ZENITH

AT&T

TECHNICAL DETAILS

DIGITAL SPECTRUM COMPATIBLE

Cover: "Carousel" by Don Mitchell of AT&T Bell Laboratories — computer-generated synthesized video images designed to challenge the Zenith-AT&T high-definition video compression system with very complex motion simulations.

September 23, 1991

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS

SECTION 1

INTRODUCTION AND SYSTEM OVERVIEW

1.1 Introduction

Zenith and AT&T have developed an all-digital high-definition television (HDTV) simulcast system that combines powerful video compression technology and a unique simulcast transmission system. The system, called "Digital Spectrum Compatible" (DSC), provides full high-definition resolution -- perceived to be equal to the studio original -- even after compressing the wide bandwidth signal into a 6-MHz channel. The system digitally transmits the compressed signal on currently unusable or "Taboo" television channels with only minimal interference to or from NTSC channels.

The DSC-HDTV system's high-performance compression and RF transmission technologies provide transparent picture quality utilizing a maximum video data rate (up to 17 megabits per second, Mb/s) and provide noise-free and ghost-free terrestrial broadcast reception throughout a broad service area. This area will be comparable to or, in many cases, greater than the current NTSC service area, even though transmitter power can be reduced by more than 90% from the power required for NTSC transmissions.

In addition to its unique performance capabilities for terrestrial broadcasting, the DSC-HDTV system is compatible with other media, including cable, satellite, studio video tape recorders, home video cassette recorders, video disc and fiber. Because the complete high-definition television signal (including video, chroma, audio, ancillary signals, decoder address, and encryption information) is encoded into one 6-MHz bandwidth signal, the system can be simply adapted to other media using current electrical and mechanical technologies. (See Section 8.)

The system's 787.5-line progressive scanning format eliminates artifacts of interlaced systems, provides full motion rendition (critical for sports and other fast-action programming) and promotes HDTV compatibility with current and future computer and digital communications technologies. Progressive scanning plus square pixels and unique compression technology make the DSC-HDTV system ideally suited for interconnectivity, extensibility, scalability and other computer-related considerations. (See Section 9.)

1.2 Extended Coverage

The DSC-HDTV system's unique four-level vestigial sideband (4-VSB) modulation technique assures noise-free and interference-free reception throughout the DSC-HDTV service area. To maintain picture information (plus audio and data) in far-fringe areas, Zenith and AT&T have created an extended coverage coding system. The 4-VSB coding is complemented by a two-level digital data system (2-VSB) which significantly expands the service area of the Digital Spectrum Compatible signal.

The resulting intelligent *bi-rate coding system* identifies and selects the most important picture information on a scene-by-scene basis and automatically transmits that data in a two-level (binary) mode. The remainder of the picture information is transmitted as 4-level digital data. Two-level digital coding makes the system far more tolerant of noise and other interference at greater distances from the transmitter. This allows extended reception of the signal beyond the traditional NTSC service area and eliminates the so-called "cliff effect," or complete and abrupt loss of picture, associated with some other all-digital approaches.

This unique coding system will be important in congested service areas in and around major cities. In areas where two co-channel transmitters operating on the same channel are close together (about 100 miles, for example), HDTV transmitter power needs to be reduced to prevent interference into the NTSC service area. In an ordinary system, interference from the relatively high-powered NTSC channel could result in excessive uncorrectable errors that would make the 4-level data unusable at certain times and certain locations in the far-fringe region. Ordinarily, a brief interruption of the 4-level digital data would cause a cliff effect, but the 2-level data of the DSC-HDTV system will sustain video and compact-disc-quality audio.

In most areas of the country where the distance between co-channel transmitters can be greater than 125 miles, the DSC-HDTV transmitter may operate at a higher power level than is possible in the more congested service areas. That higher power level will assure error-free reception over even larger service areas than today's NTSC broadcasts offer.

1.3 Video Encoding & Decoding

The powerful compression technology developed for the DSC-HDTV system handles high-detail, complex motion video such as sports or complex computer-generated images. The encoder/decoder system adapts to scene changes. When the television receiver's channels are changed, a special decoder mode within the TV receiver's circuitry is used to present a clean HDTV picture within a fraction of a second. The system is especially designed to reduce the complexity of the receiver, both in terms of amount of processing and memory. In addition, the DSC-HDTV data format supports VCR special functions such as speed-search and freeze-frame. A detailed description of the encoder and decoder is provided in Section 3.

The inherent modularity of the video encoder, as well as the use of progressive scan and square pixels, greatly facilitates the use of scaled or extended versions of the DSC-HDTV encoder in other applications. For example, computer graphic workstations can display the 787.5 format directly with space left for text or control windows. Simpler versions of the encoder can be used without modifying the compressed video format or the decoder. In particular, the same algorithm has been used to efficiently compress complex scenes at NTSC and CCIR-601 resolutions. With a small change in some parameters, the same decoder can decode compressed data at HDTV, CCIR-601 or NTSC resolutions. Interoperability with alternate media can be easily provided because of the packet-like compressed video format and flexible data rate.

1.4 Interference Rejection

Simulcast HDTV requires the use of channels that are currently a buffer between active NTSC channels in the VHF and UHF television broadcast bands. In the Zenith/AT&T system, spectrum compatibility is achieved primarily because of a unique rejection filter that prevents NTSC signals from interfering with HDTV signals. The digital filter technology (described in detail in Section 6) is a comb-type analog/digital processor which rejects interference from NTSC signals. The NTSC signal is rejected by the filter in the DSC-HDTV receiver, while a complementary preprocessor is used in the DSC-HDTV transmitter.

The digital VSB modulation system uses a band-edge pilot that assures HDTV reception even in heavy interference. The precise high-performance filter readily rejects co-channel interfering signals that are even stronger than the desired DSC-HDTV signal. Equally significant, the system accomplishes this without taxing the DSC-HDTV error correction system. In addition, the system will acquire the HDTV signal even under conditions of simultaneous multiple interfering signals, such as noise and co-channel interference.

1.5 Audio and Ancillary Data

The DSC-HDTV system provides four independent channels of CD-quality audio. The prototype hardware will use the Dolby AC-2 process for compression/decompression (see Section 4).

Provisions for closed captioning, teletext, encryption, addressing, program identification headers and other data services also are provided (see Section 5). The data is packaged in packet-like format with a substantial number of bits allocated for yet to be defined and evolving header functions. Such packaging can be easily encapsulated into packet formats standardized for different computer networks.

The 2-level data of the *bi-rate coding system* is used for two audio channels and a portion of the ancillary data to also extend coverage in these complementary functions.

1.6 Economy for the Broadcaster

The DSC-HDTV source signal's simple ratios to NTSC horizontal and vertical scan rates result in easy up and down conversion between formats. Addition of a simulcast transmitter and antenna would allow broadcasters to "pass through" network HDTV programming. The format up-converter allows broadcasters to telecast locally originated programs at low cost because all NTSC equipment remains usable for transmissions on the HDTV channel. Therefore, the upgrade of all studio equipment to HDTV equipment can be deferred.

Up-converted signals will be of NTSC studio quality resolution when derived from component source signals. However, the received up-converted signals will be significantly better than analog NTSC signals. When transmitted they will have all the advantages of digital transmission -- without cross-color, cross-luminance, interlaced scan and ghost artifacts of NTSC.

1.7 Economy for the Consumer

High speed signal processing technology and algorithm design have minimized the complexity of the DSC receiver/decoder to help make HDTV receivers more affordable to consumers. In fact, a DSC-HDTV receiver requires only enough memory to store slightly more than one high-definition frame (a total of about 1.6 megabytes of memory). The overall semiconductor requirements are moderate because the system has minimized receiver digital signal processing and other logic functions.

Economical circuit applications will be possible in the DSC-HDTV system because of the NTSC-related horizontal scanning rate, which is three times the NTSC rate and the frame rate which is identical to the 59.94 Hz field rate of NTSC. HDTV receivers are expected to be capable of handling both HDTV and NTSC broadcasts.

1.8 Conclusion

The Zenith and AT&T Digital Spectrum Compatible system solves the fundamental problems associated with digital HDTV, namely compression of wide-bandwidth signals with full high-definition resolution into a 6-MHz channel and transmission of those signals to a broad service area without detrimental interference.

The system's progressively scanned lines in a digital format offer numerous advantages over interlace scanning and analog formats. In addition, DSC-HDTV offers a unique *bi-rate coding system* that extends the HDTV reception area and eliminates the complete and abrupt loss of television reception. The *bi-rate coding system* allows uninterrupted service even in congested areas such as major metropolitan markets.

Other advantages include rapid recovery from channel changes, minimal interference into or from NTSC signals (as well as from other DSC-HDTV transmissions) and economical HDTV service for both the broadcaster and the consumer.

1/r = .15625

SECTION 2
VIDEO SOURCE SIGNAL

47202.797

2.1 Timing, Bandwidth, Image Format

Digital Spectrum-Compatible High-Definition Television (DSC-HDTV) images are progressively scanned at 787.5 lines/frame, 59.94 frames/second. The aspect ratio is 16:9 and the horizontal line rate is 47.203 kHz, three times that of NTSC. The nominal video baseband signal bandwidth is 34 MHz. (With a Kell factor of 0.9 for sampling in the horizontal direction and an approximate sampling rate of 75.3 MHz, the nominal video bandwidth equals 0.9 x 75.3/2 = 33.9 MHz.)

Transmitter and receiver signal processing are performed on square pixels in a 720 line by 1280 pixel array. In the studio, an additional guard band of pixels at all four edges is provided to allow for transient effects of processing, analog rise times, production related edge effects and timing tolerances.

75.288 MHz / 2.4 = 31.370 MHz

Square pixels are chosen to facilitate computer interface and special effects processing. The particular numbers offer easy conversion to/from NTSC for simulcast purposes. Conversion to 525 line CCIR 601 requires only a 4:3 interpolation horizontally and 3:2 vertically. In addition, the simple relationship to NTSC provides economical means for designing dual purpose HDTV/NTSC receivers.

Progressive scanning has advantages in video compression and display of motion without line pairing or resolution loss. The 1280 by 720 format is simply related to the Common Image format by a linear factor of 2:3 and is easily extensible to higher-line-number progressive formats when they become practicable. The data rate generated by this format is within the capability of current commercial high definition tape recorder technology.

2.2 Studio Compression

The compression techniques used for the broadcast of DSC-HDTV are easily simplified to produce a 200 Mb/s signal for use in the studio. This signal uses only the intraframe processing, and thus is suitable for all editing and special effects processing. Initial indications are that the quality is also suitable for multiple decoding/encoding as required.

The bit rate is suitable for serial data interfaces and also for video tape recording on D-2-type of VTRs.

2.3 Future Sources

The 1280 x 720 pixel 1:1 progressive format is designed to take advantage of the bit rate and compression techniques available for HDTV broadcasting. However, maximum performance will be attained when the source is at a higher pixel density, so that optimum prefiltering can be applied. Such a source would preferably have a simple ratio of pixel density to the 1280 x 720 format, for example the "common image format" of 1920 x 1080 pixels, scanned progressively, or 1920 x 960 pixels (1050 line progressively scanned format). If it is desirable in the future to maintain the higher pixel numbers in the production studio, this can also be accommodated in an extension of the studio compression format, by decompresing the higher-resolution signal rate into the 200 Mb/s studio compression standard plus a high-frequency residual signal. The final output of editing or special effects can still be recorded using the 200 Mb/s portion of the compressed signal.

SECTION 3
VIDEO CODING

3.1 Introduction

The Zenith/AT&T Digital Spectrum Compatible (DSC)-HDTV system uses a video compression system which is optimized for the terrestrial broadcast environment. A very high compression ratio is achieved in such a way that robust transmission is possible without sacrificing image quality. The system is designed for practical implementation and results in a Decoder which is realizable in a small number of VLSI integrated circuits.

3.2 Pre/Post Processing - Constant Luminance

3.2.1 Encoder Preprocess

The Video Encoder to be supplied to the Advanced Television Test Center (ATTC) is designed to accept RGB signals with SMPTE 240M colorimetry and transfer function. The input RGB signals are first anti-alias filtered and quantized to eight bits per sample. The digital RGB signals are passed through a de-gamma circuit which takes out the SMPTE 240M transfer characteristic, providing linear RGB signals. The linear signals are matrixed to luminance (Y) and color difference signals. The hardware delivered to the ATTC may optionally further preprocess the luminance and color difference signals in general accordance with CIE "$L^*u^*v^*$" equations to provide a uniform-perceptibility-space set of signals to the video compression and transmission circuits [1]. The visual magnitude of coding errors is thus uniformly distributed over the range of possible hues, saturations, and brightnesses, such that the visibility of any coding error is not excessive for any particular color. (In non-uniform systems, errors are typically much more visible in saturated reds than in other colors.) Slight deviations from CIE $L^*u^*v^*$ equations are used to optimize performance where reduced bandwidth of the color difference signals would result in divide-by-zero conditions or excessive quantization error sensitivity. Usage of the CIE $L^*u^*v^*$ equations will be decided before delivery to ATTC.

The color difference signals are decimated by two both vertically and horizontally, resulting in a totally uncompressed bit rate of 994 Mbits/s.

3.2.2 Decoder Postprocess

In the postprocessor, the preprocessor operations are essentially performed in inverse and reverse. The color difference signals are interpolated by two both horizontally and vertically, and linear signals are extracted from the $L^*u^*v^*$ signals (if used). In the test hardware supplied to the ATTC, the linear signals are dematrixed to SMPTE 240M RGB signals. There is a final transfer function stage to restore the 240M transfer characteristic to R, G, and B. Because the luminance and color difference signals are linear at the matrix and the gamma correction is applied afterwards, the decoder is essentially display independent and the design may be modified to match any display primaries or transfer function.

3.3 Video Compression Approach

Three basic types of redundancy are exploited in the video compression process. Motion compensation removes temporal redundancy, spatial frequency transformation removes spatial redundancy, and perceptual weighting removes amplitude redundancy by putting quantization noise in less visible areas. [2]

Temporal processing occurs in two stages. The motion of blocks from frame to frame is estimated using hierarchical block matching. Using the motion vectors, a displaced frame difference (DFD) is then computed which generally contains a small fraction of the information in the original frame. The DFD is transformed using a two dimensional discrete cosine transform (DCT) prior to removal of the spatial redundancy. Each new frame of DFD is analyzed prior to coding to determine its rate versus perceptual distortion characteristics and the dynamic range of each coefficient. Quantization of the DCT coefficients is performed based on the perceptual importance of each coefficient, the precomputed dynamic range of the coefficients, and the rate versus distortion characteristics. The perceptual criterion uses a model of the human visual system to determine a human observer's sensitivity to color, brightness, spatial frequency and spatial-temporal masking. This information is used to minimize the perception of coding artifacts throughout the picture. Parameters of the Encoder are optimized to handle the scene changes that occur frequently in entertainment/sports events, and channel changes made by the viewer. The motion vectors, compressed DCT coefficients and other overhead bits are packed into a format which is highly immune to transmission errors. In case of transmission errors, the Decoder uses a recovery technique which masks the errors. If a complete loss of signal is detected or the channel is changed, the Decoder switches to a special mode which quickly builds the picture to full quality.

The Video Encoder takes full advantage of the transmission system's ability to switch between the 1 bit/symbol and 2 bit/symbol modes. Depending on scene complexity, an improvement in error performance is achieved by adapting the ratio of 1 to 2 bit symbols in the Encoder. The Encoder will automatically select optimum error performance for each scene.

In choosing the video compression algorithm, the need for a modular architecture and low cost was considered. Particular attention has been given to minimizing the Decoder circuits that will be part of every DSC-HDTV receiver. Some elements of the encoding algorithms that affect picture quality can be altered without requiring modifications to the Decoder. This feature provides an opportunity for future improvement without affecting the installed base of equipment. Some applications may also require an inexpensive Encoder. The DSC-HDTV system has a modular architecture which allows a subset of the Encoder to be used at a much lower cost without requiring a different transmission format (see Section 9.).

3.4 <u>Encoder</u>

The Encoder is shown in Figure 3-1. Motion from frame to frame is estimated using a hierarchical block-matching Motion Estimator. The Motion Estimator produces motion vectors, which are compressed and sent to the output buffer for transmission. Each frame is analyzed before being processed in the Encoder Loop. The motion vectors and control parameters resulting from the forward analysis are input to the Encoder Loop which outputs the compressed prediction error to the Channel Buffer. The Encoder Loop control parameters are weighted by the

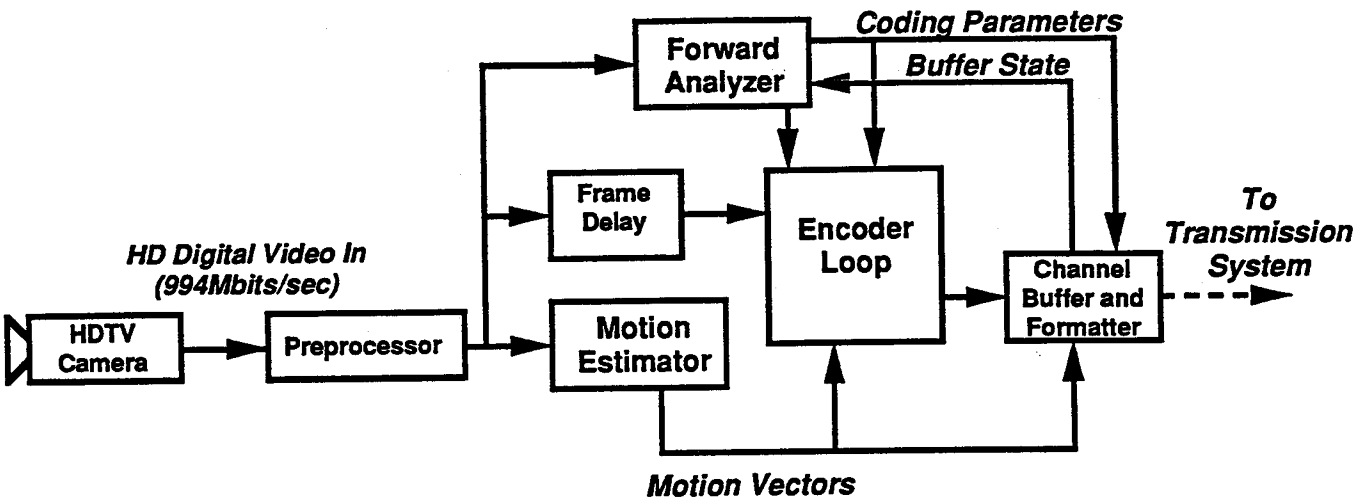


FIG. 3-1 ENCODER

buffer state which is fed back from the Channel Buffer.

In the predictive encoding loop, the generally sparse differences between the new image data and the motion-compensated predicted image data are encoded using adaptive transform coding. The parameters of the encoding are controlled in part by forward analysis. The data output from the Encoder consists of some global parameters of the video frame computed by the Forward Analyzer and DCT coefficients that have been selected and quantized according to a perceptual criterion.

Each frame is composed of a luminance frame and two chrominance frames which are half the resolution of the luminance frame horizontally and vertically. The compression system produces a chrominance bit-rate which is generally a small fraction of the total bit-rate, without perceptible chrominance distortion.

The output buffer has an output rate varying between 9 and 17 Mb/s and has a varying input rate that depends on the image content. The buffer history is used to control the coding parameters so that the average input rate equals the average output rate. The feedback mechanism involves adjustment of the allowable distortion level, since increasing the distortion level (for a given image or image sequence) causes the Encoder to produce a lower output bit rate.

The encoded video is packed into a special format before transmission which maximizes immunity to transmission errors by masking the loss of data in the Decoder. The duration and extent of picture degradation due to any one error or group of errors is limited.

3.4.1 Motion Estimation

Motion is estimated in stages on a block by block basis using the luminance frames only. At each stage the best block match is defined to be that which has the least absolute difference between blocks. The results from one stage are used as a starting point for the next stage to minimize the number of block matches per image. The Motion Estimator is capable of handling large frame-to-frame displacements typical of entertainment and sport scenes. Finally, the block size of motion estimation is adapted spatially to those places in the picture which could have the most benefit within the limit of the compressed motion vector bit rate.

The Motion Estimator compares a block of pixels in the current frame with a block in the previous frame by forming the sum of the absolute differences between the pixels, known as the prediction error. Each block in the current image is compared to displaced blocks at different locations in the previous image and the displacement vector that gives the minimum prediction error is chosen as being the best representation of the motion of that block. This is the motion vector for that block. The end result of motion estimation is to associate a motion vector with every block of pixels in the image.

The Motion Estimator is shown in Figure 3-2. To reduce the complexity of the search, hierarchical motion estimation is used in which a first stage of coarse estimation is refined by a second, finer estimation. The first stage matching is performed on the images after they have been decimated by a factor of two both vertically and horizontally. This reduces both the block size and the search area by factors of four and greatly reduces the size of the motion estimator. A block size of 16H X 8V pixels in the decimated image is used with 1 pixel accuracy. This is equivalent to 32H X 16V pixel blocks and 2 pixel accuracy in the original image. The motion vectors that are generated are passed to the second stage which performs a subpixel accuracy search centered around this coarse estimate. The total search area is 96H x 80V pixels which provides highly accurate motion estimation of very fast moving objects.

The second Motion Estimator stage generates the prediction errors of the 8 x 8 pixel blocks for each location within the search area. The prediction errors of the coarse blocks are derived from the sums of the appropriate small block prediction errors and the coarse motion vectors are refined to subpixel accuracy.

The final stage of the Motion Estimator uses the prediction errors to generate the motion vectors by finding the minimum prediction error for all blocks in every location. The resulting 8 x 8 motion vectors are then passed on to the Motion Vector Selector stage with the associated prediction errors. In addition, 32H X 16V pixel block motion vectors and associated prediction errors are sent to the Motion Vector Selector.

Half pixel motion is deduced by extrapolation of the prediction errors around the region of minimum error. A simple scheme used is to derive the half pixel motion independently horizontally and vertically. A parabola is fit to the three points around the minimum, and the resulting equation is solved to find the position of the minimum of the curve. This process simplifies to solving the

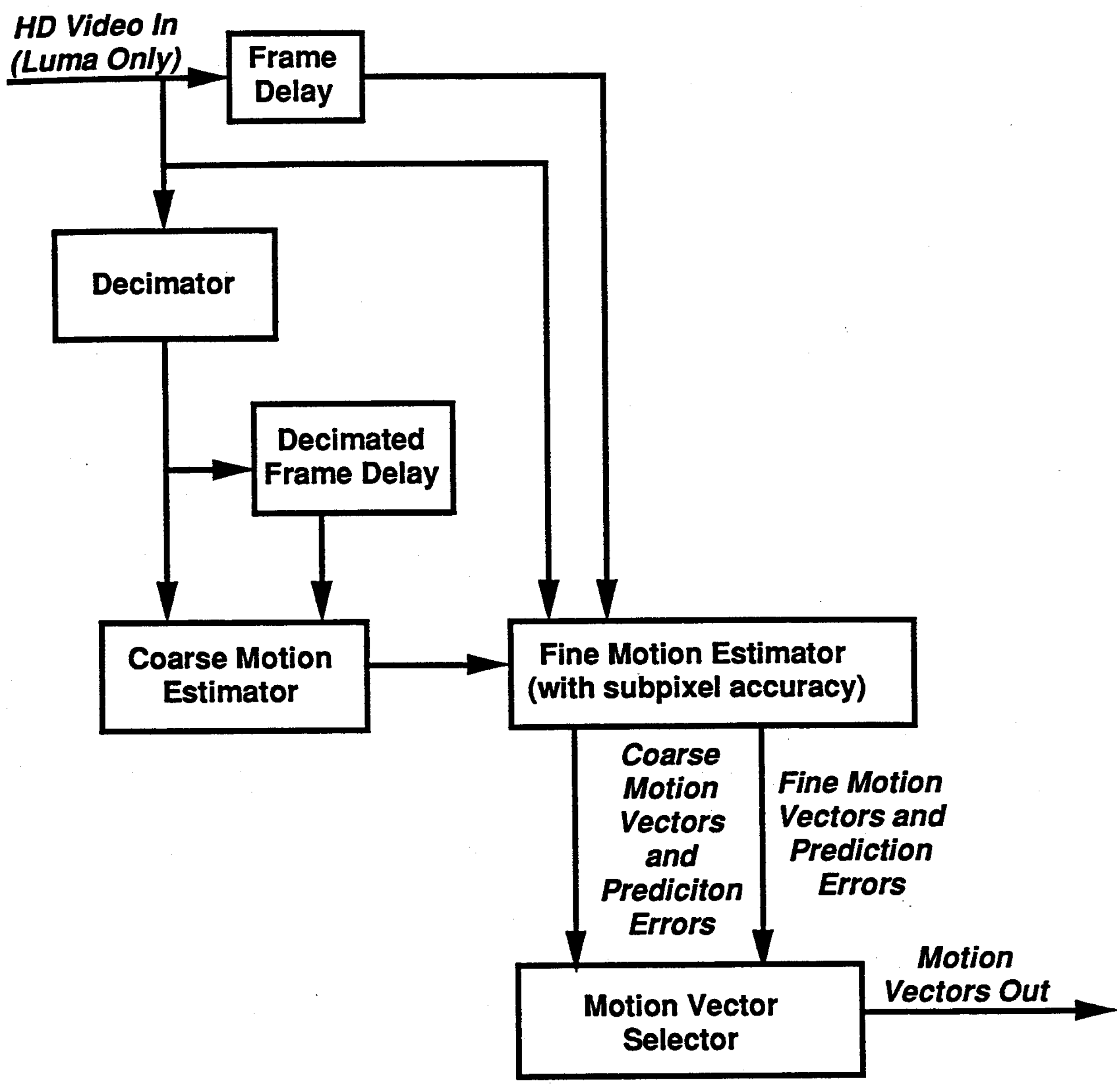


FIG. 3-2 MOTION ESTIMATOR

following

$$x' = x-1/2; ((3p_{x+1} - 2p_x - p_{x-1}) < 0)$$

$$x' = x+1/2; ((3p_{x-1} - 2p_x - p_{x+1}) < 0)$$

$$x' = x; \text{otherwise}$$

where p_x is the prediction error at x, and x' is the deduced half pixel motion vector. The use of this approximation results in a great reduction in the size of the Motion Estimator implementation.

3.4.2 Motion Vector Encoding

Given the motion vectors from the Motion Estimator, the Motion Vector Selector must select the set of motion vectors that will give the best prediction of the next frame while limiting the bit rate of the compressed motion vector data within a range.

This is achieved by sending two resolutions of motion vectors, the first set representing the motion vectors of the 32H x 16V blocks which is unconditionally transmitted, and the second set which represents the 8 x 8 motion vectors. However, not all of the 8 x 8 motion vectors are transmitted, but only those which can be sent within the bit budget remaining after the 32H x 16V motion vectors have been sent.

3.4.2.1 32H x 16V Block Motion Vector Encoding

40 45

The coarse block size of 32H x 16V has been chosen to ensure that, even in the worst case, these 1800 motion vectors can be transmitted within a reasonable budget. This block size is also useful because an integer multiple of them is contained in the 64H x 48V blocks defined as slices for the data transmission format (see Section 3.4.5). Hence for each slice, six motion vectors are sent. In order to improve coding efficiency, five of the six motion vectors are sent as the difference between itself and an adjacent motion vector. If the six motion vectors constituting the slice are numbered as shown in Figure 3-3, then the motion vectors which are

✓ 20 x 15 x 6 = 1800

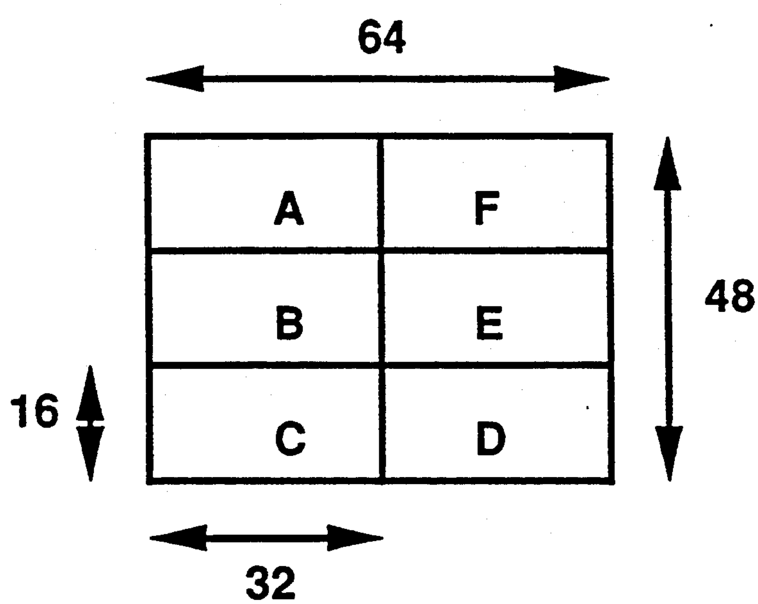


FIG. 3-3 MOTION VECTORS IN A SLICE

sent represent the values of the motion vectors A, B-A, C-B, D-C, E-D, F-E.

3.4.2.2 8 x 8 Motion Vector Encoding

The next stage in the coding scheme is the refinement of some of these larger 32H x 16V block motion vectors by sending the differences between the motion vector of the larger block and those of eight 8 x 8 block motion vectors. As long as this process improves the prediction of the next frame, it is repeated until the remaining bit budget has been consumed.

The criterion for deciding which blocks should be subdivided is the improvement in prediction error using smaller blocks compared with that of the single large block. That is, the improvement, I, is defined as the difference between the prediction error of the 32H x 16V block and the sum of the eight prediction errors of the 8 x 8 blocks. The values of I are sorted to give a list which is ordered by the relative importance of subdividing a motion vector block. Concurrently with the calculation of the improvement, the number of extra bits associated with the subdivision of this 32H x 16V block must be calculated. This consists of the bits required for the eight new 8 x 8 motion vectors. The 8 x 8 motion vectors are encoded using a variable length code word. Once the ordered list of improvements and the bit information has been generated, it is a simple task to select the motion vectors which give the greatest improvement in prediction error until the motion vector bit budget is exhausted.

In addition to the six motion vectors representing the 32H x 16V motion vectors of the slice, information is transmitted indicating which of the 32H x 16V motion vectors have been subdivided into eight 8 x 8 motion vectors. In its simplest form, this would require an extra six bits per slice to represent this information. However, because of the spatial correlation of the regions that require improvement, this information can be compressed by providing one extra bit which indicates whether any of the motion vectors are subdivided, and the other six bits are sent only if this first bit is set.

3.4.3 Encoder Loop

The Encoder Loop is shown in Figure 3-4. The Encoder Loop generates a transformed and quantized displaced frame difference (DFD) using the motion vectors, perceptual thresholds and loop control parameters. The Motion Compensated Predictor applies the motion vectors to the predicted frame which is stored in a buffer. The displaced frame (DF) is scaled by the DF-factor and subtracted from the input frame resulting in a DFD which contains a fraction of the original image. Fast recovery from channel errors and channel changes is facilitated by DF scaling. The DFD is then spatially transformed using DCT and adaptively quantized by the Quantizer Vector Selector. The quantized DCT coefficients and coded selection vectors are passed to the Channel Buffer and Formatter. The DFD coefficient quantization and spatial transform are inverted resulting in a prediction of the DFD as received by the Decoder. This reconstructed DFD is added to the DF resulting in the next reconstructed or predicted frame which replaces the current reconstructed frame in the buffer. [3,4]

The mean of each frame is calculated in the Forward Analyzer and subtracted from each frame before differencing with the zero-mean DF. This results in a zero-mean DFD for greater DCT efficiency. The mean is added back to the reconstructed frame in order to accurately estimate the quantization effects of mean addition prior to display in the receiver.

3.4.3.1 Adaptive Postprocessing

Under certain circumstances, coarse quantization of the transformed DFD coefficients can produce random noise from frame to frame resulting in increased visibility. To avoid this potential flickering of quantized information, postprocessing is applied at the receiver and therefore also in the Encoder prediction loop. At each pixel a weighted average of temporally successive pixels is stored in the reconstructed frame buffer if it differs from the current pixel by an amount less than a threshold. If the pixel differs by an amount larger than the threshold, the new pixel is used without averaging.

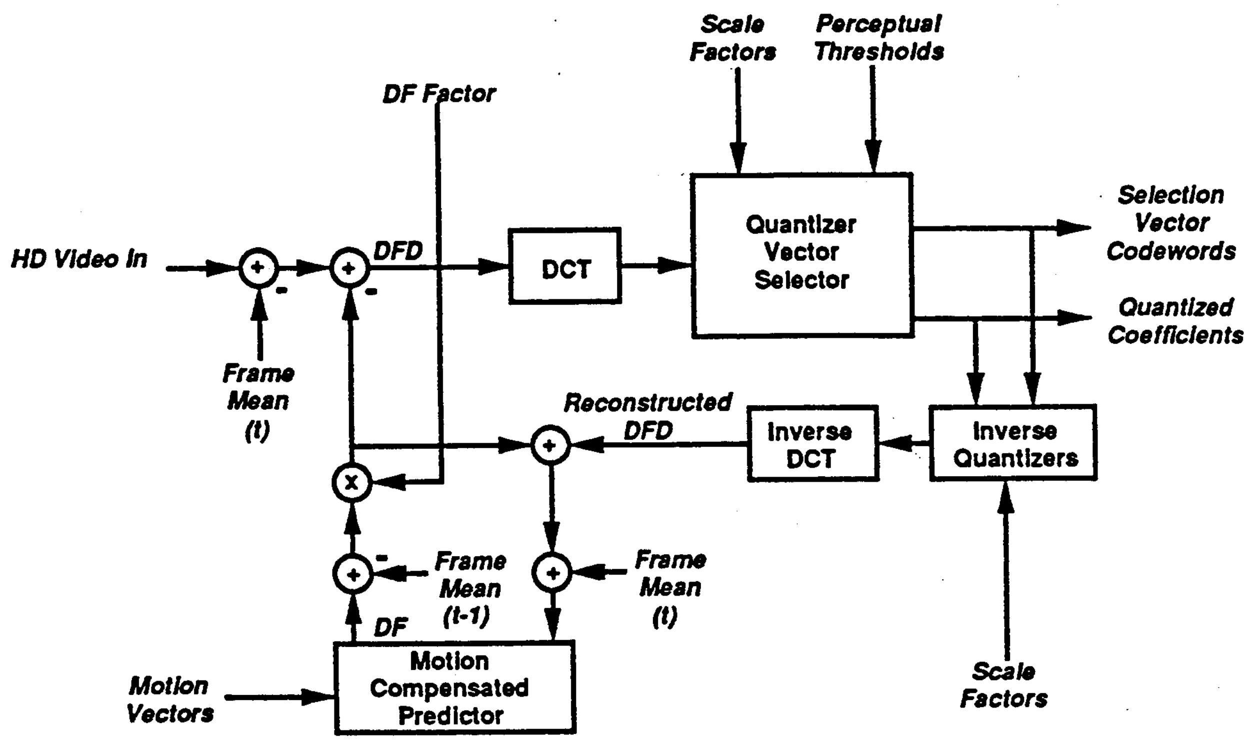


FIG. 3-4 ENCODER LOOP

3.4.3.2 Spatial Transform

8 x 8 blocks of pixels containing the DFD are transformed using a 8 x 8 two dimensional DCT. The DCT implementation is based on an algorithm that minimizes the number of multiplications required and results in an economical implementation.

3.4.3.3 Adaptive Quantization of DCT Coefficients

A principal means of reducing the number of bits needed to represent an image is to control the number of bits used to represent the individual DCT coefficients. The DSC-HDTV system recognizes that DCT coefficients vary in importance based on the limitations of human vision. By adjusting the coarseness of quantization of individual coefficients in local regions of the image, the system can minimize the amount of transmitted information, while retaining the flexibility to apply better coefficient precision where needed. [5]

A set of non-uniform quantizers is used to quantize the coefficients. Each quantizer has a zero level and is adaptively scalable to maximize efficiency given the amplitudes of coefficients. A coefficient may also be dropped or forced to zero. The quantized coefficients are variable length encoded using a codebook matched to each quantizer.

3.4.3.4 Vector Quantization of Selection Patterns

A variety of quantizers is available for each coefficient. However, constraints on the selection of quantizers is needed to accommodate the quantizer selection overhead. Within the bit budget, Vector Quantization is used to represent the possible combinations or patterns of quantizers which can be applied to a given 8 x 8 block of coefficients (i.e. a set of coefficients from a block of 8 x 8 pixels). Coding efficiency is achieved by variable length coding and transmitting the index associated with a given quantizer selection pattern instead of the pattern itself.

The Quantizer Vector Selector is shown in Figure 3-5. Unquantized DCT coefficients are input to the Quantizer Selector and buffered to accommodate Quantizer Selector Delay. Each entry in the Quantizer Selection Vector Codebook

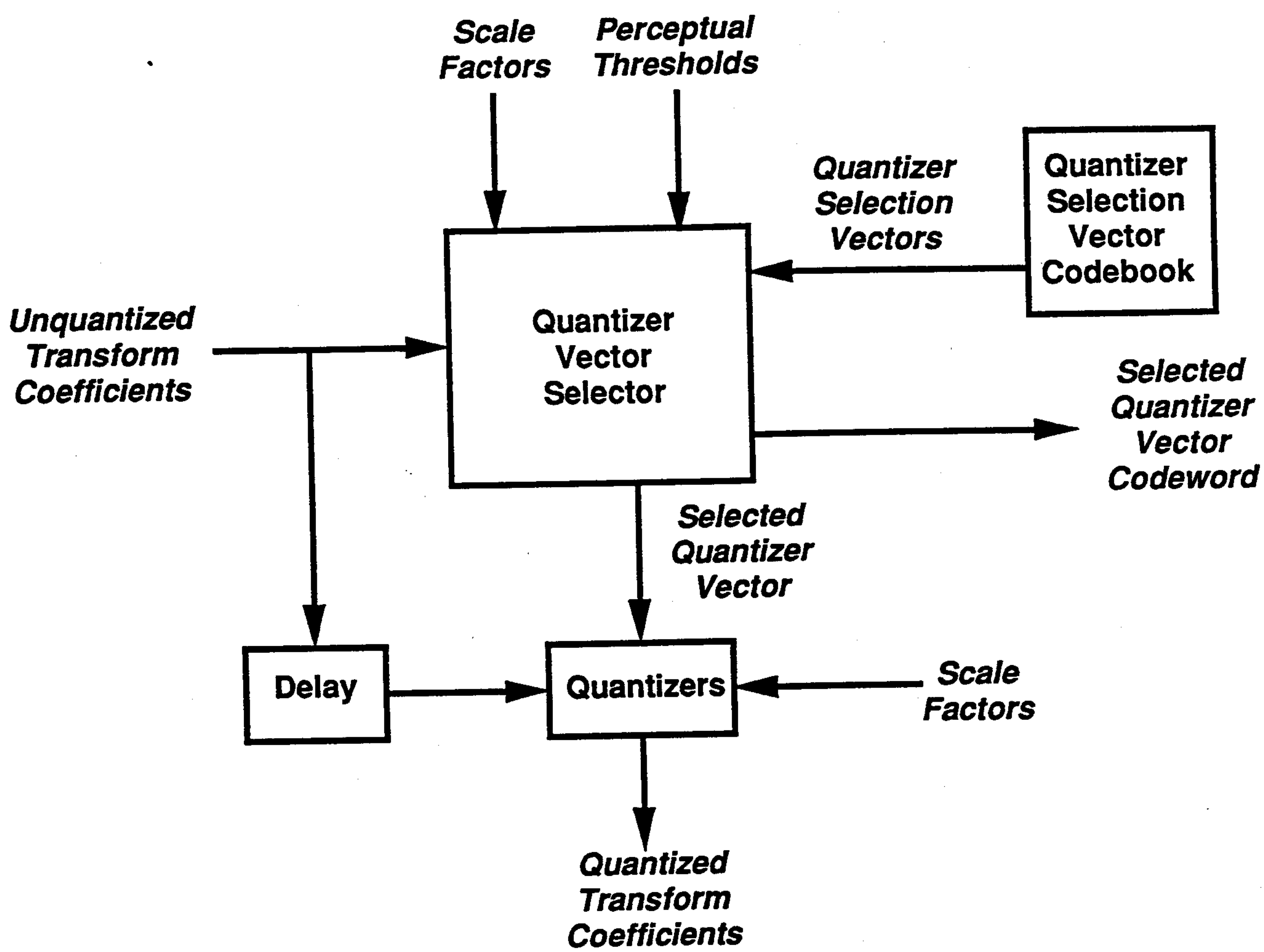


FIG. 3-5 QUANTIZER VECTOR SELECTOR

is retrieved in sequence and applied to a given block of coefficients. Each of the 64 elements of a selection vector contains a quantizer code which selects one of three quantizers (or drop) which are applied to a set of coefficients. A given set of luminance coefficients is composed of the four coefficients from a 16 x 16 superblock (2 x 2 DCT blocks) which have the same coefficient number (0, 0...7,7) where 0,0 is the DC coefficient). A given set of chrominance coefficients is composed of 6 coefficients from a 16H x 24V superblock (2H x 3V DCT blocks) which have the same coefficient number. In other words, a 16 X 16 pixel or coefficient sample super block is the domain of a given selection vector for luminance images and a 16H X 48V super block is the domain of a given selection vector for chrominance images. The luminance and two chrominance frames are quantized using separate codebooks and quantizers. The luminance codebook contains less than two thousand vectors and each chrominance codebook contains less than five hundred vectors. The codebooks are organized differentially such that only differences between successive vectors are computed. This results in a great reduction in the number of computations and the size of the implementation. In the Quantizer Vector Selector each coefficient set is either quantized or dropped and variable length encoded in order to compute the number of bits per coded coefficient set. The quantization error is compared to a perceptual error threshold, and the result is summed to produce the selection error for the vector. The optimal vector is selected by considering both selection error and bit-rate.

