NVOD
THE PREMISE BEHIND THE PROMISE

Dom Stasi
Request Television

ABSTRACT

This paper will detail one programmer's approach to digital, multichannel origination and transmission, and the technical implications of such highly processed signals to CATV reception and distribution.

As a preparation to that discussion, the text will include an overview of digital compression and coding philosophies. 4:2:2, 4:2:0, and 2:1:0 video sampling will be compared. Coding, storage, compression, filtering, decoding and distribution of compressed digital video signals will be described. Serial, and parallel studio interface standards will be explained. And finally, MPEG-1, MPEG-2, distribution protocols will be discussed with an emphasis on operational effects.

INTRODUCTION

Overview

With 1997, the much anticipated implementation of digital video from source to sink will be a CATV reality. Digital compression will bring the capability to distribute hundreds of video channels to those cable homes equipped to decode them.

The hype surrounding this transition emphasizes that the new digital signals will pass relentlessly through your existing physical plant with few, if any, modifications to traditional cable architecture and equipment. As such, programmers are readying their origination facilities, while cable operators await the windfall of impervious signals.

But despite that the transition to digital modulation is being greeted as an industry-wide opportunity to increase revenue and compete more effectively, to those of us responsible for distributing these signals, the reality is, of course, more complex.

For example, in origination, programming will not simply be digitally encoded, but often stored and distributed via non-linear devices such as virtual recorders or file servers to the exclusion of traditional videotape. In distribution, NTSC analog video signals will be supplanted by MPEG-2 compressed digital packets. Ku band satellites will be utilized alongside traditional C-band for distribution to headends. In reception, macroblocks will replace sparklies in a compromised satellite signal. Phase linearity will supplant noise temperature as the most important parameter of your LNB. Rain will cause received signals to fade as much as -13dB, routinely. Even the ubiquitous channel modulator will be nudged aside to make room for the digital upconvertor. While at your subscribers' homes, concatenation error will appear in poorly processed video signals, and entropy-loss will cause as many subscriber complaints as triple beat.

If it is to compete effectively with recent entrants into the video to the home market such as telco and the myriad DBS entities, the CATV industry understands that it must re-evaluate its position as the multichannel video provider of choice. It has done this in the past with great success. And, if it is to maintain its dominant position, CATV will simply continue to do what it has traditionally done: deliver more channels of high-quality video programming than would be available through other reasonable means.

Of course "... more channels of high quality video programming" is a subjective phrase at best. Filling this wealth of new channels with appropriate fare posses a challenge to service providers that extends across artistic, commercial and technical lines. But one thing seems clear, high-quality-video-programming will not mean - as we might be led to believe - video programs whose audio consists primarily of little beeps and booms, and whose pictures get by on 50 pixel resolution. If we're going to rise to the competitive challenge, we'll have to do a lot better than that.

For hard won experience has shown us that high quality video programming means what it has always meant: traditional entertainment fare. So, for purposes of this paper and to better understand the network it describes, entertainment fare means programming born of major effort. The stuff requiring scripts, talent, crews, production and distribution: comedy, drama, documentary and athletic endeavor - things that are hard to create on demand, are costly, must be delivered in real-time without substantive impairments, and that eat bandwidth. Additionally, virtually all of this high-end video product is accessible by our subscribers elsewhere. When a subscriber decides to purchase a pay-per-view movie or event, he is saying in effect that he will forego seeing this product live or larger than life on a 50 foot screen in surround sound. Therefore, what we deliver better be as good as we can make it.

This is not reason to lament. Entertainment programming and the bandwidth it devours have served CATV well. Further, given that CATV bandwidth is limited only by our collective imaginations and not by skin effect or government regulation, we in the CATV industry enjoy a substantial advantage over aspirants to our market. So, however fondly we might wish to emulate other industries, programming-and-bandwidth remain CATV's stock-in-trade. System operators provide the latter; programmers the former. And, if this year is remembered for anything at all, it will be remembered as that in which no effort was spared to increase the effective bandwidth of CATV systems through rebuilds, upgrades, and of course, digital compression. What remains to be determined is how well we *programmers* will respond to the need to fill this new space.

