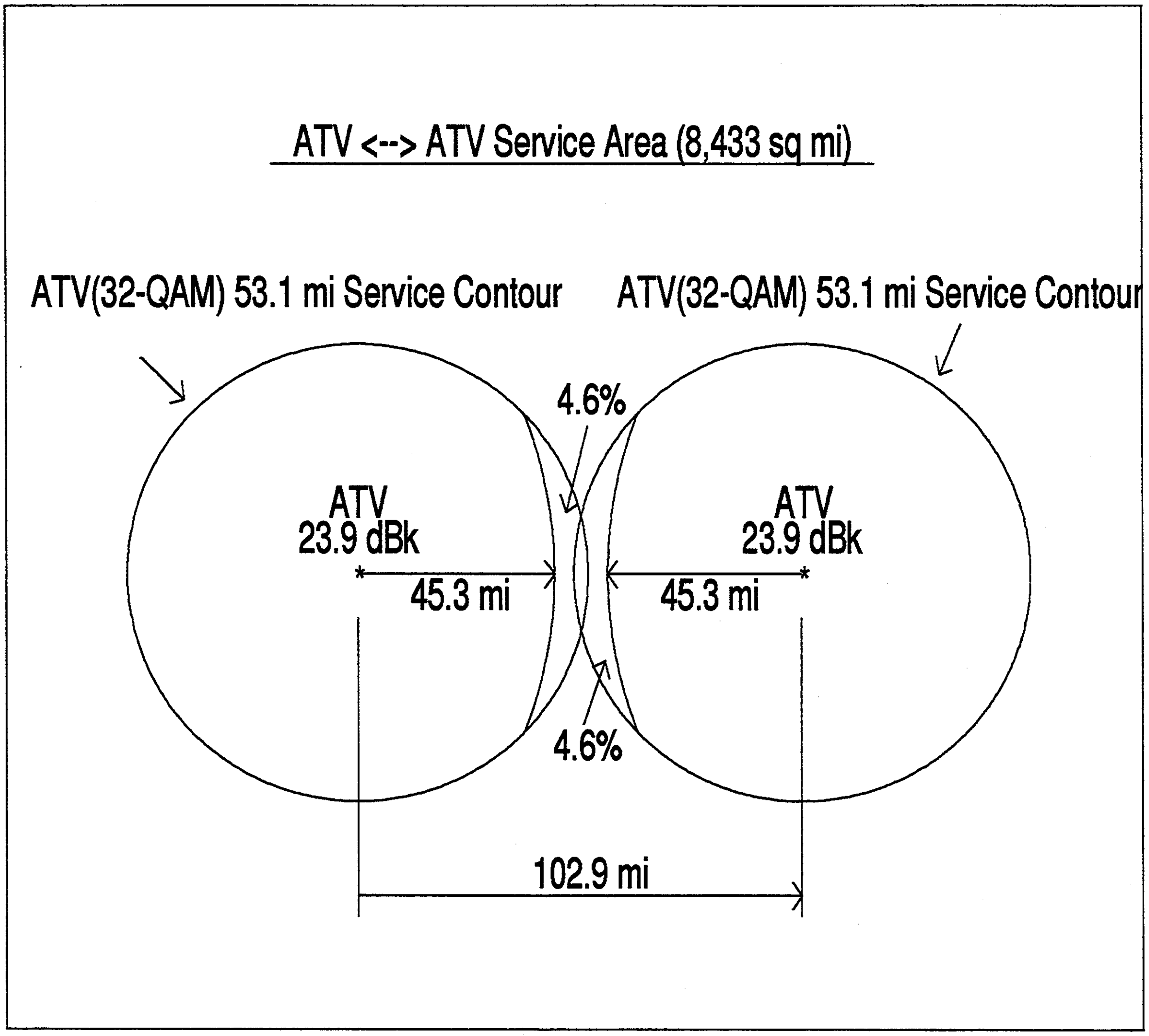

Figure 8.4: ATV-ATV (32-QAM) Equal Service Areas.

