S-N

ATVA-Progressive System

submitted by

Advanced Television Research Program
Research Laboratory of Electronics
Massachusetts Institute of Technology

on behalf of

The American Television Alliance

February, 1991

Contents

1 Introduction

The ATVA-Progressive system is an all-digital advanced television system submitted by the Massachusetts Institute of Technology on behalf of the American Television Alliance. This document is submitted for re-precertification of the previously precertified MIT system.

The ATVA-Progressive system has a number of important features. The system is channel-compatible and will fit within the channels now being used in terrestrial transmission. The system is very efficient in using the given channel spectrum. A high resolution video signal (720×1280 picture elements, 60 frames/sec, 16:9 aspect ratio) can be transmitted within a single 6 MHz channel. The system is resistant to channel impairments. Very high quality pictures are delivered to the home in the presence of substantial channel degradation including noise, ghosts, and frequency distortion. In part because of the high resistance to noise and interference, high picture quality is achieved at low transmitter power, making feasible the use of taboo channels.

To achieve the features discussed above, the ATVA-Progressive system uses a variety of sophisticated modern signal processing methods. To achieve a high degree of data compression, motion compensation and transform/subband coding are used. In this method, a motion-compensated residual is represented using a transform/subband analysis, and only transform/subband coefficients with significant energy are transmitted. Motion compensation exploits the temporal redundancy of the video signal to reduce the energy of the signal to be encoded. Transform/subband analysis exploits the spatial redundancy of the motion-compensated residual signal. Elimination of low-energy transform/subband coefficients reduces the data rate requirement without significantly affecting the picture quality. The ATVA-Progressive system uses an all-digital format for terrestrial transmission. A single carrier with double-sideband suppressed-carrier quadrature modulation (DSB QM) is used.

In the next section, we describe the details of the ATVA-Progressive system. We are currently improving the performance of the ATVA-Progressive system, and the final system submitted to the ATTC testing may have additional improvements incorporated into our current design.

2 ATVA-Progressive System

The overall system block diagram is shown in Figure 1. The encoding part of the system consists of conversion of RGB components to YUV components, source adaptive encoding, motion estimation and compensation, transform/subband analysis of motion compensated residuals, adaptive selection of high-energy transform/subband coefficients, quantization of the selected coefficients, entropy coding of the quantized coefficients, and data multiplexing/modulation. The audio is also digitally encoded. The decoding part of the system consists of data demultiplexing/demodulation, transform/subband synthesis of motion compensated residual, and synthesis of video.

2.1 Source Coding

2.1.1 Video Source

The source material is RGB 787.5 line 59.94 fps progressive scan. Only 720 lines are active. Each line has 1280 square picture elements (pixels). The resulting aspect ratio is 16:9. For simplicity of arithmetic, we will assume 60 frames/sec.

2.1.2 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. These sources span a wide range of parameters such as different frame