This paper will outline one programmer's efforts and their effect on CATV physical plant performance and subscriber/viewer perception. Given that it is the most extensive and technically prescient effort launched in 1997, it covers most of the origination and distribution permutations that might be found to vary from the conventional.

Given that this session is aimed at operating system personnel, and in order that its content be of relevant interest to a broader range of readers, I have included a review of those aspects of coding and compression that to date have been the domain of manufacturers and program networks.

Multichannel PPV

On January 1, 1997, Request Television, the ten year old pay per view network, began distribution of some 40 channels of high-quality feature film and original production sports and variety video programming to cable systems nationwide. In a conventional fm analog modulation scheme, this level of output would require about 300 videotape players feeding signals adequate to occupy 1440MHz of C-band satellite bandwidth.

Of course, dedication of such resources would be economically unfeasible at best. So, in order to accomplish its objective, Request would find it necessary to radically alter and enhance its networks' origination and distribution architecture.

An early adapter of video encoding and compression, Request has been illuminating two C-band transponders with five video signals since early 1993. More recently, with the availability of MPEG-2 compression, it became viable to encode, process, store and distribute an additional 35 video channels in a wholly non-linear, tapeless environment, and to transmit those incremental channels to five Ku-band transponders representing 120mhz of satellite bandwidth.

An evaluation of the operational specifics follows. All of the concepts considered in this paper are currently in operation in the Request TV multichannel distribution architecture.

DIGITAL VIDEO ENCODING

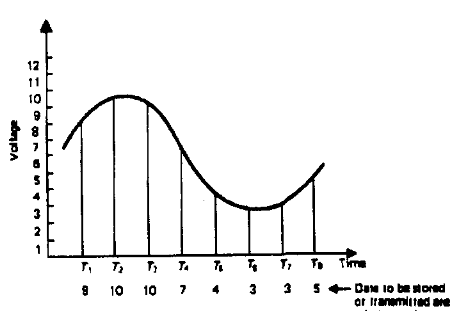
Fundamentals

Television signals nearly always begin and end as analog entities. Digital encoding of these signals at some point or points along their journey from the image target of a telecine or studio camera to the CRT of your subscribers' home TV receivers does nothing to alter that. Further, digital signals require far more bandwidth than do analog signals in order to convey comparable information rates. These are substantive liabilities. Their perception is compounded by a subscriber base that has been conditioned to expect *improvements* to picture quality with the introduction of digital signals. Yet, in reality even the most conservatively coded digital television frame will be no better than the sum of its native analog impairments. Digital encoding can do nothing to *improve* a signal.

What digital encoding *can* do is limit a signal's susceptibility to further degradation. Efficient digital processing will allow signals to be produced and stored on non-linear media, be copied and post-produced with impunity to generational degradation, and be randomly accessed for distribution. Digital encoding can also virtually eliminate contributions from many transmission channel impairments, impairments that would visibly and perhaps objectionably affect many analog signals of comparable strength. Further, a correctly coded non-linear signal will improve receiver transfer function dramatically over f.m. These are substantive *assets*. They will enable us, if we implement them intelligently, to effectively halt our signal's degradation. If we maintain a disciplined approach to the design of our encoding and distribution systems, we will, for the first time, have at our disposal the wherewithal to limit a signal's impairments to those peculiar to its creation, those inherent to its analog form when presented to the digital encoder and those contributed by the analog medium on which it is exhibited. Simply stated, digital processing will allow programmers to deliver video signals to the CATV headend in a condition similar to origination quality. Further, it will allow a cable operator to pass that quality through to the terminals of his digital subscribers' television receivers at cascades well beyond his physical plant's analog design limits. But these extraordinary performance advantages come with a price: bandwidth. And it is how we employ that most valuable of all transmission and distribution system resources that will determine whether the implementation of digital signals improves our product, or simply imposes another, more objectionable, set of signal impairments.