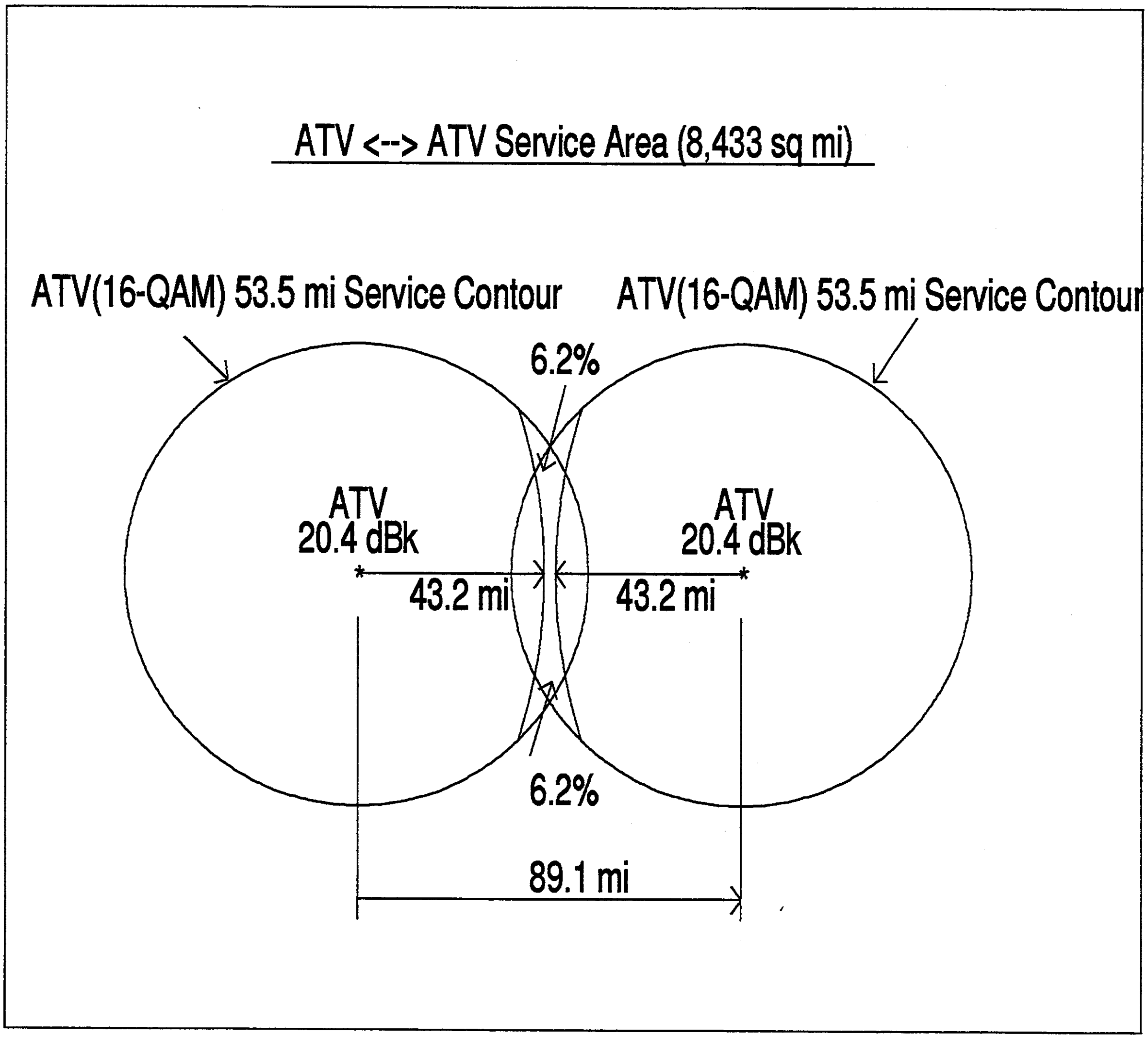

Figure 8.5: ATV-ATV (16-QAM) Equal Service Areas.