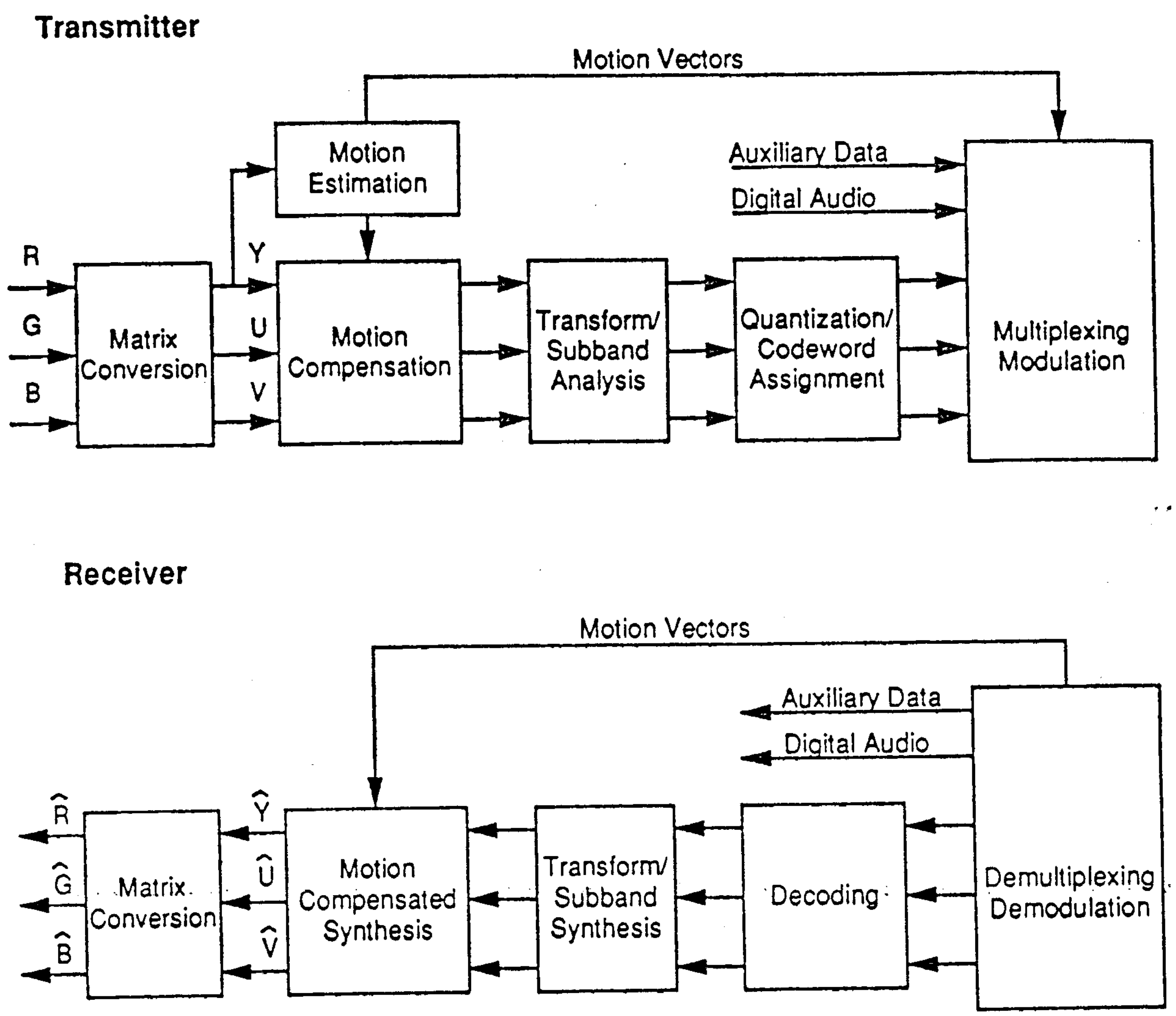

Figure 1: System Block Diagram

rates. In the ATVA-Progressive system, we exploit the differences that may exist in the sources to improve the performance of our video compression method.

As an example, consider the source material originating from film at 24 frames/sec. Instead of converting 24 frames/sec to 60 frames/sec using methods such as 3:2 pull-down and then encoding 60 frames/sec at the transmitter, only the original 24 frames/sec are encoded and transmitted. At the receiver, the 24 frames/sec signal is constructed and used to create 60 frames/sec using methods such as 3:2 pull-down method. This improves the coding efficiency significantly and essentially leads to perfect reconstruction of the film frames.

The simplest method to indicate the source format would be to insert the format information at the source origination point. If this information is not available, then a method to identify the source format is required. In the case of film, if 24 frames/sec have already been converted to 60 frames/sec by 3:2 pull-down, a simple method that exploits the characteristic that three frames are the same, the next two frames are the same, and so on can be used.

In subsequent discussions, we will assume 720 $\times$ 1280 pixels/frame and 60 frames/sec. For a different source format, adjustment has to be made in the number of available bits.

2.1.3 Matrix Conversion

The matrix conversion operation converts RGB components to YUV components. Each of the YUV components is 720 $\times$ 1280 pixels/frame with 60 frames/sec.