The selected vector is then applied to the buffered unquantized coefficients via the quantizers as shown in Figure 3-5. Scale factors are computed in the forward analyzer and applied to the quantizers to increase quantization efficiency. The vector selects which quantizer to apply to each coefficient. The quantized coefficients are sent to the inverse quantizers and inverse transform in the encoding loop, and to the Channel Buffer and Formatter for transmission after variable length encoding.

3.4.4 Forward Analyzer

The Forward Analyzer is shown in Figure 3-6. The original frame is transformed and analyzed by the Perceptual Weight Calculator which passes weights to the Buffer Controller. The perceptual thresholds are computed by the Buffer Con-

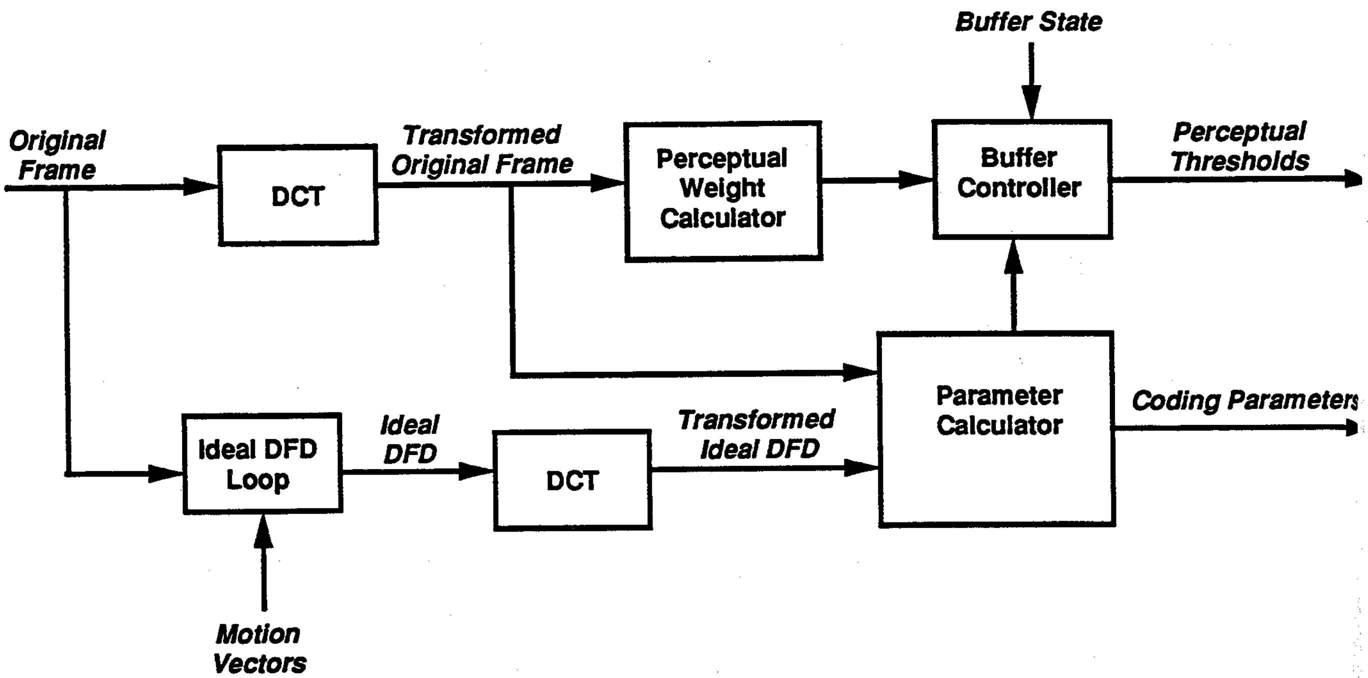


FIG. 3-6 FORWARD ANALYZER

AT&T

ZENITH

troller using the weights and the buffer state.

The Forward Analyzer also approximates the operation of the Encoding Loop in order to optimize Encoding Loop control parameters. These parameters depend on the results of analysis of the entire original frame and ideal DFD and would be available one frame too late if computed in the loop. Examples of control parameters are the image mean and the quantizer scaling factors.

3.4.4.1 Perceptual Criterion

The concept of perception-based coding depends on matching the coding to the characteristics of the human visual system (HVS). Considering the coding artifacts and the ways in which coding can affect their distribution, it is apparent that if artifacts are concentrated in localized regions of the image, the coded image distortion will be more visible. Conversely, the visibility of coding artifacts will be minimized if the coding distortions are uniformly distributed across the image. The use of perceptual thresholds results in an allocation of coding distortions so that the visibility of distortions is minimized.

The Perceptual Weight Calculator is shown in Figure 3-7. The unpredicted picture information represented by the DFD is transformed into DCT coefficients that are transmitted with varying precision. The precision needed for a particular coefficient is determined by a local perceptual threshold. For every coefficient the Perceptual Weight Calculator coupled with the Buffer Controller produces a threshold value. The coarseness of quantization (number of bits used for sending the coefficient value to the Decoder) depends on the local perceptual weights and on a global target distortion related to the time history of the channel buffer fullness. A lower perceptual threshold value implies that a low quantization error is allowed and more bits will be allocated for the corresponding DCT coefficient.

Separate luminance and chrominance (both U and V) perceptual thresholds are generated for every coefficient in each frame. The perceptual thresholds are not transmitted, but are used to ensure that the information transmitted is allocated in an optimum manner, minimizing perceptible artifacts and maximizing picture quality.

The set of perceptual thresholds, one for each coefficient sample, provides an estimate of the relative visibility of the coding distortion. Note that these

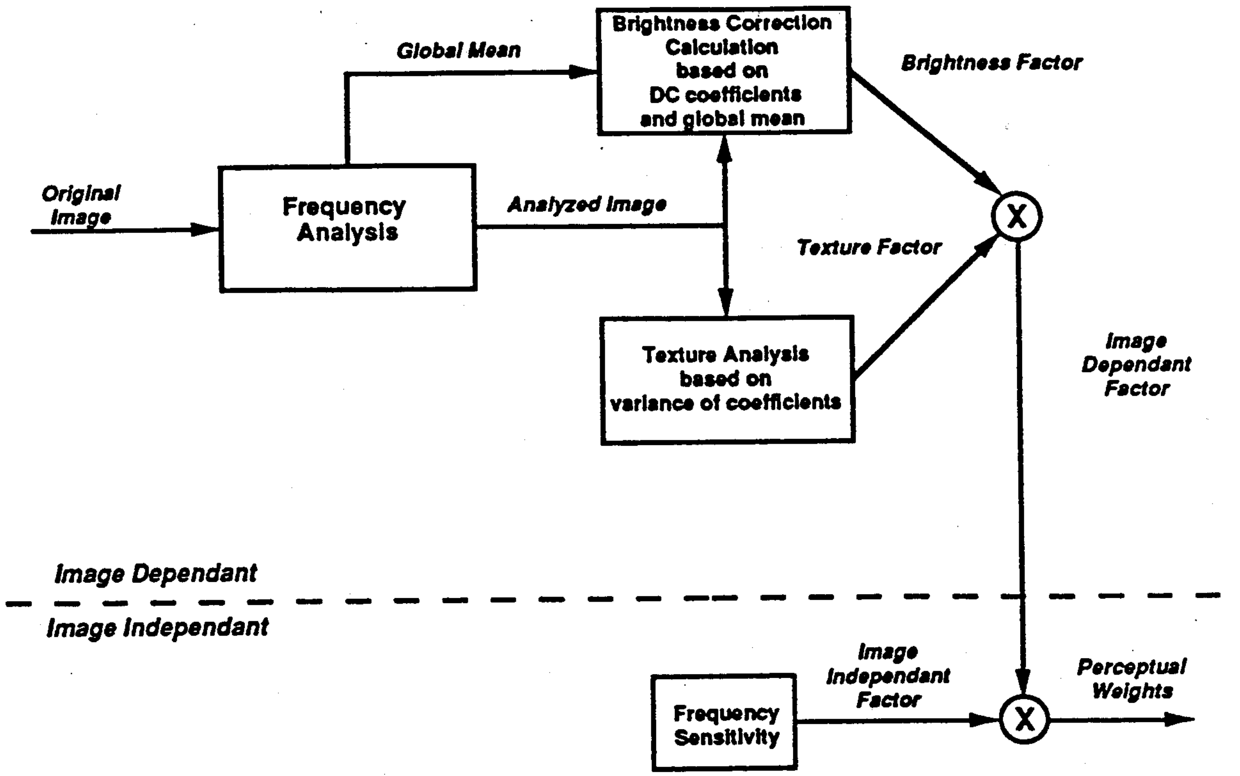


FIG. 3-7 PERCEPTUAL WEIGHT CALCULATOR

AT&T

ZENITH

thresholds depend on the content of the original image and, therefore, the bit allocation algorithm can adapt to variations in the input. The Perceptual Weight Calculator implements a model of the HVS that has been optimized to work with this encoder architecture.

Frequency sensitivity in the HVS model exploits the fact that the visual system modulation transfer function (MTF) is not flat. The MTF indicates that more quantization error can be tolerated at high frequencies than at low frequencies. Therefore, the perceptual thresholds for higher frequency coefficients are larger than those for low frequency coefficients.

For a flat field stimulus, the HVS has varying sensitivity to distortion which is related to the brightness of the flat field. This is called Contrast Sensitivity. By examining the local brightness, the perceptual thresholds can be locally adjusted to account for this property.

Up to this point, the HVS response to flat field inputs of varying brightness levels has been modeled. Since most images of interest are not flat fields, a model of spatial masking has been incorporated which further adjusts the perceptual thresholds based on the amount of local texture present at each location in the input. Our definition of texture is the amount by which the input deviates from a flat field. This results in Spatial Masking being proportional to the AC energy at each location.

3.4.4.2 Channel Buffer Control

In general, the buffer controls the coding algorithm by feeding back the history of the buffer state to the Forward Analyzer. The effect of the buffer state signal is to slow the information rate out of the Encoder when the buffer fullness is high, and to allow an increase in information rate from the Encoder when the buffer is less full. The buffer state is controlled by the global target distortion level. If the target distortion level is increased (to decrease the buffer fullness), more distortion is allowed.

The history of the buffer state and target distortion is stored and is also used in the current target distortion calculation. The Forward Analyzer maintains a certain average buffer fullness without increasing the target distortion beyond perceptible limits. If a scene change is detected, the buffer fullness and the target distortion are allowed to increase for a few frames and the DF factor is decreased to reduce the amount of erroneous prediction (DF) of the DFD.

The buffer state is transmitted to the receiver for each frame. The size of the Decoder Channel Buffer and latency in the Decoder are minimized by resetting the buffer state upon receiver startup caused by channel changes or severe loss of transmitted data.

3.4.5 Channel Buffer and Formatter

As shown in Figure 3-8, the Channel Buffer and Formatter receives an assortment of partially encoded video data from the rest of the Encoder and packs it into the compressed frame buffer. Since an error at a given point in a variable length coded data stream causes data after that point to be lost, the variable length codes are packed into slices which correspond to fixed 64 H X 48 V pixel regions in the original image. To avoid channel error propagation beyond a slice, the slice boundaries are periodically marked in the data stream to allow a restart of variable length decoding. The encoded video data stream is divided into fixed length data segments for transmissions(see Section 6.1.2). Since the variable length slices do not correspond to the fixed length segments, each segment contains header information which marks the first slice boundary in the segment.

The motion vectors, quantizer selection vectors, and quantized coefficients are variable length encoded. Various other coding parameters are fixed length encoded and put into special global segments by the segment formatter. The following coding parameters are contained in the global segments:

- scale factors
- luminance and chrominance mean values
- DF-factor
- buffer fullness
- frame number
- frame number of frame following this global segment (for synchronization)

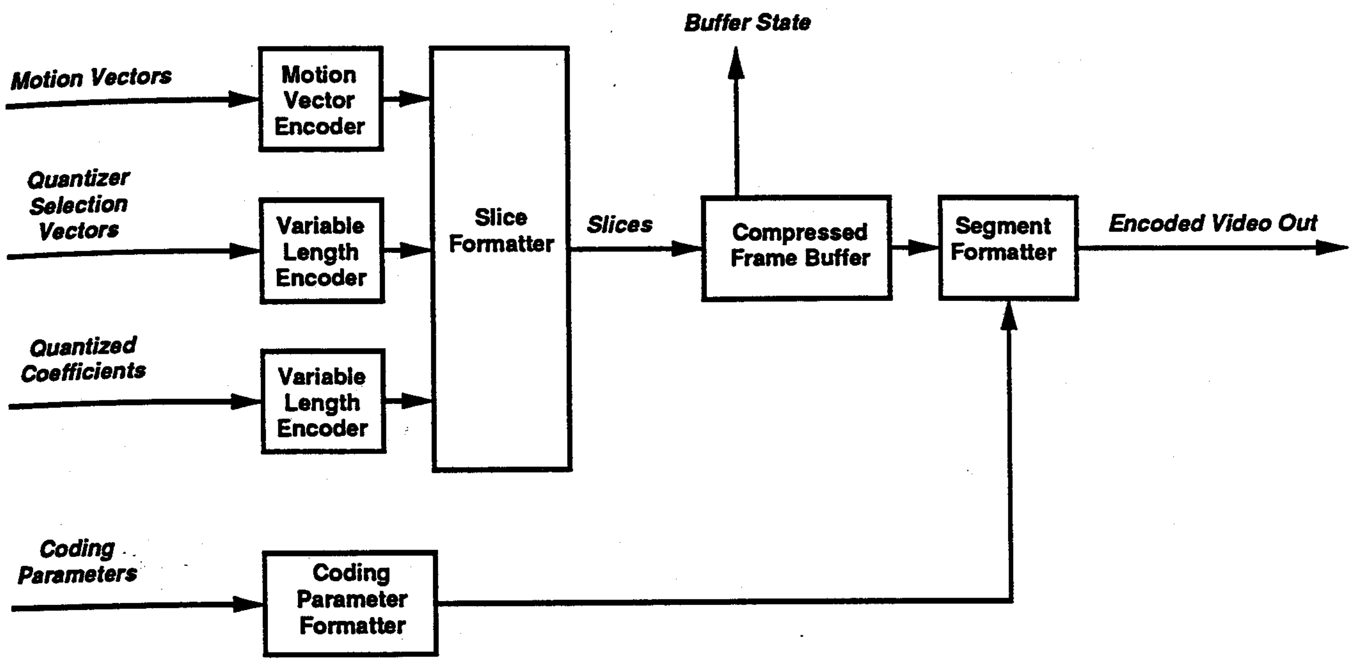


FIG. 3-8 CHANNEL BUFFER AND FORMATTER

The state of the compressed frame buffer is calculated periodically and relayed back to the Forward Analyzer. The Forward Analyzer alters the perceptual thresholds to prevent overflow or underflow of the buffer which is being emptied into the channel via the transmission system. In case underflow is unavoidable, special underflow segments are transmitted which contain pseudo-random data.

3.4.5.1 Variable Length Coding

The data representing motion vectors, quantizer selection patterns, and DCT coefficients, are statistically non-uniform. Usually, the data are clustered, and the probability distributions can be estimated from analysis of real scenes. The use of variable length codes takes advantage of this statistical non-uniformity by assigning short code words to the most frequent values, and assigning longer words to less frequent values.

3.4.5.2 Bi-Rate Coding

The main criticism of digital transmission systems is that the error rate as a function of carrier-to-noise ratio (C/N) increases sharply at the noise threshold resulting in the complete loss of transmitted data. HDTV viewers near the limits of the service area would suffer from abrupt loss of picture resulting from a small decrease in C/N. This situation is considered to be unacceptable.

The DSC-HDTV transmission system time division multiplexes between 1 bit per symbol (binary) and 2 bits per symbol transmission resulting in more robust transmission (or lower noise threshold) for the binary portion of the video data (see Section 6.1.3). The ratio of 1 versus 2 bits/symbol transmission is adaptively selected in the Encoder as part of the buffer control system. The majority of high definition video source material can be compressed and transmitted using a moderate percentage of binary transmission without any loss of picture quality. The most challenging source material may result in a slight reduction in picture quality to achieve more robust transmission. However, protection of the highest priority video segments results in a significant improvement in error masking capability.

Although the Encoder automatically optimizes between picture quality and robustness, interactive control of this balance will be possible at the point of encoding, such as at the network source.

Each Data Segment within a given Data Field (see Section 6.1.2) is assigned a priority. The buffer control determines the number of segments which will be transmitted using the 1 bit/symbol (binary) transmission mode. These highest priority segments are selected for binary transmission for each Data Field. The binary segment selection information is packed into the global segment as a "Transmission Bit Map" (see Section 6.1.3) with the coding parameters for each frame. The global segment heads each Data Field, is always binary, and is repeated to ensure correct transmission. The receiver uses the "Transmission Bit Map" for decoding the mixture of binary and 4-level video Data Segments.

3.5 Decoder

The Decoder is shown in Figure 3-9. The compressed video data enters the Buffer which is complementary to the compressed video buffer at the Encoder. The Decoding Loop uses the motion vectors, DCT coefficient data, and other side information to reconstruct the HDTV images. The Postprocessor converts the luminance and chrominance frames into RGB for display. Channel changes and severe transmission errors are detected in the Decoder causing a fast picture recovery process to be initiated. Less severe transmission errors are handled gracefully by several algorithms depending on the type of error, as described below.

The fraction of the original image in the DFD enables the reconstruction of a usable quality image from a single compressed frame for fast recovery from channel changes. This also facilitates a simple implementation of VCR special modes such as forward and reverse speed search.

Processing and memory in the Decoder are minimized. Processing consists of one inverse DCT and a variable length decoder which are realizable in a few VLSI IC's. Memory in the Decoder consists of one full frame and a few compressed frames (about 1.6 Mbytes).

3.5.1 Channel Buffer and Deformatter

After Reed-Solomon error detection and correction, the encoded video data stream enters the segment deformatter as shown in Figure 3-10. The error correct-

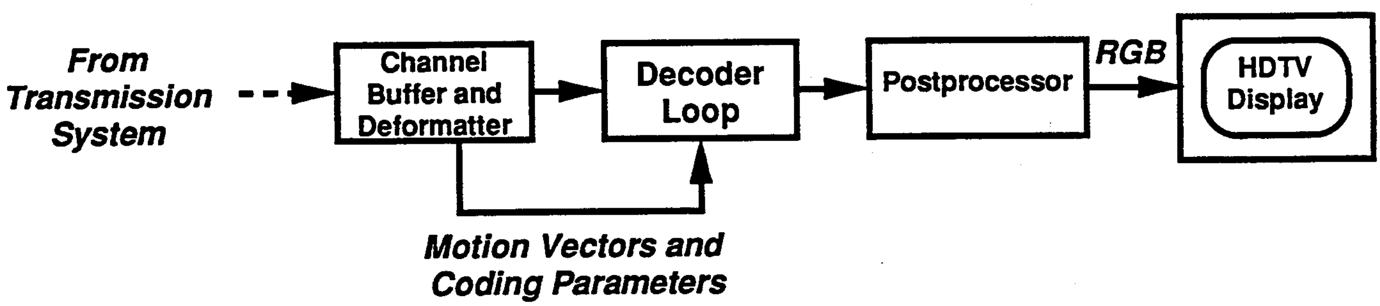


FIG. 3-9 DECODER

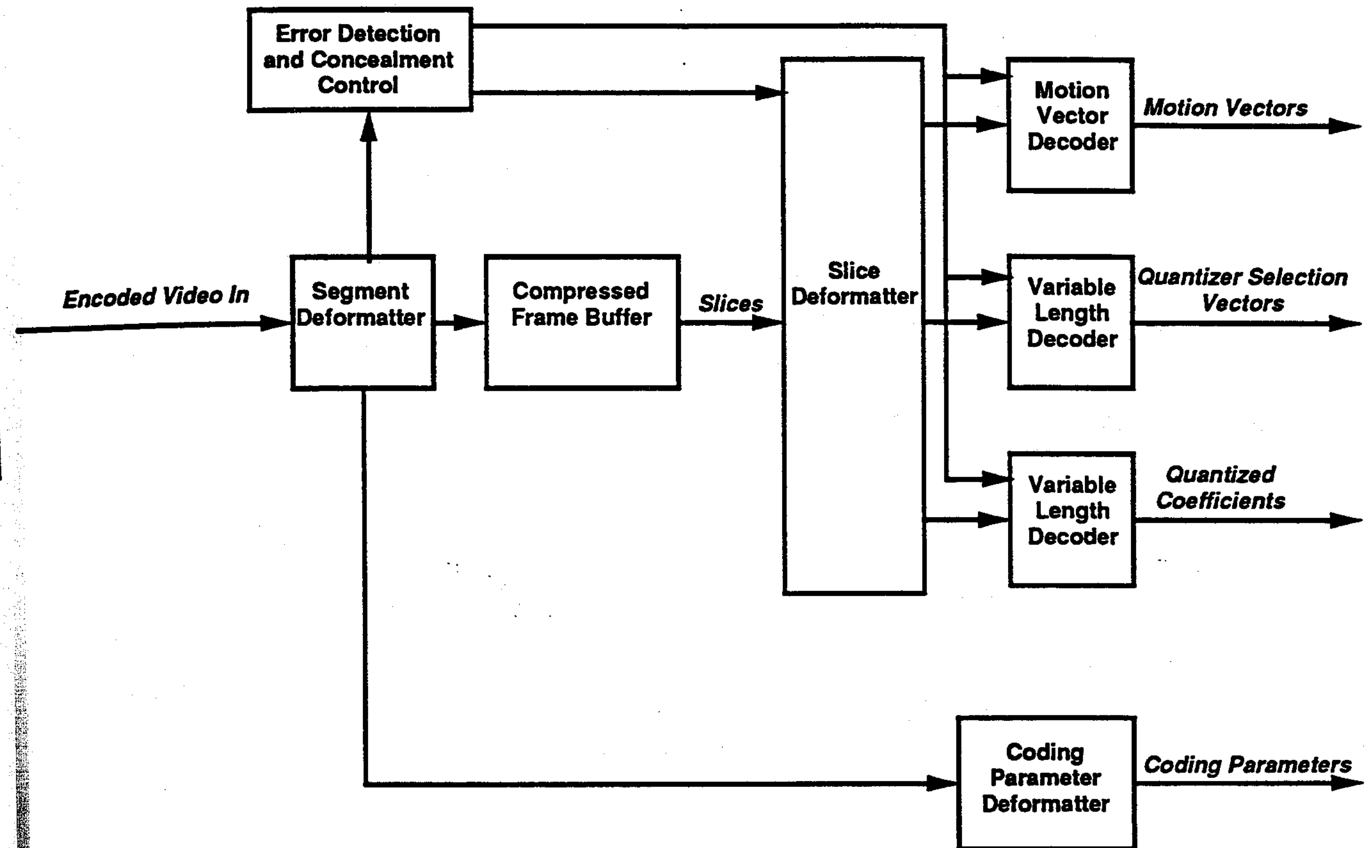


FIG. 3-10 BUFFER AND DEFORMATTER

ing system indicates the presence of detected and corrected errors in each segment. The error detection and concealment controller causes erroneous data to be replaced by estimated values during variable length decoding. Entire slices are replaced if a slice marker is lost or an insufficient portion of a particular slice is recovered. If one of the global data segments is lost, the duplicate global segment is used. by what

From where? After the beginning of each slice is determined in the slice deformatter, the motion vectors, quantizer selection vectors and quantized coefficients are variable length decoded. The coding parameters are extracted from the global segments for use in the Decoder Loop.

3.5.2 Decoder Loop

The Decoder Loop is shown in Figure 3-11. Quantized coefficients, quantizer selection vector codewords, motion vectors and loop control parameters are received from the Channel Buffer and Deformatter. Each quantizer selection vector codeword is used to look up a quantizer selection vector which is an array of quantizer selection codes. One of the inverse quantizers is applied to each quantized coefficient as selected by the corresponding selection code in the vector. The quantized coefficients are received in a fixed order which corresponds to the selection code order in the vectors.

The fully decoded coefficients are inverse transformed producing the DFD. The DFD is added to the DF resulting in the frame which is displayed. The frame mean is added prior to display and storage in the frame buffer. After some portion of the frame is written into the buffer, the motion vectors are applied to generate the next DF.

3.6 Film Mode Optimization

The Encoder buffer control automatically detects the presence of 24 frame/sec or 30 frame/sec scene material by analysis of ideal DFD energy in the Forward Analyzer. In this case, an alternate buffer control algorithm is used which takes full advantage of repeated frames in the source and minimizes variations in target distortion between repeated frames. If film is detected, all video segments

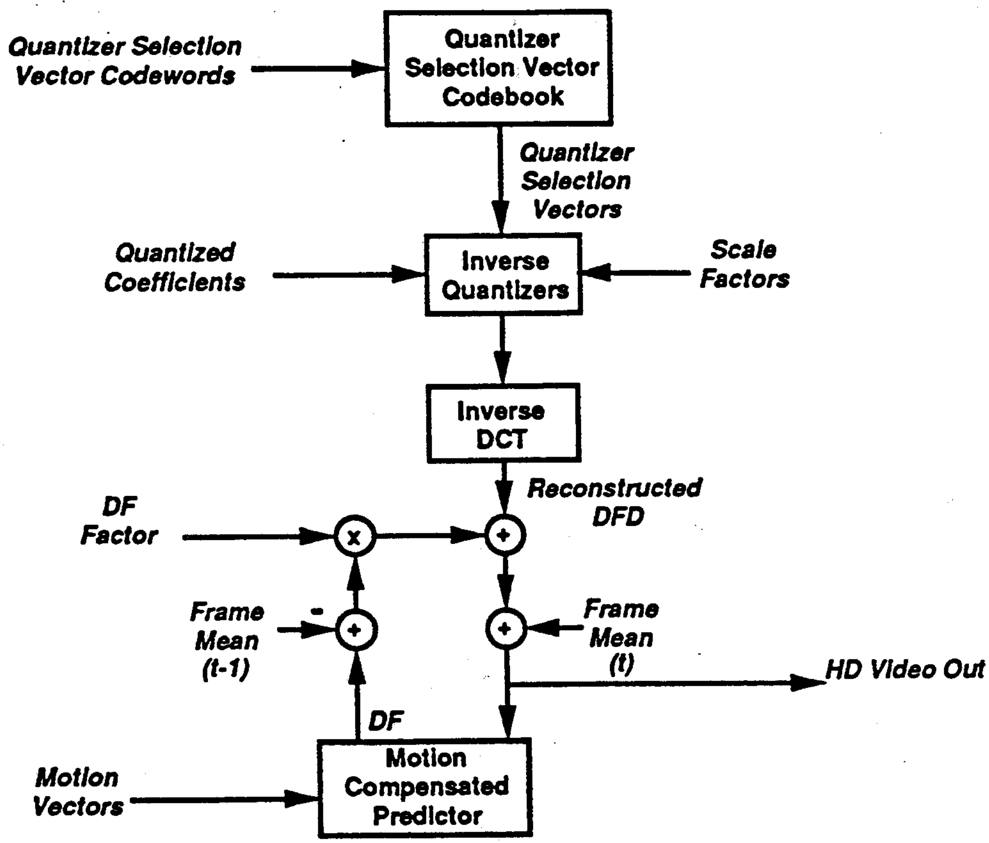


FIG. 3-11 DECODER LOOP

will undergo binary transmission for maximum coverage area.

3.7 Video Coding Summary

The video compression algorithm includes a unique set of features that maximizes picture quality within the available channel bandwidth. The principal features are:

- Hierarchical, subpixel motion estimation within a motion vector budget
- Forward analyzer to determine coding parameters of each new frame
- Handling of scene changes with no perceptible distortion
- how quick? Quick buildup of picture after channel change by the viewer
- Adaptive post-processing of reconstructed images to improve picture quality
- Perceptual quantization and dropping of DCT coefficients
- Vector quantization of quantizer selection patterns
- Smooth control of quantization based on buffer fullness history
- Automatic film mode
- Support for VCR special functions
- Randomizing and limiting effects of transmission errors
- Adaptive bi-rate channel coding for maximum coverage area
- Simple receiver

Transparent image quality is achieved using motion compensated transform coding coupled with a perceptual criterion to determine the quantization accuracy required for each DCT coefficient. The combination of a sophisticated encoded video format, advanced bit error protection techniques and adaptive channel coding results in a highly robust reception and decoding of the compressed video signal.

The DSC-HDTV Video Encoder is optimized to simplify the Decoder. The Motion Estimator, Forward Analyzer, Quantizer Vector Selector, and buffer state control are all functions which exist only in the Encoder. An additional attribute of the DSC-HDTV Video Encoder is the capability to improve the Encoder algorithm in the future without modifying the Decoder. The Decoder is realizable in a small number of VLSI IC's. Frame memory and processing in the receiver are minimized while maintaining transparent image quality.

SECTION 4
AUDIO CODING

Audio capacity is provided by four independent 125.874 kbit/second channels multiplexed with the video data. An audio sampling rate identical to the DSC-HDTV horizontal scan rate of 47.203 KHz has been chosen to prevent low frequency beats in the receiver. The channels are independently coded. This provides a main stereo pair which is appropriate for carrying matrix-coded surround sound as well, plus two more channels which are suitable for high-quality Second Audio Program or any other desired usage.

The main stereo pair is carried entirely by two-level transmission (by W1 data, as defined in Section 6.1.2), so that it is a robust transmission receivable under even extreme conditions of signal impairment. The secondary channels are carried as four-level data (W2 data), receivable under all but the most unusual of conditions, during which reception can revert to the main pair.

Each data field includes 1050 error corrected bytes of audio data. Half of this data is sent as two-level, 1-bit per symbol signals for the main stereo pair (using 525 x 8 = 4200 symbols per data field). The data for the second program is sent as four-level signals, requiring an additional 2100 symbols per data field.

The audio data is compressed by the Dolby AC-2 process, which is described in detail in the Appendix. The Dolby process has its own error correction and masking capabilities, which are then concatenated with the Forward Error Correction of the DSC-HDTV transmission system.

SECTION 5
ANCILLARY DATA

The DSC-HDTV system format provides two separate channels for ancillary data. The first channel of 126 x 8 x 29.97 = 30.21 kbits/second is sent as two-level (W1) data and the second channel of 1596 x 8 x 29.97 = 382.7 kbits/second is sent as four-level (W2) data. (See Section 6.1.2 for definitions of W1 and W2.) The sum capacity of the two ancillary data channels is 412.867 Kbits/sec. The two-level data channel has a 6 dB better noise margin than the four-level data channel and should be assigned to ancillary services which can benefit most from the extra robustness. Currently, bit assignment of data to specific services is undefined. For transmission, the ancillary data channels are time-multiplexed, error-protected, and interleaved along with the video and audio data.

In addition to captioning for the hearing-impaired and Teletext, other possible applications for this data are testing, control and cuing, channel/program identification, alert/emergency, header and descriptor information, customer addressing and encryption.

One of the scenarios identified for the early stages of HDTV broadcasting is the "pass-through" stage. In this stage a video coded program is supplied to the local broadcaster, for example, by satellite. The local broadcaster has a HDTV transmitter but only a minimum of other HDTV equipment. Ancillary data transmission can be made a part of the pass-through operation and the broadcaster can then insert messages, for example alert/emergency information. Data insertion can be realized with very modest equipment cost.

Due to hardware constraints, the prototype DSC-HDTV system will not allow testing of all the ancillary data. For ATTC testing, the prototype hardware will supply a full-capacity two-level (W1) channel of 30.21 Kbits/sec and a four-level (W2) channel of 347.4 Kbits/sec.

SECTION 6
DSC-HDTV TRANSMISSION SYSTEM

6.1 Transmission Format.

6.1.1 Data Frame.

The Video Encoder generates a maximum of approximately 17 Mb/s which along with audio data, ancillary data and synchronization are transmitted in a 6-MHz television channel. This is achieved in an NTSC-like transmission signal format, shown in Figure 6-1.

To avoid confusion between DSC-HDTV video source and display format on the one hand, and DSC-HDTV transmission signal format on the other, some new terms are used for the latter. These terms are summarized in the immediately following subsection 6.1.2.

Note in Figure 6-1 that one Data Frame corresponds to one NTSC frame, one Data Field to one NTSC field and one Data Segment to one NTSC horizontal line.

The interrelationships between the numbers in Figure 6-1 are easily verified with the help of Table I in subsection 6.1.2.

Note also in Figure 6-1 that all symbols are protected by the Reed-Solomon (RS) Code except the Synchronizing (Sync) Interval symbols and the Test Data symbols. The latter carry system test data. The Sync Interval symbols are not protected by the RS Code because sync detection must take place before error correction. Both Data Segment Sync and Data Field Sync have their own error preventing redundancy as explained below and in Section 6.3.

6.1.2 Transmission Signal Parameter Definitions.

Data Segment = A sequence of 684 symbols of total duration of 63.56 microseconds (corresponding to the duration of one NTSC horizontal line period). The terms "Horizontal Line" or "Line" are reserved for the DSC-HDTV display parameter and have a duration of 1/3 of a Data Segment.

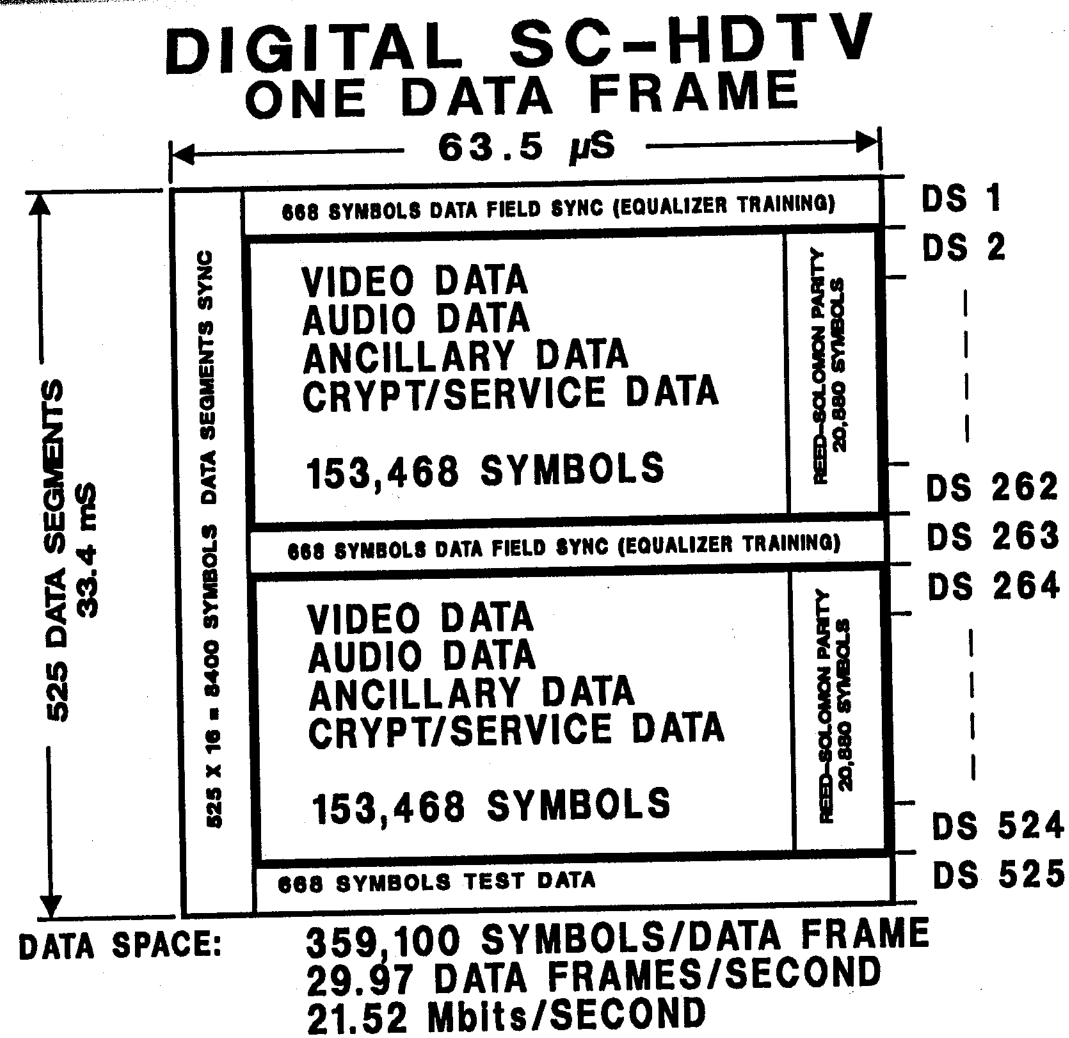


FIG. 6-1

AT&T

ZENITH

Data Segment Sync Interval = A group of 16 (out of 684) symbols of a Data Segment. The first four symbols constitute Segment Sync proper; the next 12 encode DC offsets. The Data Segment Sync Interval is 1.49 microseconds.

Data Field = A group of 262 or 263 data segments (averaging 1/59.94 second duration, corresponding to one NTSC "field").

Data Frame = Two successive Data Fields (corresponding to one NTSC "frame"; duration 1/29.97 seconds).

Total Symbol Rate = 10.76 Msymbol/sec.