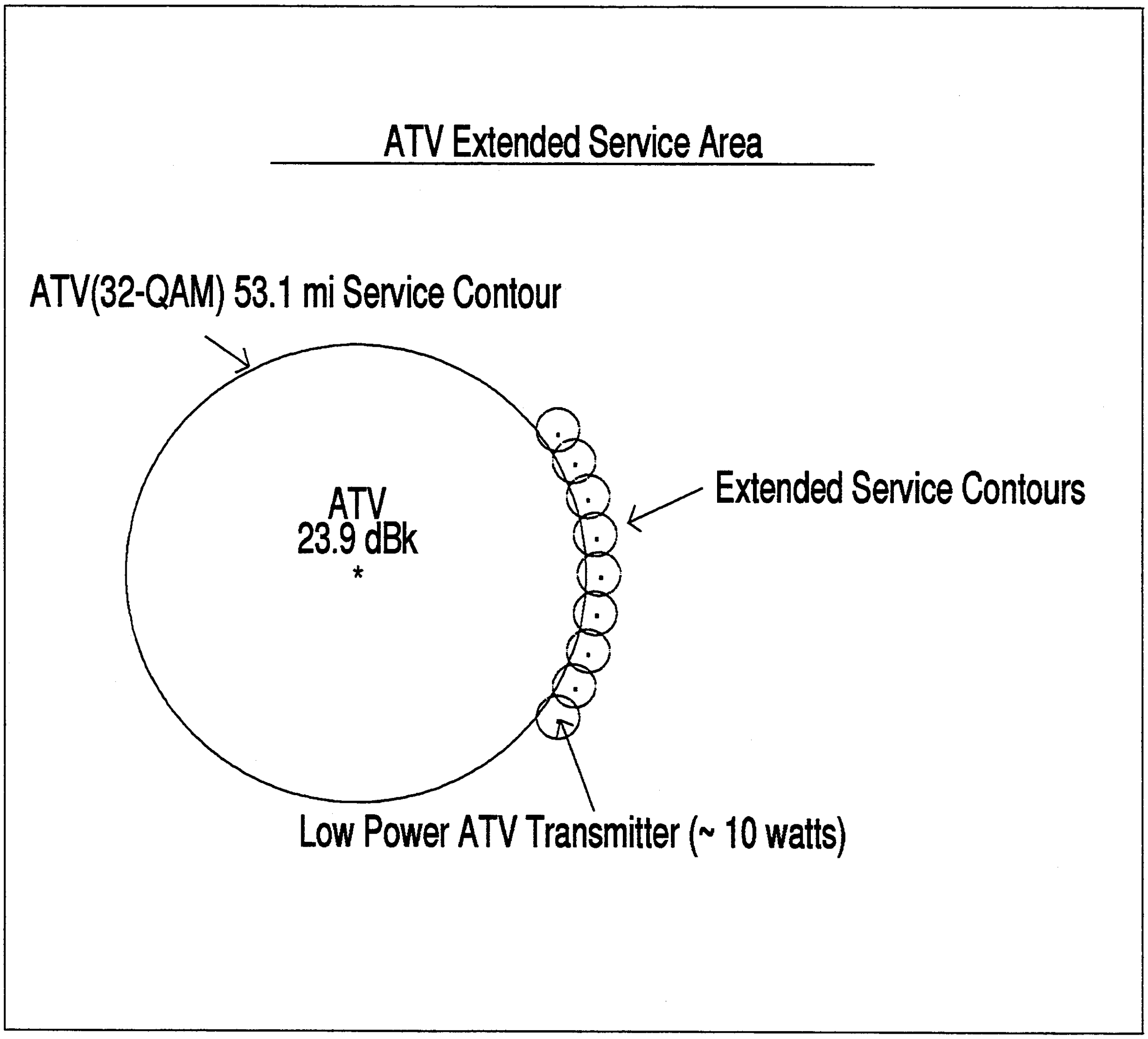

Figure 8.6: ATV Extended Coverage Area.

9 Extensions

9.1 Cable Transmission

The CCDC HDTV system is compatible with cable transmission. The CCDC HDTV signal can be placed in a 6 MHz cable channel adjacent to other CCDC HDTV signals or NTSC VSB-AM signals. Features of the CCDC HDTV System for cable applications include:

- Pass through of satellite or broadcaster-delivered signals to the cable subscriber without signal decompression and recompression at the cable head end.
- Lower power requirements than VSB-AM NTSC will result in an unimpaired HDTV signal delivered to the subscriber with all of the advantages of lower system power loading.
- Channel transmission compatibility with the DigiCipher™ multi-channel NTSC system allows reception/access control of both signals with the same cable converter.