2.1.4 Motion Estimation

A motion picture or television broadcast is a sequence of still frames that are displayed in rapid succession. The frame rate necessary to achieve proper motion rendition is usually high enough to ensure a great deal of temporal redundancy among adjacent frames. Much of the variation in intensity from one frame to the next is due to object motion. The process of determining the movement of objects within a sequence of image frames is known as motion estimation. Processing images accounting for the presence of motion is called motion-compensated processing.

Motion estimation methods can be classified broadly into two groups known as 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. The error criterion often used is the squared error. Spatio-temporal constraint methods use an error criterion defined in the spatio-temporal constraint equation domain.

The method that we use in the ATVA-Progressive system is based on the spatio-temporal constraint equation, given by

$$v_x \frac{\partial f(x,y,t)}{\partial x} + v_y \frac{\partial f(x,y,t)}{\partial y} + \frac{\partial f(x,y,t)}{\partial t} = 0 \quad . \tag{1}$$

In equation (1), (v_x, v_y) is the motion vector representing the horizontal and vertical velocity, and $f(x,y,t)$ represents the video signal as a function of two spatial variables x and y and a time variable t. Minimization of $\int (v_x \frac{\partial f(x,y,t)}{\partial x} + v_y \frac{\partial f(x,y,t)}{\partial y} + \frac{\partial f(x,y,t)}{\partial t})^2$ with respect to the motion vector leads to two linear equations for the two parameters v_x and v_y.

The spatio-temporal constraint equation used is based on the assumption of translation with uniform velocity, which is highly restrictive. For example, the assumption is not valid for object rotation, camera zoom, regions uncovered by translational object motion, or multiple objects moving with different motion vectors. However, by assuming uniform translational motion only locally within a small region and suppressing motion compensation in the regions where motion estimates obtained are not accurate, significant reduction in temporal correlation can be achieved by motion estimation and compensation based on the spatio-temporal constraint equation. In comparison with region matching methods, the method we use performs well for both noisy and noise-free video frames with computational reduction by an order of magnitude.

The motion vectors are estimated only from the luminance component and the same motion vectors are used for both the luminance (Y) component and chrominance (U & V) components. One motion vector with resolution of $\frac{1}{2}$ pixel is obtained every 32×32 pixels. Each motion vector (both vertical and horizontal displacements) is represented by 12 bits. The bit rate requirement for the motion vectors, therefore, is given by 12 bits/vector $\times$ 23 $\times$ 40 vectors/frame $\times$ 60 frames/sec = .6624 Mbits/sec.

2.1.5 Motion Compensation

From the previously encoded frame and the motion vectors, a prediction is made for the current frame to be encoded. The difference between the current frame to be encoded and the prediction is computed. This difference is called the motion compensated residual. The motion compensated residual is obtained for each of the three components (Y, U, V).

When the motion compensated residual has sufficiently large energy relative to the image frame, motion compensation is disabled and the image frame itself is encoded. This will be the case when there is a scene change.

For channel acquisition and for recovery from bit errors, some leakage is allowed in motion compensation. A scene change also helps channel acquisition and recovery from bit errors. When a scene change occurs, synchronization between the transmitter and the receiver is immediate.

2.1.6 Transform/Subband Representation

Each of the YUV components is analyzed by a transform/subband analysis filter. The analysis filter is a 2-D separable filter that divides the motion compensated residual into 8 × 8 bands. The region of support size of the 2-D separable filter used is 16 × 16. Unlike the Discrete Cosine Transform (DCT) which is often used, our method reduces artifacts such as artificial blocking effects significantly. The same filters are used for each of the three components.

2.1.7 Adaptive Selection

The transform/subband coefficients are weighted according to the frequency band and luminance/chrominance components to exploit the variation in sensitivity of the visual system.

The weighted transform/subband coefficients are selected based on the energy. The coefficients with the highest energy are selected until the required number of bits to encode the selected coefficients reaches the available number of bits. An efficient method was developed to perform this operation rapidly.