Maximum Total Bit Rate = 21.00 Mbit/sec.

W1 Data = Data transmitted at 1 bit/symbol

W2 Data = Data transmitted at 2 bits/symbol

TABLE I

1 Symbol = 1 or 2 Bits
1 Byte = 8 Bits
Symbol Rate = 10.76 Msymbol/second
= 179,550 Symbols/Data Field
= 684 Symbols/Data Segment
Symbol Interval = 92.9 nanoseconds

NTSC Horizontal Line Frequency, f_H = 4.5/286 MHz = 15.734 kHz*
Video Data Clock Frequency $f_d = 4 \times 7 \times 9 \times 19 \times f_H$ = 75.34 MHz
Transmission Data Clock Frequency $f_t = f_d/7$ = 10.76 MHz

*f_H always only refers to NTSC horizontal line rate.

6.1.3 Bi-Rate Transmission.

As previously mentioned in Section 3, DSC-HDTV has a self-adapting variable transmission bit rate. Some data are transmitted at 1 bit/symbol (defined to be of Weight-one or "W1 data") and other data are transmitted at 2 bit/symbol ("W2 data").

W1 data are received with a 7 dB lower carrier-to-noise (C/N) threshold and a 6 dB lower carrier-to-interference (C/I) threshold than W2 data; W1 data are thus more robust than W2 data.

Some data are always transmitted as W1 data, and some always as W2 data but the data representing the video signal are generally transmitted partly as W1 data and partly as W2 data.

TABLE II

SIGNAL	Weight	Symbols/Field Field 1	Field 2
Transmission Bit Map (TBM), Video Control	W1	2,352	2,352
Audio	W1	4,200	4,200
Ancillary	W1	504	504
Parity (Error Correction)	W1	960	960
Video, 240 Segments	W1/W2	141,120	141,120
Parity (Error Correction)	W1/W2	19,200	19,200
Audio	W2	2,100	2,100
Ancillary	W2	3,192	3,192
Test (One Segment/Field 2)	W2		668
Parity (Error Correction)	W2	720	720
Data Segment Sync	W1	4,192	4,208
Data Field Sync	W1	668	668
TOTAL		179,208	179,892

Space for 682432 symbols
− (179,208 + 179 842) = 323,332 − w/o def
320 640
⇒ @ 21 MHz send all video as W2
∴ Can send 1/3 as W1 2/3 as W2
141,120 + 19,200 = 160,320 = 9 609 590 mbps
61.
= 20,752 207 mbps
371 792 ⇒ 11, 142,617 Mbps

Table II shows how the data are allotted over Data Field 1 and Data Field 2 and indicates whether the data are W1, W2 or either.

The first line in Table II, Transmission Bit Map (TBM), refers to what in Section 3.4.5.2 is identified as "Binary Segment Selection Information".

The symbol rate is a constant 10.76 Msymbol/sec but the bit rates are variable. The maximum bit rate is found from the data in Table II by assuming that all variable data are W2. Partial addition of the columns yields 170,866 symbols/field (average of Field 1 and Field 2) of W2 data and 8,684 symbols/field of W1 data. This totals to a bit rate of:

$$(170{,}866 \times 2 + 8{,}684) \times 59.94 = 21.004 \text{ Mb/sec}$$

A similar calculation yields a minimum bit rate of:

$$(173{,}204 + 6346 \times 2) \times 59.94 = 11.14 \text{ Mb/sec.}$$

A detailed Byte Assignment for Field 1 and Field 2 before interleaving (Section 6.1.7) is shown in Figure 6-2a and 6-2b, respectively. Each of the six narrow cross-hatched segments near the top represent a second segment, needed to accommodate the indicated bytes immediately above, which are W1 data. W1 data thus inherently always occur in data segment pairs. The variable W2 data, however, also always are sent in data segment pairs to simplify processing. The nine data segments at the end of each data field are W2 but are not paired.

The maximum number of video data bytes per field (W2) is:

$(141{,}120 \times 2)/8 = 35{,}820$ bytes/field, corresponding to 16.92 Mb/sec.

The minimum number of video data bytes per field (W1) is:

$141{,}120/8 = 17.640$ bytes/field, corresponding to 8.46 Mb/sec.

Transmission Bit Map (TBM) data and Video Control data are directly supplied by the Video Encoder. TBM data identify the variable data (W1 or W2) for the next field.

Although not explained in detail until Section 6.2.2, the modulated data signal as transmitted includes a pilot carrier to aid the receiver in tuning to a desired channel in the presence of strong NTSC cochannel interference.

6.1.4 Data Segment Sync.

Each Data Segment starts with four symbols (1/2 byte of W1 data) dedicated to synchronization of the receiver Video Data Clock. These four symbols are the only ones always recurring in the same pattern and are shown in Figure 6-3a. The periodic identical recurrence of these four symbols makes possible their reliable detection even under severe noise and/or interference conditions and even without parity bits from the RS Code.

Bi–Rate Data Format
Byte Assignment
Data Field 1

171 Bytes/63.5uSec

Data Segment Sync

4 x 262 = 1,048

			Segment
167 Data Field Sync (Equalizer Training)			1
15 TBM	132 Video Control	20 Parity	2
			3
15 TBM	132 Video Control	20 Parity	4
			5
15 Anc	132 Audio	20 Parity	6
			7
15 Anc	132 Audio	20 Parity	8
			9
15 Anc	132 Audio	20 Parity	10
			11
18 Anc	129 Audio	20 Parity	12
			13
(MIN) 120 x 147 = 17,640 If all sent W1 (MAX) 240 x 147 = 35,280 If all sent W2		2,400 Parity or 4,800 Parity	14
			253
15 Anc	132 Audio	20 Parity	254
15 Anc	132 Audio	20 Parity	255
15 Anc	132 Audio	20 Parity	256
18 Anc	129 Audio	20 Parity	257
147 Anc		20 Parity	258
147 Anc		20 Parity	259
147 Anc		20 Parity	260
147 Anc		20 Parity	261
147 Anc		20 Parity	262

FIG. 6-2a

Bi-Rate Data Format
Byte Assignment
Data Field 2

171 Bytes/63.5uSec

Data Segment Sync

4 x 263 = 1,052

			Line
167 Data Field Sync (Equalizer Training)			263
15 TBM	132 Video Control	20 Parity	264
			265
15 TBM	132 Video Control	20 Parity	266
			267
15 Anc	132 Audio	20 Parity	268
			269
15 Anc	132 Audio	20 Parity	270
			271
15 Anc	132 Audio	20 Parity	272
			273
18 Anc	129 Audio	20 Parity	274
			275
(MIN) 120 x 147 = 17,640 If all sent W1 (MAX) 240 x 147 = 35,280 If all sent W2		2,400 Parity or 4,800 Parity	276
			515
15 Anc	132 Audio	20 Parity	516
15 Anc	132 Audio	20 Parity	517
15 Anc	132 Audio	20 Parity	518
18 Anc	129 Audio	20 Parity	519
147 Anc		20 Parity	520
147 Anc		20 Parity	521
147 Anc		20 Parity	522
147 Anc		20 Parity	523
147 Anc		20 Parity	524
167 Test Data			525

FIG. 6-2b

AT&T

ZENITH

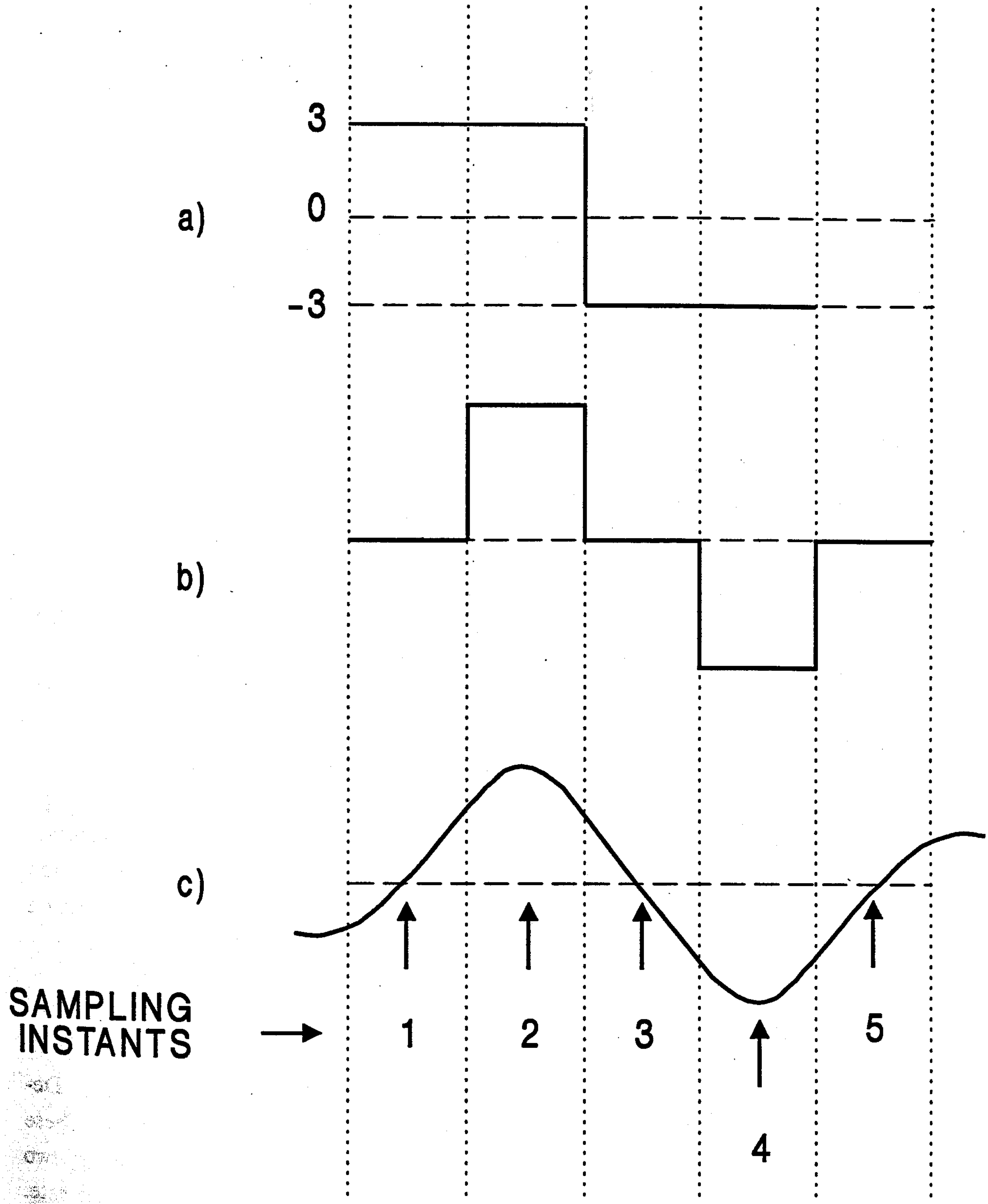


FIG. 6-3 DATA SEGMENT DETECTION

The details of Segment Sync detection in the receiver are described in Section 6.3.3. Suffice it to mention here that the sync is detected in an integrate-and-dump circuit that rejects random input (all other data are random and can be positive or negative) but that enhances the bipolar Segment Sync. Segment Sync controls video data clock synchronization. If the data clock does not coincide with the bipolar sync, a correction is applied to the data clock generator until coincidence is reached.

The 12 remaining symbols of Data Segment Sync are used for DC Offset Control and their use is explained in Sections 6.1.8 and 6.1.9, below. The 16 symbols constituting Data Segment Sync are not interleaved.

6.1.5 Data Field Sync.

Each Data Field starts with one Data Segment of Data Field Sync. It consists of 668 symbols of pseudo-random data of maximum level.

The receiver compares the data from a local lookup table to the incoming pseudo-random sequence data in order to reliably establish synchronization.

To differentiate between Data Field 1 and Data Field 2, the pseudo-random sequences change polarity from field to field.

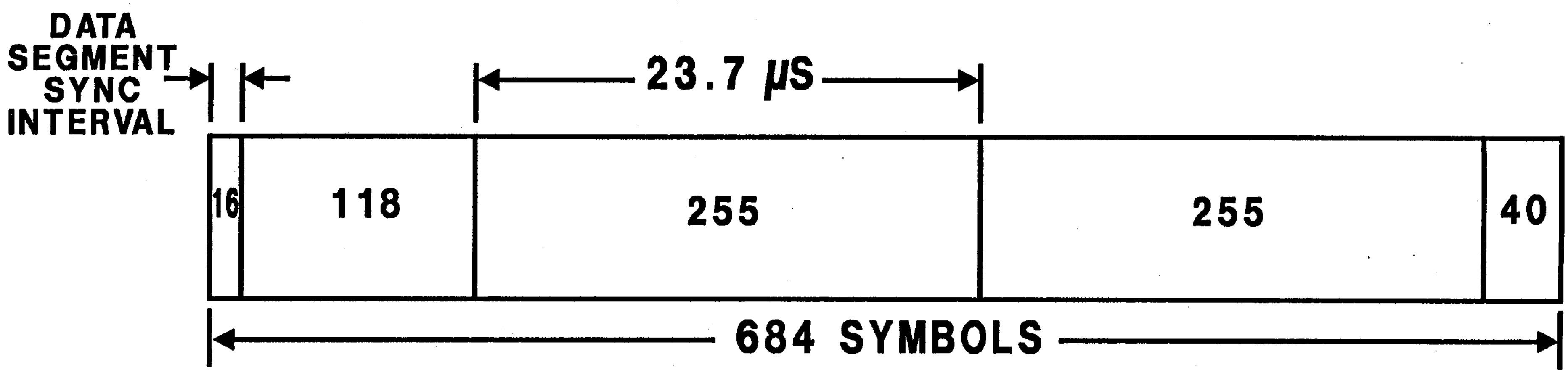
The arrangement of the pseudo-random sequence in the data segment is illustrated in Figure 6-4.

Data Field Sync is also used as the reference signal for Ghost-Cancellation/Channel-Equalization and for data error calculation as part of the decision process whether to use the Post-Comb in the receiver or not (Section 6.1.8).

6.1.6 Error Control.

A Data Segment contains 684 data symbols as was shown in Table I. Deducting 16 Segment Sync symbols leaves 668 symbols for information data. These symbols represent 668/4 = 167 bytes for W2 data; 167 bytes of W1 data occupy two Data Segments. The RS Code protects a block of 167 bytes for either data rate. Twenty bytes out of 167 are set aside as RS parity bytes so that the code is characterized by RS (167,147) t=10. Twenty parity bytes can correct 10 byte errors. The

DSC–HDTV
DATA FIELD REFERENCE SIGNAL



SYMBOL RATE = 10.76 MHz

FIG. 6-4

DSC-HDTV system further increases the error-correcting capability of the RS Code by detecting "erasures". An erasure is detected by sensing when a piece of data received is ambiguous (e.g., halfway between two nominal levels).

With the additional erasure detection the RS (167,147) t=10 code can correct twice as many errors provided that the sum of half the number of erasures and the number of errors does not exceed 10.

There is no direct proportional relationship between byte error rate and bit error rate because a byte error can be caused by 1, 2...up to 8 bit errors in the same byte. Only at very low error rate do the two converge but the bit error rate is never expected to be smaller than the byte error rate for the same C/N ratio.

6.1.7 Interleaving.

The RS Code is particularly powerful in protecting against burst errors. A strong NTSC co-channel may cause an interference into the DSC-HDTV transmission channel which has a burst-like character at every rise and fall of the NTSC sync and which falls somewhere on a data segment. The RS Code can easily correct the errors caused by this interference. It is a common kind of interference due to the fact that the NTSC sync is always present and is always stronger than other parts of the signal at RF. Strong NTSC cochannel interference, however, activates the Post-Comb (see next section). Its differencing property makes it especially apt to pass a high interference output on jumps in the NTSC signal, for example, strong vertical edges.

DSC-HDTV transmission uses interleaving to gain extra protection against strong vertical edges in an NTSC cochannel and against burst-type interference in general. Interleaving is especially useful against interference caused by high-contrast text in the NTSC image.

Interleaving is provided to a depth of one-half data field (inter-segment interleaving over 130 segments). Additional interleaving takes place within one data segment (intra-segment interleaving). Interleaving is so arranged that every twelfth symbol is always of the same data weight (W1 or W2). This is important for proper Precoder operation and is explained in Section 6.1.9.

6.1.8 Interference Rejection System

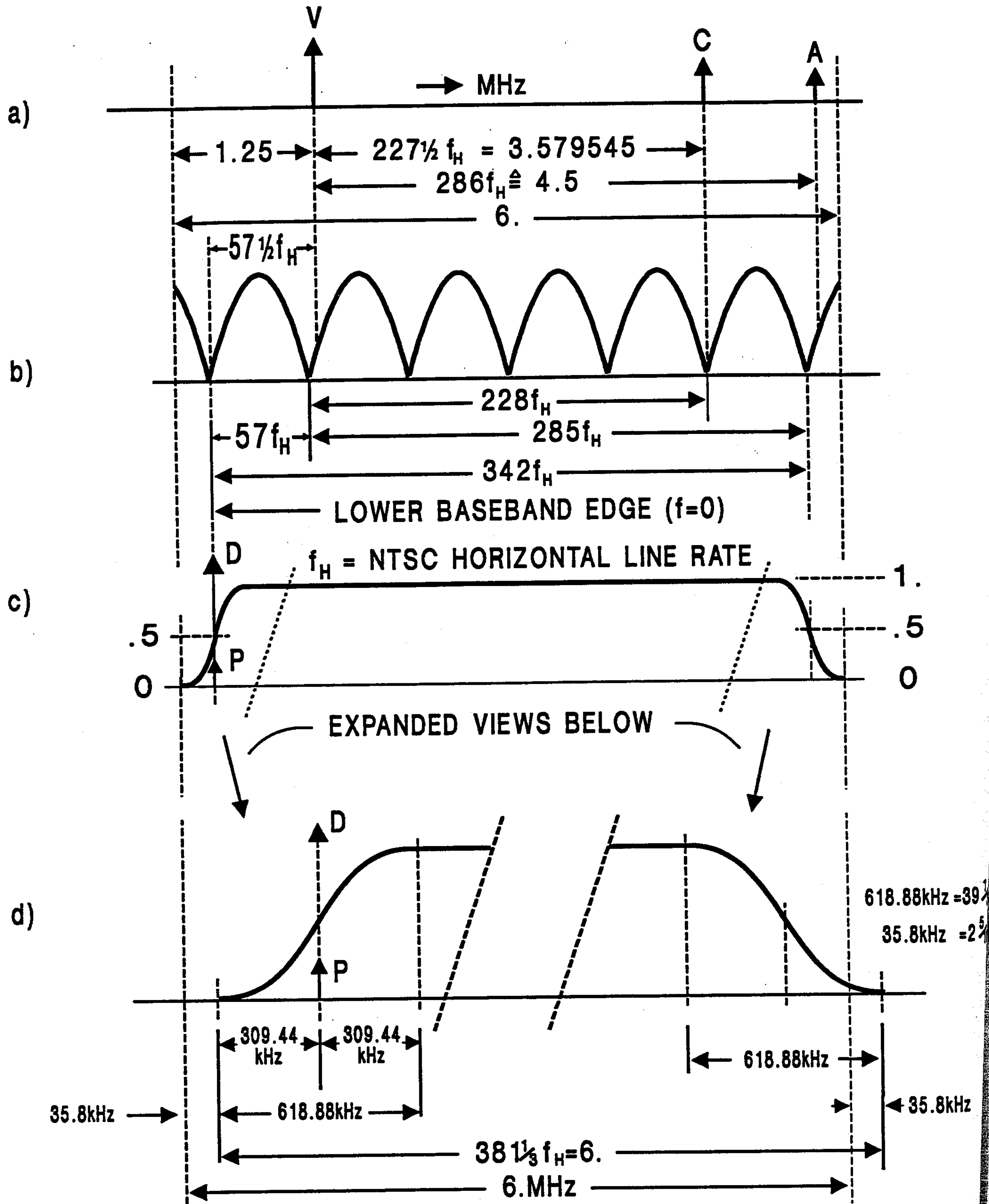
The interference rejection properties of the DSC-HDTV receiver are based on the frequency location of the principal components of the NTSC cochannel interfering signal within the 6 MHz TV channel and the periodic nulls of a suitably chosen DSC-HDTV receiver baseband comb filter.

Figure 6-5a shows the location and approximate magnitude of the NTSC principal components which are: 1) the visual carrier (V) located 1.25 MHz from the lower band edge, 2) the chrominance subcarrier (C) located 3.58 MHz higher than the visual carrier frequency, and 3) the aural carrier (A) located 4.5 MHz higher than the visual carrier frequency.

Figure 6-5b shows the chosen receiver comb filter (Post-Comb) which provides periodic spectral nulls spaced at a baseband integer submultiple of the symbol rate which is related to NTSC horizontal scanning rate. The specific numbers are: $f_{symbol} = 684\ f_H = 10.76224$ Msymbols/sec. where a symbol is either four-level (W2) or two-level (W1). The integer submultiple of the symbol rate chosen for the receiver comb filter is 12, thus making the frequency spacing between nulls equal to 10.76224/12, or, 0.89685 MHz ($= 57f_H$). Such a comb filter will have 7 nulls across the 6 MHz channel. The NTSC visual carrier frequency falls close to the second null into the channel from the lower band edge. Fortuitously, the 6th null from the lower band edge is correctly placed for the NTSC chrominance subcarrier, and the 7th null from the lower band edge is almost correctly placed for the NTSC aural carrier.

Comparing Figure 6-5a and Figure 6-5b shows that the visual carrier falls $1/2\ f_H$ (= 7.867 kHz) above the second comb filter null, the chroma subcarrier falls exactly at the 6th null and the aural carrier falls $3/2\ f_H$ (= 23.6 kHz) above the 7th null. The lst and 7th nulls are arranged to be located at the -6 dB point of the cascade of the transmitter and receiver SAW channel pulse shaping filters.

The detail at band edges for the overall channel is shown in Figure 6-5c and Figure 6-5d. Figure 6-5d shows that the chosen frequency relationship (of a difference of $57.5\ f_H$ between the DSC-HDTV carrier and the NTSC cochannel carrier) results in a DSC-HDTV spectrum shift with respect to the nominal channel. The shift equals +35.8 kHz ($41/18\ f_H$) or +0.6%. This is slightly higher than currently applied channel offsets and reaches into the upper adjacent channel at a level of



DSC-HDTV INTERFERENCE FILTER

POST COMB
SPECTRUM OCCUPANCY

AT&T

ZENITH

FIG. 6-5

-41.7 dB. If that is another DSC-HDTV channel, the nominal situation is restored. If it is an NTSC channel, the shift is below the (RF equivalent of the) Nyquist slope of an NTSC receiver where there is high attenuation and slightly above its customary lower adjacent channel sound trap. No adverse effects of the shift have been found nor are they foreseen.

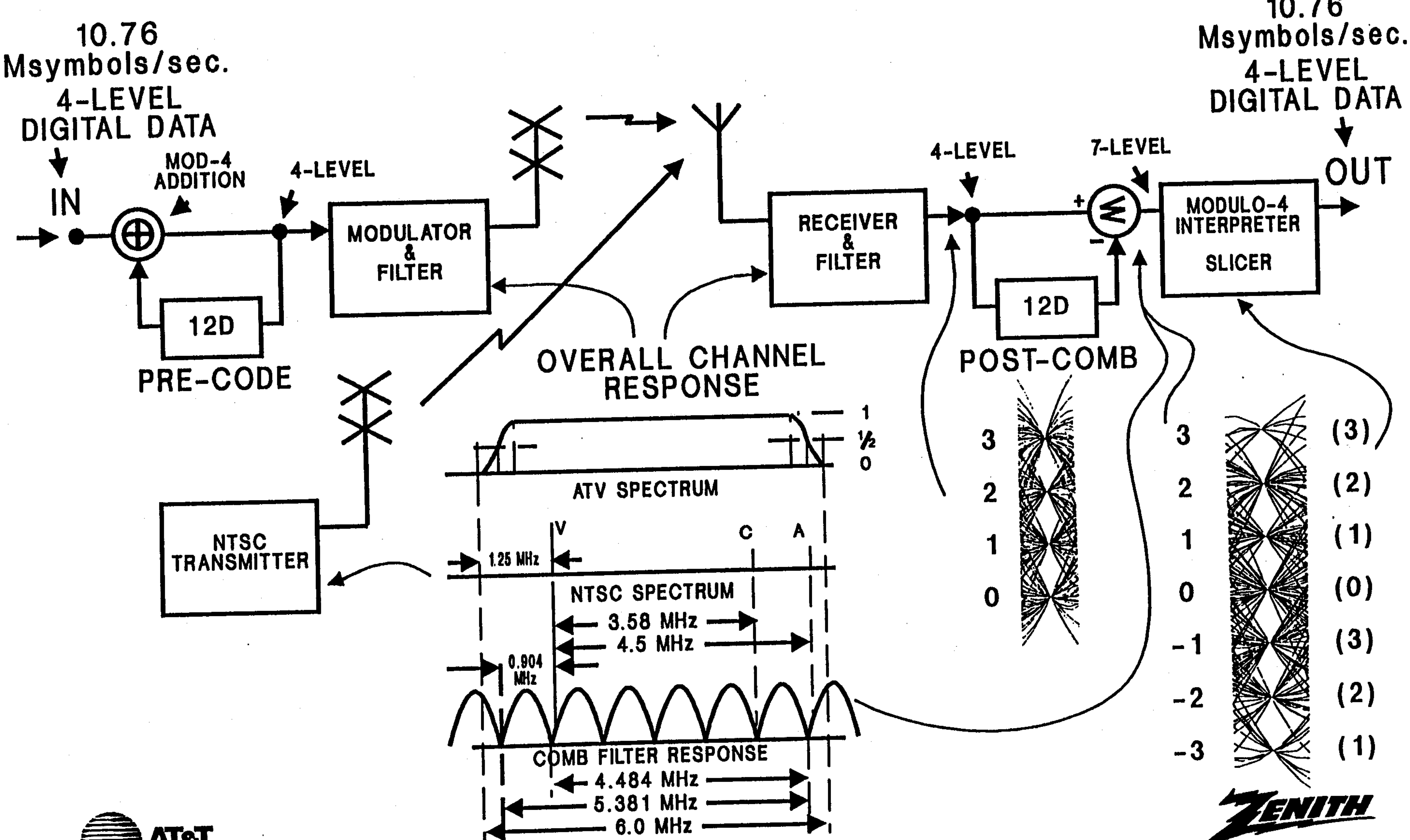
The question of what happens to the eye pattern of the digital signal naturally arises when a comb filter is interposed between the received digital data signal and the decision circuits. The three stacked eyes representing the four levels of the received signaling waveform are converted into six stacked eyes representing seven levels. Such a process in the literature is called partial response (or duobinary) [6]. When the four-level signal is pre-coded into a new four-level signal, interpretation of the seven-level signal is possible using a modulo-4 process.

Figure 6-6 shows the arrangement of receiver Post-Comb filter (partial response filter), modulo-4 interpreter, transmission path possibly corrupted by NTSC cochannel interference and random noise, and transmitter Precoder.

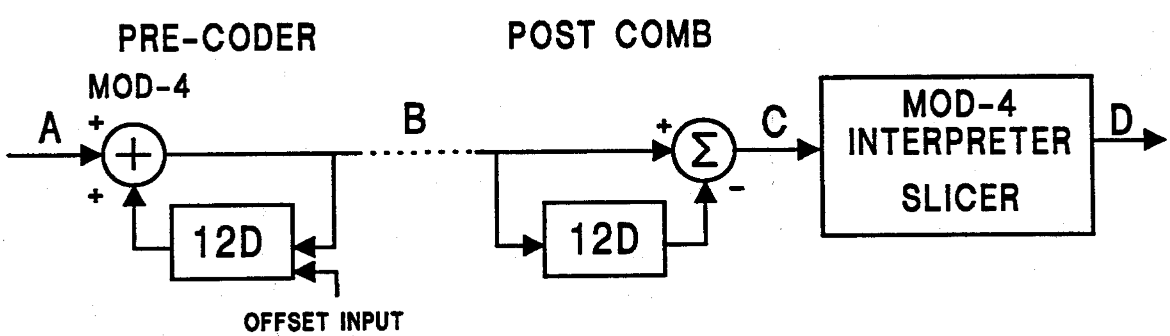
It is important to note that the Post-Comb filter, which uses a feed-forward topology, functions as an linear circuit using arithmetic subtraction whereas the precoder, which uses a feed-back topology, functions in the digital domain using modulo-4 addition. (Note that if the Precoder were implemented as an linear circuit it would be unstable because infinite gain would be required at the receiver comb filter null frequencies. However, when modulo arithmetic is used, the feedback network is stable, and furthermore, the number of levels coming out of the network is the same as entering the network, thus not altering the transmitted power.)

Figure 6-7 gives an example of a 4-level sequence entering the Precoder. Notice that the 12D delay element (delay = 12 symbols) has an extra "Offset input". In this input are added the extra bits for the purpose of "DC offset", which is explained in Section 6.1.9. The extra bits can be 0, 1, 2 or 3 for W2 data, and 2 for W1 data. Figure 6-7 illustrates the W2 case where the DC offset bit is zero. The second column in Figure 6-7, marked "B", shows the precoded sequence followed by the seven-level sequence, column C, leaving the receiver comb filter. Note the seven different levels from -3 to +3 in column C of Figure 6-7. The final sequence is the modulo-4 interpreted seven-level sequence which can be seen to be identical to the original four-level sequence entering the Precoder, thus illustrating the

REDUCTION OF NTSC CO-CHANNEL INTERFERENCE WITH COMB FILTER



A	B	C	D
0	0	0	12 SYMBOLS IGNORED
1	1	1	
2	2	2	
3	3	3	
1	1	1	
3	3	3	
0	0	0	
3	3	3	
3	3	3	
2	2	2	
0	0	0	
1	1	1	
3	3	3	3
3	0	-1	3
1	3	1	1
3	2	-1	3
3	0	-1	3
0	3	0	0
3	3	3	3
1	0	-3	1
2	1	-2	2
3	1	-1	3
0	0	0	0
2	3	2	2
1	0	-3	1
1	1	1	1
2	1	-2	2
2	0	-2	2
1	1	1	1
1	0	-3	1
3	2	-1	3
3	3	3	3
2	3	2	2
3	0	-1	3
3	3	3	3
0	3	0	0



4-VSB CODING EXAMPLE

FIG. 6-7

transparency of the system.

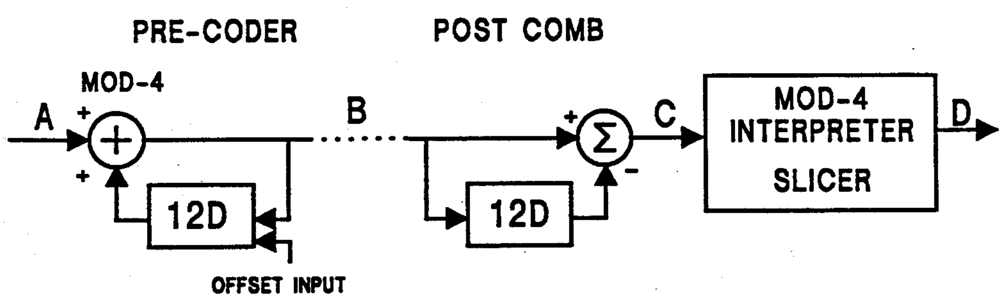
Figure 6-8 shows an example of W1 data entering the Precoder. The two binary levels of W1 data are 0 and 2. The DC offset bit in this example is assumed to be 1. As before, column A represents W1 input. Column B represents the output of the Precoder with and without the one bit offset. The Post-Comb output is in Column C; the first 12 symbols when equipment is first turned on are ignored and not used. Only when the 13th symbol arrives at the Post-Comb is a reliable output obtained. (The Post-Comb is a predictor and can be surprised.) Mod-4 interpretation converts -2 to +2 and Column D is identical to Column A.

6.1.8.1 Comb Filter Noise Consideration

The comb filter, while providing rejection of steady-state signals located at the null frequencies, has a finite response time of 12 symbols (1.115 usec). Thus, if the NTSC interfering signal has a sudden step in carrier level (low to high or high to low), one cycle of the beat frequency (offset) between the DSC-HDTV carrier frequency and the NTSC carrier frequency will pass through the comb filter at full amplitude as instantaneous interference. Examples of such steps of NTSC carrier are: leading and trailing edge of sync (40 IRE units), and leading and trailing edges of text (up to 92.5 IRE units). The magnitude of the interference is determined by the desired DSC-HDTV carrier to undesired-NTSC-carrier ratio at the receiving location. Interleaving will spread the interference and will make it easier for the RS Code to handle. The RS error-correction code can correct a maximum of 10 remaining transients in a data segment period (= 63.5 usec).

The receiver comb filter, while providing needed cochannel interference benefits, degrades the noise performance of the system by 3 dB. As a consequence, an alternate path around the comb filter is provided by using a feed-forward network operating with modulo-4 arithmetic (the exact inverse of the Precoder) which avoids the 3 dB noise penalty in those reception areas where cochannel NTSC interference is not a problem. The switch-over is automatically accomplished in the receiver as a function of measured error rate of the two alternatives.

A	B (NO OFFSET)	B (+1 OFFSET)	C	D
0	0	1	1	
2	2	3	3	
2	2	3	3	
0	0	1	1	
0	0	1	1	12 SYMBOLS IGNORED
0	0	1	1	
2	2	3	3	
0	0	1	1	
0	0	1	1	
2	2	3	3	
2	2	3	3	
2	2	3	3	
2	2	3	2	2
2	0	1	-2	2
0	2	3	0	0
2	2	3	2	2
0	0	1	0	0
0	0	1	0	0
0	2	3	0	0
2	2	3	2	2
2	2	3	2	2
0	2	3	0	0
2	0	1	-2	2
2	0	1	-2	2
0	2	3	0	0
0	0	1	0	0
0	2	3	0	0
0	2	3	0	0
2	2	3	2	2
2	2	3	2	2
0	2	3	0	0
2	0	1	-2	2
0	2	3	0	0
2	0	1	-2	2
0	0	1	0	0
2	2	3	2	2



2-VSB CODING EXAMPLE

FIG. 6-8

6.1.9 Pilot Signal and DC Offset Control.

The pilot signal (see 6.2.2) could be added in analog form in the modulator by adding a slight amount of unmodulated carrier. Control over the level is maintained more accurately, however, by adding a (digital) DC level at baseband. The nominal transmission signal levels of W2 data are -3, -1, +1 and +3, and of W1 data they are -2 and +2. This can be very accurately changed into -2, 0, +2 and +4 for W2 and 0, +4 for W1. The levels are sketched in Figure 6-9 in equivalent analog form.

(The choice of signal levels for the W1 data is based on: keeping the average signal power of W1 less than W2 data and commonality of receiver slicing hardware. See Sections 6.1.10 and 6.3.7.) The pilot transmission signal power is analyzed in Section 6.1.10.

There is some risk involved with adding slight amounts of pilot. The very long-time average of the data signal is zero but over shorter time spans the average may be negative and so decrease or even cancel the pilot. This is not harmful if the pilot cancellation is of shorter duration than the carrier recovery system time constant. The following approach is used to assure that cancellation is of minimal duration.

A data segment comprises 684 symbols, 12 groups of 57. Deducting the first 4 sync symbols yields 8 groups of 57 and a final 4 of 56 symbols. In each of the 12 groups the average DC is measured and additional offset bits are mod-4 added until the average level is positive. Before interleaving, all data in a segment are of the same weight, since inter-segment interleaving is performed by interleaving whole groups, after inter-segment interleaving all symbols in a group of 57 (or 56) are still of the same weight. The number of bits added in each of the 12 groups is encoded in a corresponding symbol that is part of the 12 remaining sync symbols. In this manner the possible duration of pilot cancellation is kept considerably shorter than 63.5/12 = 5.3 microseconds on the average (twice that long worst-case), short enough to prevent unlocking of carrier recovery which takes several hundred microseconds.

Data Field sync symbols and Data Segment sync symbols are neither interleaved nor precoded and are not subjected to DC offset operation. However, pilot is added to these symbols as well as to every other symbol.

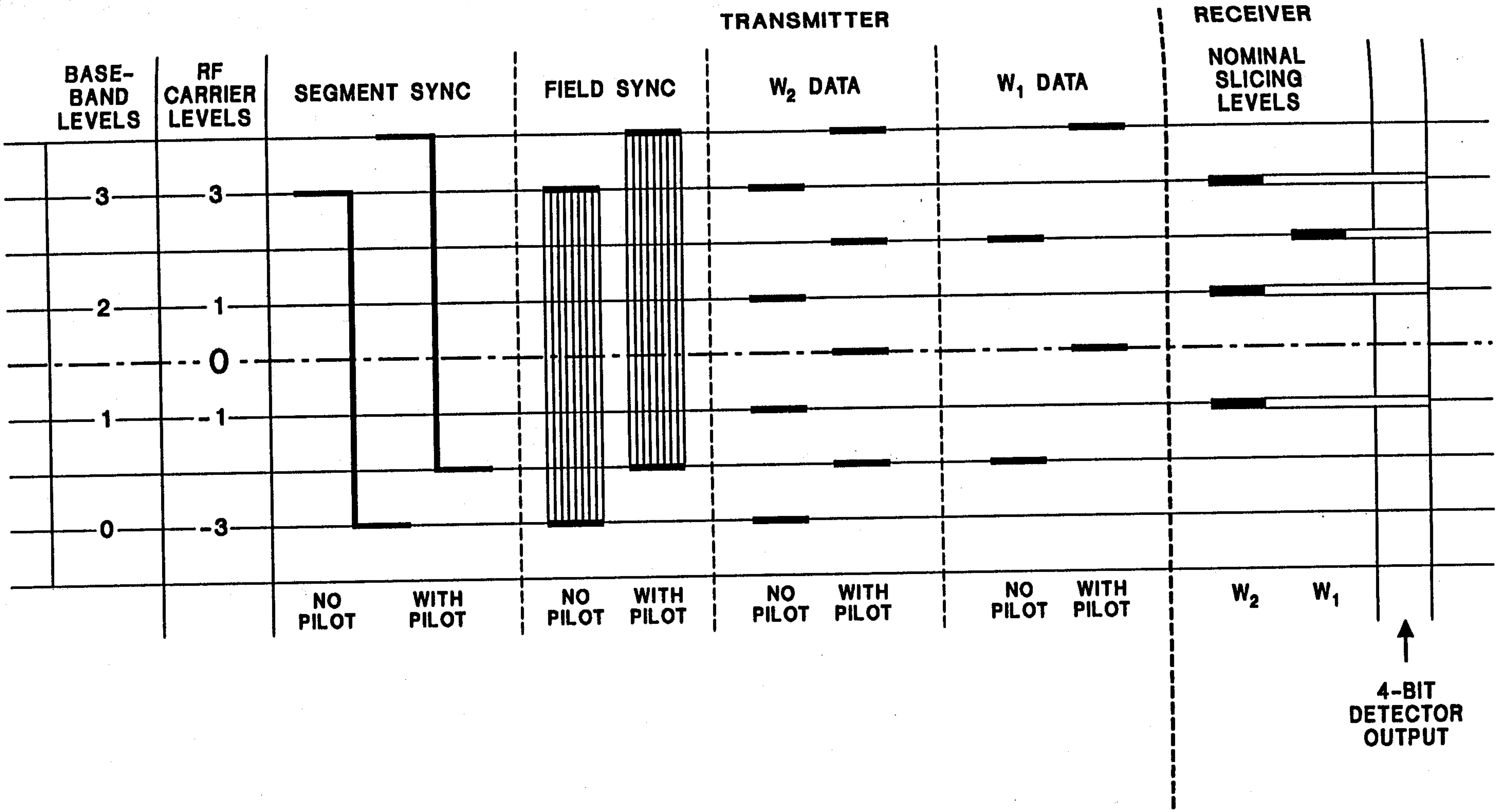


FIG. 6-9
DSC-HDTV DATA LEVELS

AT&T

ZENITH

Data Field symbols are transmitted as maximum level data; they come from a pseudo-random (PR) generator at basband as "0"s and "3"s. The 668 symbols in the data field sync segment consist of two PR sequences of 255 symbols each and two partial PR sequences of 118 and 40 symbols, respectively. (See Figure 6-4.) A true PR sequence has a number of 0's and a number of 1's that differ by one. The DC average is thus very close to zero (not counting offsets). Partial PR sequences may have a non-zero DC average. The present ones, however, do have DC averages close to zero.