9.2 Satellite Transmission

The CCDC HDTV system can be transmitted over C-band or Ku-band satellite channels using QPSK modulation.

The system can support both FSS and BSS satellite transponders. The threshold C/N is 7.5 dB measured over a 24 MHz bandwidth. Therefore the CCDC HDTV System allows the use of smaller dish sizes compared to other analog or hybrid HDTV systems.

9.3 Other Terrestrial Distribution

Since the CCDC HDTV system is an all-digital system, it can be readily applied to other transmission media such as microwave distribution service (MDS), multi-channel (MMDS) and fiberoptic cables (FO).

An inherent characteristic of the all-digital system is that the HDTV service is free from transmission artifacts caused by various transmission media. Also, the complexity of the interface equipment between various transmission media is substantially lower.

9.4 VCR and Video Disc Recorders

All-digital recording and playback of HDTV signals using the signal format of the CCDC HDTV system is within the reach of current technology for consumer use since the total data rate is on the order of 20 Mb/s. The cost and performance benefits of digital recording will be significant compared to analog recording.

9.5 Extension to Production Standard

The intra frame encoding mode for the whole frame can be used for a production standard. In this case every frame is encoded without ME/MC. At a rate of 3 Mb/frame, production quality video can be obtained for a spatial resolution of 1280×720 pixels. At 60 fps rate, the bit rate required for production quality video is 180 Mb/s. At this rate, the frame can be decoded and reencoded many times with little degradation.

10 Hardware Description

The CCDC HDTV hardware has been designed to fully comply with the ATTC interface specifications. The system consists of two 6', EMI shielded racks, one for the encoder and one for the decoder. The encoder and the decoder draw approximately 20 amps each at 120 volts AC.

10.1 CCDC Encoder and Transmitter Hardware

The hardware on the transmitter end consists of 12 multi-layer printed circuit boards in a rack mountable cage for the video and transmission portions. The audio encoding portion is installed in a separate rack and will be described separately. The video encoder and transmitter rack also has a power supply, a fan, video distribution amplifiers, a control panel, and an interface panel mounted in a rack as shown in Figure 10.1.

10.2 CCDC Decoder and Receiver Hardware

The hardware consists of 17 multi-layer printed circuit boards in a rack mountable cage for the video and transmission portions. As in the encoder, the audio portion is installed in a separate rack and will be described separately. The video decoder and transmitter rack also has a power supply, a fan, a VHF/UHF tuner, video distribution amplifiers, 2 digital audio decoders, a control panel, and an interface panel mounted in a rack as shown in Figure 10.2.

10.3 Audio Hardware

The MIT-AC prototype system is implemented with Motorola DSP96002 floating-point digital signal processors. These processors provide the computation needed for the compression and decompression algorithms. The floating-point arithmetic is a major software convenience. Part of the system is written in a high-level language for development flexibility.

Each channel requires one processor for encoding and one for decoding to allow considerable room for algorithm development. A custom VLSI implementation would be much more compact and would not require a floating-point ALU, provided careful attention is paid to precision and scaling issues.

Communication with the video/audio/data multiplexer is through a generic synchronous link, allowing the audio system to be evaluated independently in a simple end-to-end configuration.

A PC-compatible computer, which is used for convenience, serves as a host for all four audio channels, providing power, code, and display/configuration information. Each stereo pair consists of an MM-96 dual-96002 board manufactured by Ariel Corporation and a custom interface board supporting the AES/EBU and channel interfaces.

Data conversion is a necessary part of the prototype hardware, given the analog interface to the ATTC signal, although we expect that a broadcast environment would provide final audio in digital form directly to the coder. We chose the AD-500 A/D converter and the DA-1000E D/A converter, both from Apogee Electronics Corporation.