The number of bits available per frame is given by .24956 Mbits. The video bit rate required is given by

$$0.24956 \text{ Mbits/frame} \times 60 \text{ frames/sec} = 14.99 \text{ Mbits/sec.}$$

2.1.8 Quantization and Buffer Control

The location and amplitude of each chosen transform/subband coefficient is encoded jointly using a Huffman encoding method. The statistics required for the Huffman encoding method are obtained from video that is representative of a number of different scenes. The quantization step and the adaptive selection step operate jointly to ensure that the total number of bits required per frame is less than 0.24956 Mbits.

The allocation of bits to transform/subband coefficients results in a locally variable-bit rate system, which is used in a fixed-rate channel environment. A typical solution is a buffer at the coder quantizer. As the buffer fills, the quantizer is made more coarse. This reduces the bit rate and avoids overflow. As the buffer empties, the quantizer is made finer. This increases the bit rate and fills up the buffer. Some care is required to ensure stability and to ensure that the buffer can never overflow or underflow.

Buffers add complexity. More importantly, the buffer is only locally adaptive. When it encounters busy-detail regions, it begins allocating more bits. When it encounters even busier-detail regions next, these regions are not given enough bits, because the buffer may be full at that time.

In the ATVA-Progressive system, we use a fixed amount of storage for one complete frame (0.24956 Mbits) and enough coefficients are selected to just fill the buffer storage. This allows the buffer to adapt to the scene change on a global basis rather than on a local basis, significantly improving the system performance.

2.1.9 Digital Audio

The audio is also represented digitally. Transform/subband coding and exploitation of the human auditory perception allow us to encode one channel of CD quality audio at 0.125 Mbits/sec. The total number of bits allocated for audio is 0.5 Mbits/sec. The ATVA-Progressive system will support four audio channels, with 0.125 Mbits/sec available for each channel.

2.1.10 Auxiliary Digital Data

The system also supports a 0.126 Mbits/sec auxiliary data stream for data services such as closed-end captioning, teletext, program guides, etc.

2.1.11 Access Control

The system supports 0.126 Mbits/sec for conditional access information such as authorizations and keys.

2.1.12 Error Correction

Forward error correction is allocated 3.042 Mbits/sec of the data stream.

2.1.13 Summary of bit allocation

A summary of the bits used is shown in Table 1.

	Bits	
Motion Vectors	0.6624	Mbits/sec
Selected Coefficients	14.9736	Mbits/sec
Audio	0.5000	Mbits/sec
Auxiliary Data	0.1260	Mbits/sec
Access Control	0.1260	Mbits/sec
Error Correction	3.0420	Mbits/sec
Total	19.4300	Mbits/sec

Table 1. Allocation of Bits

2.1.14 Multiplexing/Modulation and Demultiplexing/Demodulation

The digital bits from the various sources are multiplexed and modulated for transmission. At the receiver the received signal is demodulated and demultiplexed. The multiplexing/modulation and demultiplexing/demodulation technology is discussed in Sections 3 and 4.

2.2 Source Decoding

The decoding process at the receiver is the inverse of the coding process discussed in the previous sections. The received signal is demodulated and demultiplexed. Using the data representing the location and amplitude information, the transform/subband coefficients are properly identified. The result is then used by the transform/subband synthesis system to obtain the motion-compensated residual signal. The residual signal is combined with the previously reconstructed frame to reconstruct the current frame.

The display is a progressively scanned display at 59.94 frames/sec with 787.5 lines/frame. The aspect ratio used is 16:9. The digital bits representing the audio information are used to support 4-channel audio.

Further details on motion compensated video compression can be found in the following reference: Jae S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, 1990.

3 Channel Transmission Format

The multiplexed bits are converted to modulated waveforms in the modulator and are recovered in the demodulator. Figures 2 and 3 illustrate the modulator and demodulator components which accomplish this. Each component is described in more detail in the following sections.