The difference between Field Sync in Field 1 and Field 2 is a change in polarity (but not in offset). There is thus a negligible change in DC average in the polarity change, and pilot cancellation is prevented.

6.1.10 Transmission Data Signal Power and BER

This section refers to the error probability curves of Figures 6-10 to 16-12. Figure 6-10 shows the unprotected 16-QAM symbol error probability curve showing the familiar 20 dB S/N ratio (average power) needed for 10^{-5} symbol error probability [7].

It is important to note the distinction between S/N and C/N. S includes only the AC signal (the data eye) power while C includes the added pilot power. The graphs in this section and in the technical literature should be read with this distinction in mind. In Section 7 it is noted that C/N, when discussing coverage and the like, is in reality S/N - that is, minus the pilot power. Since the difference between S/N and C/N is 0.8 dB for W2 data, the consequences are insignificant.

Also shown in Figure 6-10 are the symbol error probabilities for 4-VSB. At high signal-to-noise ratio the curve is identical to the 16-QAM curve. At low signal-to-noise ratios the 4-VSB curve is superior. In the 4-VSB the two outer states have a maximum probability of error of 50% no matter how large the noise. The outer states in 16-QAM have a higher asymptotic error probability. The 2-VSB signal has a theoretical improvement of 7 dB at high signal-to-noise ratios. The high noise asymptotic behavior shows an even larger improvement. To minimize the complexity in the receiver, it is advantageous to use the same states for the two and four-level system. If the two outer states were used, the average signal power would be higher in a 2-level system than in a 4-level system. In an interference-limited envi-

DSC-HDTV System

Unprotected Error Probabilities

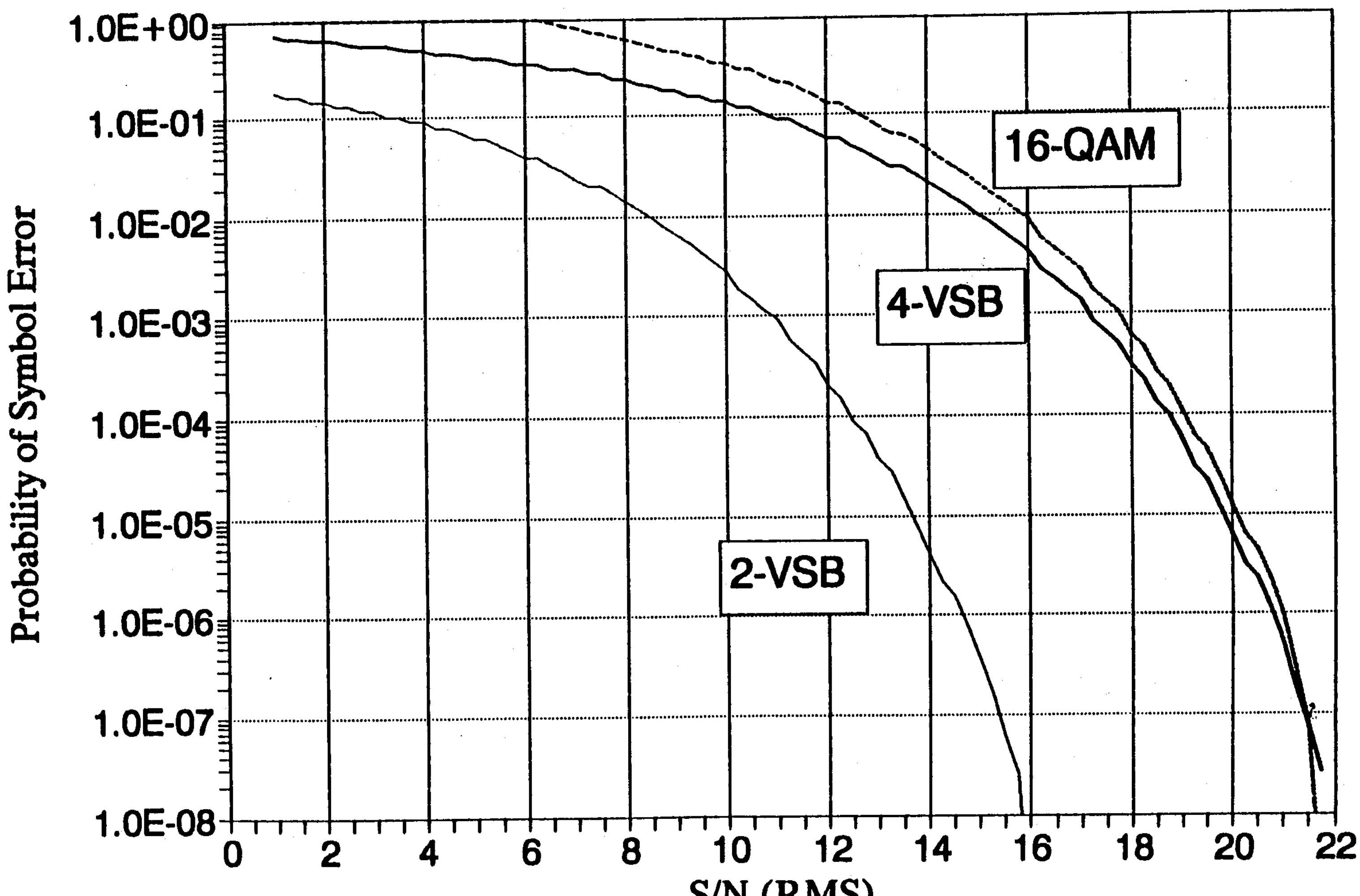


FIG. 6-10

AT&T

ZENITH

ronment this is not desirable. Therefore, the first and third states are used and the average signal power is approximately 1 dB lower than the 4-level system. The improvement obtained by using the two level system in signal-to-noise ratio and signal-to-interference ratio is therefore 6 dB (as shown in Figure 6-10).

Figures 6-11 and 6-12 give the error probabilities after Reed-Solomon error correction. The RS blocks are 167 bytes long. For W2 data the block is transmitted in one Data Segment. For W1 data the block is transmitted over two consecutive Data Segments. The curves also give the byte error probability, which is an intermediate step used to calculate the RS block error probabilities.

The video compression system has been designed to be error tolerant and uses a masking mode to cover uncorrected errors. The threshold for visible errors is approximately 5×10^{-3}. The curves indicate for W2 data this threshold is reached at a signal-to-noise ratio of 16 dB. For W1 data the threshold is reached at 10 dB signal-to-noise.

A rugged system must be able to acquire a signal and maintain lock condition in the presence of very heavy noise. The present prototype receiver can acquire a signal and maintain lock at a signal-to-noise of approximately 0 dB.

6.1.10.1 Pilot Power

The pilot or average data level is different for W1 and W2 data. The 4-levels of W2 are -2, 0, +2 and +4. All data states are equally probable. Therefore, the average is +1. The average signal power is +5. The total carrier power (average signal power plus pilot) is equal to +6. The pilot represents an increased transmitter power of .8 dB. In the interference-limited environment the pilot is not a contributing factor. The pilot was placed near band edge where the NTSC Nyquist filter will remove it.

The 2-levels for W1 data are 0 and +4. The average or pilot is equal to +2. The mixture of W1 and W2 data is a function of the video content. Therefore, the pilot level changes with video content. The total carrier power will vary a few dB. For accurate measurements at the ATTC, a reference signal will be used to measure power. Two such signals are available. The Data Field sync segment which is made up of a pseudo-random sequence which repeats exactly once each Data Field, or the once per frame Test Segment could be used. This segment is made up a

DSC-HDTV System

Reed-Solomon for 4-VSB Transmission

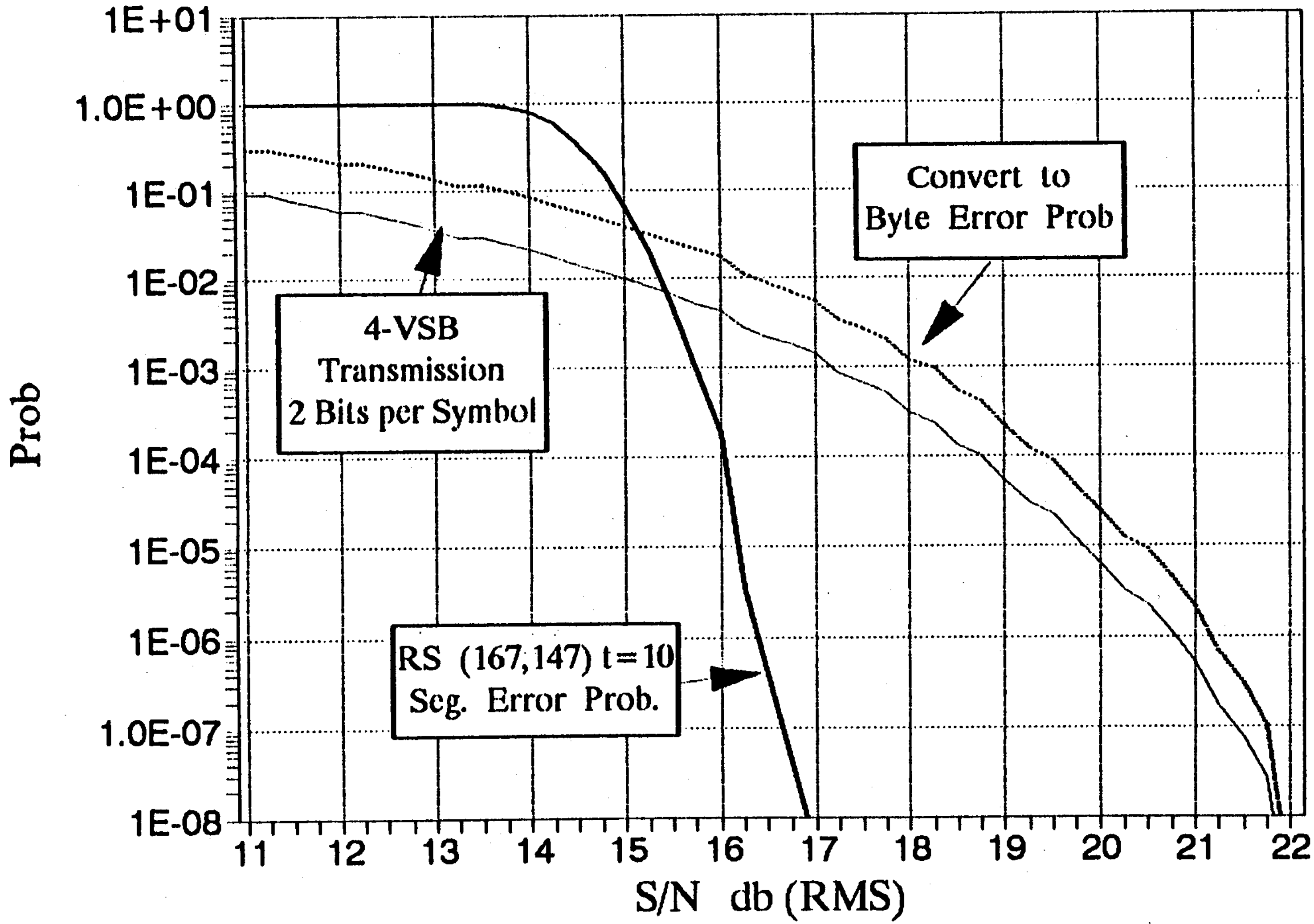
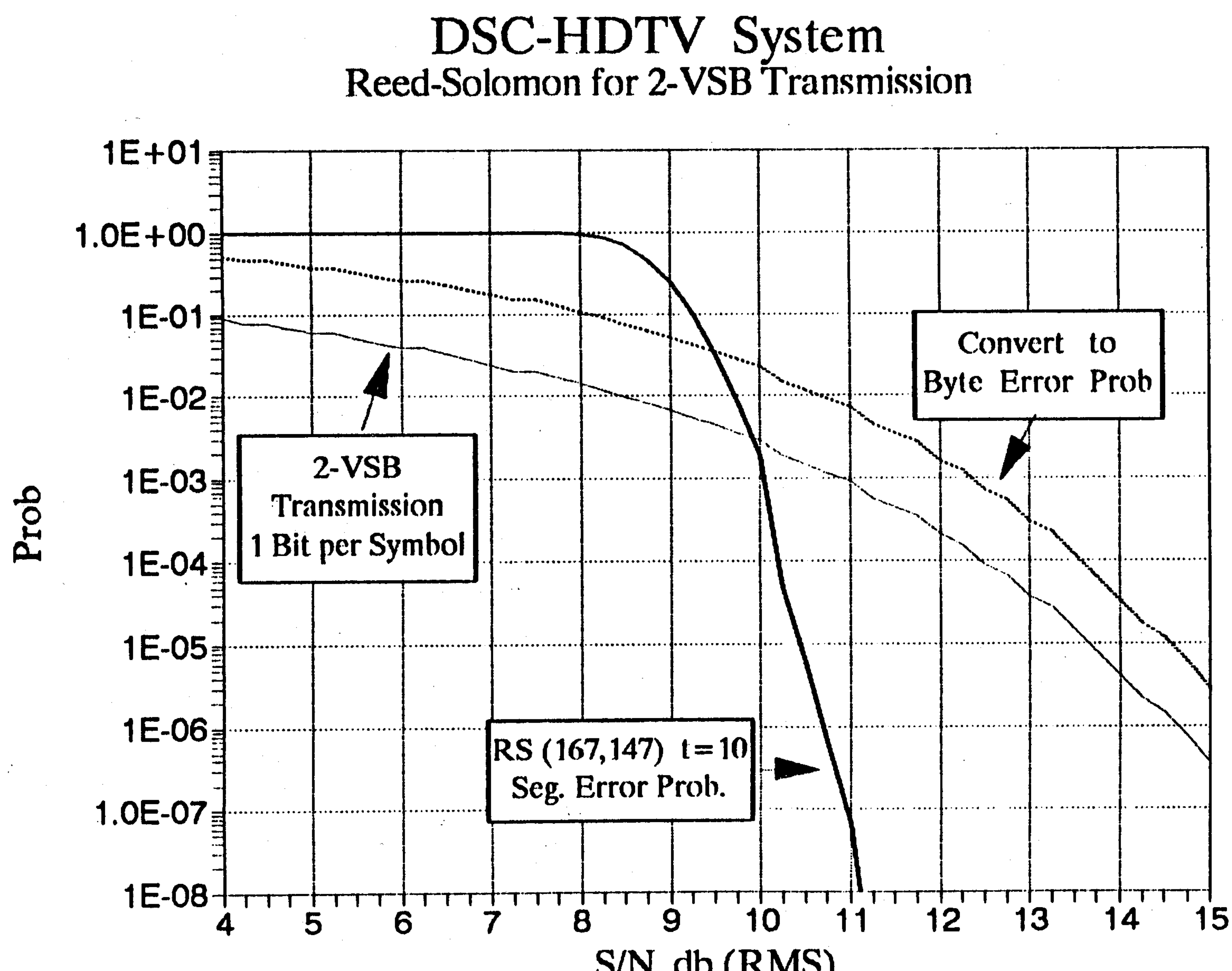


FIG. 6-11

AT&T

ZENITH



DSC-HDTV System
Reed-Solomon for 2-VSB Transmission
Prob
1E+01
1.0E+00
1E-01
1E-02
1E-03
1E-04
1E-05
1E-06
1.0E-07
1E-08
4
5
6
7
8
9
10
11
12
13
14
15
S/N db (RMS)
2-VSB
Transmission
1 Bit per Symbol
Convert to
Byte Error Prob
RS (167,147) t=10
Seg. Error Prob.

FIG. 6-12

3-step signal. A gating signal for either of these signals will be provided.

6.2 TRANSMITTER

6.2.1 Transmitter Block Diagram

Figure 6-13 shows the transmission side of DSC-HDTV.

The video source signals are A/D converted and transformed to the YUV or "$L^* u^* v^*$" format (see Section 3.2.1).

The video data clock signal of frequency $f_d = 4788 f_H = 75.3357$ MHz is the basic clock signal (f_H refers to NTSC; see Table I, Section 6.1.2). Transmission data clock signals of frequency $f_t = f_d/7 = 10.76$ MHz (see Table I) are distributed for baseband processing of the transmission data.

Frequency f_d is chosen between 72 and 76 MHz, the spectrum gap between Channel 4 and Channel 5.

As mentioned in Section 4, each Data Field includes 4200 symbols of W1 data for audio channels 1 and 2 and 2100 symbols of W2 data for audio channels 3 and 4. This represents 2100 bits/field for each of the 4 audio channels or 125.874 kb/s. This corresponds to the exact 8th multiple of f_H, NTSC horizontal line rate which equals the Data Segment rate. The audio bits are protected by the RS Code of DSC-HDTV. The 125.874 kb/s themselves also include error protection provided by the Dolby AC-2 encoding system.

The blocks of Multiplexer (MUX) through Precoder perform the functions previously explained in Sections 6.1.6 through 6.1.9. The data rate in that part of the system is 2.69 Msymbols/s on 4 parallel bus systems of 8 bits width. In the Mapper the data is converted to a single bus at 4 x 2.69 = 10.76 Msymbol/s.

Test data are for self-diagnostic purpose, specific to the equipment used for the ATTC tests. They are not interleaved and their DC average is kept positive.

The pre-equalizer is a digital filter that precorrects for transmitter IF, RF, transmission line and antenna deviations from nominal response. The sinx/x compensation for the D/A converter is also performed here.

After D/A Conversion, the analog form of the signal is applied to the Modulator.

The modulation system, "4-VSB", developed for DSC-HDTV, is discussed in

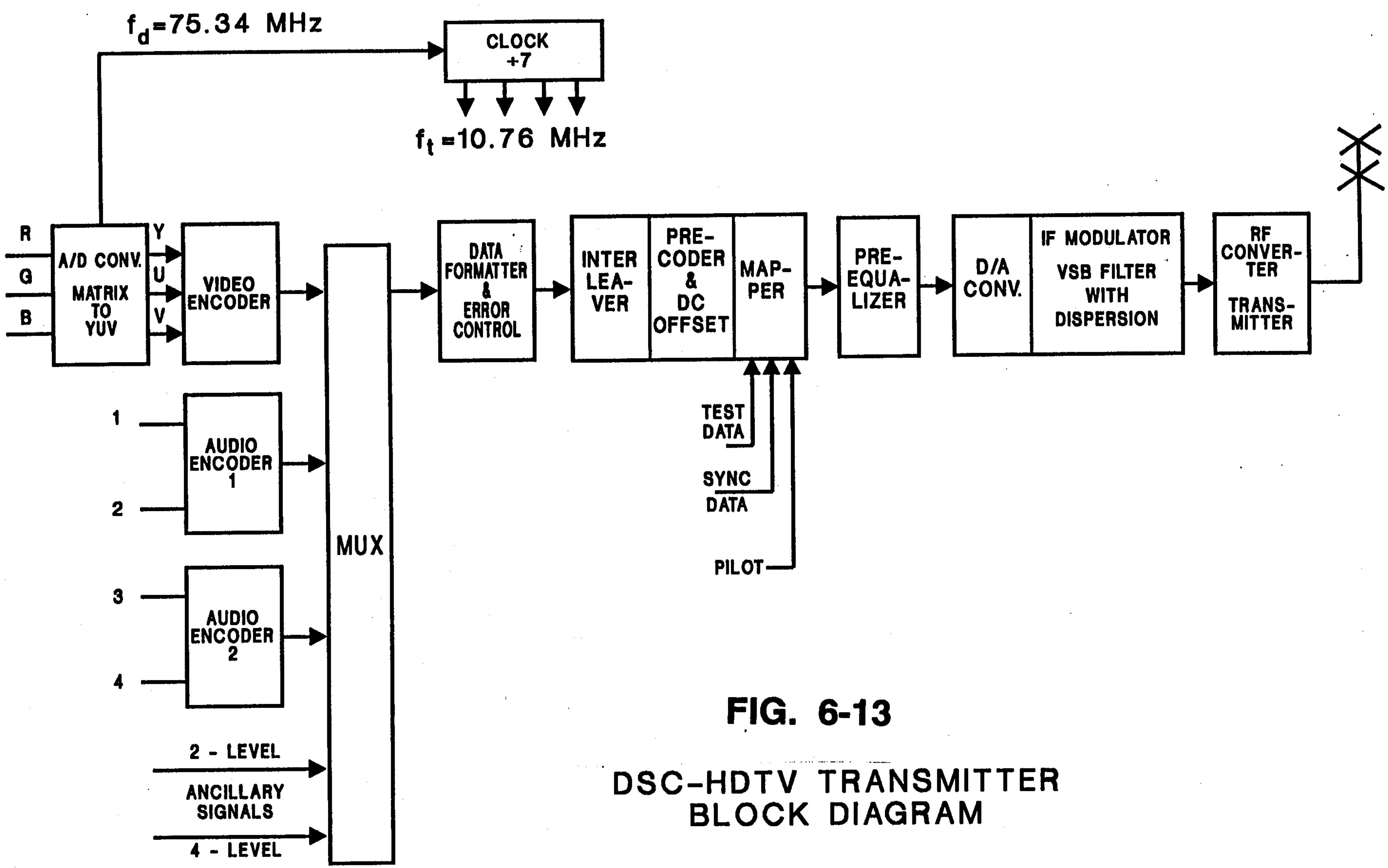


FIG. 6-13

DSC-HDTV TRANSMITTER
BLOCK DIAGRAM

AT&T

ZENITH

the next section.

6.2.2 4-VSB/2-VSB Modulation

A common way to accommodate a maximum of 21 Mb/s in a 6 MHz channel is by 16-QAM. The bitstream is split into two and each is transformed into symbol sequences of 5.38 Msymbol/sec at 2 bits/symbol. Each sequence of 4-level symbols modulates one of a set of two carriers in quadrature by suppressed carrier, double-sideband AM.

Such a method poses receiver carrier regeneration problems under circumstances of strong NTSC cochannel interference. The data-directed detection method fails with an NTSC carrier 1.75 MHz away and only a few dB below the desired signal which is the situation in which a simulcast HDTV receiver must function.

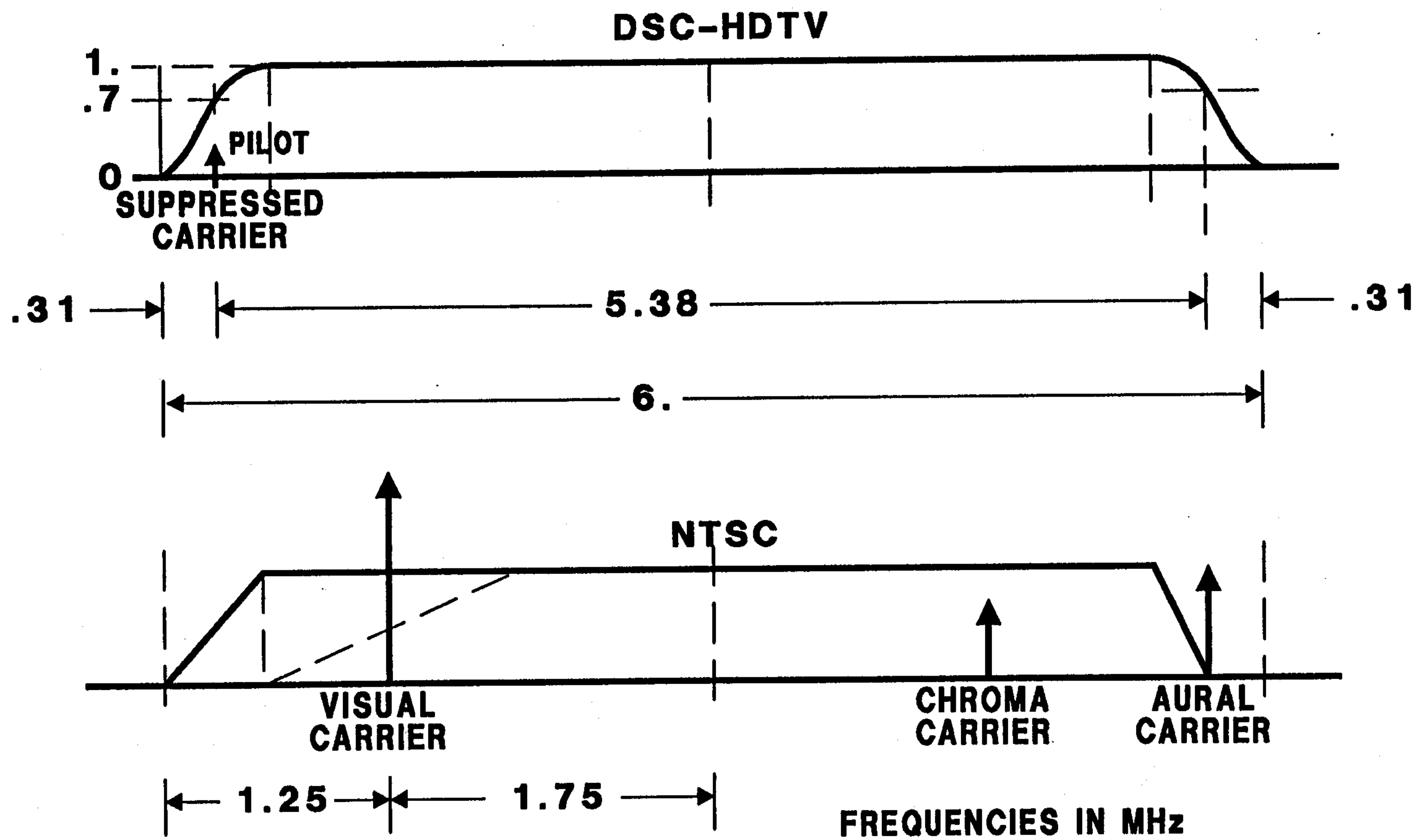
Using non-suppressed carrier would aid detection but requires extra power at a place in the spectrum where it would cause interference into NTSC receivers. The carrier's mid-band location is where its visibility would be at its worst. Even a much attenuated carrier, a pilot signal, would be a risk for interference into NTSC receivers in that frequency region.

A frequency region in which an NTSC receiver has strong attenuation is near the low end of the (RF) band. A pilot in this region will not cause interference into NTSC receivers.

It now becomes a natural choice to place both carrier and pilot near band edge and apply vestigial sideband modulation. The data signal of maximum 21 Mb/s is partially converted into four-level symbols at 10.76 Msymbol/s, "4-VSB", and partially into two-level symbols, "2-VSB". The spectral placement of the DSC-HDTV RF signal in relation to an NTSC cochannel is shown in Figure 6-14.

It was shown in Section 6.1.10 that the theoretical symbol error-probability of 4-VSB is almost identical to that of 16-QAM. 4-VSB has a quadrature component which is the Hilbert transform of the in-phase component over the single sideband region, which is almost the whole band. In 16-QAM the in-phase and quadrature components are, of course, unrelated. There are other, differences between 4-VSB and 16-QAM.

DIGITAL SC-HDTV
DSC-HDTV/NTSC RF COCHANNELS



AT&T

ZENITH

FIG. 6-14

FIG. 6-14

Computer simulation has shown some significant differences in detected eye height between 4-VSB and 16-QAM when group delay tilts or data detecting carrier phase differences are introduced. 4-VSB shows somewhat greater eye height than 16-QAM when these imperfections are applied. Since 4-VSB's symbol rate is twice that of 16-QAM, the eye width in absolute time measure of 4-VSB is half that of 16-QAM. The prototype receiver will use a phase-locked crystal oscillator. This, in combination with the channel equalizer, provides for a stable, accurate data detection system.

Some eye diagrams obtained by simulation are shown in Figures 6-15 through 6-17.

The first two eye diagrams show the case of nominal channel response, Figure 6-15a for 16-QAM in-phase detection, and Figure 6-15b for 4-VSB. Note the difference in horizontal scale, referred to above.

The next two eye diagrams show the result of a linear group delay tilt of 80 nanoseconds across the 6 MHz channel; again comparing 16-QAM (Figure 6-16a) to 4-VSB (Figure 6-16b).

The last two eye diagrams show the influence of a demodulating carrier deviation from ideal by 13 degrees; Figure 6-17a for 16-QAM and Figure 6-17b for 4-VSB.

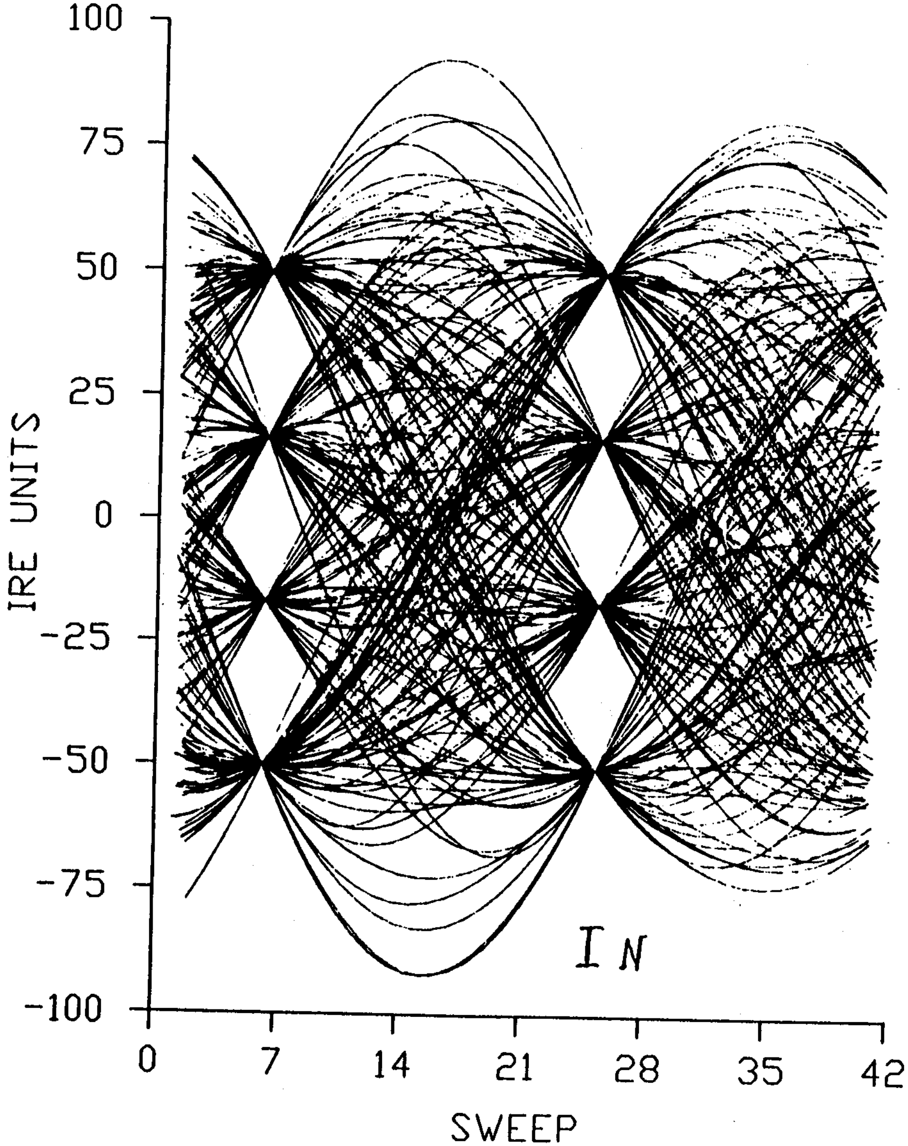
In summary, even though carrier-to-noise ratio, data rates and channel occupancy are equal for 4-VSB and 16-QAM, in the given conditions of HDTV and NTSC in the same band, 4-VSB has distinct advantages over 16-QAM.

6.2.3 SAW Filter and Dispersion

The overall nominal system characteristic is shown in Figure 6-18. The optimum arrangement is to divide the rolloff equally over transmitter and receiver. This yields the characteristic of Figure 6-4 top, implemented as SAW filters for transmitter as well as for receiver. The overall characteristic does not require the same skirt selectivity on both sides, although the system is so implemented.

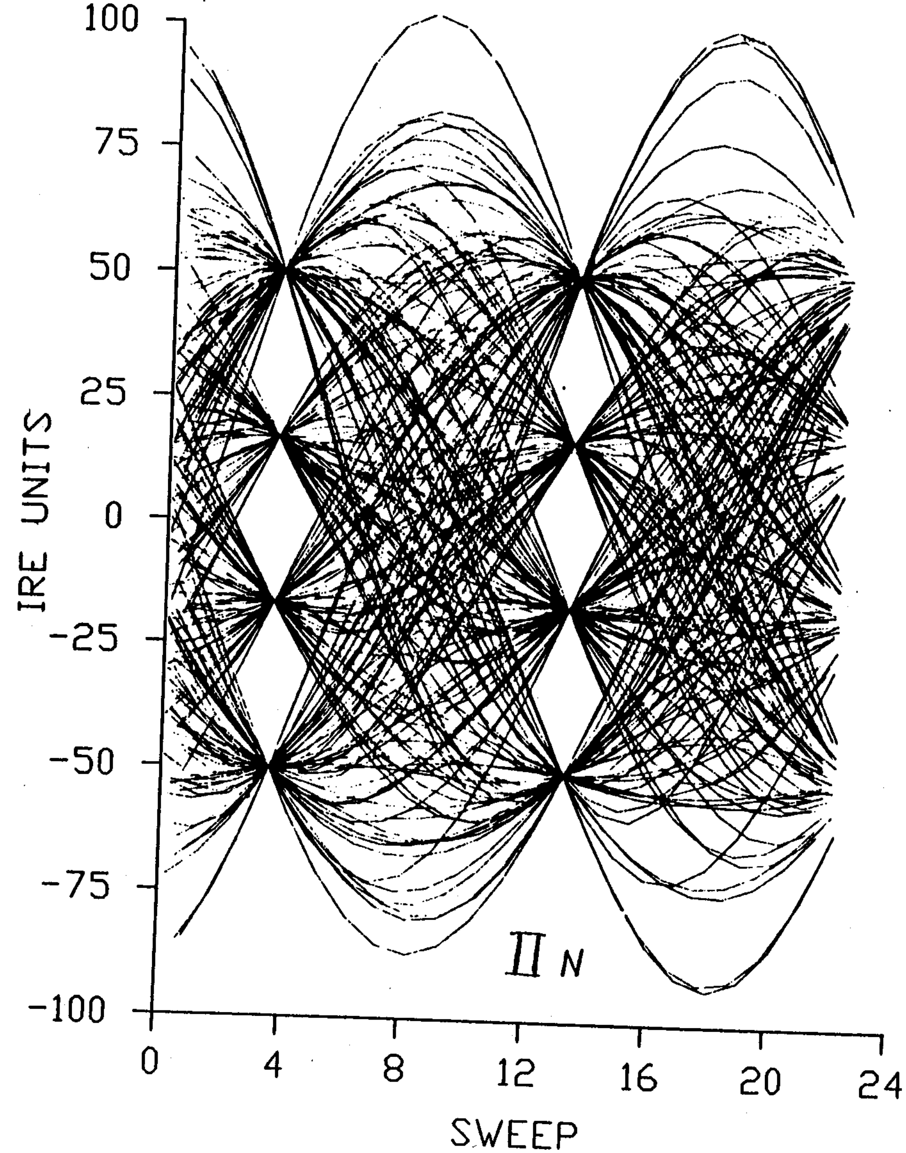
Transmitter and receiver SAW filters have complementary sloped group delay characteristics. This imparts a dispersion to the transmitted data signal which is intentional.