10.4 Consumer HDTV Receiver

The CCDC system has been designed to provide optimum performance while the hardware complexity is kept low. This allows the production of low cost consumer HDTV sets as well as low cost broadcasting equipment. Consumer HDTV sets can be designed using a total of 12 custom VLSIs as shown in Table 10.1. In addition, 2 Mbytes of memory are required for the implementation of the digital video decoder.

VLSI Chip	Qty Per HDTV Receiver
16/32-QAM Demod	1
Adaptive Equalizer	4
FEC Decoder/Sync	1
Decompression	4
Video Mux/Filter/OSD	1
Digital Audio Decoder	1
Total	12

Table 10.1: VLSI Chip count.

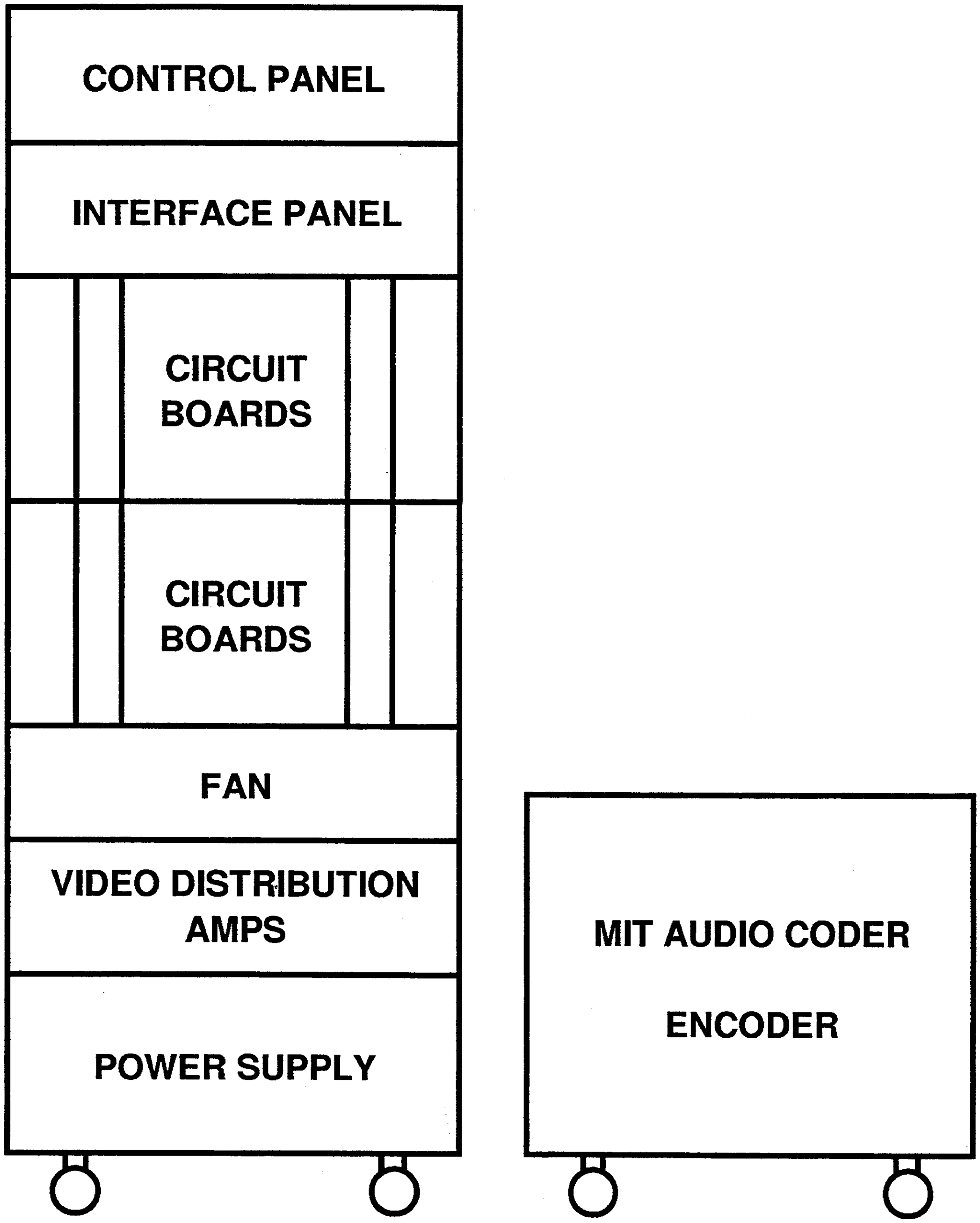

Figure 10.1: CCDC HDTV Transmitter.