<u>Sampling & Bandwidth</u>

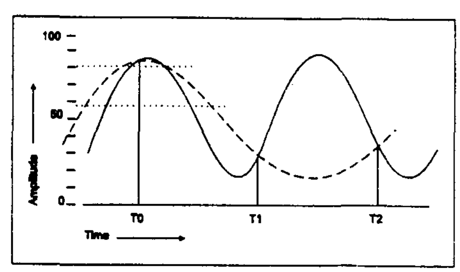
A review of sampling theory will remind us that if an analog signal is to be digitally encoded and reproduced accurately, it must be sampled at a rate not less than twice its highest analog-baseband frequency (f_b). This is the well known Nyquist limit. Therefor, as was stated earlier, digital encoding imposes a bandwidth penalty equal to NLT 2X that of its analog counterpart. However, in practice, to counteract effects such as aliasing, sampling is usually carried out at several times the baseband frequency (fig-1).



{fig-1}

Sampling is merely a mechanism to convert signals from the continuous time and voltage domain characteristic of analog representations, to the periodic. As such, sampling signals are clocked at regular intervals, and their amplitude assigned an integer value. If we know the sampling frequency (f_s) and are able to read the sampled values for each point in time, we can transmit and receive periodic representations of any signal as a series of bits, and upon reception restore that sampled signal to its original form. This requires that sampling signals be pulses of limited duration, occurring at discrete frequencies, thus yielding the channel's fundamental data rate.

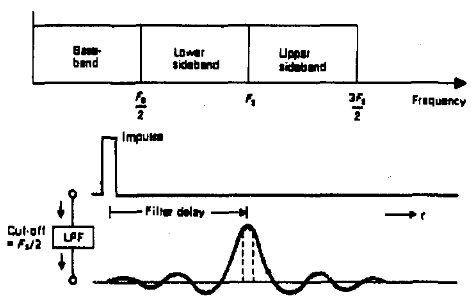
Figure-1 shows an analog wave sampled at $8f_b$. For interval T-1, the analog-to-digital converter would record an integer value of 8. For period T-3, the instantaneous recorded value would be approximately 10, and so on. Upon recovery, this coarse integration would yield a stepped-function waveform. In practice, however, received signals are restored to their original form through filtering. The higher the sampling rate, the more accurately the reproduced waveform will resemble its original counterpart.



{fig-2}

Conversely, when a signal is sampled at too low a rate, there are simply too few samples to accurately reconstruct the original analog waveform (dashed line, fig-2).

Therefor, when encoding video whose highest baseband frequency is (for example) 4.2MHz, the sampling frequency, and thus the initial data rate, must be at least 8.4Mbps. A simple interpretation of Carson's bandwidth rule reveals that a channel must be at least twice that frequency, yielding a minimum channel bandwidth in excess of 16MHz (fig-3).

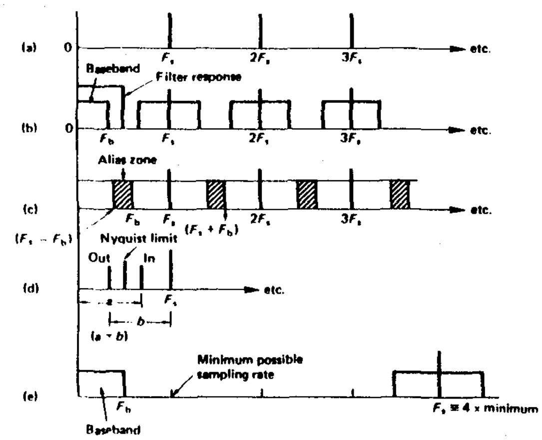


{fig-3}

Again, in reception, the periodic data signal can be reconstructed and returned to the continuous time domain by passing it through a low-pass filter. Practical filters, of course, have imperfect forms and roll off according to a slope rather than exhibiting brick-wall cut-off response. In the sampling process, sum and difference products are generated across a range equal to the sampling frequency plus the highest baseband frequency sampled and the sampling frequency minus the highest baseband frequency:

$$[(f_s + f_b)+(f_s - f_b)]$$

Consider, then, how when an f_s equal to $2f_b$ is heterodyned with that baseband, its highest frequency excursions will produce difference products that fall within the reconstruction filter's passband (fig-4). These difference signals will be reproduced as aliasing artifacts, visible in the reproduced analog video.