As long as the DSC-HDTV RF data stream is truly random, the cochannel



IN-PHASE DETECTED 16-QAM

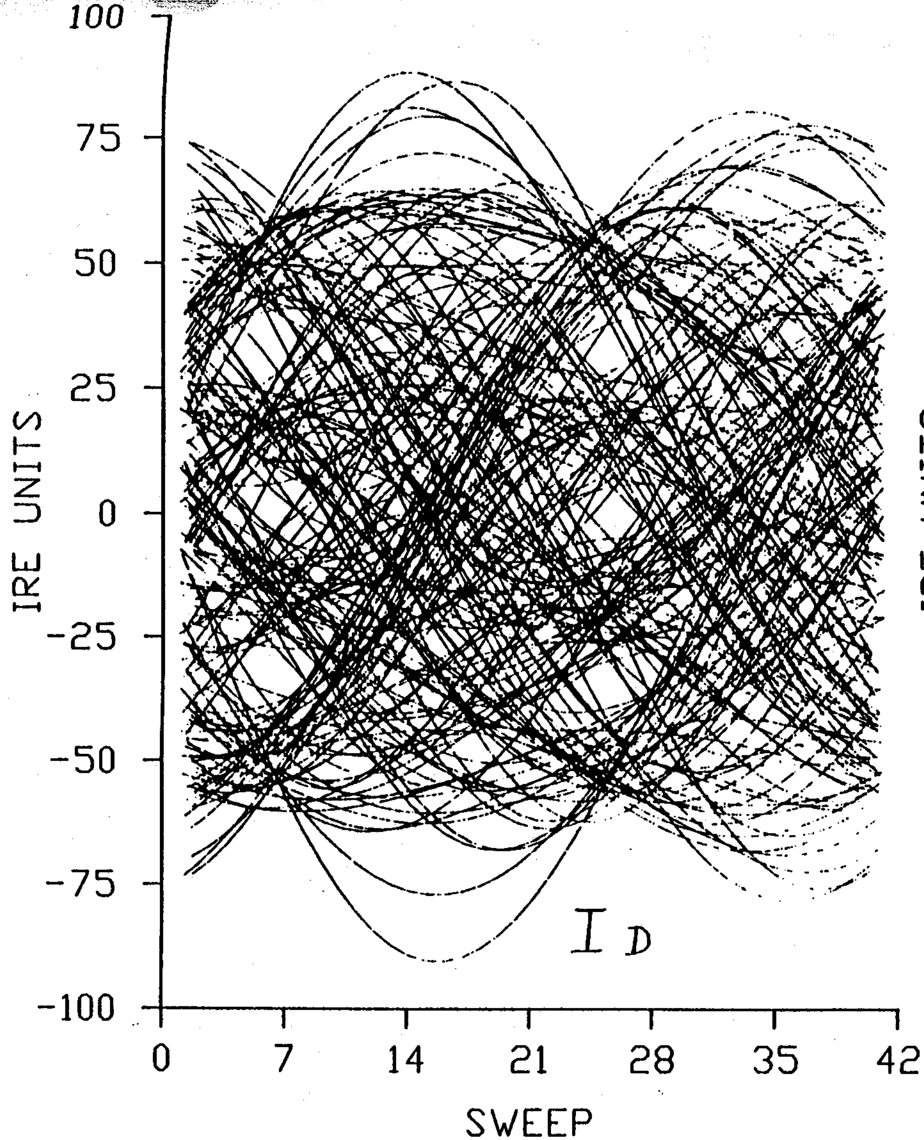
FIG. 6-15a



4-VSB

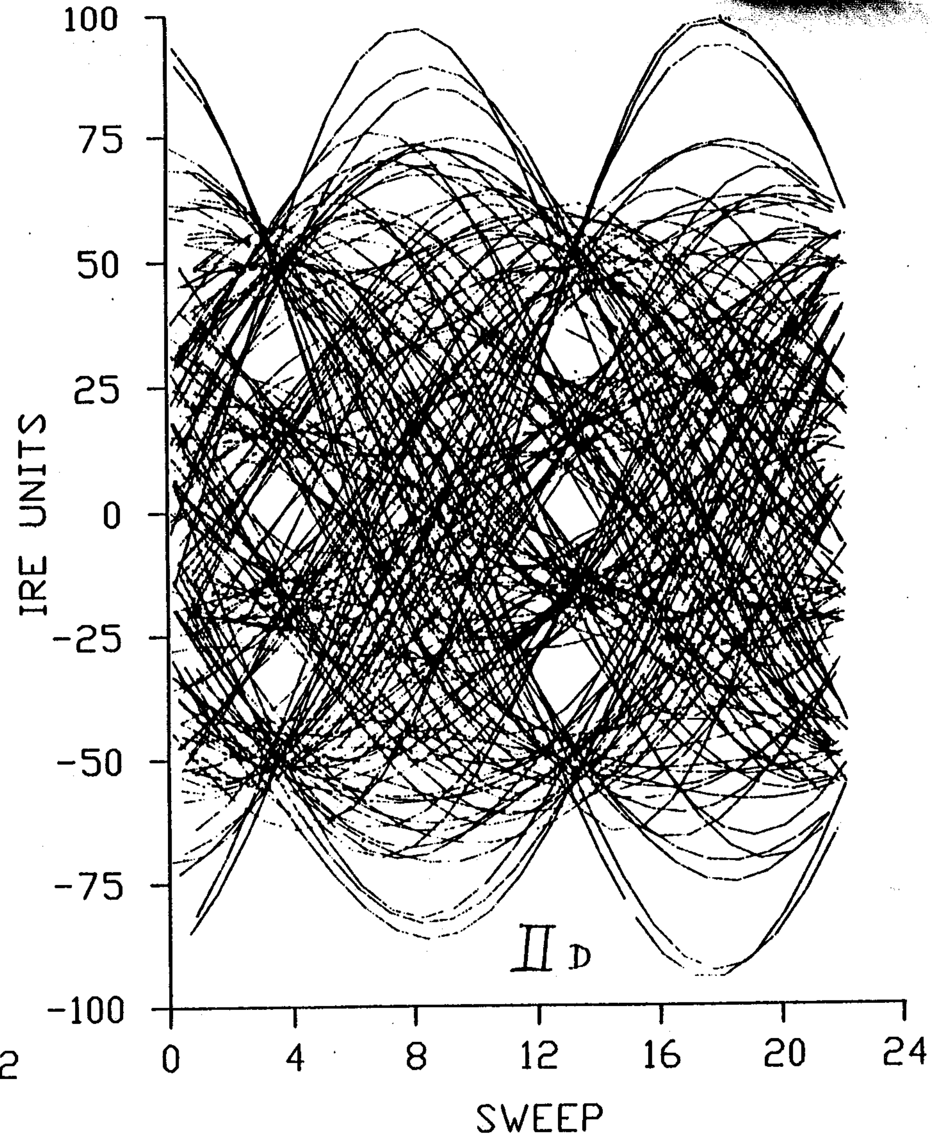
FIG. 6-15b

NOMINAL SYSTEM



IN-PHASE DETECTED 16-QAM

FIG. 6-16a



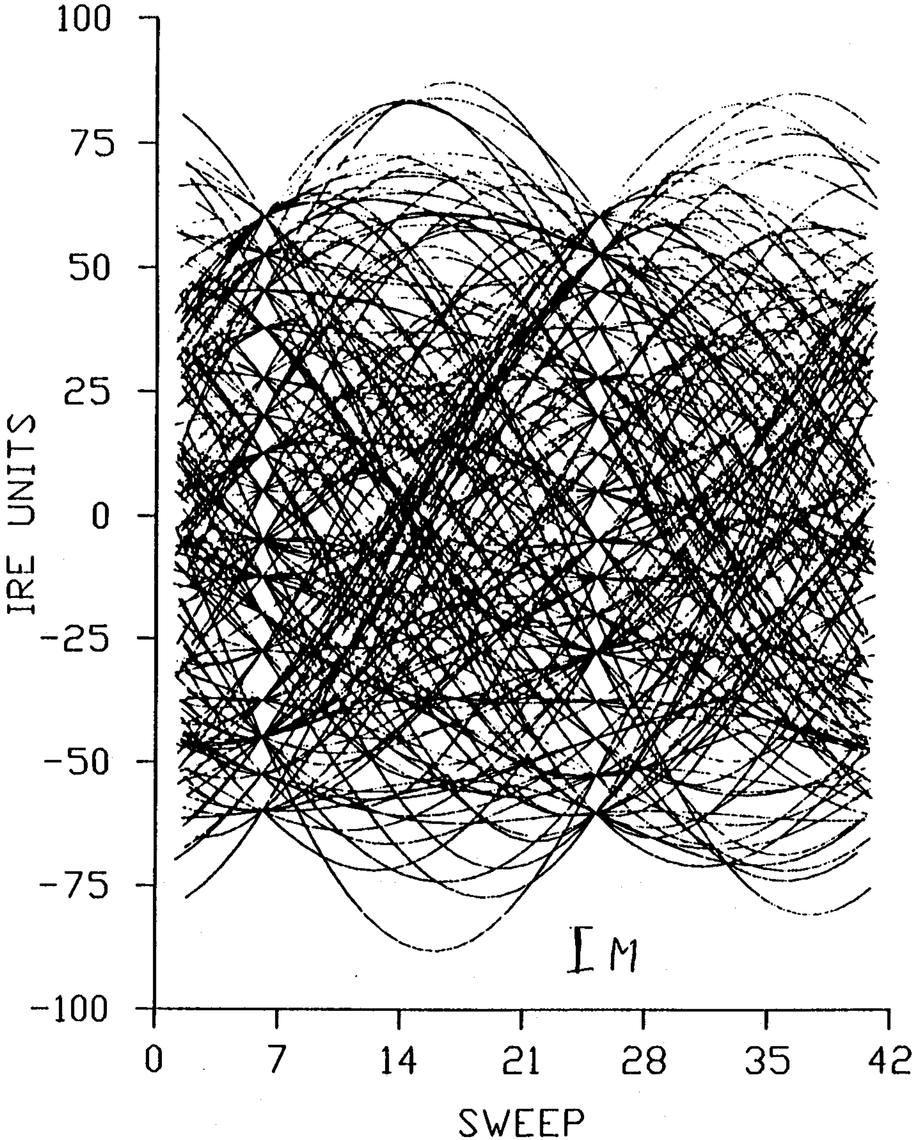
4-VSB

FIG. 6-16b

80 NANOSECONDS GROUP DELAY CHANNEL TILT

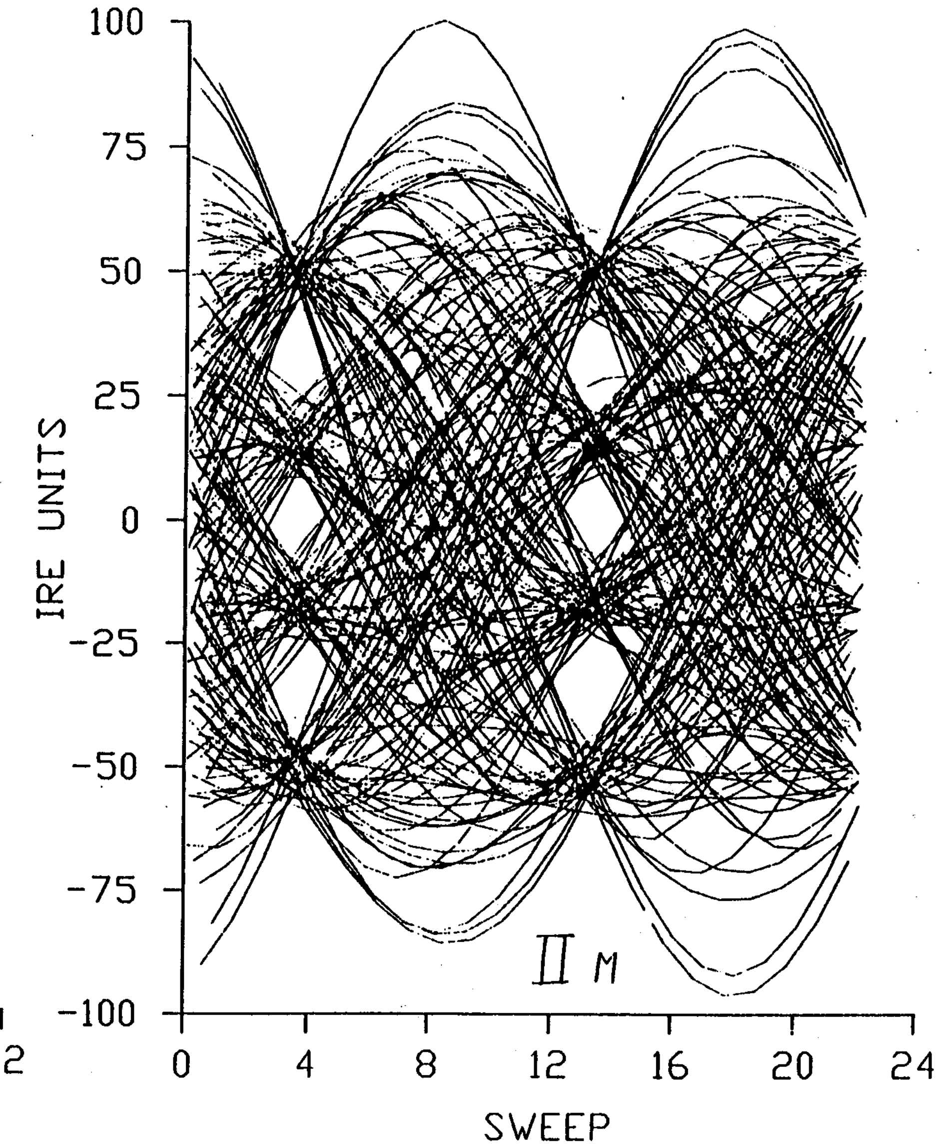
AT&T

ZENITH



IN-PHASE DETECTED 16-QAM

FIG. 6-17a



4-VSB

FIG. 6-17b

13 DEGREES DEMODULATING CARRIER PHASE ERROR

AT&T ZENITH

DIGITAL SC-HDTV
NOMINAL 4-VSB CHANNEL

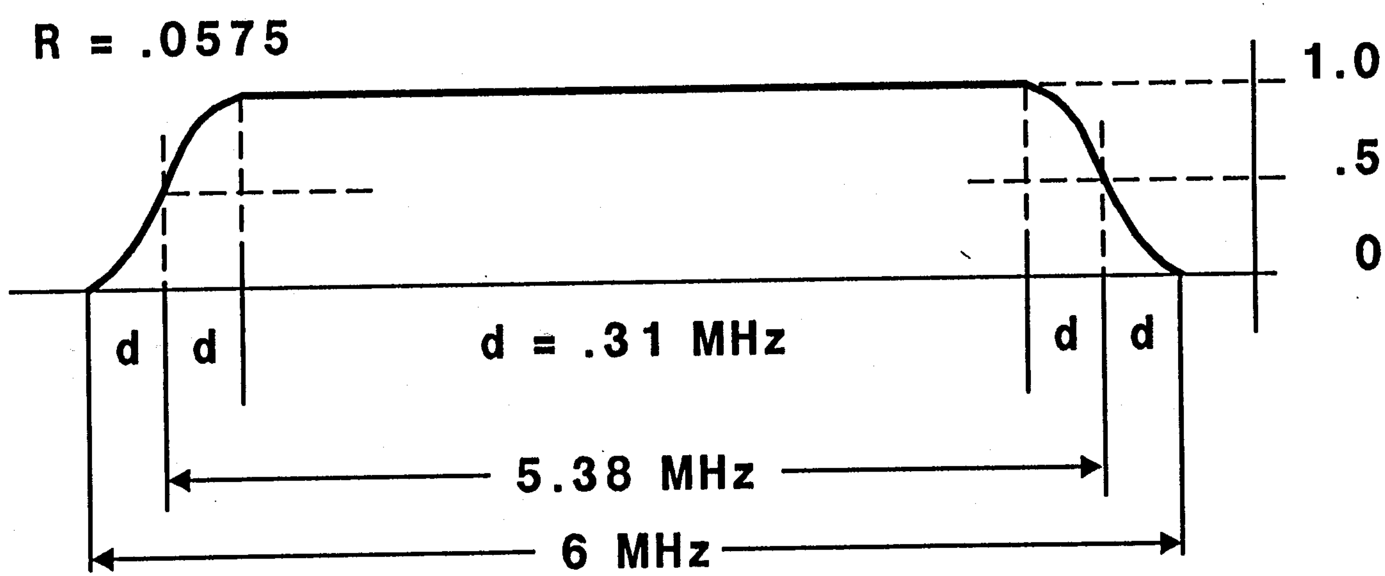


FIG. 6-18

AT&T

ZENITH

interference into an NTSC channel has a random noise-like character and is of minimum visibility. The one exception to randomness of the data is the four symbols of Data Segment Sync which occur every NTSC line at the same place and in the same format. As interference into the NTSC victim channel, they stand out because of their regularity. They cannot be randomized because reliable Data Segment Sync detection depends on exact repetition for a number of data segments.

Dispersion is introduced to soften the repetitive character of the sync symbols. Since the sync is preceded and followed by random pulse signals, dispersion causes a portion of the preceding (or following) random signal to be superimposed on the non-random sync signal and so gives it a measure of randomness.

Receiver dispersion cancels the transmitter dispersion and spreads (disperses) any sharp transition of the NTSC cochannel interference into DSC-HDTV and so contributes to reducing the effect of the interference.

6.2.4 Low-Power Transmitter

Modern NTSC TV transmitters use a two-step modulation process. The first step usually is modulation on an I.F. carrier (which varies with manufacturer) which is the same for all channels. The RF Converter translates the SAW filtered output of the first stage to the final spectrum position.

This efficient method will be applied for DSC-HDTV as well.

It is shown in Section 7 that a DSC-HDTV transmitter with coverage comparable to an NTSC transmitter has 14.5 dB less power.

6.2.5 RF Carrier Frequency Offset

In extreme cochannel situations, the DSC-HDTV system is designed to take advantage of precise RF carrier frequency offset of the DSC-HDTV carrier to an NTSC cochannel carrier. This will probably only be necessary in three stations in the U.S. (Section 7.3.1).

Since the DSC-HDTV signal sends synchronizing information in an NTSC-like format, precise offset causes NTSC cochannel interference into the DSC-HDTV receiver to phase alternate from sync-to-sync. The DSC-HDTV receiver circuits average successive syncs to cancel the interference and make sync detection more reliable.

For DSC-HDTV cochannel interference into NTSC, the interference is noise-like and does not change with precise offset.

Although it might be postulated that a DSC-HDTV transmitter be located so as to experience equal interference from two worst-case cochannel NTSC stations (for example, in a three-way triangle), such a situation is so unlikely that the DSC-HDTV signal is assumed to have only one dominant NTSC cochannel. The DSC-HDTV cochannel pilot should be offset from the dominant NTSC picture carrier by an odd multiple of half NTSC-line rate. For ATTC testing, we propose a 57.5 f_H = 904.72 kHz offset.

For DSC-HDTV to DSC-HDTV cochannel interference, precise carrier offset prevents possible misconvergence of the adaptive equalizer. If perchance the two HDTV Data Field sync signals should fall within the same line time, the adaptive equalizer could misinterpret the interference as a ghost. To prevent this, for close HDTV-to-HDTV cochannel situations, we propose to use a carrier offset between two HDTV cochannels of $f_H/2$ = 7.867 kHz which causes the interference to have zero net effect in the adaptive equalizer.

6.3 RECEIVER

The Receiver block diagram is shown in Figure 6-19. The Tuner, IF and Detector are analog devices and are followed by the A/D converter.

At this point, syncs are recovered and the clocks are synchronized. The data signal is applied to the Interference Filter (Post-Comb filter) first and next to the automatic channel equalizer which receives the Data Field sync signal as a reference.

At this point the optimum data signal response is obtained and the signal is ready to be detected. The Slicer establishes the decision levels and the Demapper splits the single bitstream into four at 10.76/4 = 2.69 Mb/s.

The Digital Post Coder is the device that compensates for the mod-4 translation in the Transmitter Precoder if the reception is sufficiently free of NTSC cochannel interference so that the Interference filter (Post-Comb filter) is bypassed. The decision is based on actually measured error rates on Field Sync data.

The De-interleaver performs functions complementary to those of the

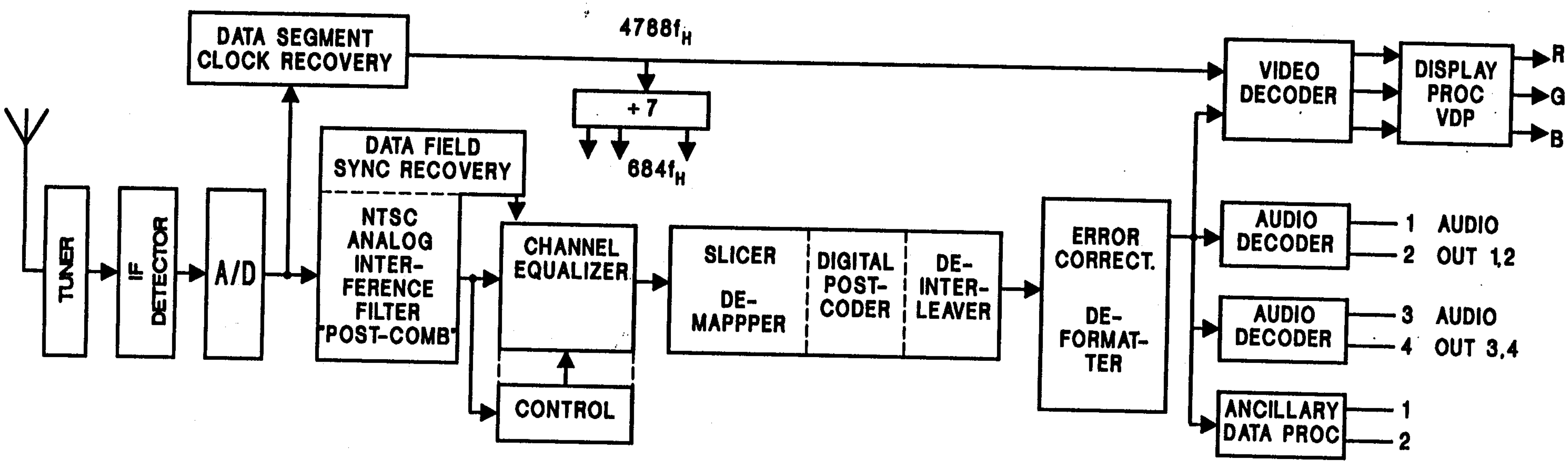


FIG. 6-19 RECEIVER BLOCK DIAGRAM

AT&T

ZENITH

transmitter in reverse order. First intra-segment de-interleaving is done, and inter-segment de-interleaving is next.

Error correction is the next function to be applied to all relevant data. Then the continuous bitstream has to be divided into the constituents of video, audio and ancillary data.

6.3.1 Tuner

The Tuner block diagram is shown in Figure 6-20. It is of the double-conversion type. The first IF frequency is 920 MHz, high enough above the band to provide skirt space for a bandpass filter to prevent the local oscillator from up-stream leaking and low enough for second harmonics of UHF channels to fall above of the first IF bandpass. Harmonics of cable channels could possibly occur in the first IF passband but are not a real problem because of the signal levels used in cable systems. Details are found in Section 6.3.1.1.

The first mixer is driven by a synthesized local oscillator (L.O.) signal. Its frequency is above the band; the mixer uses high-side injection.

The mixer is followed by an L-C filter functioning in tandem with the subsequent narrow 920 MHz bandpass ceramic resonator filter. The L-C filter provides selectivity against the harmonic and subharmonic spurious responses of the ceramic resonators. A 920 MHz IF amplifier is placed between the two filters. Delayed AGC is applied immediately following the first L-C Filter.

The second mixer is driven by the second L.O., which is controlled by the frequency and phase-locked loop (FPLL) [8]. The second mixer drives a constant gain 44 MHz amplifier.

The 920 MHz bandpass filter has a -1 dB bandwidth of 7 MHz. The SAW filter has the characteristic shown in Figure 6-14, top, and has the earlier mentioned group delay slopes to undo the dispersion (Section 6.2.3).

The Noise Figure of the tuner ranges from 7 to 9 dB over the entire VHF, Cable and UHF frequency bands.

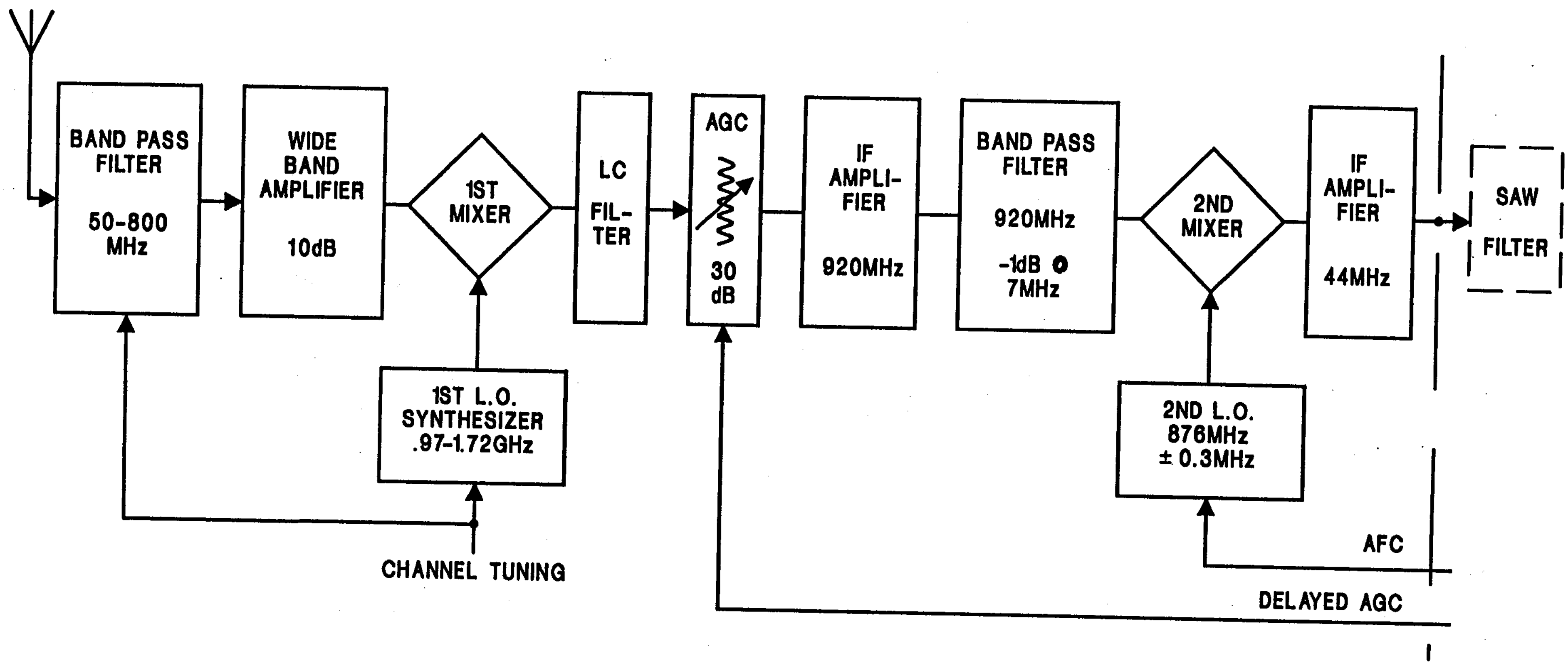


FIG. 6-20

DSC-HDTV RECEIVER TUNER
BLOCK DIAGRAM

AT&T

ZENITH

6.3.1.1 Interpretation of the Interference Chart

The main diagonal, 1,1 in Figure 6-21, represents the relationship between incoming RF and L.O. frequency which two frequencies have a constant difference of 920 MHz, the IF frequency.

The intersection of any other marked line and the main diagonal, 1,1, represents a potential 920 MHz beat which cannot be defeated by RF selectivity. For example, the line marked 2,4 represents an interference situation by the second harmonic of the L.O. and the fourth harmonic of the RF. The RF harmonic number in this case, however, is high enough not to present a serious problem for a double-balanced mixer with good linearity and with good cancellation of even order products. The same can be stated for the interference represented by the line marked 1,4.

The horizontal dashed lines represent higher harmonics of incoming RF. Note that up to the 4th harmonic none come from broadcast TV signals. They may come from other signals but those are mostly intermittent. Higher than the 4th harmonic the level is so far down as to not cause problems.

6.3.2 Carrier Recovery

The frequency of the first L.O. is synthesized and the third L.O. comes from a constant reference oscillator. Any drift or deviation from nominal has to be compensated in the second L.O.

The control for the second L.O. comes from the FPLL. The frequency loop gives the second L.O. a pull-in range of $\pm$ 300 kHz and the phase loop bandwidth is approximately 4 kHz.

The FPLL circuit operation can be summarized as follows, referring to Figure 6-22. The upper loop, containing the AFC Lowpass loop filter and the Amplifier/Limiter, acts on the frequency difference between the VCO and the incoming pilot, tending to reduce the frequency difference. Mixer M3 has a positive or negative average DC output depending on the polarity of the frequency difference. Filtered in the APC Lowpass Filter, the DC adjusts the VCO, which is the second L.O.

DSC-HDTV
INTERFERENCE CHART

1ST IF FREQUENCY = 920MHz

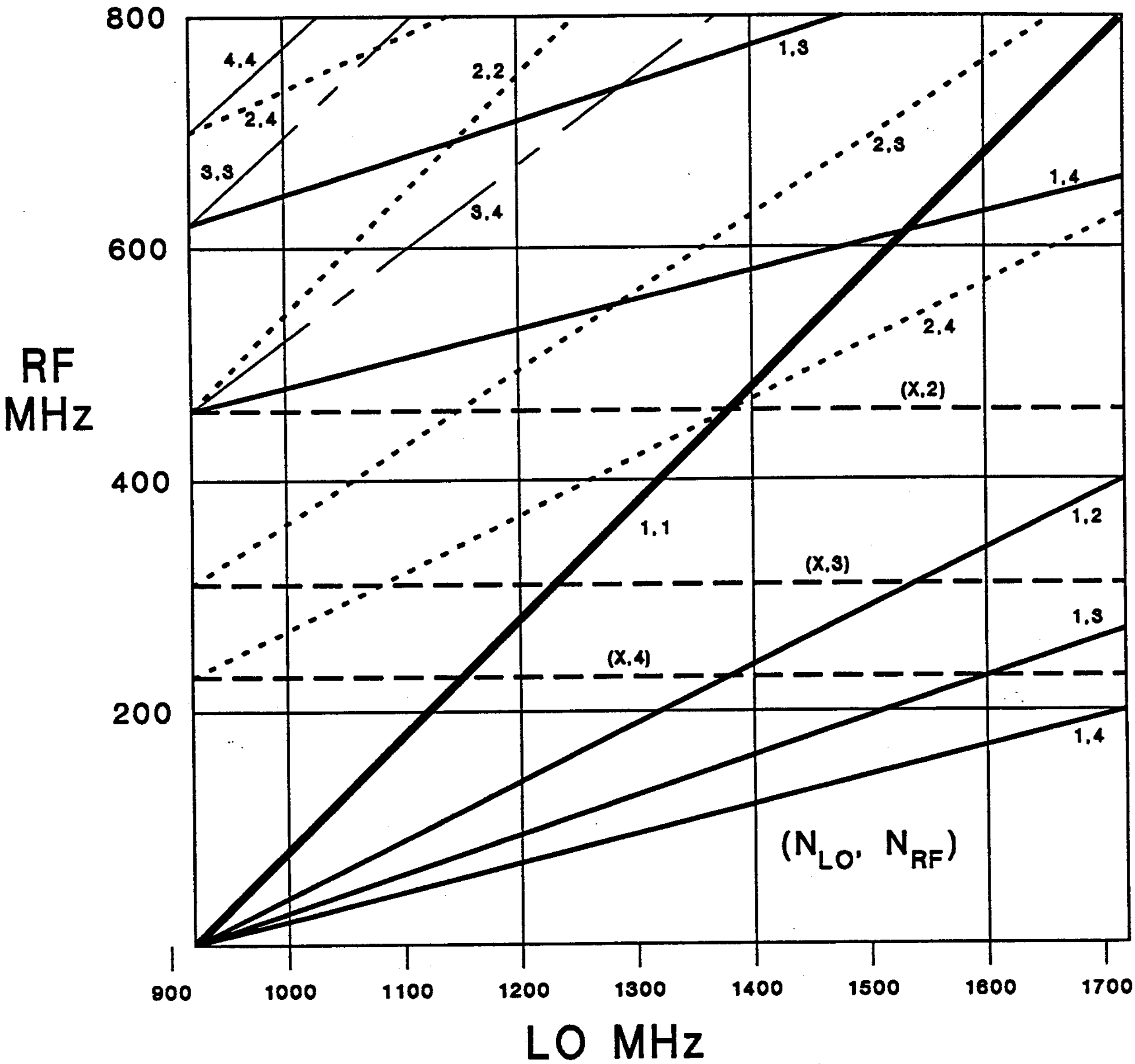


FIG. 6-21

DSC-HDTV RECEIVER
ANALOG PROCESSING

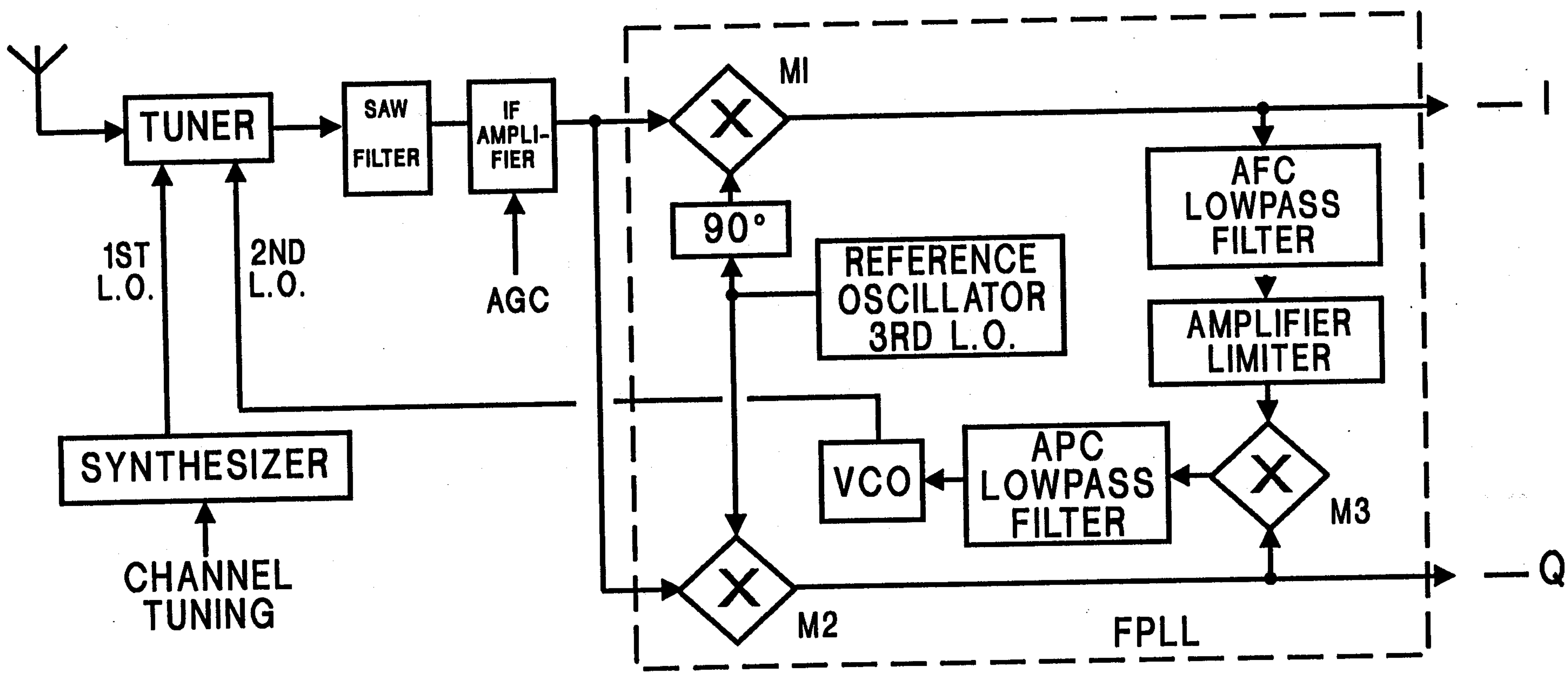


FIG. 6-22

AT&T

ZENITH

When the frequency difference comes close to zero, the APC (lower) part of the loop takes over. This is a normal phase-locked loop circuit. When frequency lock is reached, Mixer M1 yields a DC output which reaches M3 as a bias allowing the Quadrature (Q) output to pass to the APC Lowpass Filter. At that stage the AFC Lowpass Filter effectively limits the spectrum of the non-pilot part of the In-phase (I) signal in the loop. The Limiter output at phase-lock is a maximum, and since it is amplitude limited, noise and interference are eliminated. Now, only the APC portion of the circuit, the phase-locked loop, controls the local oscillator.

The FPLL combines a wide pull-in range with a narrow bandwidth range and stable operating points at 90 degrees and at 270 degrees - all necessary features for reliable operation under strong cochannel interference conditions.

6.3.3 Segment Sync Recovery and AGC

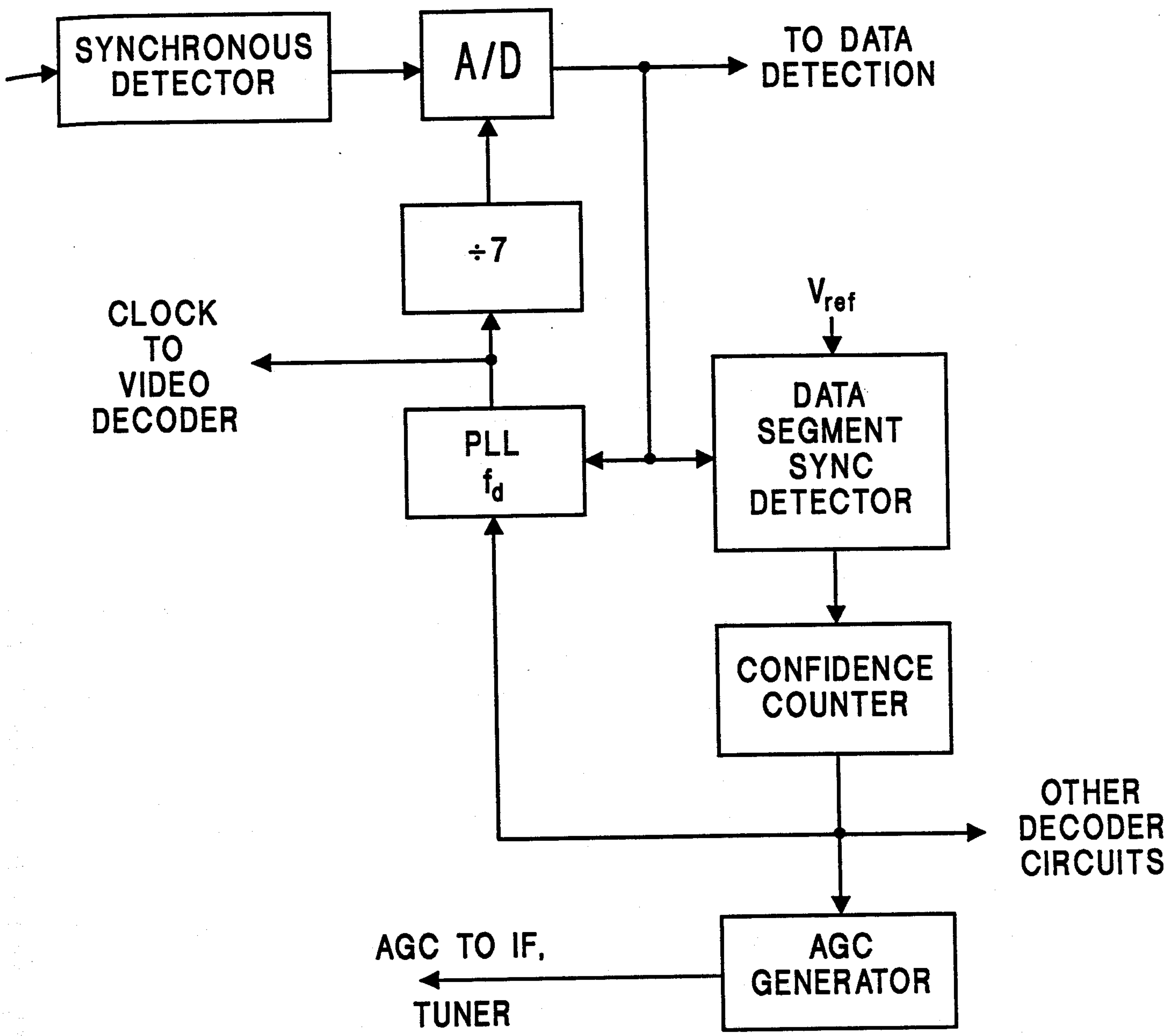
A block diagram of these circuits is shown in Figure 6-23. The A/D converter is the same as shown in Figure 6-19, the DSC-HDTV Receiver block diagram, as the block following the IF and Detector.

The A/D output feeds two separate circuits, one to synchronize the video data clock generator and the other to generate Data Segment sync.

The phase-locked loop circuit, PLL, has a controlled crystal VCO that nominally generates the video data clock of f_d = 75.3357 MHz. The frequency f_d is divided by 4788 which yields exactly f_H. The phase-locked loop circuit includes the digital equivalent of a discriminator. If the divided-down f_d does not equal the frequency of the incoming Data Segment sync, a correction is applied. The regenerated f_d is used in the Video Decoder.

The other receiver circuits are supplied with the transmission data clock of frequency f_t = 10.7622 MHz (= $f_d/7.$ = 684 f_H), as shown in Figure 6-19.

The second circuit fed by the A/D output regenerates the Data Segment sync. As mentioned in Section 6.1.4, referring to Figure 6-3, the repetitive nature of the data segment sync builds up the voltage in an integrate-and-dump circuit while all other random data do not build up. Once syncs begin to be detected, a confidence counter starts running. After a preset delay during which Data Segment pulses must have been appearing uninterruptedly, pulses are passed to the AGC circuit and back to the phase-locked loop circuit.



DSC-HDTV RECEIVER

VIDEO DATA CLOCK RECOVERY
SEGMENT SYNC RECOVERY
AGC GENERATION

FIG. 6-23

AT&T

ZENITH

The AGC voltage is applied to the second IF directly and "delayed" e.g., after passing a preset threshold, to the first IF.