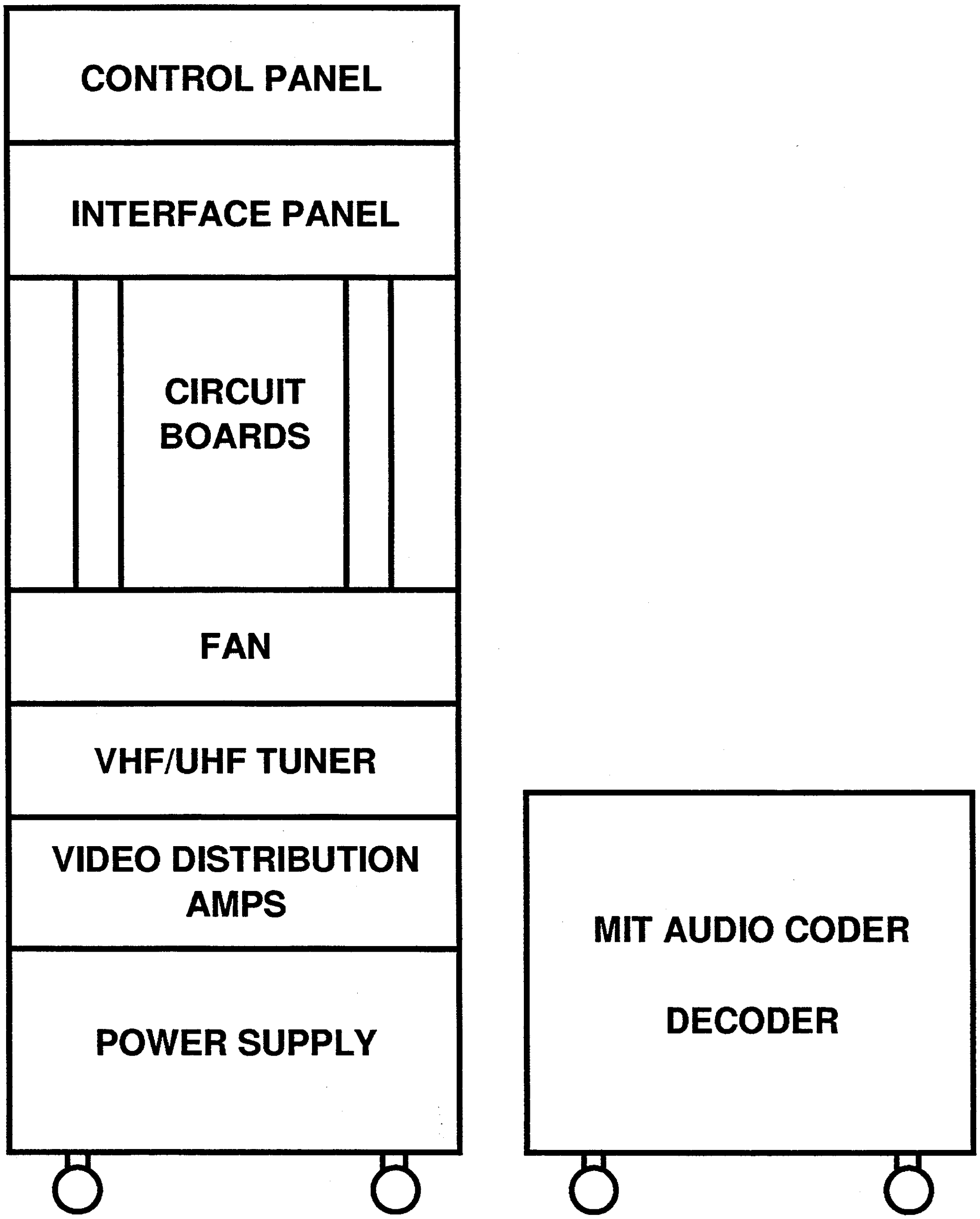

Figure 10.2: CCDC HDTV Receiver.