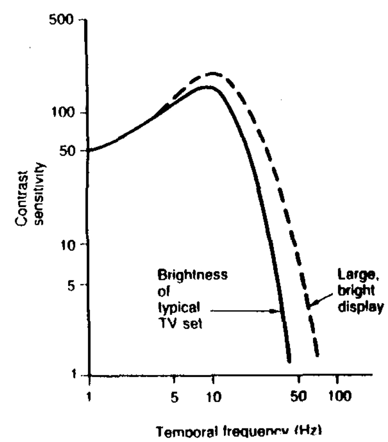


{fig-4}

A commonly employed solution to the problem of aliasing effect is achieved through over sampling. By raising the sampling rate to a frequency well above twice the highest baseband we not only achieve finer integration of the sampled waveform, but place $(f_s - f_b)$ correspondingly outside the reconstruction filter's -3db response. This, however, widens the required channel bandwidth farther still. Add to this the necessity of including overhead bits such as forward error correction, control, and ancillary data to the transmitted digital signal, and both data rate and bandwidth requirements quickly obviate the transmission advantages of digital signals. These disadvantages are overcome through recently developed techniques of digital video compression. However, before embarking on a discussion of compression as applied to Request TV's multichannel origination and distribution model, a discussion of some of the variations of digital video encoding and the specific sampling structures employed is perhaps appropriate.

Component Video Sampling

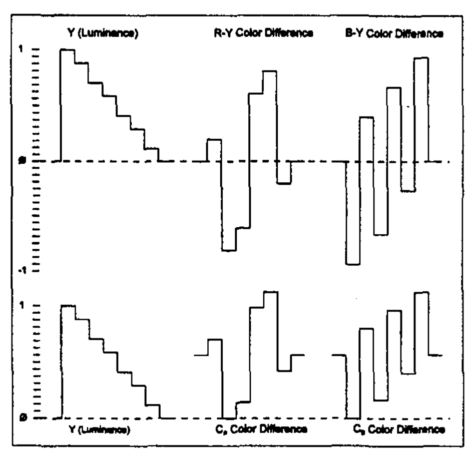
In addition to signaling components, CCIR-601 sets forth the protocol for separating an analog video waveform into its respective luminance (Y) component, which varies directly with a recorded scene's instantaneous brightness, and its two chroma difference signals (R-Y, B-Y) which convey the color information.



{fig 5}

Since the human eye is sensitive to light level changes of about 1% (see fig-5), a television system must accommodate an end-to-end gray-scale of approximately 100:1. If a digital system is to reproduce a visually linear luminance scale that accords with the commonly used IRE (Institute of Radio Engineers) brightness scale, while remaining at least comparable to the analog system's minimum limits, it would seem that it must be capable of quantizing a sampled video signal to 100 discreet values. This is accommodated, in fact exceeded, in professional systems through eight-bit or ten-bit sampling. Eight bit (2^8) sampling yields 256 discreet binary levels, while ten-bit (2^{10}) yields 1024. Both are found in common practice since linear quantization is used in most applications. In Linear quantization each sampling step is of identical size. Results, however, depart from the linear when we consider again that the human eye can detect light changes of 1%. A 1% change at a luminance level of 2 equals 0.02 IRE, while a 1% change at a luminance level of 90 corresponds to 0.9 IRE units.

Each of the component signals (Y, R-Y, B-Y) is sampled at either eight or ten bit levels. Since the color difference signals are bipolar (fig-6), they are either D.C. offset or twos-compliment coded and assigned values which are always positive. To differentiate between encoded color difference components and their bipolar counterparts, color difference signals are called C_r and C_b as opposed to R-Y and B-Y, respectively.

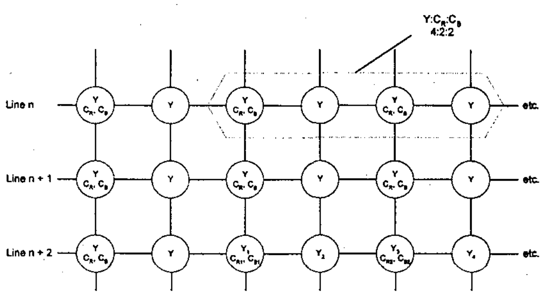


{fig-6}

Luminance is quantized such that values between 0 and 100 IRE units correspond to eight-bit values ranging from 16 to 235, and ten-bit values ranging from 64 to 1023.

Chroma values extend from 16 to 240, and from 64 to 960, respectively.