6.3.4 Data Field Synchronization

The Data Field Reference signal timing is illustrated in Figure 6-4. The signal consists of a two-level pseudo-random sequence of 255 symbols repeated four times with the first and the last sequence truncated. Data Field sync detection is achieved by comparing the incoming signal data to the local lookup table data, data segment by data segment. A large output is produced on the data segment containing Field Sync signifying synchronization. Even under strong ghosted conditions of the incoming sequences, the correlation is high and Data Field sync detection is secure.

The polarity of the pseudo-random sequence determines whether field 1 or field 2 is detected.

The Data Field reference signal is also used as the reference for channel-equalization/ghost-canceling (Section 6.3.6) and during receiver lock-up in the process of deciding whether the Interference Filter (Post-Comb filter) is needed (Section 6.3.9).

6.3.5 Interference Filter and Digital Post-Coder

The interference rejection system is described in Section 6.1.8 with reference to Figures 6-5 through 6-8. As mentioned there, the option of an alternate path around the Post-Comb filter is provided in case cochannel NTSC interference is not a problem. This recovers the 3 dB loss in noise performance. In that case, the modulo-4 precoding (Figures 6-6 to 6-8) is complemented by modulo-4 digital post-coding. This post-coding is done in a mod-4 feed-forward network which uses mod-4 digital arithmetic rather than linear addition in the summer (Σ).

Which of the two circuits is used in any given situation is determined by another confidence circuit. If error-free decoding is indicated for a preset period of time, the Post-Comb is bypassed and the mod-4 Post-Coder is inserted.

If the Post-Comb is bypassed, the W1 data are 2-level and the W2 data are 4-level, as transmitted. Post-combing translates 2-level data into 3-level and 4-level into 7-level (see Figure 6-9).

6.3.6 Channel Equalizer

The Equalizer/Canceler is used to compensate for linear channel distortions, such as tilt and ghosts. The equalizer delivered for testing will use a Least-Mean-Square algorithm and adapts on the transmitted Data Field sync.

The equalizer filter consists of two parts, an 80-tap feedforward section followed by a 200 tap feedback section. The feedforward section operates as a linear FIR filter while decision feedback is used in the feedback section. The equalizer operates at the symbol rate of 10.76 MHz.

To aid convergence, the equalizer also includes an extra adder which is used to add or subtract a DC value to compensate for DC errors which can be caused by circuit offsets or non-linearities.

6.3.7 Slicer and Demapper

The incoming symbols, after cochannel interference filtering in the Post-Comb (if necessary) and after equalization, are analyzed in the Slicer. The De-Mapper expands the single bus with symbols at 10.76 Msymbol/s to 4 parallel buses at 2.69 Msymbol/s. The slicing process is complicated by three factors. One, the data come in different weight (W2 data require three slicing levels; W1 only one but not coinciding with a W2 slicing level), and secondly, the Post-Comb, if in action, requires double the number of slicing levels; six for W2 data and two for W1 data. Thirdly, the detection of erasures replaces every decision level by an erasure band. This effectively splits every decision level in two. This is shown on the right-hand side of Figure 6-9.

When the data reaches the Slicer, it is known whether or not the Post-Comb is in action, but, due to interleaving, it is not known whether the current symbol is W1 or W2. This is resolved by "soft slicing". Soft slicing treats all symbols as W2 and also automatically uses the doubled number of slicing levels when the Post-Comb is used, interpreting the output modulo-4. The signal levels in this part of the circuit are represented by multiple-bit binary numbers. By reducing the bit-count after slicing to four, however, enough information is passed to interpret the four bits as belonging to W1 or W2 data after de-interleaving.

The first data to reach the Slicer after a Data Field sync are known to be W1 data representing the Transmission Bit Map (TBM) and Video Control (Figure 6-2, Table II). Both constitute the "Global Segment" of Section 3.4.5. TBM identifies the weight of every segment pair of the field for de-interleaving. After de-interleaving, therefore, the final interpretation can be made of the video data and the four bits are now reduced to two.

6.3.8 Error Correction and De-Formatting

The data delivered by the De-Interleaver are binary data and (disregarding errors) are in the same form as delivered by the Data Formatter and Error Control section of the Transmitter (see Figure 6-13). At this point the data are in an appropriate form for a digital recording device.

The RS decoder performs the byte-error correction on a segment-by-segment basis, processing both byte errors, recognized as such, and erasures.

The Deformatter apportions the bits to their respective decoders.

The Video Decoder (Section 3) uses the recovered video data clock signal (f_d = 75.3357 MHz) and the detected and corrected digital video data. The latter is converted to digital luminance and chrominance signals. The Display Processor of Figure 6-19 (the Video Post-Processor in Figure 3-9) converts these digital signals into analog RGB signals for picture tube display.

6.3.9 Sequencing of the Receiver

The receiver incorporates a "universal reset" which initiates a number of "confidence counters" and "confidence flags" involved in the lock-up process. A universal reset takes place, for example, when tuning to another station or turning on the set.

A confidence counter counts the number of times a specific function has been successfully completed. If the count reaches a preset lower limit, a confidence flag is set. The confidence flag is a signal for the system to proceed to the next function. Build-up of the confidence count proceeds much faster than release; thus random signals will not prevent reaching confidence.

When initially tuning to a particular channel, the synthesizer in the tuner sets the approximate first L.O. frequency for that channel. A universal reset is also activated at this time.

Next, the DC adjustment is enabled. This matches DC of the A/D converter pulse sequence output with the DC offset needed for centering the signal samples in the digital dynamic range of the A/D converter. This range is larger than the nominal signal level to accommodate noise and/or interference.

At the same time, the frequency and phase-locked loop (FPLL) comes into action. The frequency loop acts to decrease the difference between the incoming pilot and the local reference frequency. The discriminator characteristic is so designed that sidelock to a strong NTSC cochannel carrier is not possible.

When frequency-lock is reached, the circuit switches to phase-lock operation. Phase-lock sets a confidence flag for the next operation.

The next operation is Data Segment sync detection. When properly detected, a confidence counter starts and a confidence flag is raised after a preset count. This enables AGC and video data clock recovery as explained in Section 6.3.3 and Figure 6-22. Proper AGC is not essential for Data Segment sync detection because the system is designed for gain to be on the high side without AGC. Data Segment sync may be limited in level but it is always recognized.

With AGC activated, the data symbols (as A/D converted) have the proper levels and the slicers can operate accurately. The next flag is set.

The Data Field sync detector is then activated. The data signal can be detected and the pseudo-random (PR) sequences of the Field Sync can be recognized. They are continually compared to the local sequence in a ROM. When lock is found, another confidence counter is started and, after a preset period, the next flag is set.

The Data Field sync (PR) sequence is used to compare the error before and after the Post-Comb. The Post-Comb input is compared to the two-level nominal sequence in ROM, and measured errors are accumulated. The Post-Comb output is compared to the three-level nominal sequence in ROM and again the error is accumulated. Whichever error is smallest at the end of a preset period determines whether the Post-Coder or the Post-Comb is active. If the three level error is smallest, the Post-Comb is active and the Post-Coder is bypassed. If the two level error is smallest, the Post-Comb is bypassed and the Post-Coder is activated.

Next, the Channel-Equalizer is enabled. The Field Sync sequence in ROM (either two-level if the Post-Comb is not active or three-level if it is) is used as a training signal for adjusting the taps. If no more tap adjustments take place or after a preset maximum time period, the last confidence flag is raised. Actual data detection and error correction start.

SECTION 7
SPECTRUM AND INTERFERENCE ISSUES

7.1 Introduction

In this section the noise and interference limitations of digital HDTV systems are examined in general and with particular application to the new DSC-HDTV system.

It is generally agreed within the ATS Advisory Committee that the new High Definition Television Service will be an interference-limited service:

- Simulcast HDTV demands cochannel spacing as low as 100 miles because of the 6-MHz of extra spectrum required for each present Television Broadcast Station [9].
- The maximum power levels for HDTV will be determined by the tolerable cochannel and Taboo channel interference penetrations into the NTSC service areas.
- The usable HDTV service area in turn will be determined by the visibility of cochannel and Taboo channel interferences from NTSC and other HDTV channels. Of this the cochannel interference from NTSC will be the dominant factor.
- This usable HDTV service area will have to be large enough to be commercially attractive to broadcasters and at least comparable to NTSC service areas.

(All examples in Section 7 are restricted to UHF reception for ease of understanding.)

7.2 Purpose of ATTC Interference Tests

The end objective of the Advanced Television Test Center (ATTC) generating interference D/U (ratio of Desired signal level to Undesired signal level) data at various desired HDTV receiver input levels is to determine interference and noise limited service contours for the proposed HDTV systems under common spectrum and interference constraints.

The maximum HDTV Effective Radiated Power (ERP) needs to be determined from the allowable cochannel interference into the NTSC service and the cochannel spacing. The conversion from D/U interference data (taken at various desired HDTV signal levels) to geographic terms of service and interference contours can be done after the interference criterion and other "HDTV Planning Factors" are agreed to.

A similar conversion has to be done for Adjacent and Taboo channel interferences. PS/WP-3 of the Advisory Committee has under study how to process this conversion of ATTC data. Since this task is not yet completed, only preliminary projections can be made based on certain assumed criteria and known or projected systems' performance.

7.3 Available Service Area Determinants

7.3.1 Minimum Cochannel Spacing

Analysis of the PS/WP-3 data shows that for 100% accommodation of all present NTSC stations with an additional 6 MHz simulcast channel three stations in the U.S. will have a cochannel spacing between 100 and 112 miles while more than 99.8% of all stations will be spaced more than 112 miles.

We thus have used a minimum cochannel spacing of 112 miles to determine HDTV transmitted power and available HDTV service area.

7.3.2 Derivation of Maximum HDTV ERP

Lacking an input from PS/WP-3, the following approach is assumed:

Two NTSC cochannel transmitters at 155 miles (Zone I) minimum spacing, 37 dBk ERP and 1250 feet HAAT reach each other's 28 dB D/U linear penetration point at 15 miles inside the Grade B (57 mile) contour. This corresponds to 14.2% of the service area not having the minimum 28 dB protection. F(50,50)[1)] propagation for the Desired, F(50,10) for the Undesired signal and 6 dB front-to-back ratio (F/B) for the receiving antenna is assumed.

Given that the above cochannel interference condition into the NTSC service is also tolerable (in the worst case) for digital HDTV at 112 mile spacing, then it can be shown that the ERP of the HDTV transmission would have to be reduced by 14.5 dB to 22.5 dBk.[2)]. For cochannel spacings larger than 112 miles the allowable ERP could, of course, be increased, as would the service area.

7.3.3 Receiving Antenna Characteristics

Unlike what was done for the NTSC Planning Factors, it is assumed, particularly in the fringe region of HDTV reception, that a certain advantage can legitimately be taken from the front-to-back ratio of receiving antennas.

In absence of directives from PS/WP-3 we have assumed for the examples in this document a variable F/B ratio of 0 to 15 dB depending on the location of the receiving installation with respect to the Desired HDTV and the Undesired NTSC transmitters as shown in the examples. This can be looked upon as a first order approximation of antenna directivity. An Antenna Gain of 10 dB and a Downlead Loss of 4 dB were used to yield a net gain of 6 dB.[3)]

[1)] F(1,t) refers to the FCC NTSC Propagation Characteristics exceeded at 1% of the locations for t% of the time.

[2)] While NTSC ERP is expressed as peak power during synchronization pulses, the HDTV ERP differs from the 2-1-91 Technical Description of DSC-HDTV by 2.5 dB only because a peak rather than average power reference was then used for DSC-HDTV.

[3)] These values are tentative and somewhat different from those used in the Pre-Certification document because of more recent input from PS/WP-3.

7.3.4 Definition of HDTV Service

The grades of service in NTSC (City Grade, Grade A and Grade B) are defined by the NTSC Planning Factors, including minimum fieldstrengths expected at 50% of receiving locations for 50% of the time.

Absent a firm recommendation, the following propagation characteristics have been assumed for the examples in this document:

for all Desired NTSC Signals: F(50,50)
for all Desired HDTV Signals: F(50,90)
for all Undesired Signals: F(50,10).

A recommended definition of HDTV service will be established by the Advisory Committee on Advanced Television Systems (ACATS), with inputs from the results of the Non-Expert Subjective viewing tests to be conducted by Advanced Television Evaluation Laboratory (ATEL) in Canada. This definition must take into account some criterion of impairment that can be considered "passable but not objectionable".

Inasmuch as this criterion will be a function of both Desired and Undesired program content and will lie somewhere between the 4-VSB and 2-VSB Carrier-to-Noise ratios and Desired-to-Undesired ratio thresholds, only the range between these thresholds and the corresponding service areas can be presented.

7.3.5 Receiver Noise Figure

A 10 dB (UHF) receiver noise figure is assumed for the examples in this document. This is readily achievable with present day tuner technology.

7.3.6 Carrier-to-Noise Ratio Threshold

The DSC-HDTV system does not exhibit the steep carrier-to-noise ratio (C/N) threshold typical for digital systems where within a very small input C/N range the output changes from perfect to unusable. Instead, as explained in more detail in the previous sections, the DSC-HDTV system is a bi-rate system that encodes the most important parts of the HDTV image as W1 symbols and the remainder as W2 symbols.

This results in extension of the noise threshold by up to 6 dB depending on the content of the desired DSC-HDTV image.

Inasmuch as this will result in a variable threshold and service area, the examples are based on the conservative C/N and D/U thresholds measured for 4-VSB with the extension effect becoming an extra margin of service. The measured 4-VSB C/N threshold is 16 dB while the 2-VSB C/N threshold is 10 dB.

7.3.7 Noise Limited DSC-HDTV Service Area

The Noise Limited Service Area (NLA) can be determined from the C/N Threshold, the receiver Noise Figure, the F(50,90) Propagation Characteristics, and the net Antenna Gain.

For the above values, this results in a primary Noise Limited (Service) Contour (NLC) of 53 miles, only slightly less than the Grade B contour (at 57 miles) of a 14.5 dB higher ERP NTSC emission at the same antenna HAAT (1,250 feet).

A secondary Noise Limited (Service) Contour results at 59 miles due to the 2- VSB threshold.

7.3.8 NTSC-to-HDTV Cochannel Interference Ratio Threshold

The bi-rate coding feature of the DSC-HDTV system mentioned above also greatly extends the D/U interference threshold. Again, the degree of extension is dependent on the desired HDTV program as well as the undesired NTSC modulation and may vary from 0 to 6 dB.

The Service Area examples are based on the measured D/U ratio for 4-VSB, which is near 0 dB, while the 2-VSB D/U threshold is as low as -6 dB.

7.4 Analysis of DSC-HDTV Interference Data

As explained in detail in Section 6.1.10.1, the pilot level in DSC-HDTV varies depending on the relative 4-VSB and 2-VSB data content. Experiments have shown that the presence or absence of the pilot does not affect the perceived interference into NTSC or other DSC-HDTV transmissions. The D/U and C/N ratios used in this section thus are based on data signal power and exclude the pilot.

Figure 7-1 shows the NTSC reference condition of two UHF 37 dBk transmitters at 1250 feet HAAT and 155 mile (minimum) cochannel spacing. The Grade B contour is 57 miles, Grade A is 44 miles. With a receiving antenna front-to-back ratio of 6 dB, the 28 dB D/U contour penetrates 15 miles (two miles inside the grade-A contour at 42 miles) and results in 14.2% of the Grade B service area having less than 28 dB protection.

Figure 7-2 shows the condition of an undesired DSC-HDTV transmitter of 22.5 dBk and at 1250 feet HAAT, located 112 miles from a 37 dBk 1250 feet HAAT NTSC cochannel transmitter.

The interference caused by DSC-HDTV into NTSC reception is noise-like. An ERP of 22.5 dBk results in the 30 dB D/U[1)] contour crossing the axis between transmitters at 42 miles from the NTSC transmitter, again assuming a 6 dB front-to-back ratio. This is the same as the NTSC-to-NTSC reference of Figure 7-1 and corresponds to 12.7% of the grade-B service area having less than 30 dB protection.

The cochannel interference from NTSC into DSC-HDTV is shown in Figures 7-3 and 7-4.

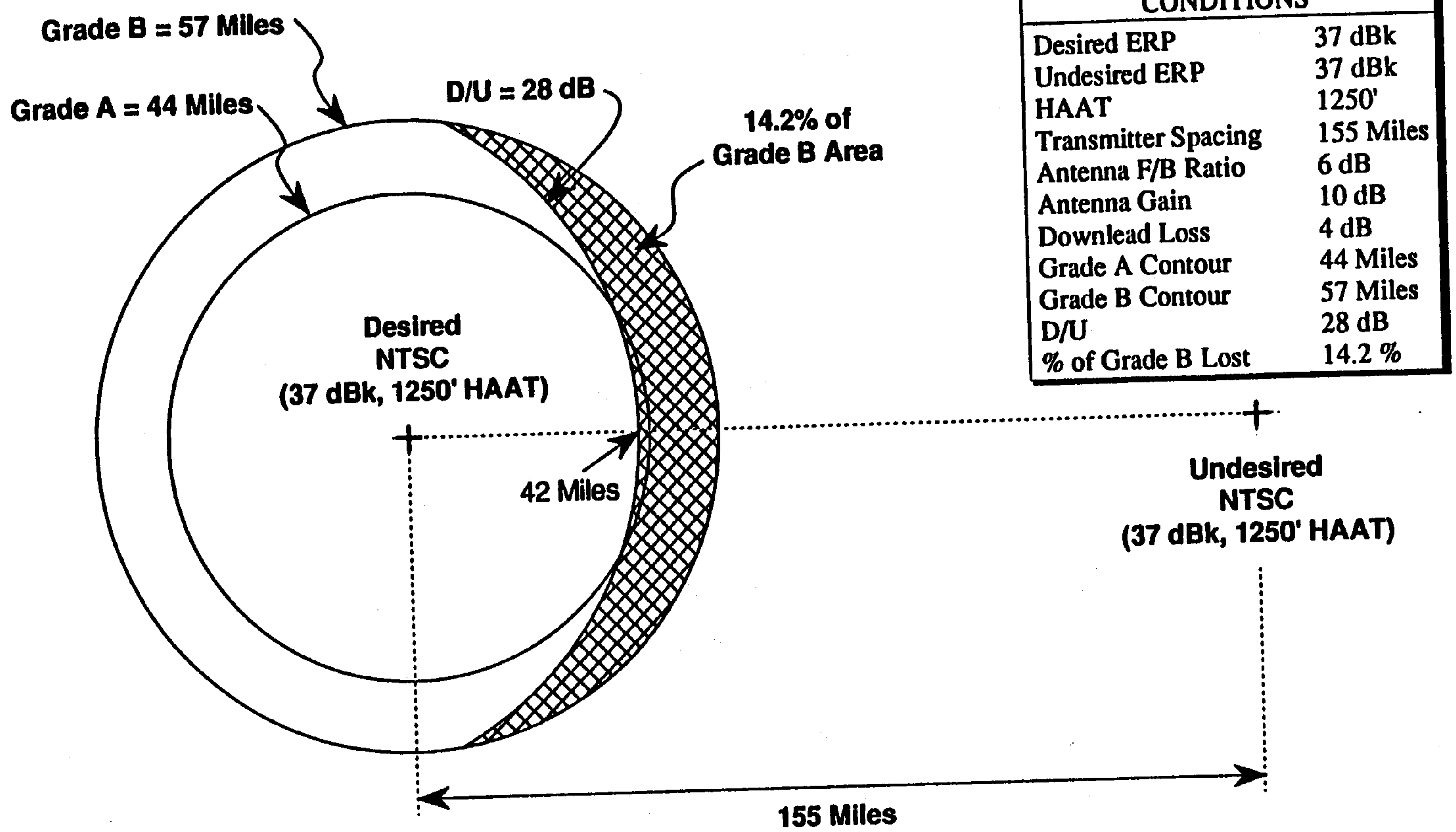
The Noise and Interference limited service contours in Figure 7-3 pertain to the 4-VSB performance parameters of:

C/N = 16 dB

D/U = 0 dB

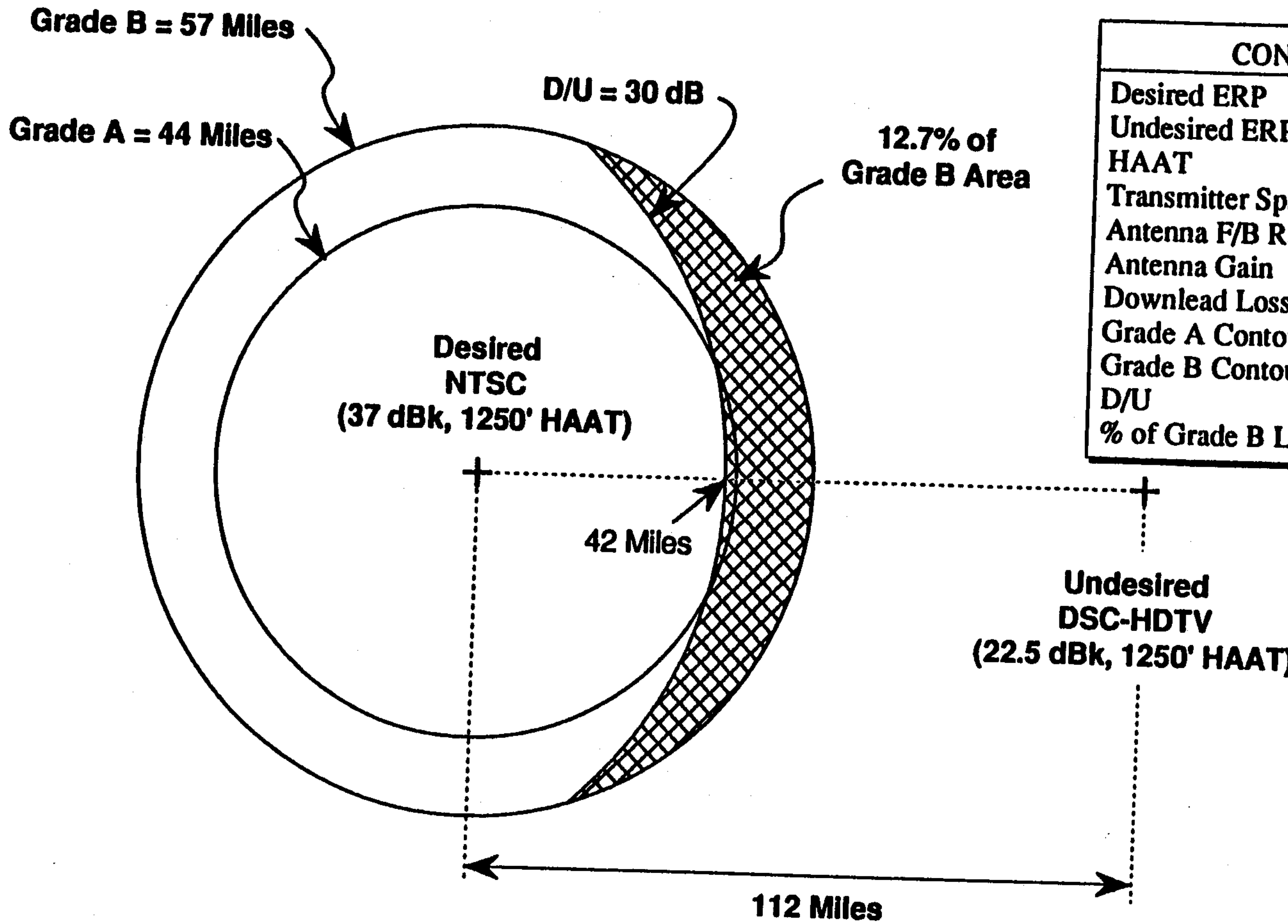
The 4-VSB Noise Limited Service Contour (NLC) is 53 miles (only slightly less than the NTSC Grade B contour).

[1)] The 30 dB D/U criterion is essentially the same as the 28 dB criterion used in the 2-1-91 Technical Description of DSC-HDTV because of 2.5 dB difference in the chosen DSC-HDTV power references (peak versus average).



CONDITIONS	
Desired ERP	37 dBk
Undesired ERP	37 dBk
HAAT	1250'
Transmitter Spacing	155 Miles
Antenna F/B Ratio	6 dB
Antenna Gain	10 dB
Downlead Loss	4 dB
Grade A Contour	44 Miles
Grade B Contour	57 Miles
D/U	28 dB
% of Grade B Lost	14.2 %

FIG. 7-1
COCHANNEL INTERFERENCE
NTSC into NTSC at 155 Miles



CONDITIONS	
Desired ERP	37 dBk
Undesired ERP	22.5 dBk
HAAT	1250'
Transmitter Spacing	112 Miles
Antenna F/B Ratio	6 dB
Antenna Gain	10 dB
Downlead Loss	4 dB
Grade A Contour	44 Miles
Grade B Contour	57 Miles
D/U	30 dB
% of Grade B Lost	12.7 %

FIG. 7-2

COCHANNEL INTERFERENCE

DSC-HDTV into NTSC at 112 Miles

CONDITIONS	
Desired ERP	22.5 dBk
Undesired ERP	37 dBk
HAAT	1250'
Transmitter Spacing	112 Miles
Antenna F/B Ratio	15 to 0 dB
Antenna Gain	10 dB
Download Loss	4 dB
Desired Propagation	F(50,90)
Undesired Propagation	F(50,10)
4-VSB C/N	16 dB
4-VSB D/U	0 dB
4-VSB NLC	53 Miles
4-VSB ILC Minimum	45 Miles
4-VSB % Lost	5.9 %

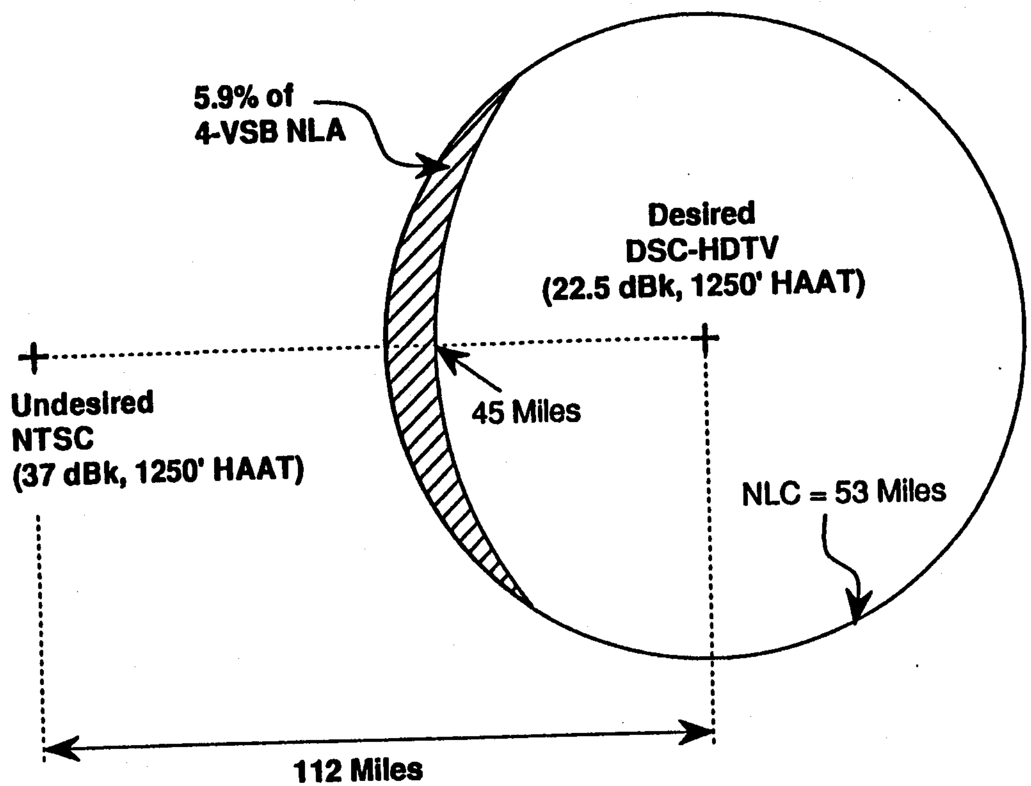


FIG. 7-3
COCHANNEL INTERFERENCE
NTSC into DSC-HDTV at 112 Miles (4-VSB)

CONDITIONS	
Desired ERP	22.5 dBk
Undesired ERP	37 dBk
HAAT	1250'
Transmitter Spacing	112 Miles
Antenna F/B Ratio	15 to 0 dB
Antenna Gain	10 dB
Download Loss	4 dB
Desired Propagation	F(50,90)
Undesired Propagation	F(50,10)
4-VSB C/N	16 dB
2-VSB C/N	10 dB
4-VSB D/U	0 dB
2-VSB D/U	-6 dB
4-VSB NLC	53 Miles
4-VSB ILC Minimum	45 Miles
4-VSB % Lost	5.9 %
2-VSB NLC	59 Miles
2-VSB ILC Minimum	49 Miles
2-VSB % Lost	9.9 %

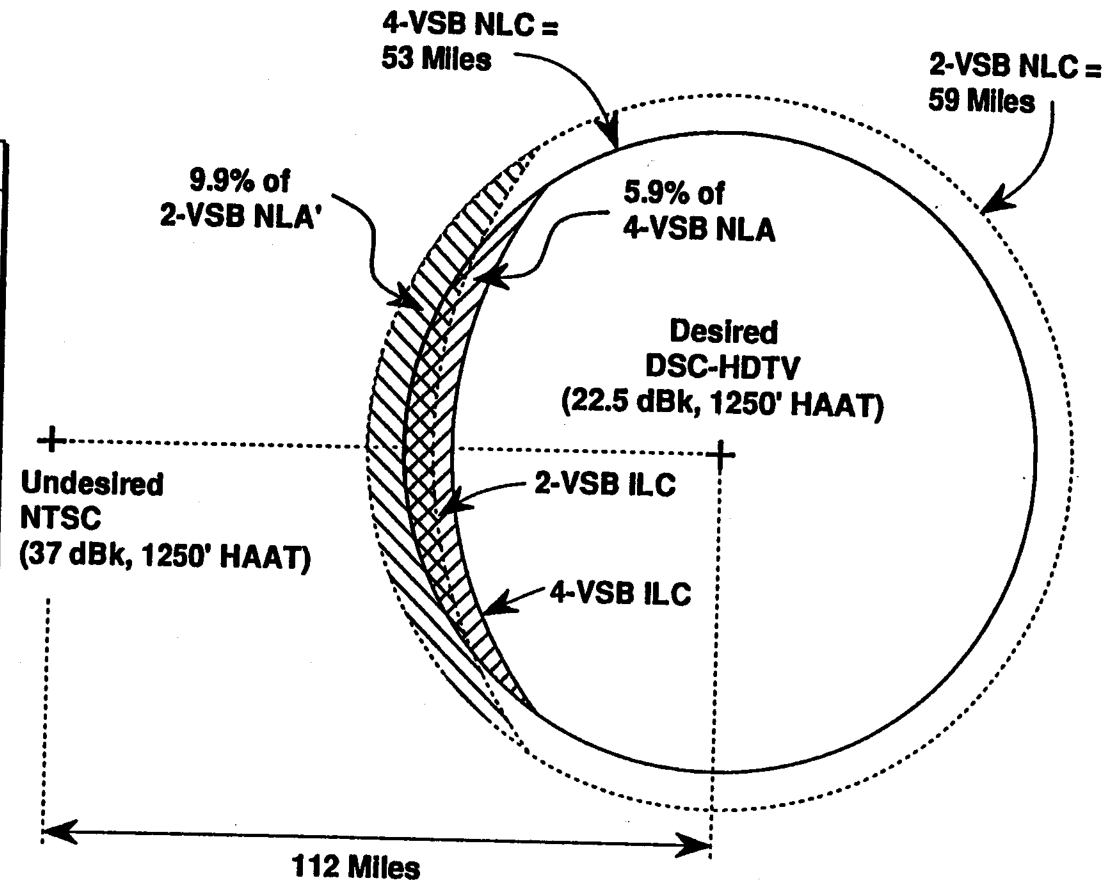


FIG. 7-4
COCHANNEL INTERFERENCE
NTSC into DSC-HDTV at 112 Miles (2-VSB and 4-VSB)

The 4-VSB Interference Limited Service Contour (ILC) due to the NTSC cochannel station 112 miles away, penetrates to 45 miles from the DSC-HDTV transmitter and results in approximately 6% of the primary NLA having less than 0 dB protection.[1)]

Figure 7-4 shows the same transmitter conditions as Figure 7-3, but the NLC and ILC due the 2-VSB mode are also shown for performance parameters of:

C/N = 10 dB

D/U = -6 dB.

The NLC is now extended by 6 miles to 59 miles and the minimum ILC to 49 miles from the DSC-HDTV transmitter.

Although perfect DSC-HDTV reception cannot be guaranteed to always occur between the 4-VSB and 2-VSB threshold contours, complete loss of video, audio or data does not occur. Error free audio (2 channels) will be received. Noise bursts or momentary signal fades causing the input signal to momentarily fall below the 2-VSB threshold result in a momentary signal freeze rather than a complete loss of video. It is clear that the birate feature of the DSC-HDTV system will go a long way to allay fears on the part of some with regard to the brick-wall loss of service character of digital systems.

7.5 Adjacent Channel and Taboo Interference into DSC-HDTV

Other than the predominant NTSC cochannel interference, the interference from NTSC adjacent channels is expected to be the most limiting factor on HDTV TV service area.

[1)] As the F/B ratio varies around the NLC perimeter from 15 to 0 dB, the shape of the ILC will cease to be a simple crescent, depending on the D/U threshold. For the extremely low DSC-HDTV D/U thresholds of 0 dB (4-VSB) and -6 dB (2-VSB) at 112 miles or larger spacing, this effect is not yet significant. It can be shown that, everything else being equal, a D/U threshold of 4.5 dB will result in the ILC completely encircling the Desired transmitter.

Interference from NTSC transmissions into HDTV channels with a Taboo relationship will be a function of the HDTV tuner design as well as the system's Carrier-to-Interference-Ratio Thresholds. Satisfactory reception of DSC-HDTV can be achieved in the presence of NTSC Taboo channels with economical tuner designs.

Adjacent NTSC channel transmitters should be colocated within 10 miles with HDTV transmitters or be located approximately 10 miles outside the NLC. Requirements on the control of out-of-band emission by the NTSC transmitter may become somewhat more severe in some installations but are expected to be feasible

7.6 Adjacent Channel and Taboo Interference into NTSC

Interference into NTSC from DSC-HDTV transmitters on Taboo and adjacent channels can be expected to be significantly less than interference from NTSC into NTSC. This is the result of the absence of picture and sound carriers and the chroma subcarrier in DSC-HDTV, and of the reduced power requirements for the same coverage. In addition the visibility and objectionability of DSC-HDTV interference into NTSC due to non-linearities in the NTSC tuner will be greatly reduced because of the absence of modulated beats or sync and picture envelopes in the caused interference.

7.7 DSC-HDTV to DSC-HDTV Cochannel Interference

Spectrum Scenario Studies of PS/WP-3 indicate that the minimum HDTV to HDTV spacing will be around 125 miles.

The cochannel D/U threshold for DSC-HDTV into DSC-HDTV is approximately the same as the random noise C/N threshold for both the 2-VSB and the 4-VSB modes.

The example in Figure 7-5 shows this worst-case situation. At the indicated power levels and thresholds, only 2.9% of the 4-VSB noise-limited service area and 4% of the 2-VSB NLA is interference limited.

CONDITIONS	
Desired ERP	22.5 dBk
Undesired ERP	22.5 dBk
HAAT	1250'
Transmitter Spacing	125 Miles
Antenna F/B Ratio	15 to 0 dB
Antenna Gain	10 dB
Downlead Loss	4 dB
Desired Propagation	F(50,90)
Undesired Propagation	F(50,10)
4-VSB C/N	16 dB
2-VSB C/N	10 dB
4-VSB D/U	15 dB
2-VSB D/U	10 dB
4-VSB NLC	53 Miles
4-VSB ILC Minimum	48 Miles
4-VSB % Lost	2.9 %
2-VSB NLC	59 Miles
2-VSB ILC Minimum	52 Miles
2-VSB % Lost	4.0 %

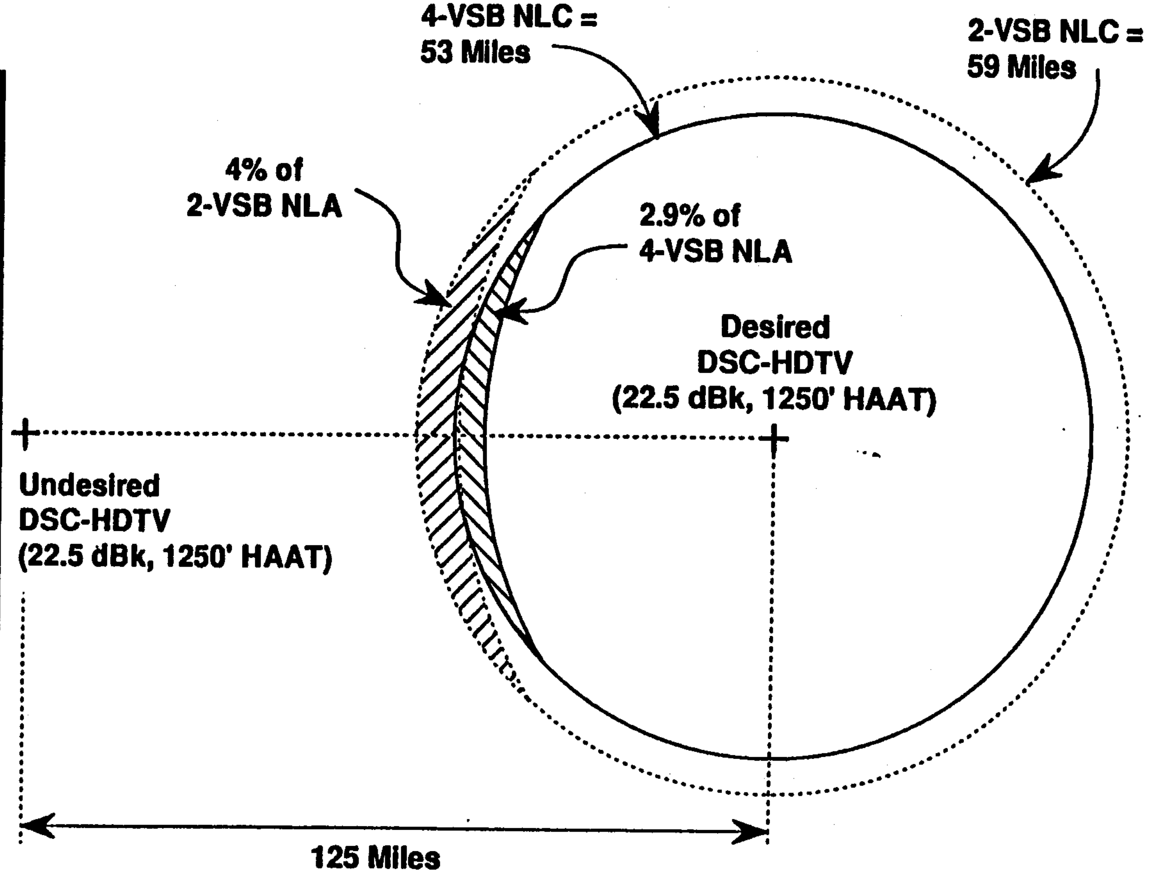


FIG. 7-5
COCHANNEL INTERFERENCE
DSC-HDTV into DSC-HDTV at 125 Miles

SECTION 8
OTHER DISTRIBUTION MEDIA

8.1 Cable Delivery System

8.1.1 Channel Capacity

There are considerably fewer unusable ("Taboo") channels in cable than there are in terrestrial broadcasting. The low power of DSC-HDTV signals, however, facilitates procuring the few extra channels initially needed for simulcasting DSC-HDTV and NTSC versions of the same program. Some extra channels may be found in the cable channels overlapping sensitive aeronautical channels. The power in those channels is often limited due to the common problem of cable leakage. A few extra channels may also be found immediately above the highest channel usable for NTSC signal delivery. Even though cable systems have the lowest C/N ratio in that frequency range, DSC-HDTV can operate there due to a lower C/N ratio requirement than NTSC.