11 Summary

The Channel Compatible DigiCipher HDTV system is an all-digital system which has a number of important features. The system is channel compatible and will fit within the channels currently being used for terrestrial broadcast transmission. A high resolution progressively scanned baseline video signal of 1280×720 picture elements, 59.94 fps, and 16:9 aspect ratio, and four channels of compact disc quality audio can be transmitted within a single 6 MHz channel.

The system is source-adaptive. The system can recognize and adapt itself to the particular characteristics of the source format so that the highest video quality can be reconstructed. The system is also extensible and scalable, so that future improvements can be accommodated and receivers of different price/complexity/performance classes can be accommodated.

The system is resistant to channel impairments. Very high quality video can be delivered to the home despite substantial channel degradations. Because of this resistance to noise, high picture quality can be achieved with the same coverage area as the current NTSC service at substantially lower power, thereby making the use of taboo channels possible.

The system has been designed to be modular. Video processing, audio processing, and transmission systems are separated and largely independent of each other. This has the advantage that the video and audio processing systems can be used for other applications and will not require significant changes when other media are used for transmission. Modularity also leads to reduced hardware complexity which will lead to significant reduction in the development and production cost of VLSI chips.

In designing the system we have not modified the modulator to account for the existing NTSC service. This ensures that inefficiency does not persist after the HDTV service becomes widely available and the NTSC service discontinues at some point in the future. We believe that the performance of our current system is sufficient to provide adequate HDTV service. We also believe that interference from the NTSC service can be accounted for by proper receiver design in a standard-independent manner.

To achieve the above features, state-of-the-art signal processing and communication technologies have been used. To ensure success, we used well-established engineering principles for the overall design. Specifically, we have used motion compensated transform coding for video processing, adaptive transform coding for audio processing and quadrature amplitude modulation with Reed-Solomon coding, trellis coding and adaptive equalizer for transmission. Within this framework, we have then added in all three areas a considerable amount of new technology. We believe that the adoption of the proposed system will result in a high performance, robust, flexible and cost effective HDTV standard.

For basic principles on video processing, see [1, 2]. For basic principles on digital communications, see [3, 4]. For details on the Digicipher™ system, which is the first system submitted by the American Television Alliance and which has already been evaluated by the ATTC, see [5, 6].

References

[1] J. S. Lim, *Two-Dimensional Signal and Image Processing.* Englewood Cliffs, NJ: Prentice Hall, Inc., 1990.

[2] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression.* New York, NY: Plenum Press, 1988.

[3] R. M. V. Bhargava, D. Haccoun and P. Nuspl, *Digital Communications by Satellite.* New York, NY: John Wiley and Sons, Inc., 1981.

[4] K. Feher, *Advanced Digital Communications.* Englewood Cliff, NJ: Prentice-Hall, Inc., 1987.

[5] W. Paik, "Digicipher™ - all digital, channel compatible, HDTV broadcast system," *IEEE Trans. on Broadcasting*, vol. 36, December 1990.

[6] G. I. Corporation, *DigiCipher™ HDTV System Description.* G.I., August 1991.