Given that human visual response to color or hue changes is coarser (less sensitive) than it is to brightness variations of comparable magnitude (see fig-5), the chroma difference signals are sampled at some fraction of the luminance rate. This is commonly 1/2 the luminance sampling rate, but several variations have been introduced and are growing more common as video compression systems proliferate.



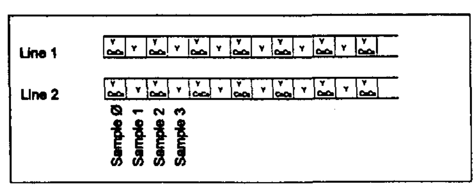
{fig-7}

Conventional CCIR-601 component video encoding yields the sampling profile of fig-7. In this example of a sampled horizontal scan line sequence, we see that Line Segment-1 has a luminance sample (Y) and a pair of chroma samples (C_r , C_b). Segment-2 also has a luminance sample, but no chroma is sampled during its period. The cycle repeats for segments three, four, etc. This cadence is known as 4:2:2 sampling and is the most common sampling pattern encountered in digital video processing: for each 4 luminance samples, 2 each of chroma are taken for a total of eight samples.

Signal processing in the digital domain represents an opportunity to eliminate the vexing interpolation errors common to analog scan conversions. Yet in digital component post-production, a practical difficulty arises from the line (word-length) differences encountered between 525/60 (in actuality the NTSC field rate is 59.94) and 625/50 systems.

Both circumstances have been accommodated in CCIR-601 by choosing a sampling frequency that is an even integer multiple of each format's line period. The H line period of 625/50 systems is 64^{-6} seconds, while that of 525\59.94 = 63.555^{-6} sec. Both of these periods yield whole integer numbers of samples when clocked at multiples of 3.375MHz. Applied in the 4:2:2 sampling protocol , this rate is well above Nyquist, yielding the sampling pattern and frequencies illustrated in fig 7.

Exactly 858 luminance samples are taken in per each 63.555 microsecond period and 864 per 64 microsecond line in the 625/50 system. Chroma is sampled at 1/2 the Y rate. {fig-8}



{fig-8}

These simple sub-multiple patterns contain samples of timing and reference signals such as sync pulses, which are discarded, reducing the number of relevant samples (those representing active picture) to 720 each (Y, C) per active line, hence full CCIR-601 horizontal resolution is equal to 720 pixels.

Composite Video Sampling

In its distribution architecture Request TV makes use of not only component digital video, but composite as well. Described in Society of Motion Picture and Television Engineers (SMPTE) standard 244M, this serial digital interface standard allows for direct digitization of the signal in its composite (intercarrier) form. The distribution standard is produced by sampling a composite baseband at 4fsc, or:

$$4X3.58^6 \text{ Hz} = 14.3^6 \text{ Hz.}$$

Samples are taken by D.C. offsetting the bipolar video signal, setting sync-tip (-40IRE) equal to digital 16, and sampling the entire horizontal line including sync pulses, color burst, and vertical serrations. Nine-hundred-ten (910) samples are taken per line.

The technique is profoundly simple, allowing for the use of compact video tape formats such as D-3. As with component coding, the D-3 digital video tape format enjoys virtual impunity to multi-generation degradations, making it attractive for modest-cost post production use. But more important, in a PPV environment, D-3 permits fitting feature length (up to 180 minutes) program material easily onto a single video cassette, thus minimizing VTR machine-count by at least half.

Composite digital video conveys few advantages, and is not well suited to digital compression, thus limiting its practical use as a transport protocol.

Serial Digital Interface Protocol (SDI)

For a digital signal to be more than a post-production convenience, it must lend itself to distribution. Intra-plant distribution between the various Request TV post-production, control room, file server and compression facilities is carried out utilizing both conventional analog, or wherever feasible SMPTE 259M serial digital transport.

The 259M interface protocol was developed by Sony Corp. and allows for full bandwidth CCIR-601 digital signals to be time multiplexed at manageable data rates. The Y, C_r, C_b samples are integrated serially at 27MHz. For ten bit serial transmission this results in a 270Mbit data stream, easily transported over coaxial cable for intra-facility distribution.

It is this 270 megabit data stream that represents but a single digital video signal at the input to an encoder.