3.1 Digital Data Transmission

The digital video information is multiplexed together with 4 digital audio channels, an auxiliary 0.126 Mbits/sec data stream, and 0.126 Mbits/sec access control to form the composite 19.43 Mbits/sec digital stream. The auxiliary data stream is made available for transmission of closed captions or other digital data. Sync bits are inserted into the transmitted bit stream in order to mark the frame boundaries. These unique bit sequences are used by the receiver to establish frame synchronization.

The encoder performs encoding of one frame at a time. The digital data in each frame is transmitted in the order shown in Figure 4. This basic frame structure is repeated for every frame.

3.1.1 Data Randomization

Prior to transmission, the digital data is processed with a sequence randomizer which randomizes the transmitted bits. The sequence randomizer is implemented as a Linear Recursive

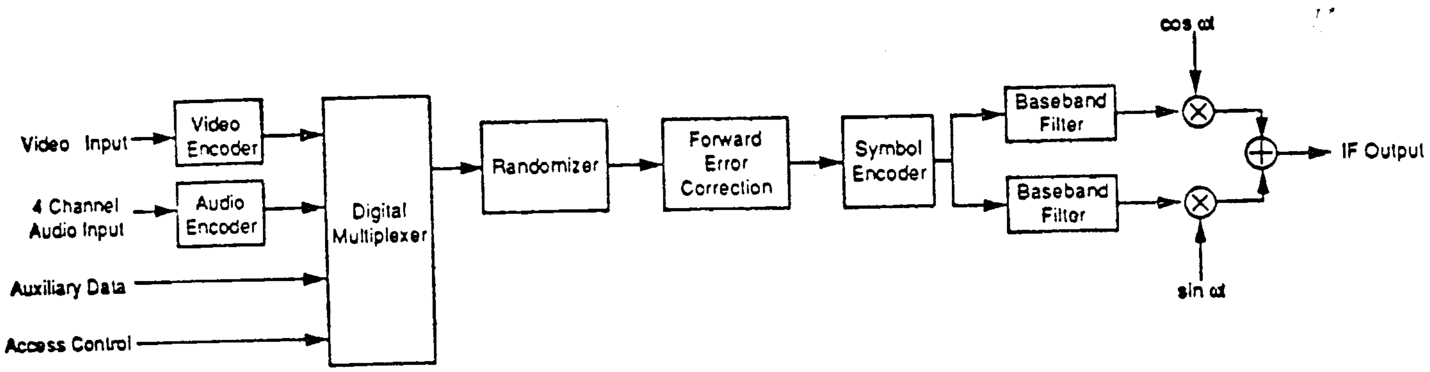

Figure 2: Modulator Components

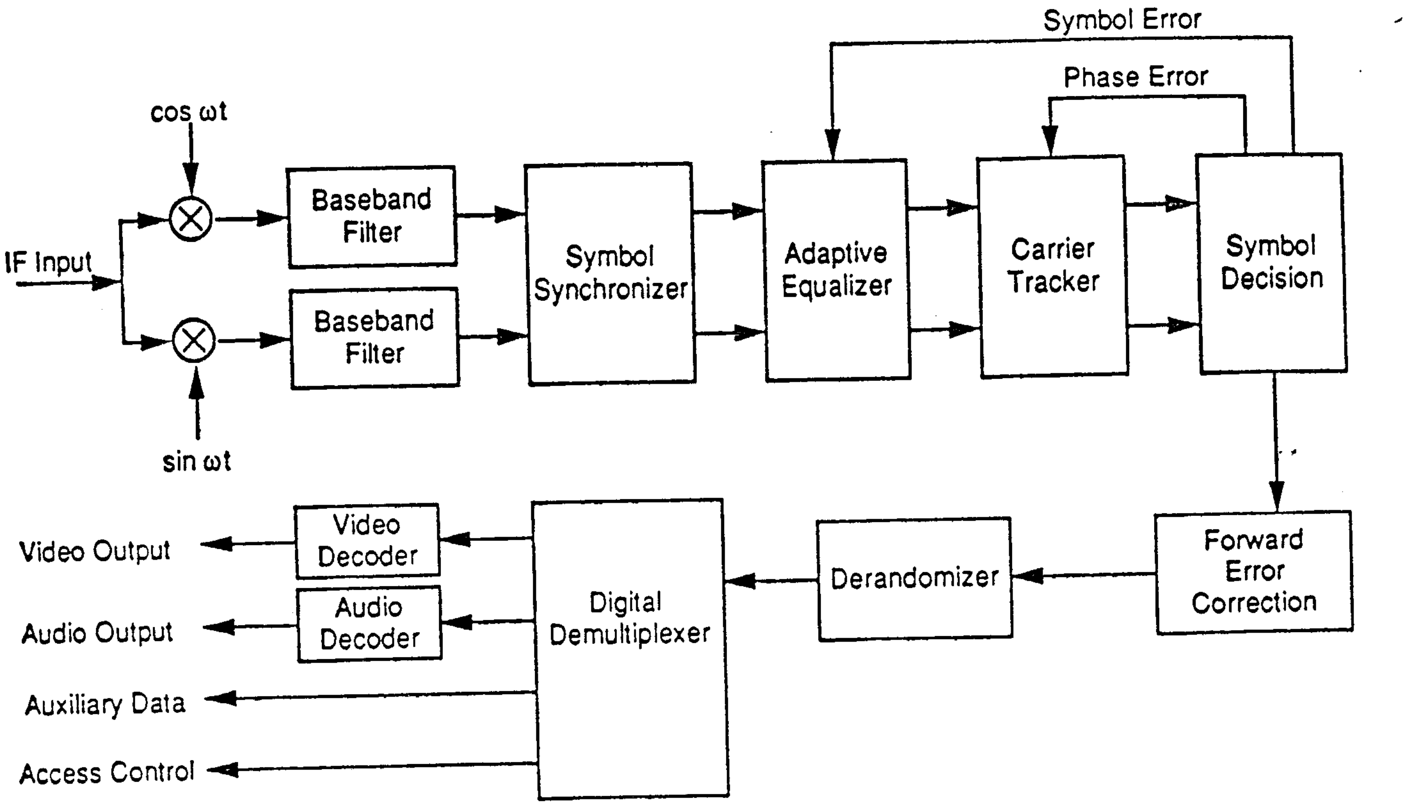

Figure 3: Demodulator Components

Frame Sync	Motion Vectors	Audio Channels	Auxiliary Data	Modulation	Video Coefficients

Figure 4: Digital Data Transmission Frame

Generator with a generator polynomial of $1 + x^{-18} + x^{-23}$. Randomizing is performed on the digital data in order to guarantee that the transmitted sequence of bits is sufficiently random so the adaptive equalizer in the receiver remains properly converged during periods of idle transmission. Figure 5 illustrates the transmit randomizer implementation.

3.1.2 Forward Error Correction

Reed-Solomon coding of rate 130/154 (t=12) is employed to correct transmission errors caused by noise/interference. The system threshold is 19 dB C/N including 2.5 dB of implementation margin. At 19 dB C/N there will be one undetected error event per day. The threshold C/N is much lower than the C/N required for satisfactory reception of analog VHF/UHF signals.

3.1.3 Symbol Encoding

The bits in the randomized data stream are grouped into four bits (pairs of dibits). These dibits are encoded to generate a complex I and Q symbol sequence at a sampling frequency of 4.86 MHz. Symbols are generated at a rate of 4.86 MHz, corresponding to 4.86 million I and Q samples per second.