At a later, second, stage more channels may become available through the use of "Fiber Backbone". It is noted that fiber-optic television transmission will also benefit from the low power and the absence of carrier and subcarriers of DSC-HDTV.

8.1.2 Intermodulation

The upper signal level in current NTSC cable systems is limited by composite triple-beat intermodulation products. A few DSC-HDTV channels add negligible amounts of intermodulation and no triple-beat due to the lower power and the absence of carriers. It is estimated that the intermodulation contribution of a few DSC-HDTV channels is insignificant.

8.1.3 Encryption

The Cable operator may decide to deliver DSC-HDTV as a premium service, which requires encryption of the digital data signal. Digital encryption, when properly decrypted, has the advantage of retaining signal quality. Repeated encryption/decryption cycles (concatenation) are possible without loss of quality.

The Conditional Access box illustrated in Figure 8-1, performs the decryption. Note that the box is placed subsequent to the Video Cassette Recorder (VCR) branch points. The advantage of this arrangement is discussed in the VCR section, below. Since a new service is involved, connectors for decryption (and for VCR) can be made integral parts of the receiver, obviating the need for RF-type of decoders (and RF-type of VCR's).

8.2 The Satellite Delivery System

The satellite link may use frequency modulation or QAM with its own overhead.

Regardless of whether the satellite link is used for network or cable feed or for home delivery of TV programs, the DSC-HDTV signal format has two important advantages over the NTSC format.

As a first advantage, synchronization signals and conventional video frequencies are replaced by a noise-like signal which results in a narrower and a more symmetrical spectrum. Reduced FM deviation is possible which causes less distortion and improved signal-to-noise ratio and extends the threshold. Secondly, noise addition is not noticeable in DSC-HDTV but is in NTSC. In NTSC it is the color in particular that suffers.

8.3 VCR Recording and Playback

The VCR requires and delivers FM signals for NTSC; for DSC-HDTV it is FM or QAM or baseband. As mentioned in relation with Figure 8-1, the DSC-HDTV receiver can supply baseband connectors for the VCR and no RF version is needed.

The VCR unit is of a complexity somewhat comparable to that of a current

DIGITAL SC-HDTV RECEIVER
MULTI-MEDIA OPERATION

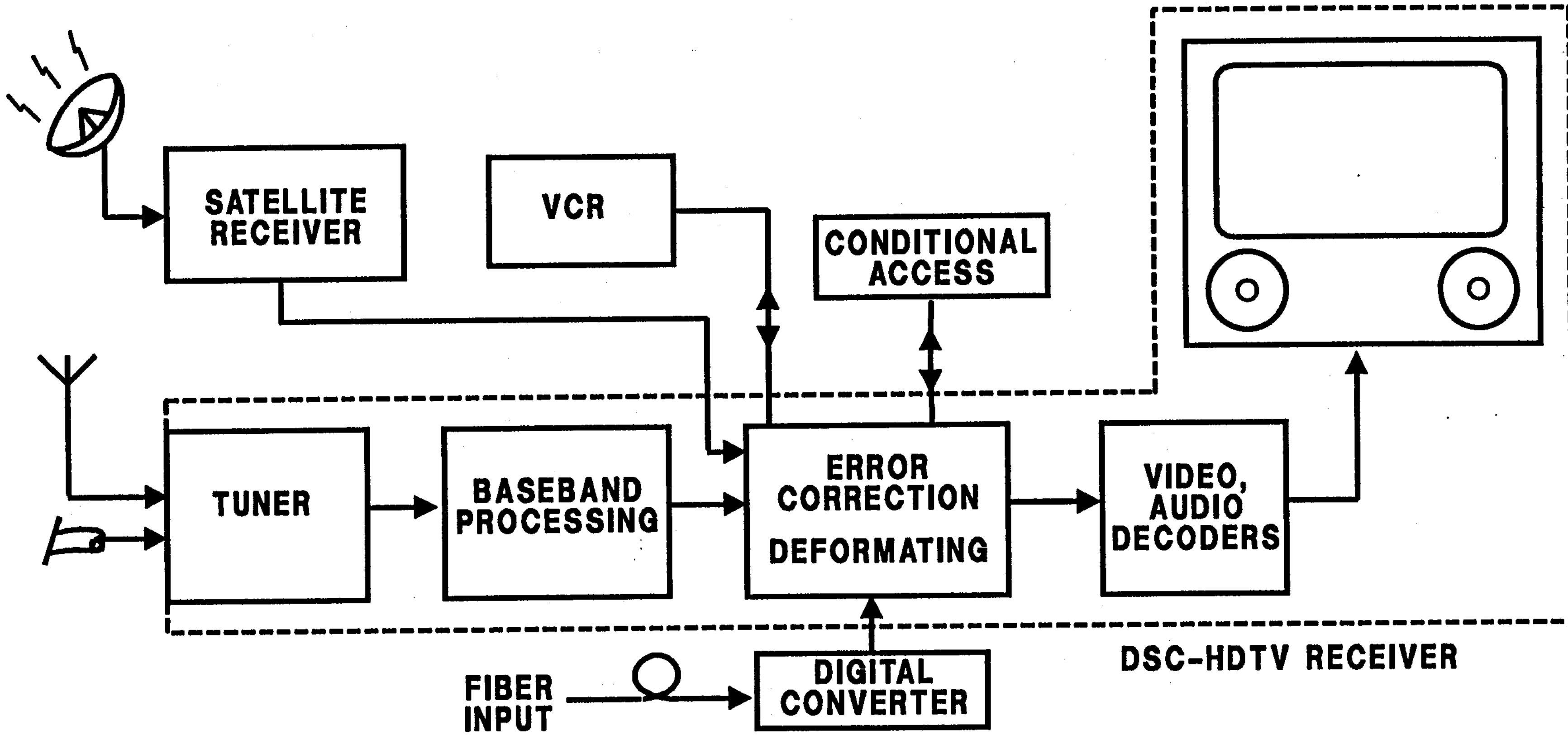


FIG. 8-1

AT&T ZENITH

S-VHS unit. Electronically, it can be simpler since no subcarriers have to be processed and neither encoding nor decoding processing is required.

Reference was made above to the relative placement of VCR and Conditional Access box, as illustrated in Figure 8-1. The consequence of this placement is that no decrypted recordings can be made of encrypted signals regardless of the source of the encrypted program. This deterrent against piracy is an incentive to program producers to simultaneously release programs for VCR and for pay-cable. Current practice is to delay pay-cable release to the dismay of the pay-cable operator.

SECTION 9
COMPUTER AND ALTERNATE MEDIA APPLICATIONS

9.1 Computer Friendliness

The rapid proliferation of video coding and communication applications is creating a demand for an HDTV standard which is computer friendly and can be easily adapted to meet the requirements of alternate media. In the future, improvements in source and display resolution will result in a need for extensions of the DSC-HDTV system. The DSC-HDTV system source specifications and encoder were designed with extensibility and interoperability in mind.

The use of progressive scanning and square pixels facilitates resampling to other scanning formats and computer workstation image manipulation. Computer graphics workstations can display the 787.5 format directly with no need for pixel resampling. Extra space on the workstation display is conveniently available for text or control windows. The basic display format itself is useful for a wide range of computer applications. Where up/down conversion is required, filtering and resampling are greatly eased by progressive scan and square pixels. For applications which require frame rate conversions or still frame capture, progressive scan is again ideal. And, of course, progressive scan provides flicker-free display of computer generated text and graphics even when the common practice of displaying single-line-width objects is employed.

9.2 Extensibility and Scalability

The video encoder and decoder architectures are inherently modular and the compressed video data format is designed to accommodate virtually any ratio of intra-frame to inter-frame data. As a result of this high degree of flexibility, the video encoder can be easily scaled or extended for:

- higher quality pictures given higher bit-rate transmission media
- lower cost encoder given higher bit-rate media and/or less complex

source material

- compressed video transmission over packet networks

Extension to a high definition studio compression standard has already been discussed in Section 2. This case involves removing all frame-to-frame processing, thus providing high-quality individual frames for special effects processing, with a bit rate on the order of 200 Mb/s.

Less demanding business video and camcorder applications could use a simpler encoder than the broadcast system. Complexity of the encoder can easily be reduced in applications which have either lower source bandwidth or higher transmission or storage capacity. In all these cases, the DSC-HDTV encoder can be easily scaled to provide the appropriate combination of encoder simplicity and picture quality.

Another simple extension of the DSC-HDTV video encoder can meet the growing need for high quality video transmission over packet networks. The packet-like format used in the DSC-HDTV video coder is easily converted to standard packet formats by a simple redefinition of the ample header space used for terrestrial broadcasting and cablecasting. The priority ordering of data segments and the existing segment error concealment algorithms in the DSC-HDTV video encoder then provide packet loss concealment.

Eventually, the demand for ever higher picture quality will produce a need for additional video compression standards which, ideally, would be compatible extensions of the HDTV standard. A higher resolution extension of the DSC-HDTV system can be conveniently implemented by decomposing the higher resolution frames into HD frames and high frequency residual frames by straightforward filtering techniques. The standard DSC-HDTV system codes the HD frames and a simple augmentation encoder codes the residual signal. The decomposing, merging and filtering are greatly simplified due to progressive scan and square pixels in the DSC-HDTV system.

9.3 Summary

The DSC-HDTV source standard and video compression system include a unique set of features that maximize picture quality within the available broadcast channel bandwidth, but also allow simple trade-offs or extensions for other applications. These applications can be accommodated by:

- Direct import of the square-pixeled DSC-HDTV picture into workstation displays.
- Use of the flexible packet-like format of the DSC-HDTV signal.
- Dropping inter-frame coding for applications demanding transparent quality of individual frames in a compressed signal.
- Trade-offs among inter-frame data, intra-frame data and over all bit rate and quality.
- Compatible extension of the basic DSC-HDTV coder.

SECTION 10
EXTENSION TO NON-NTSC BROADCASTING SYSTEMS

The Simulcast principle for HDTV is also applicable in other countries and regions of the world where PAL and SECAM standards are used instead of NTSC, and where channel bandwidths and intercarrier sound frequency spacings are different from CCIR system M as used in North America.

The desirability of terrestrial HDTV transmitter colocation with an adjacent existing television broadcast allocation is true for other countries outside of North America.

10.1 Interference Filter

The application of a receiver NTSC into HDTV cochannel interference filter (Post-Comb) is feasible for terrestrial HDTV broadcasting in other countries even though PAL/SECAM color subcarrier frequencies are different than used in NTSC and even though channel bandwidths and intercarrier sound frequency spacings are different than those used in the U.S.A.

For most of the countries not in North America, the channel bandwidths are 7 MHz to 8 MHz thus allowing proportionally higher digital bit-rates for almost the same error-rate performance and potentially somewhat higher displayed picture resolution.

10.2 Bi-rate Transmission

The variable transmission bit rate feature is directly applicable to HDTV standards for other countries to allow for extended coverage into fringe areas. And, of course, the delivery of a noise and interference free picture and sound throughout the service area when all information is transmitted digitally applies equally well in other countries and regions.

SECTION 11
REFERENCES

[1] Hunt, R. W. G.: The Reproduction of Colour, 4th Ed., Fountain Press, 1987, Chapter 8.

[2] Netravali, Haskell: Digital Pictures, Plenum Press, 1988.

[3] Netravali, Stuller: "Motion-Compensated Transform Coding", Bell System Technical Journal, Vol. 58, No. 7, September 1979, pp. 1703-1718.

[4] Netravali, Robbins: "Motion-Compensated Television Coding, Part I", Bell System Technical Journal, Vol. 58, No. 3, March 1979, pp. 631-670.

[5] Netravali, Prasada: "Adaptive Quantization of Picture Signals Using Spatial Masking", IEEE Proceedings, Vol. 65, No. 4, April 1977, pp. 536-548.

[6] A. Lender, Correlative (Partial Response) Techniques and Applications to Digital Radio Systems, Chapter 7 in Digital Communications, Microwave Applications, K. Feher, editor; Prentice Hall Inc. 1981.

[7] Thomas, Weiduc, Dunani, "Digital Amplitude-Phase Keying with M-ary Alphabets", IEEE Transactions on Communications, Vol. CM-22, No. 2, February 1974, pp. 168-179.

[8] R. Citta, "Frequency and Phase Lock Loop", IEEE Transactions on Consumer Electronics, Vol. CE-23, No. 3, August 1977, pp. 358-365.

[9] Robert Eckert, et. al: Interim Report: "Further Studies on the Availability of Spectrum for Advanced Television", FCC/OET TM89-1.

APPENDIX

AC-2: A FAMILY OF LOW COMPLEXITY TRANSFORM BASED MUSIC CODERS*

Louis D. Fielder
Grant A. Davidson
Dolby Laboratories, Inc.
San Francisco, California 94103

ABSTRACT

Two high-quality data rate reduction music coders from a family of TDAC transform based coders are discussed. An overview of the psychoacoustic principals used in their design is given and their limitations discussed. The use of psychoacoustics and DSP technology are combined to yield a low complexity approach to music coding. Issues of complexity, word length requirements, and memory usage are examined for both general-purpose DSP and custom IC implementations.

0. INTRODUCTION

The use of data rate reduction coders for digital audio applications shows great promise for a large variety of storage and transmission applications. Since Compact Disc digital audio employs a data rate greater than 1.4 Mbits/sec., this type of digital audio has been limited only to areas that can maintain a high data rate. Fortunately, the development of high quality data rate reduction technology for music applications has changed this situation. Now lower data rates may be used for audio in radio and television broadcast, computer hard disk storage, and telephone line connections. This paper will describe two coders from the Dolby AC-2 family, developed for different applications, that have the desired characteristics of data rate reduction, excellent sound quality, and computational simplicity.

The need to reduce the data rate for the practical application of digital audio into many areas has resulted in much work in the field of data rate reduction for music, as typified by Brandenburg et al. [1990], Johnston [1988], Schroeder et al. [1987], Stoll and Dehery [1990], Davidson et al. [1990], and Fielder [1989]. The fundamental approach of these

*Published at the AES 10th International Conference, Sept. 1991, London.

techniques is to divide the audible frequency range into sub-bands which approximate auditory critical bands. Crucial elements in the design of these coders are the bit allocation and quantization schemes in which perceptually relevant sub-bands are identified, and the appropriate fraction of the available bit rate assigned to their representation. Many of these algorithms require a great deal of processing power to perform the frequency division and quantization operations (e.g., multiple DSP chip implementations for a single audio channel). Furthermore, they all extrapolate published models of human hearing and masking to a broader class of signals than those upon which the models were based.

This paper builds on the work described by Davidson et al. [1990] and Fielder [1989] which described 15 kHz bandwidth coders with resultant data rates between 128 and 192 kbits/sec. per channel. The two coders described here have 20 kHz bandwidth, require less than one programmable DSP chip to implement one stereo pair, and possess excellent sound quality. In particular, one coder, which will be called the low-delay coder, achieves excellent subjective and objective quality at 4:1 compression, exhibits robust tandem coding performance (i.e., where a number of encode/decode processes occur in series) and has a coding/decoding delay less than 9 msec. This low-delay feature is essential for applications requiring that announcers monitor their own coded voice signals. The other coder trades coding delay for a lower bit-rate (6:1 compression) and will be called the moderate delay coder. The coding systems described here can be applied for either 44.1 k or 48 ksamples/sec., however the remaining discussions will center on 48 ksample/sec. results.

A general overview of the psychoacoustics of masking as it effects the design of data rate reduction music coder technology will be given. Next, the details of the two coding systems resulting from this psychoacoustic examination will also be presented. Issues of implementation will also be discussed and the use of 24-bit and 16-bit DSP chips will be examined and processor speed/memory requirements determined. The use of custom DSP chips will also be considered. It will be shown that the two systems described are quite low in complexity while at the same time providing excellent sound quality.

1. APPLICATION OF PSYCHOACOUSTIC MODELS TO CODER DESIGN

The basis of all good rate reduction music coders is the application of the psychophysics of the human auditory system. As a result, a discussion of the present state of knowledge in this area is essential for the understanding of coders of this type. Masking effects for simple signals will be extended to the development of the filter bank design and quantization technology used in music coders. It will be seen that the targeted application will greatly influence the way the psychoacoustic principals are utilized. Next, these principles will be extended to more complex signals and discussed for AC-2 coding. An indication of the effectiveness of the AC-2 coding system in controlling the amount and frequency characteristics of the errors due to the reduction of word-lengths for data rate reduction will be given by a spectral comparison between both coder's performance and frequency characteristics of auditory masking.

1.1 Critical-Band Model of Hearing

Central to the development of a workable model of the auditory system is the critical-band concept and its relationship to the masking characteristics of the ear. The critical-band model of the human auditory system was first developed by Fletcher [1940] to explain why masking experiments showed that signals covering a frequency range less than a certain threshold bandwidth produced the same masking and detection properties as other signals with smaller bandwidths. The fundamental approximation is that the ear acts as a multi-channel real-time analyzer with varying sensitivities and bandwidths throughout the audio range. Despite the intrinsic simplicity of the model, it has been shown to be very enduring. Effective data rate reduction coders for music rely heavily on this model.

The critical-bandwidth represents the minimum frequency bandwidth resolvable for masked signals. For example, the masking of a low level error signal caused by a larger level tone nearby in frequency is maximal and continues at a constant level until the frequency separation between them exceeds this bandwidth. Detection of a signal component takes place based on the entire energy within a critical-bandwidth, whether it is tonal in nature, noise-like, or a combination of the two. Later workers have further refined this concept; Zwicker et al. [1957] examined this resolution bandwidth via various detection and masking experiments. Later Zwicker [1961] established 24 fixed critical-bands over the 20 Hz - 15 kHz frequency range.

1.2 The Use of Single Tone Masking Curves

Information on the masking effect of signal components is available primarily for single tones or bands of noise. As a result, coder design depends greatly on principles derived from these simple masking experiments. These typically generate masking curves of single high level component masking the presence of another smaller component and are quite useful because they can be used to derive an upper bound on the levels of permissible error signals due to the data rate reduction process. Since the masking effect varies significantly depending on whether the large level component or masker is tone-like or noise-like in character, the more demanding situation of sinewave masking curves are shown in Figures 1, 2, and 3. The figures present various 1/3 octave hearing thresholds when subjects are subjected to various levels of 100 Hz, 500 Hz, and 4 kHz sinewave maskers, as described by Fielder [1987]. For more information on the variation of the masking effect for tonal or noise signals, see Ehmer [1959].

The most appropriate way to examine masking phenomena is to perform a spectrum analysis based on critical-bandwidths. Since critical-band analyzers are not common, a good approximation can be made with the use of 1/3 octave bands; see Fielder [1987] for further details. These spectral analyses of masking are then used as a basis for the design of the coder filter bank structures and the methods to reduce the bit rate via word-length reduction.

The first observation from Figures 1-3 is that masking is generally greatest at the masker's frequency. This indicates that the coder design should concentrate error energy directly adjacent to the signal frequency. The next property the figures have in common is that the masking effect slowly decreases with increasing frequency separation, if the smaller signal is higher in frequency than the masker. The masking effect for signals at a 70 dB acoustic level may extend only a few octaves upward in frequency while higher level situations may produce six upward octaves of significant masking.

Looking at masking of signals lower in frequency than the masker shows a very different situation. For these signals, the masking effect falls off much more quickly. This is particularly evident for frequencies between 500 Hz - 2 kHz when evaluated in a dB per Hz fall-off from the masker frequency; in this frequency region the slope can be as steep as 100 dB per 350 Hz below 500 Hz (i.e. 90 dB/octave) and drop as deep as 40 dB within 1/2 octave. This rapid decrease in masking for components lower in frequency than the masker has significant consequences in coder design, and has been one of the primary reasons that data rate reductions of 4:1 or greater have awaited the practical availability of powerful DSP architectures which can practically implement the necessary complementary filter structures with sharp frequency characteristics that are suitable for music coders.

The differences in the masking characteristics versus frequency are also significant. In Figure 1 the masked threshold falls off only for frequencies above 100 Hz. The upward frequency fall-off in masking above 100 Hz is rapid on a dB per Hz basis, with a slope that is as much as 100 dB per 400 Hz. In the case of 100 Hz masking curves, it is important to note that a ratio of as much as 100 dB may be necessary between the 100 Hz masker and a resultant error component, if the error is to be inaudible. This means that any filter bank used by a rate reduction coder is most effective if its ultimate attenuation spans this 100 dB range. The masking curves of 100 Hz are typical for masking situations for maskers at or below 200 Hz.

The masking curves for 500 Hz, depicted in Figure 2, show a different situation. In this case there is a rapid reduction of downward frequency masking of up to 100 dB per 360 Hz, while having a much slower reduction at higher frequencies. In addition, high sound levels between 90-110 dB cause a very large masking effect at the second harmonic, causing the masking effect to be significantly extended upward in frequency. These 500 Hz curves are typical for the masking properties of midrange signals in the 500 Hz - 2 kHz region. Although not shown, at 2 kHz the slope of the masking curves have only 1/2 - 1/3 the slope of masking curves at 500 Hz, but the total fall-off has increased to 60 dB.

Figure 3 shows masking that is typical for high frequency signals. Masking for lower frequency error components falls off fast but not as fast a dB per rate as in the case of midrange signals. However, the total may exceed 70 dB for maskers at 8 kHz and above. As in the case of midrange signals, upward frequency masking reduces slowly with frequency but covers a more extended frequency range.

1.3 Temporal Masking and Time vs. Frequency Trade-Off

Sinewave masking experiments and the shape of masking curves derived from them indicate the requirements for the filter bank of a low bit rate coder under steady-state signals. Another requirement is the accommodation of human auditory characteristics during transient events. Although the frequency resolution for steady state sinewave signals is extremely sharp, the characteristics of auditory masking for transient events involves time resolutions on the order of a few milliseconds. The temporal characteristics of masking are important because the filter banks used for data rate reduction coders can disperse error signals in time. This spreading occurs because of the fundamental trade-off between temporal and frequency resolution of filters. For this reason, filter bank design typically involves a trade-off between these conflicting goals.

Just as in the case of the frequency characteristics of auditory masking under steady state signal conditions, there is a basic asymmetry in the characteristics of temporal masking. The masking of small signal components occurring during in time before a masker (i.e. backward masking) is substantially less than the forward masking effect in which the same small signals occur after the masker. Backward masking remains strong for about 4 milliseconds and disappears for time separations larger than 10's of milliseconds, while forward masking lasts approximately ten times as long. For further information on the temporal masking characteristics of the ear, see Carterette and Friedman [1978]. The temporal resolution characteristics of a filter bank used for data rate reduction of music signals should maximize the masking effect so that the largest data rate reduction induced errors are tolerated by the ear. Since a transient event can occur anywhere within the effective time window of a particular filter, this argues strongly for filter banks with time resolutions less than 4 milliseconds.

1.4 Filter Bank Design and Auditory Masking

The filter bank of a coder is the primary element that allows rate reduction to occur with minimal audible consequences. It does that by confining the error temporally and spectrally in such a way as to allow the greatest errors to occur. This spectral and temporal confinement must satisfy the following conditions. First, the ideal filter bank should have a frequency selectivity less than one critical band in any part of the audio band, have a fall-off rate of 100 dB per 360 Hz, with an ultimate rejection of 100 dB, and finally, have a temporal spreading effect of less than 4 msec. A filter bank which is easy and efficient to implement is also desirable. Unfortunately, the attainment of all the previously mentioned goals is extremely difficult and a compromise is necessary. As a result, further discussion will concentrate on the compromises and results for the low and moderate time delay AC-2 coders.

The design of the AC-2 coding technology is strongly influenced by the desire to keep the implementation as low in complexity as possible, while preserving coder effectiveness. For this reason, the AC-2 coders use Time Domain Aliasing Cancellation (TDAC), as developed by Princen and Bradley [1986]. This transform has the computational complexity advantages of an FFT and has excellent frequency selectivity characteristics. Unfortunately, the resultant filter bank is constant bandwidth, rather than having the varying bandwidths of the auditory system. This disadvantage of the TDAC can be overcome by approximating the nonuniform bandwidths of the human auditory system by grouping transform coefficients together to form sub-bands with bandwidths approximately that of the auditory system.

Consider first the TDAC filter bank for the moderate time delay coder, useful in applications where a low data rate is more important than low time delay. In this case the transform length is chosen to be 512 samples, which is found to be the best compromise between frequency and temporal selectivity. The resultant filter bank has a frequency selectivity that is sufficient for most of the audio band, while at the same time having a time resolution on the order of 10 msec. This compromise is acceptable since limitations in the temporal or spectral resolution are minor and can be greatly improved by a quantization process that allocates additional data to mitigate the increased audibility of errors during transient circumstances.

The other AC-2 coder is targeted for applications were low time delay is important, such as disk based storage applications requiring fine time resolution editing or for broadcast applications where an announcer may listen to the transmitted signal as a verification of proper system operation. Monitoring of the transmitted voice signal is problematic for the announcer if the time delay is too long, because it interferes with the cognitive process of speaking. The time delay at which speech difficulties begin to occur is not well defined, but 10 msec. appears to be a reasonable compromise, see Gilchrist [1990] for more details. The transform block length for this coder is set at 128 samples by this requirement and the resultant encode/decode delay is 8 msec.

This restriction in the block length has important consequences in the coder design because it moves the filter bank temporal- frequency resolution trade-off away from the optimal compromise. As a result, the frequency resolution is inadequate for masking the error signals for frequencies below 3 kHz. Insufficient frequency selectivity translates to either reduced audio quality or increased data rate. For this reason, this coder uses a higher data rate of 192 kbits/sec per channel. The time resolution of the system is 2.7 msec. and the resultant coder has excellent performance under transient conditions.

The loss of frequency selectivity to satisfy time resolution or computational complexity issues is very important in coder design. Figure 4 demonstrates this point by comparing the filter bank selectivity of three filter banks used in music coders to that of a masking curve for a 100 dB S.P.L. 1 kHz sinewave. This masking curve for 1 kHz was chosen since it is nearly a worst case for the selectivity requirements of a single tone situation. Both filter banks used in the two AC-2 coder are shown, and in addition, a typical uniform bandwidth sub-band filter band, having 750 Hz bandwidth, is included.

Examination of Figure 4 shows that none of the filter banks presented have ideal frequency selectivity when compared to this most demanding requirement of the human auditory system. The consequences of this fact is that all the coders implemented with these filter banks must either have a higher data rate than ideal or have lower sound quality. Inspection of this figure shows that the moderate delay version of AC-2 has the selectivity closest to that required, implying that little additional data rate is required to preserve sound quality. Next in selectivity is the uniform sub-band filter; the sharpness of the filter roll-off is excellent but it has the limitation that the filter's bandwidth is too wide for low frequency and midrange signals. This lack of frequency selectivity will result in quantization error that is spread over a wide frequency range (i.e. 550-1500 Hz) and must be accommodated by an increase in data rate. This increase in data rate results in an additional word-length requirement because the overall level of the error must be lowered until all of its spectrum lies below the masking curve. Finally, the short time delay AC-2 filter bank frequency selectivity is considered. In this case, additional data rate is seen to be required to mitigate the insufficient frequency selectivity of the low time delay filter bank. This, along with the desire for excellent multi-generation sound quality results in a data rate for this coder of 192 kbits/sec.

In conclusion, the examination of sinewave masking shows that the frequency selectivity of the moderate delay AC-2 is somewhat less than the worst case condition of 1 kHz masking. This indicates that its computationally efficient filter bank does not significantly limit the performance. The low time delay AC-2 coder selectivity is examined and shown to be too broad for use in the lowest possible data rate system. Fortunately, this increase in data rate is modest because the selectivity of the human auditory system is poorer than this filter bank over most of the audio band (i.e. 4 kHz - 20 kHz). One additional benefit of the short time delay AC-2 coder is that it possesses a temporal resolution substantially below that at which either forward or backward masking effects occur. The disadvantage of having too wide a filter bank bandwidth was demonstrated by the 750 Hz sub-band filter example.

1.5 Extension to Complex Signals

The use of simple stimuli masking models has determined the basic requirements of frequency and time resolution. This is done because there is not a widely accepted model of hearing for more complex signals. Unfortunately, real music signals are complex, so coder design must extend these simple masking models to the complex conditions of music signals. In the case of the AC-2 coding systems, simple stimuli masking principles are extended in a very conservative manner. Although many coding systems adaptively allocate most of the available data rate in a signal dependent manner to produce errors that are just below predicted masking, this was found to be an unnecessarily aggressive approach for applications with data rates at or above 128 kbits/sec.

The conservative approach of the AC-2 coder family is as follows: The appropriate TDAC transform filter bank is first combined with a trial quantization process that has a fixed number of bits assigned to each band, which are adjusted to simultaneously satisfy the masking requirements of simple and complex signals. Once this fixed allocation scheme is properly adjusted for optimal audible effect, a modest amount of the data responsible for this representation of the audio signal is removed and replaced by a smaller amount of adaptively allocated data, resulting in 20% or less data of this type. The advantage of the largely non-adaptive nature of most of the data is that problems in the extension of simple masking models are not nearly as serious as in the case of coders that have a more adaptive allocation strategy. This prevents serious audible mistakes from occurring: in fact the audible performances of the AC-2 coders without any adaptive bits are quite good.

This method of extension to more complex signals is evaluated and optimized by both objective and subjective means. This includes comparison of computed noise spectra with psychoacoustic masking threshold data, and conducting A:B listening tests. Subjects are asked to evaluate signals coded by hardware in real-time to facilitate exposing the coder to a wide variety of instrumental, vocal, and synthetic audio signals.

Although coder performance is more rigorously evaluated using complex music signals, many important features are revealed by the sinewave error spectrum. Figures 5 and 6 are a comparison of both coder's 1 kHz error spectra with a 100 dB S.P.L., 1 kHz masking curve. The moderate delay AC-2 coder results are shown in Figure 5 and those of the low delay AC-2 coder in Figure 6. Both figures give an indication of the worst case performance of the coder because the 1 kHz auditory selectivity is the most severe. These comparisons assume a consumer playback sound level at 108 dB peak acoustic level, being limited by the maximum loudness capabilities of typical home loudspeakers and amplifiers. In both figures, the error spectra are shown for coder operation with, and without, the adaptively allocated portion of the data. Both coding systems are interfaced to 16-bit ADC's and DAC's so the noise of the conversion process also is present.

Examination of Figure 5 shows that the error signal and converter noise under normal operation is just at or below audibility. The error spectra above 2.5 kHz for both situations are limited by the ADC/DAC noise floors and indicative of the 91 dB dynamic range of typical 16-bit conversion systems. The frequency region below 2.5 kHz is a result of coder operation and a significant deviation from an ADC/DAC noise floor results. In this region, the 1 kHz error spectrum under normal operation is substantially below the masked threshold curve, except for frequencies between 400 Hz - 700 Hz, where the error spectrum is comparable to the masked threshold. This indicates that a slight modulation noise may be audible, although in practice this has not been heard. The error spectrum shown without the adaptive portion of the data shows that modulation noise is now quite audible since the error spectrum is significantly above the masked threshold in the frequency range of 400 Hz - 1200 Hz. The generation of audible modulation noise indicates that the adaptive bit allocation process is necessary to preserve excellent sound quality. Notice that the error spectrum falls off less rapidly than that of the downward frequency portion of the 1 kHz masking curve.

TDAC is applied to model the auditory system by grouping adjacent transform coefficients into sub-bands for further decomposition and analysis. The number of coefficients per sub-band is computed a priori to approximate the nonuniform critical-bands. Transform coefficients within one sub-band are converted to a frequency block floating-point representation, with one or more mantissas per exponent, depending upon the sub-band center frequency. Each exponent represents the quantized peak log-amplitude for its associated sub-band. The exponents collectively provide an estimate of the log-spectral envelope for the current audio frame, computed on a critical-band frequency scale.

From a psychoacoustic perspective, the log-spectral envelope provides an ideal framework for estimating which sub-bands of a given audio frame are perceptually most relevant, and for ranking them in relative order of importance for dynamic bit allocation. Furthermore, the nonuniform frequency division scheme offers key advantages compared to one based on uniform-width filter banks. Accordingly, the AC-2 frequency division scheme reduces the need both for relying upon a complex masking model, and for using a second, higher-resolution filter bank in the encoder.

The dynamic bit allocation routine is completely feed-forward in nature and is constrained to produce a constant bit-rate as required for transmission applications. Bits are allocated in accordance with a set of deterministic rules derived from conservative use of single-tone masking curves. A portion of the routine employs a water-filling procedure in which sub-bands are ranked and allocated bits on a band-by-band basis.

The allocation routine provides step-size information for an adaptive quantizer. Each sub-band mantissa is quantized to a bit resolution defined by the sum of a fixed allocation and a dynamic allocation. The total fixed allocation for one frame outweighs the dynamic allocation in approximately a 4:1 ratio. For a given level of error protection overhead, this approach was found to provide more robust coding and error performance, since the number of most-significant mantissa bits is known a priori in the decoder. In the final stage of the encoder, exponents are multiplexed and interleaved with mantissa bits for transmission to the decoder. Optional error correction codes may be added at this step. The amount of overhead information reserved for error control coding can be adjusted to give greater or lesser protection depending upon channel error performance for a given application.

Serial bitstream formats can be optimized for the application. In the DP501/DP502 digital audio encoder/decoder products employing AC-2, two independent channels are interleaved in a regular pattern of alternating 16-bit segments. This format allows for straightforward demultiplexing of the encoder bitstream into separate channels, and for recombining monophonic bitstreams from different encoder units. Provision is also made for the insertion of a 1200 bit/s auxiliary data stream, algorithm identification bits, ADC overload status, and other information.

In the AC-2 decoder, shown in Figure 8, the input bitstream is demultiplexed and errors, if any, are corrected. The received log spectral envelope is processed in a stage identical to the encoder bit allocation routine, which generates step-size information for the adaptive inverse quantizer. The fixed and dynamically-allocated portions of each mantissa are concatenated to regenerate compressed transform coefficients. A sub-band block floating-point expander then linearizes the compressed transform coefficients and passes them to an inverse MDCT/MDST transform stage. After the inverse transformation, a window identical to that used in the encoder is used to post-multiply the reconstructed time-domain samples for each frame. Adjacent windowed frames are overlapped by 50% and then added together to reconstruct the PCM output.

Total coding/decoding time delay is determined by the frame size N, the manner in which frames are processed, and the processor speed. In the low-delay coder, input frames are processed one-by-one, resulting in a theoretical minimum total coding and transmission delay of 2.5N samples when employing infinitely-fast encoder and decoder processors. In actual practice, the delay increases to about 3N samples for fully-utilized (finite-speed) encoder and decoder processors. With a frame size of N = 128 samples and a sample rate of 48 kHz, a delay of 8 msec. is obtained. In the moderate-delay coder, two successive frames from one channel are buffered and processed jointly. In this case, the total delay when using fully-utilized processors is about 4N samples, which results in less than 45 msec. of delay at a sampling rate of 48 kHz and with N = 512.

3. HARDWARE IMPLEMENTATION

All coders within the AC-2 family have been optimized for very low hardware implementation cost. By today's standards, cost is ultimately measured by the die size and package cost of a custom VLSI implementation. Accordingly, the cost equation must not only include such traditional complexity measures as multiply-add count and RAM/ROM memory usage, but regularity of computation and minimum word-length requirements as well. Considerable attention has been given to structuring the computations in AC-2 to minimize VLSI implementation cost and simultaneously achieve the audio performance objectives. At a sampling rate of 48 kHz, the total number of multiplies and adds per second in a stereo AC-2 encoder is about 2.7 million. The decoder complexity is slightly lower. This compares to calculations by Reader [1991], estimating a total of about 35 million multiplies and adds per second for a straightforward implementation of a current generation sub-band encoder, and 16 million multiplies and adds per second in the decoder.

The low computational complexity can be attributed to several factors. First, the computational structures employed are highly regular in nature. Second, an efficient technique has been found for implementing the evenly-stacked TDAC transform by combining a core FFT routine with pre-twiddle and post-twiddle operations. Third, the nonuniform frequency division stage and log spectral energy representation enables the use of a low-complexity dynamic bit allocation routine. Finally, the use of functions which are inefficiently implemented on programmable DSPs or in custom-ICs, such as logarithms, square roots, and divides, have been found unnecessary. The only functions required are multiply, add, integer left/right shift, normalize, and compare.

3.1 General-Purpose Programmable DSPs

Programmable DSPs provide a flexible and expedient path to real-time algorithm development, and as such provide an attractive means for a first implementation. An early embodiment of AC-2 based on the Fourier transform was implemented using six Texas Instruments TMS32010s by Fielder [1989]. This work subsequently led to an implementation employing TDAC and based on the Motorola DSP56001, as detailed by Davidson et al. [1990]. In the latter case, a single 27 MHz chip could either encode or decode two independent channels. Recent improvements in software run-time efficiency have reduced this speed requirement to 20 MHz.

3.1.1 24-Bit Fixed-Point

Since it's inception in 1987, the Motorola DSP56001 has proven to be a capable platform for implementation of a wide variety of audio processing algorithms. This general trend has been supported by several audio compression implementations, including AC-2. The DSP56001's 24-bit data path, flexible addressing modes, and dual-accumulator arithmetic logic unit (ALU) are keys to its successful application in audio.

In particular for AC-2, we found that the 24-bit word-length was sufficient for all arithmetic tasks. Furthermore, no elaborate scaling or rounding procedures were required. The dynamic range of the implementation, as measured from PCM input to output, is 108 dB. This figure greatly exceeds the theoretically-achievable dynamic range of 16-bit ADC and DAC converters, and is commensurate with next-generation 18 and 20-bit converter technologies.

One of the more time-intensive processing blocks of those shown in Figures 7 and 8 is the inverse transform, which requires about 18% of the total DSP processing time. Surprisingly, however, the most time-intensive tasks are bit multiplexing and demultiplexing. This indicates that a custom IC could save significant ALU resources compared to a DSP if dedicated logic performed the multiplexing and demultiplexing. This topic is discussed further in Section 3.2.

3.1.2 16-Bit Fixed-Point

A study was made to determine the feasibility of implementing an AC-2 decoder on a 16-bit DSP chip. The motivation for this work was to identify a lower-cost platform for the implementation of an AC-2 decoder, while maintaining the flexibility of a programmable DSP. Our results indicate that current generation 16-bit DSPs, such as the Texas Instruments TMS320C5x, Analog Devices ADSP-2105, and Motorola DSP56116, are sufficiently powerful to implement a single-chip stereo encoder or decoder.

An analysis of finite word-length effects was conducted in part by modifying the real-time AC-2 DSP56001 software to emulate a reduced word-length processor. The data word-length was selected on-the-fly with switches. Coefficient word-lengths could also be varied. This approach allowed us to independently adjust, and jointly minimize, the data and coefficient word-lengths in each processing stage of the coder. The real-time variable word-length simulation served as a valuable tool for rapid objective and subjective evaluation of finite precision arithmetic effects.