However, given that this paper presumes to discuss the practice of distributing several digital video signals within the bandwidth normally allocated to but a single 27MHz transponder, or 6 MHz analog channel, some discussion of digital compression basics is in order.

DIGITAL VIDEO COMPRESSION

Fundamental Precepts

Digital compression does not supplant sampling theory. In order to accurately encode an analog signal that signal must initially be sampled at the Nyquist rate or higher. This remains independent of the level of compression employed. In fact, accurate sampling at a high data rate will ultimately aid in optimizing compression.

What digital compression mechanisms exploit in order to reduce data rates is a combination of signal redundancy and the anomalies of human visual perception. In most encoded intelligence the sampling rate, thus the coding bit rate, is constant, while the information rate of the signal it is sampling, varies. The difference between sampling bit rate and information rate is a signal's redundancy. To an extent, redundancy is predictable. The difference between a signal's information content and that predictability is entropy. Compression signals isolate these properties of entropy and redundancy, eliminating the latter from the coding process to the extent possible.

Redundancy in a television signal on a frame by frame basis is substantial. It consists at the very least of highly predictable repetitive signaling components such as horizontal and vertical synchronizing signals, equalization pulses, and vertical serrations. Redundancy also exists in interframe scene similarities characteristic of the 60Hz field scanning rate of NTSC television. Statistically there exists a great deal of scene content correlation between successive video frames. If signaling and scene redundancies can be isolated, they needn't be coded or transmitted every frame, instead, redundancies need theoretically to be transmitted but once and memorized at the decoder for inclusion to subsequent received frames.

Further, the 60Hz field rate itself is a deliberate exploitation of the human eye's light retention characteristics. The eye does not respond instantly to light stimulus, but requires between 150 and 300 milliseconds to transfer an image along the visual cortex. The brain then retains that image for about 100 milliseconds. This combination of latency and persistence is exploited in all television signals, whether analog or digital, full bandwidth or compressed. As fig-5 illustrates, the human visual system's response to rapid changes in contrast, i.e. light level, falls off sharply above 10 Hz, with virtually zero response to flicker above 50Hz. Thus, in the frequency domain, a 60 field per second signal, interlaced at 2:1, is sampled at a rate well above the critical flicker frequency of the human eye, is displayed at a repetition rate well within its 100ms retention period, and is reproduced within a bandwidth merely half that required for a comparable *information* rate in a progressively scanned system.

This level of sampling results in acute redundancy between the majority of subsequent frames in the analog NTSC signal. When subjected to compression, such redundancy yields an advantage that NTSC developers never intended, but which remains an advantage nonetheless.

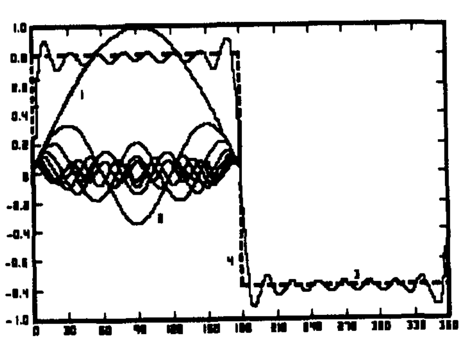
Transform Coding

One way of isolating redundancy, while performing bandwidth reduction occurs through the use of transform coding. In the context discussed here, a transform is a mathematical process for examining data in one domain (in the case of television signals that is either the continuous or periodic time domain), and expressing it in another. To distribute signals in the continuous time domain of analog, requires the bandwidth discussed previously. As such it must be assumed that a signal will traverse all frequencies within its range over time. Digital encoding merely considers such a signal in a *periodic* time domain. Therefor a video channel assumes, at the very least, the entire spectrum from d.c. to 4.2MHz.

Since under no practical circumstances do all frequencies within a given band exist simultaneously, it follows, that if intelligence can be transformed to the frequency domain, and examined periodically, only a code representing the instantaneous frequency of the sampled wave need be transmitted. A substantive coding gain is achieved through such transformation.

In sampled systems, the frequency domain is entered through the Fourier transform. Fourier analysis makes practical use of our understanding that any periodic waveform can be reproduced by adding together some arbitrary number of sinusoids, provided those sine waves are each of a specific instantaneous frequency and phase and are combined in the exact order required to produce our sampled waveform.