A Parameters of the CCDC System

CCDC System Parameters

Operating Mode	32-QAM	16-QAM
VIDEO		
Scanning	Progressive	Progressive
Aspect Ratio	16:9	16:9
Frame Rate	59.94 Hz	59.94 Hz
Bandwidth		
Luminance	34.0 MHz	34.0 MHz
Chrominance	17.0 MHz	17.0 MHz
Active Pixels		
Luminance	1280(H)×720(V)	1280(H)×720(V)
Chrominance	640(H)×360(V)	640(H)×360(V)
Sampling Frequency	75.5 MHz	75.5 MHz
Colorimetry	SMPTE 240M (Approximate)	SMPTE 240M (Approximate)
Horizontal Line Time	21.18 μs	21.18 μs
Compression Technique	MC-DCT	MC-DCT
AUDIO		
Number of Channels	4/6	4/6
Bandwidth	24 kHz	24 kHz
Sampling Frequency	48 kHz	48 kHz
Compression Technique	MIT-AC	MIT-AC
DATA		
Video Data	18.88 Mb/s	13.60 Mb/s
Audio Data	755 kb/s	755 kb/s
Async Data and Text	126 kb/s	126 kb/s
Control Channel Data	126 kb/s	126 kb/s
Total Data Rate	19.89 Mb/s	14.60 Mb/s
TRANSMISSION		
FEC Data	6.54 Mb/s	6.54 Mb/s
Data Transmission Rate	26.43 Mb/s	21.15 Mb/s
QAM Symbol Rate	5.287 MHz	5.287 MHz
Adaptive Equalizer Range	-2 to 24 μs	-2 to 24 μs
SYSTEM THRESHOLD		
Noise (C/N)	15.7 dB	11.7 dB
ATV Interference (C/I)	15.2 dB	11.2 dB
NTSC Interference (C/I)	6.0 dB	0.0 dB

B Clock and Carrier Frequencies Used in the CCDC HDTV System.

DIGITAL VIDEO ENCODER/DECODER	
Master Oscillator	528.6713 MHz
A/D & D/A Sampling Clock	75.5245 MHz
Digital Video Processing Clock	15.1049 MHz
H Sync	47.2 kHz
V Sync	60.0 Hz
DIGITAL AUDIO ENCODER/DECODER	
DSP clock	33.33 MHz
A/D & D/A Sampling Clock	48 kHz
Digital Audio Data Clock	251.75 kHz
DATA INTERFACE	
Master Oscillator	18.525 MHz
Digital Data Board Rate	9600 bps
DIGITAL MODULATOR/DEMODULATOR	
16-QAM Symbol Rate	5.2867 MHz
2 x Symbol Rate	10.5734 MHz
4 x Symbol Rate	21.1469 MHz
Local Oscillator	44 MHz (mod)/43.5 MHz (demod)
TUNER	
1st IF	1200 MHz ±3 MHz
2nd IF	43.5 MHz ±3 MHz
1st LO	1257 - 2087 MHz
2nd LO	1156.5 MHz

C Relationship between the CCDC System and the DigiCipher System

The American Television Alliance has submitted two systems to the ATTC for evaluation. The first system, DigiCipher, has already been tested by the ATTC. Many system capabilities in both systems are not progressive or interlace specific. Certain features of our CCDC system, for example, would be implemented in our DigiCipher system and likewise certain features of the DigiCipher system could be implemented in our CCDC system.

C.1 Source Adaptive Coding

Extensions of the CCDC system allow source-adaptive coding which accommodates a wide range of frame rates, resolutions and progressive/interlaced scanning. The DigiCipher system is currently based on interlaced scanning with a special 24 fps film mode, but can also allow extensions to incorporate full source adaptive coding.

C.2 Extensibility and Scalability

Through source-adaptive coding and the use of a frame header, the CCDC system has been designed to be extensible and scalable so that future improvements can be accommodated and receivers of different price/complexity/performance classes can be accommodated. The DigiCipher system can incorporate this feature.

C.3 Chrominance Resolution

The CCDC system has the extension capability to preserve full chrominance resolution. This feature is particularly significant for film where the available bit rate is high, and for computer generated graphics and text containing saturated colors. The DigiCipher system can incorporate this capability.

C.4 Vector Coding

The CCDC system uses vector coding to encode the selection information. This feature can be incorporated in the DigiCipher system.

C.5 Adaptive Codebook for Amplitude Quantization

The CCDC system encodes the selection information and the transform amplitude information separately. In addition, the amplitude information is encoded using a quantization table that adapts to the differences in the luminance/chrominance components, inter/intra-mode processing and different spatial frequencies. The DigiCipher system can incorporate this feature.

C.6 Accurate Motion Compensation

The CCDC system has $\frac{1}{2}$-pixel accuracy and adaptive block-size motion estimation. The DigiCipher system can incorporate this feature.

C.7 Error Concealment

The CCDC system incorporates additional error concealment techniques. In addition, the system limits the error to a macroblock. This feature can be incorporated in the DigiCipher system.