Figure 9 presents a plot of the spectral error between an original and a coded 100 dB S.P.L., 100 Hz sinewave as processed by both 24-bit and a 16-bit ALUs in the decoder. Results from the moderate-delay decoder are shown since arithmetic round-off noise in the inverse transform is highest for long frame lengths; round-off noise in the low-delay coder is more than 6 dB lower. The idle channel noise produced by 16-bit ADC and DAC converters is included to show when the coder is limited by the conversion process. The low frequency sinewave represents a demanding test signal since minimal masking of the 4 to 6 kHz region occurs, where the ear's hearing threshold is low.

At frequencies below 500 Hz, noise introduced by transform coefficient quantization dominates arithmetic round-off noise. This region is perceptually insignificant because both noise curves are below the masking curve. Above 2 kHz, round-off noise for the 16-bit ALU significantly exceeds the masking curve, indicating that 16-bit single-precision (SP) arithmetic is inadequate.

Most of the noise shown in Figure 9 is generated during the inverse FFT computation of the inverse MDCT/MDST transform computation. Therefore, conventional techniques for reducing round-off error in fixed-point FFTs apply, such as those described by Meyer [1989]. We found that the combination of dynamic scaling between IFFT stages, optimal rounding, and optimal placement of quantizers in the butterfly produced a significant, but still insufficient, reduction in round-off noise. Furthermore, such techniques may impose a three-fold increase in IFFT butterfly computation time within a general-purpose DSP.

Based on these results, a preferred approach is to employ an extended-precision (EP) scheme based on 16 x 32-bit multiplies, which for many 16-bit DSPs results in a fixed two-fold increase in butterfly computation time, and provides a digital noise floor which is more than 40 dB lower than that obtainable with 24-bit SP multiplies. All other processing stages of the decoder can be implemented with 16-bit SP arithmetic. The minimum required DSP clock speed using 16-bit EP is only about 18% higher than the equivalent rating for a 24-bit fixed-point or 32-bit floating-point device.

3.2 Full-Custom VLSI

In order for an audio processor to be utilized in high volume applications, the device cost must usually be low. Since programmable DSP chips frequently contain more hardware logic than required for a given application, we have considered the design of a special-purpose VLSI architecture for implementing an AC-2 decoder. The architecture is capable of implementing any of the coders in the AC-2 family with one IC.

The architecture consists of three sections: a bit demultiplexer, a quantizer step-size control, and an inverse transform and reconstruction processor. The chip inputs are a serial bitstream and data clock, and the output is one or more 20-bit PCM digital audio channels. The bit demultiplexer performs such functions as data de-scrambling and bit de-interleaving. The demultiplexer directs the unpacked exponent data to the quantizer step size control, and the unpacked fixed and adaptive mantissa bits to a dedicated state machine/barrel shifter. The quantizer step size control, composed of a simple programmable microcontroller, processes incoming exponents and directs the state machine and barrel shifter to concatenate fixed and adaptive transform coefficient mantissa bits. The reconstruction processor performs either an IMDCT or IMDST, producing one frame of PCM samples. These samples are then windowed and overlap/added with the previous windowed block of PCM data to reconstruct audio samples. Since the multiply-add rate of the audio synthesis stage is quite low, a bit-serial multiplier has been employed. The serial multiplier requires significantly less chip area than a single-cycle array multiplier of the same word-length.

4. CONCLUSIONS

Adaptive transform coding of audio signals with AC-2 technology offers a high-quality, low complexity approach for data rate reduction of professional grade audio. Two 20 kHz bandwidth examples of the AC-2 coding family have been discussed, providing 4:1 and 6:1 bit-rate compression at low and moderate time delays, respectively. The excellent sound quality and computational ease of implementation of the AC-2 technology make it a natural candidate for broadcast, computer multimedia, and digital storage applications. The 128 kbits/sec. data rate of the moderate delay coder make it very appropriate for Digital Audio Broadcast and High Definition Television applications. The low delay coder is optimized for music material contribution applications (i.e. studio to transmitter and contribution quality links) requiring excellent multi-generational sound quality and a time delay acceptable for off-air monitoring during voice announcing.

The performance of these systems has been quantified by examination of simple stimuli masking models which have been the driving force shaping the design of the employed filter bank structures. Sinewave masking models have been used because a comprehensive and complete model for complex signals is not widely agreed upon. As a result, extension of the simple models is necessary for the design of practical coding systems. It was shown that the AC-2 family used a conservative extension process which resulted a relatively small amount of adaptively allocated data. As a consequence, these coder techniques were robust with respect to difficult program material. Other benefits created by this approach were a relative insensitivity to the effects of data-stream errors and low computational complexity.

Issues of computational complexity and practical implementation were discussed in some detail. It was shown that the AC-2 coder family is straightforward to implement at 128 and 192 kbits/sec. In particular, implementation of a stereo encoder or decoder was readily accomplished in one 20 MHz Motorola DSP56001. It was also shown that a practical modification of the frequency division algorithm permitted the realization of full fidelity realizations on 16-bit fixed-point DSP chips. A custom approach was also presented. It was shown that the AC-2 algorithms lend themselves well to dedicated chip hardware because of their reliance on simple shift operations and a low-complexity bit allocation strategy.

In conclusion, the AC-2 coder family represents one of the most cost effective solutions to very high-quality music coding applications at a 4:1 to 6:1 compression ratio. Although only two coders with data rates of 128 and 192 kbits/sec. were discussed, this technology can be applied to other sample rates, lower data rates (i.e. 64 kbits/sec.), and other signal bandwidths as well.

5. REFERENCES

K. Brandenburg, H. Gerhaeuser, D. Seitzer, and T. Sporer, "Transform Coding of High-Quality Digital Audio at Low Bit Rates: Algorithms and Implementation," Proc. of IEEE Intl. Conf. on Comm., Atlanta, pp. 932-936, April 1990.

E. C. Carterette and M. P. Friedman, Handbook of Perception, pp. 305-319, Academic Press Inc., New York, 1978.

G. Davidson, L. Fielder, and M. Antill, "Low-Complexity Transform Coder for Satellite Link Applications," presented at the 89th Convention of the Audio Engineering Society, preprint 2966, Sept. 1990.

R. H. Ehmer, "Masking of Tones vs. Noise Bands," J. of the Audio Eng. Soc., vol. 31, pp. 1253-1256, Sept. 1959.

L. D. Fielder, "Evaluation of the Audible Distortion and Noise Produced by Digital Audio Converters," J. Audio Eng. Soc., vol. 35, pp. 517-535, July/August 1987.

L. D. Fielder, "Low-complexity Transform Coder for Music Applications," presented at the IEEE ASSP Workshop on Appl. of Sig. Proc. to Audio and Acous., New Paltz, New York, session 2- 1, Oct. 1989.

H. Fletcher, "Auditory Patterns," Reviews of Modern Physics, vol. 12, pp. 47-65, Jan. 1940.
N. H. C. Gilchrist, "Delay in Broadcasting Operations," presented at the 90th Convention of the Audio Engineering Society, preprint 3033, Feb. 1991.

J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE J. on Sel. Areas in Comm., vol. JSAC-6, no. 2, pp. 314-323, Feb. 1988.

R. Meyer, "Error Analysis and Comparison of FFT Implementation Structures," Proc of IEEE Conf. on Acous, Speech and Sig. Proc., Glasgow, pp. 888-891, 1989.

J. P. Princen and A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time-domain Aliasing Cancellation," IEEE Trans. on Acous., Speech, and Sig. Proc., vol. ASSP-34, no. 5, pp. 1153-1161, Oct. 1986.

C. Reader, "Comments on Audio CD and Analysis of Audio Complexity," ISO/MPEG Document ISO/IEC/JTC1/SC2/WG11/MPEG91/082, May 20, 1991.

E. F. Schroeder, H. J. Platte, and D. Krahe, "MSC: Stereo Audio Coding with CD-Quality and 256 kbits/sec," IEEE Trans. on Cons. Elec., vol. CE-33, no. 4, pp. 512-519, Nov. 1987.

G. Stoll and Y. F. Dehery, "High-Quality Audio Bit-Rate Reduction System Family for Different Applications," Proc. of IEEE Intl. Conf. on Comm., Atlanta, pp. 937-941, April 1990.

E. Zwicker, "Subdivision of the Audible Frequency Range into Critical Bands," J. of the Acous. Soc. of Am., vol. 33, no. 2, p. 248, Feb. 1961.

E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical Band Width in Loudness Summation," J. Acous. Soc. of Am. vol. 29, pp. 548- 557, May 1957.

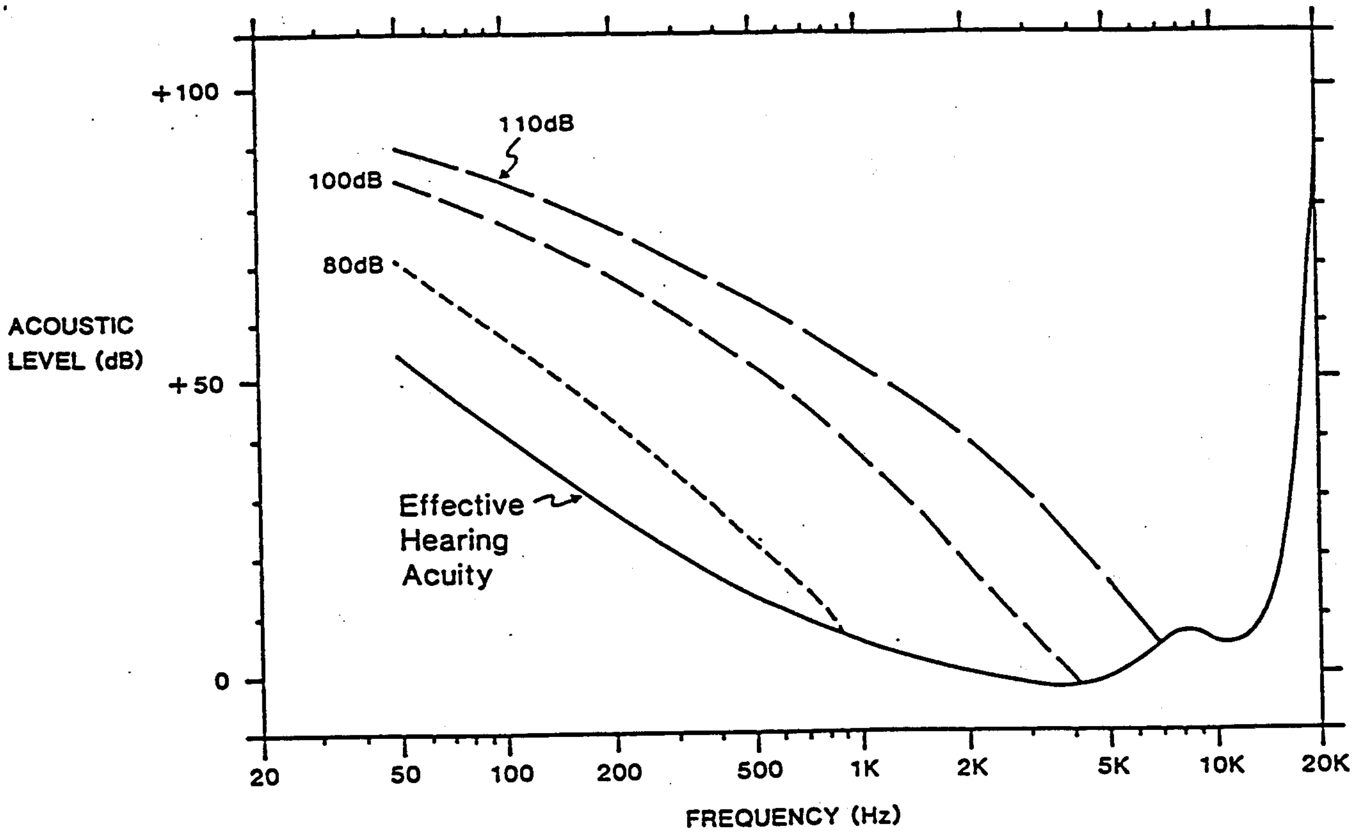


Fig. 1. 100 Hz Masked Threshold Curves (0 dB = 20 micropascals)

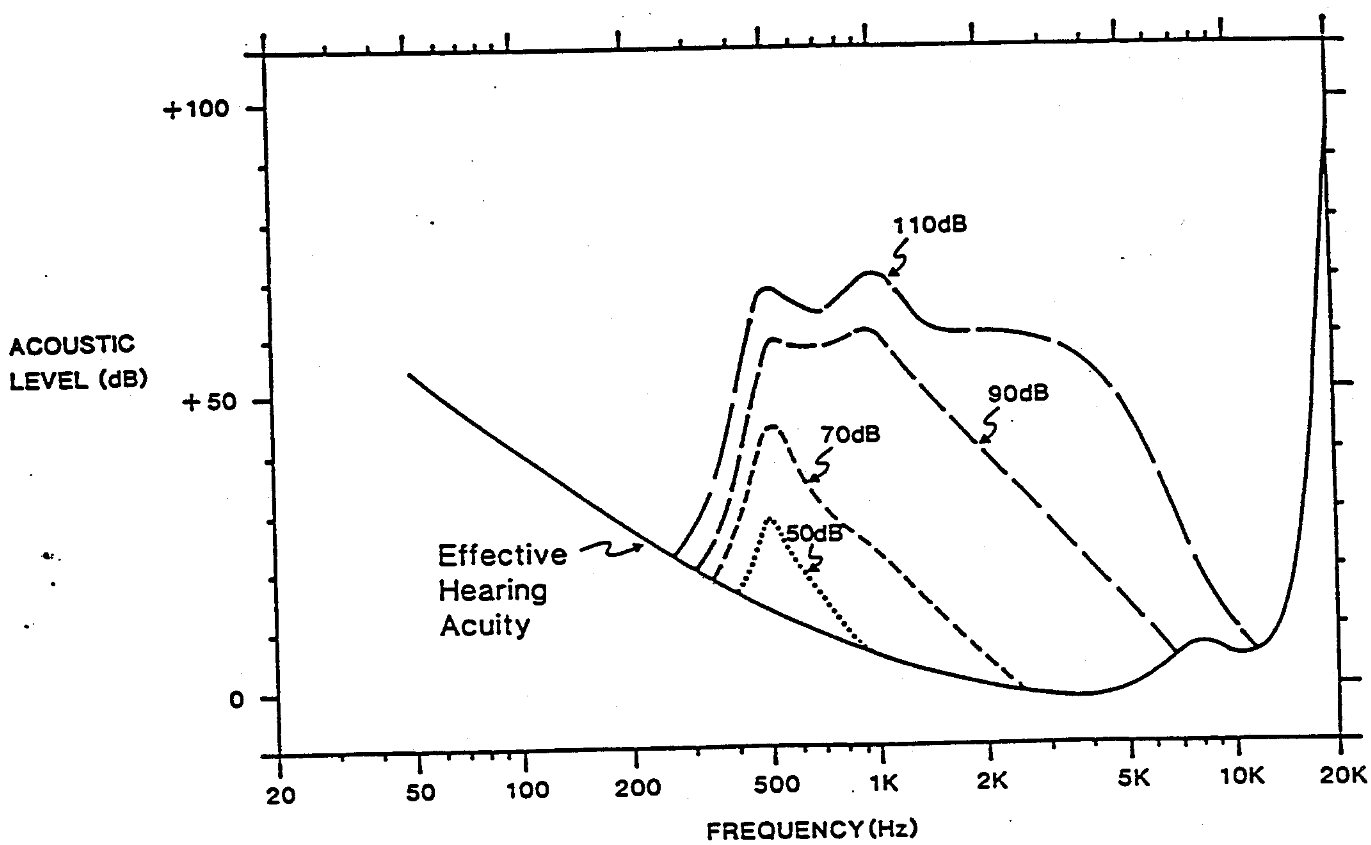


Fig. 2. 500 Hz Masked Threshold Curves (0 dB = 20 micropascals)

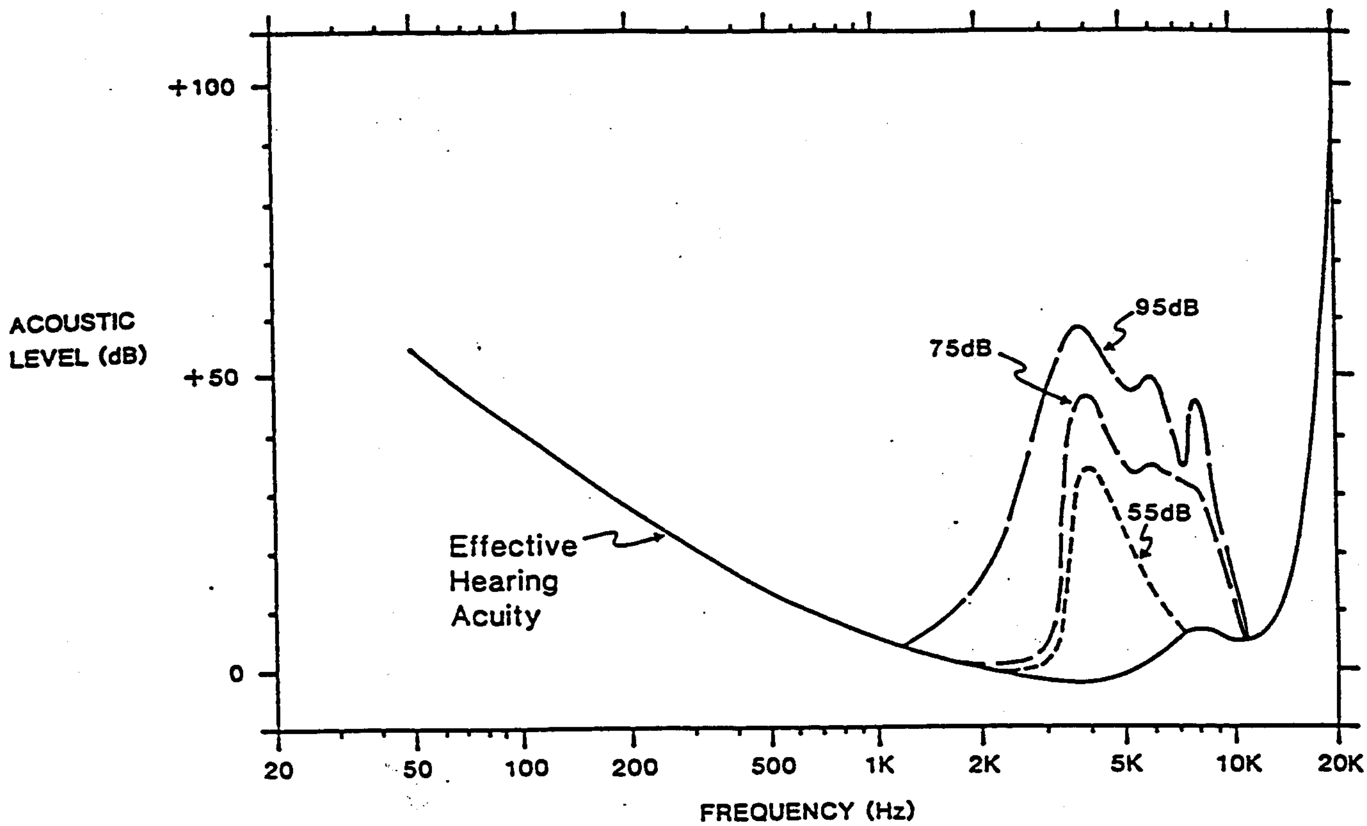


Fig. 3 4 kHz Masked Threshold Curves (0 dB = 20 micropascals)

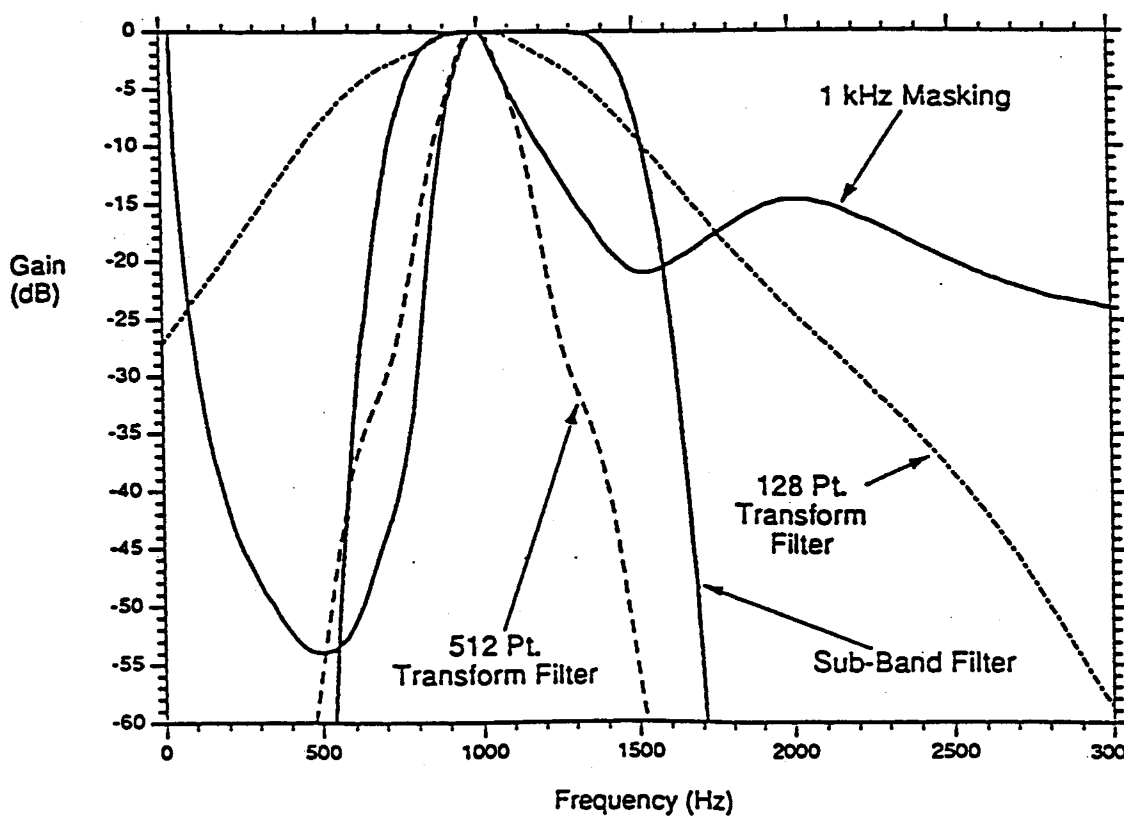


Fig. 4 Comparison Between Various Filter Banks and 1 kHz Human Auditory selectivities

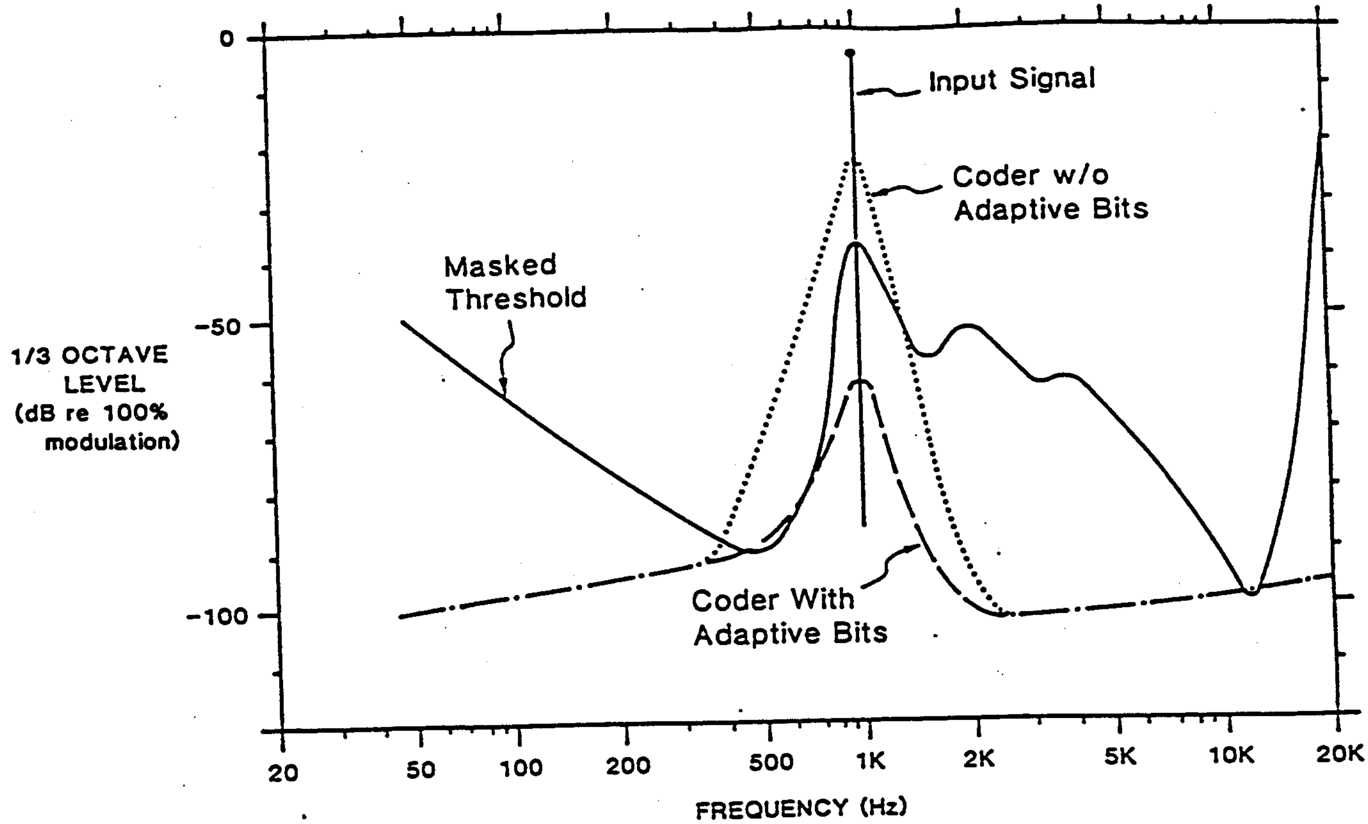


Fig. 5 Moderate Time Delay AC-2 Coder Performance with the Application of a 1 kHz Sinewave

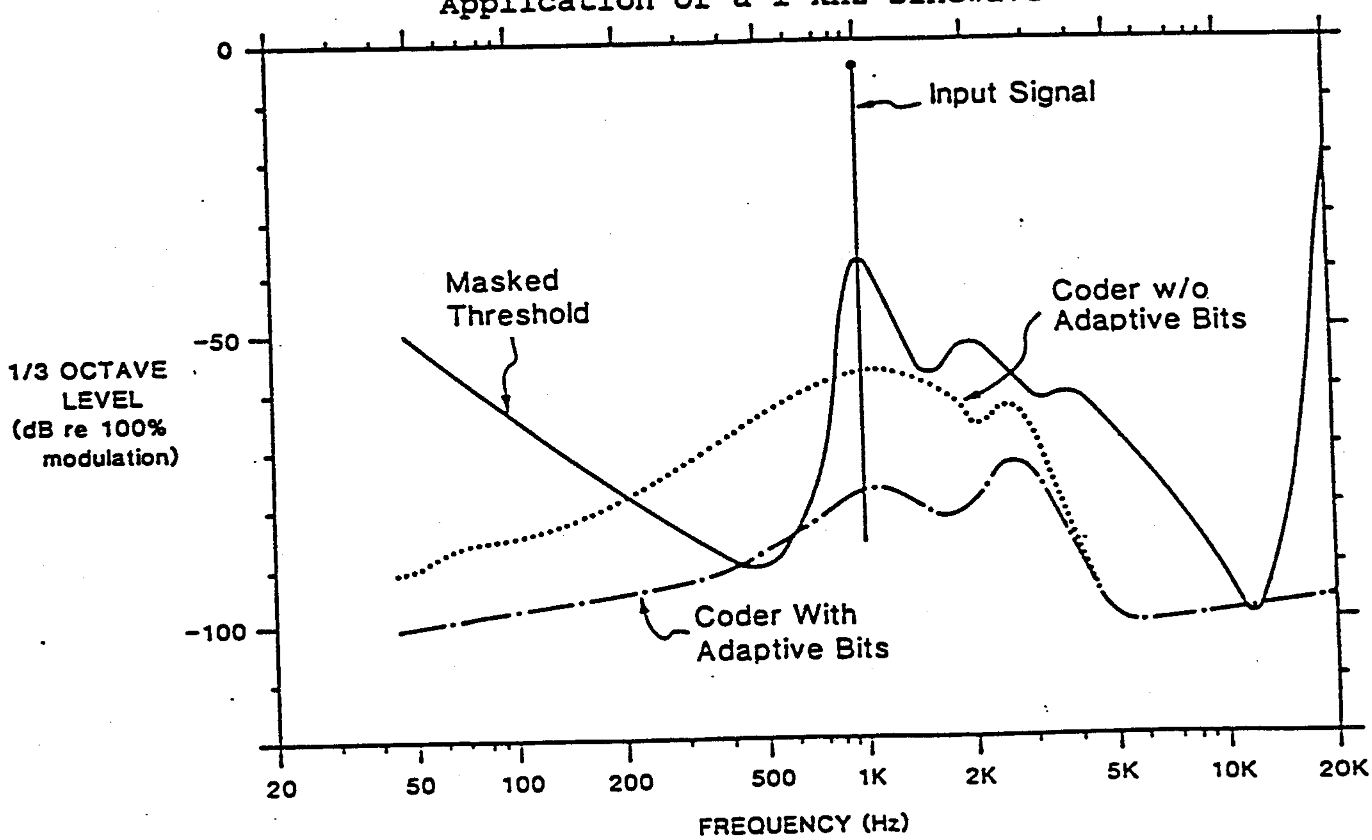


Fig. 6 Low Time Delay AC-2 Coder Performance with the Application of a 1 kHz Sinewave

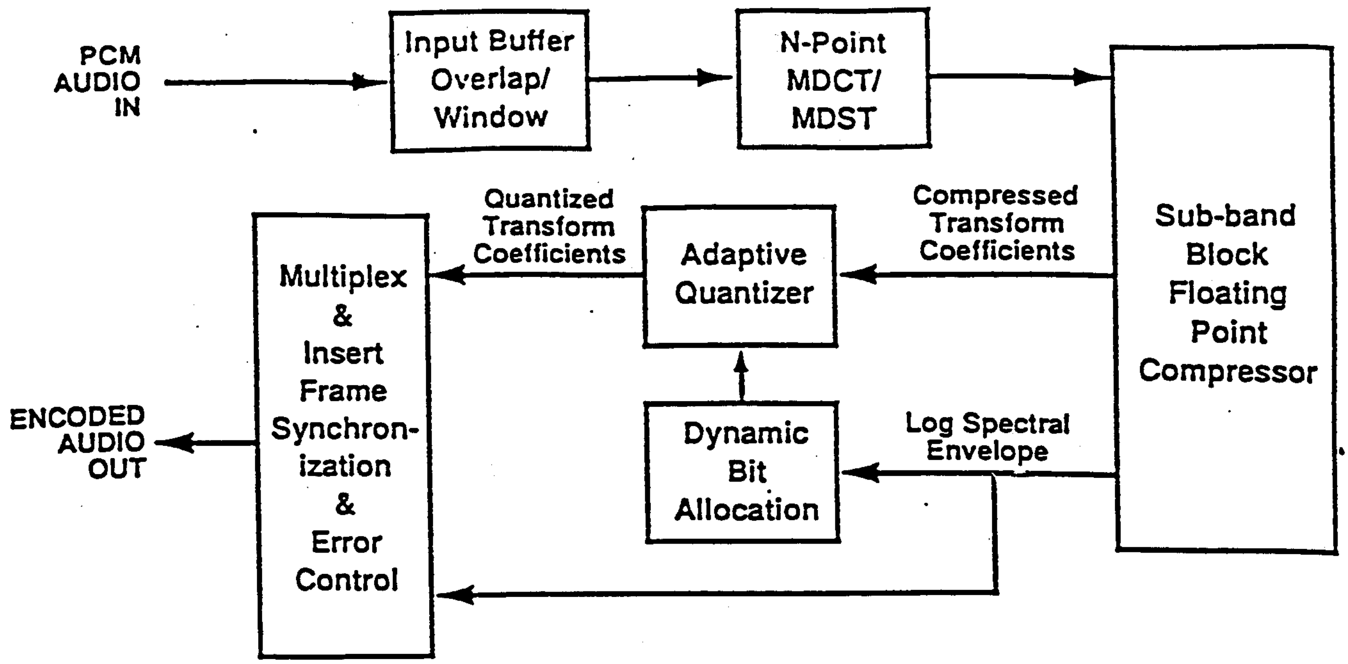


Fig. 7 AC-2 Digital Audio Encoder Family Block Diagram

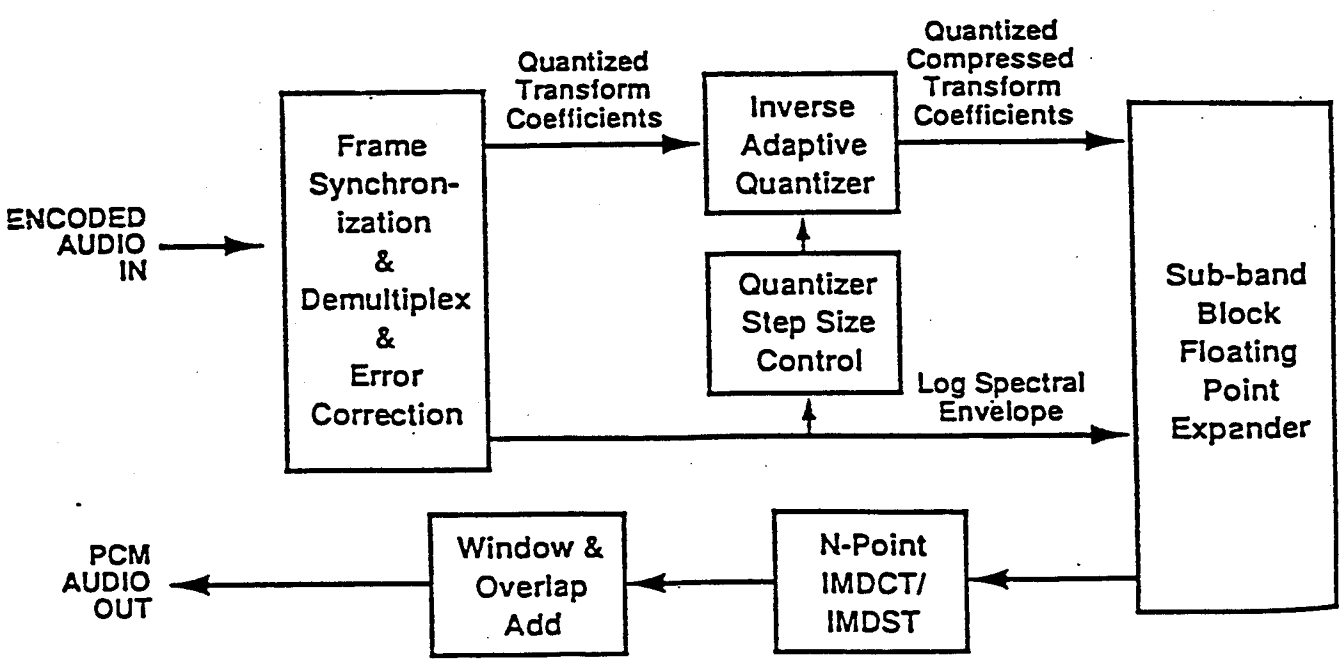


Fig. 8 AC-2 Digital Audio Decoder Family Block Diagram

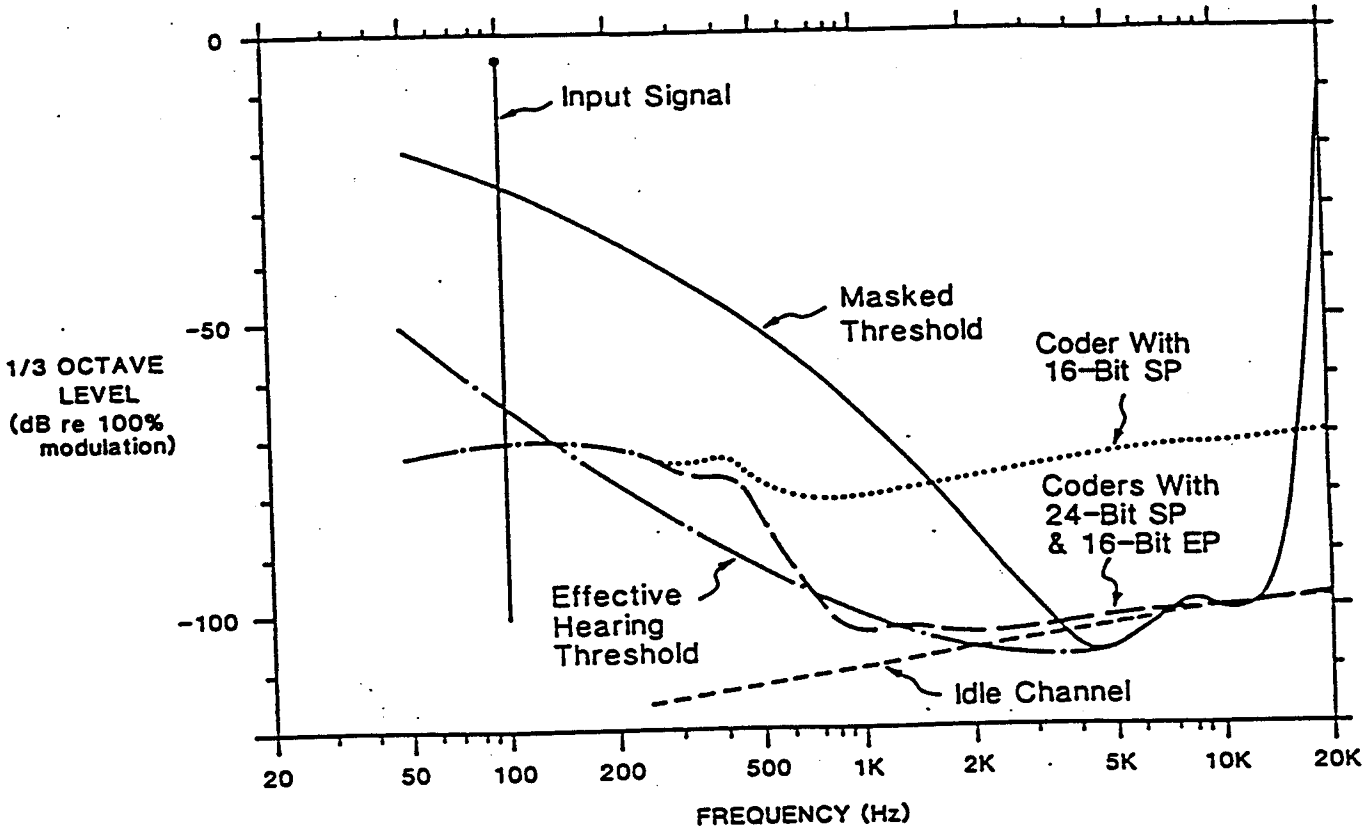


Fig. 9 Comparison of Filter Bank Arithmetic Noise of Single Precision (SP) 16-Bit Arithmetic Versus Extended Precision (EP) 16-Bit and Single Precision 24-Bit Arithmetic