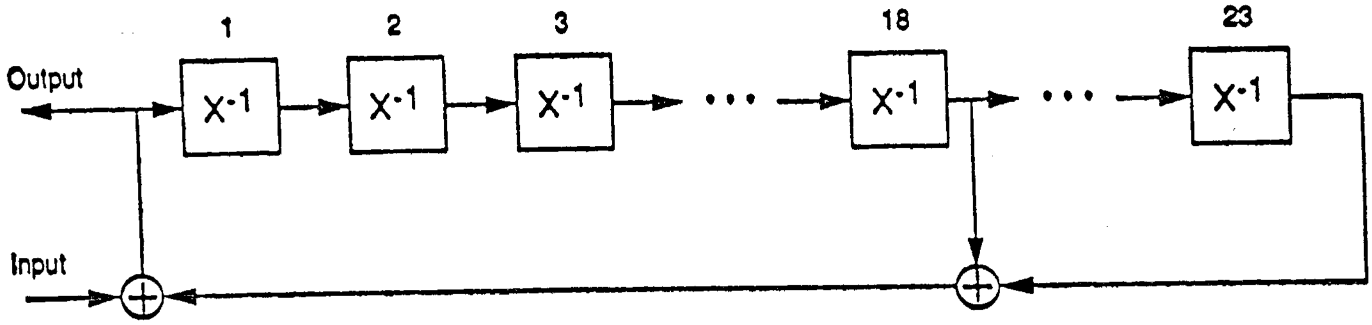

Figure 5: Randomizer Implementation

3.2 Transmit Filter and Baseband Spectral Shaping

The quadrature signal is generated at a sampling frequency of 4.86 MHz. The minimum bandwidth necessary to transmit this signal is 4.86 MHz, corresponding to the Nyquist rate. In practice, it is necessary to augment the Nyquist bandwidth with transition bands. The target channel bandwidth is 6 MHz, corresponding to 0.57 MHz transition bands on each side. Spectral shaping filters in the transmitter limit the signal to a 6 MHz double-sided bandwidth. A linear phase finite impulse response digital filter is used for spectral shaping. Figure 6 illustrates the frequency response of this filter.

3.3 Baseband to IF Signal Conversion

Following spectral shaping, the complex signal is quadrature modulated onto the 44 MHz carrier to produce the IF output. The resulting signal is transmitted suppressed carrier and appears as colored random noise.

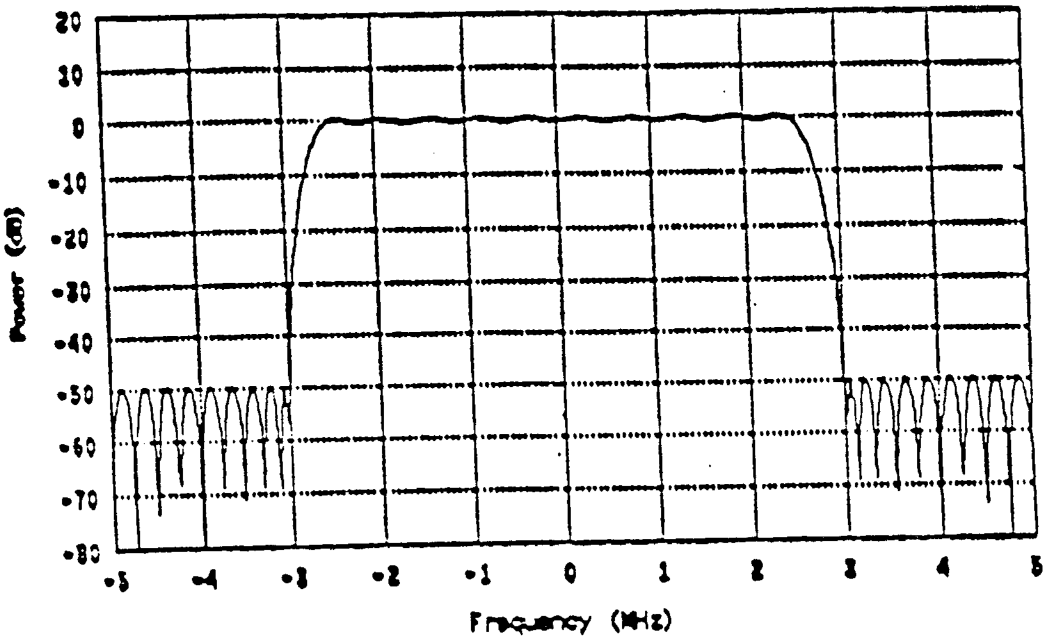

Figure 6: Spectral Shaping Filter Frequency Response

4 Receiver Architecture

The receiver consists of an RF/IF downconverter, a baseband demodulator, and decoders for the video and audio data streams. The RF/IF subsystem downconverts the RF input to IF, 41-47 MHz. The demodulator subsystem recovers and processes the quadrature data components, providing error-corrected data components at its outputs.

4.1 Downconverter

The receiver can tune the VHF and UHF bands. The received RF signal is downconverted to standard TV IF, 41-47 MHz.

4.2 Demodulator

Figure 3 illustrates the demodulator architecture. Components are:

- Quadrature Detector
- Symbol Synchronizer
- Adaptive Equalizer
- Carrier Tracking Loop
- Symbol Decoder
- Forward Error Correction
- Derandomizer

4.2.1 Symbol Synchronizer

The symbol synchronizer contains two components: a symbol timing phase locked loop and a digital resampling circuit. The symbol timing phase locked loop generates a clock at twice the symbol rate (nominal frequency 9.72 MHz). This clock drives a digital resampling circuit which resamples the 4.86 megasample/second complex signal to a rate of twice the symbol rate so as to generate two samples per symbol interval. Both samples occur at the same relative position in the symbol interval of all symbols.

4.2.2 Adaptive Equalizer

Adaptive equalization is employed to handle the reflections (multipath) found in typical VHF or UHF broadcast reception, and the microreflections generated in cable transmission. The equalizer will cancel complex multipath of up to 2 microseconds. It will also have a single echo cancellation equalizer for a long multipath of up to 32 microseconds.

4.2.3 Carrier Tracking Loop

The carrier tracking loop computes a carrier correction phase term to be applied to each complex symbol output from the adaptive equalizer. Input to the carrier tracking loop is the phase error signal from the symbol decision device.

4.2.4 Symbol Decoder

The symbol decision device determines which constellation point is closest to each complex output from the adaptive equalizer, and decodes the dibits.

5 Appendix A

1 Sync Interfacing between ATTC and Proponents

1. Encoder Sync Specifications
   The source material is 787.5 line 59.94 fps progressive scan. Only 720 of the lines are active. The active portion of each line is 17.65 μs in duration with 3.53 μs for retrace.

2. Decoder Sync Specifications
   The system will deliver an RGB/sync signal that is compatible with the input.

2 RF Interfacing between ATTC and Proponents

3. Modulator Carrier Frequency
   The modulator will have a carrier frequency of 44 MHz. This frequency corresponds to the center of the IF band from 41 to 47 MHz.

4. Up-Conversion Frequency
   A 245 MHz frequency source is required to up-convert the IF to VHF test Channel 11. This frequency will center the information band within the test channel which is assigned frequencies between 198 and 204 MHz. Note that since this corresponds to a superhet structure, the RF signal will appear spectrally inverted. A corresponding superhet structure in the receiver will re-invert the spectrum prior to demodulation.

5. Receiver Tuner Frequency
   The receiver tuner will operate at 201 MHz for receiving test Channel 11 and 527 MHz for receiving test Channel 23.

6. Multi-path Cancellation
   The system will include multi-path cancellation. It will cancel a single dominant multipath component with a maximum time delay of 32 μs, and higher order multipath components with a maximum time delay of 2 μs.

7. Signal Peak Power
   The radiated signal has a constant peak power level independent of picture content and motion.

8. System Carrier Frequencies
   The system does not employ subcarriers. The modulation is double sideband quadrature modulation with the carrier fully suppressed.

9. UHF Taboos
   We believe the system will operate over most available UHF taboos and with adjacent channel transmitters (both upper and lower) closer than presently permitted under FCC rules. We will be conducting some tests of our own to substantiate these claims.

10. Time Sharing of Augmentation Channel
    Not Applicable.

11. Maximum Energy Density Image
    The system employs randomization techniques to ensure that the radiated signal is always "random". Therefore the energy density of the radiated signal is essentially independent of image content.

12. Digital Input Port
    Two digital input ports will be provided. One port will support a stereo digital input and the other port will support a 0.1 Mbits/sec digital data stream. The 0.1 Mbits/sec stream is intended for application such as closed captions.