While this is of limited practical value, the compliment can be accomplished routinely (fig-9).



{fig-9}

Through Fourier analysis we can examine a specific *existing* waveform, and derive its instantaneous frequency and phase components. If we do this periodically and frequently we can derive enough frequency domain information to reproduce a likeness of that waveform by recombining those myriad samples in an inverse transform. Unlike the continuous time domain of analog, in digital systems the waveform is represented periodically. Through Fourier analysis this produces a finite, or discrete number of samples, resulting in a discrete number of sampled freqiencies. In sampled systems, this type of transform application is known as DFT, or Discrete Fourier Transform, and it is the DFT which forms the basis for data reduction through transform coding as implemented throughout Request Television's compression ensemble.

In order to reconstruct a periodic wave from the linear recombination of sampled frequencies, its instantaneous amplitude, frequency, and *phase* must be known. The DFT performs a spectrum analysis of the sampled input waveform by searching separately for each fundamental frequency each sample contains. This is accomplished by comparing each sample to a series of sine waves of known frequency and phase called basis functions. The instantaneous phase is determined by a wave series, shifted 90^0 to the first and of identical frequency. Sine values are canceled leaving additive cosine components relating to instantaneous amplitude, frequency and phase of the sampled wave. Known as the DCT or Discrete Cosine Transform, this form of sampling is the data-reduction principle upon which most current compression systems are built. Again, transforming from the time to the frequency domain, allows the generation of numerical coefficients representing only the instantaneous frequency and phase characteristics of a sampled wave. Frequencies not present are not transmitted.

It follows further, that if such a periodic signal can be stored and compared with its predecessors, redundancy can be reduced, yielding greater coding gains.

MPEG Coding

Of the MPEG (Moving Picture Experts Group) compression standards that exist today, MPEG-2 is that which will prevail and which has found its way into what is fast becoming widespread commercial network distribution application. However, in order to discuss MPEG-2, those aspects of its genealogy specific to MPEG-1 should be mentioned.

The isolation of redundancy between consecutive frames is the greatest contributing factor in the reduction of data rates when endeavoring to compress full motion video material. The MPEG compression techniques make optimal use of this powerful process. Additionally, MPEG signals are packetized, allowing switched distribution through asynchronous transport protocols and the manipulation of encoded video as digital files.

MPEG-1 is a compression standard developed to produce full motion video at 1.5Mbps. It does so by subsampling which results in noticeable loss of entropy (that information remaining in a video scene or digital file after all redundancy is removed) on fast motion or highly detailed scenes. Additionally, MPEG-1 operates at half the vertical and horizontal resolution of full CCIR-601.

Fig-7 recalls that in standard 4:2:2 sampling the chroma is sampled at 1/2 the H resolution of luminance, while vertical resolution is unchanged. MPEG-1 subsamples all components, ignores alternate fields, and produces a 2:1:0 sampling lattice.

In the context of a discussion of PPV or premium CATV network distribution, what is important to understand about MPEG-1 is its contribution to the evolution of its successor standard, MPEG-2. An examination of the sampling lattice reveals, however, that vertical resolution is the same for both Y and C components. This is exploited in MPEG coding.

MPEG coding reduces redundancy through the use of motion estimation, bi-directional, and both intraframe and interframe coding. MPEG-1 also drastically limits resolution in order to achieve its 1.5Mbps data rate. The protocol is best characterized by its use of I, B, and P frames. These are additional sorts of video frames that allow motion estimation through the comparison of pixel locations between previous and subsequent frames. I frames are complete pictures sent every 10 frames or so, which allow for switching.

B frames are bi-directional predictors which are transmitted and stored at the decoder.

P frames make use of motion vectors and allow the received image to be shifted spatially in reception. All of these features reduce data rate, but add to the memory requirements in the decoder.

Sampled frames are downloaded and buffered in groups that allow for non-sequential processing. Interframe attributes such as motion or redundancy are identified in this manner. The first picture in the group will be intra coded and applied to the DCT. Once transformed and quantized, it is inverse quantized, inverse transformed and buffered for reference. This forms an *I* frame. A *P* frame is then processed by reference to the stored *I* frames and motion information is detected and subtracted from the frame being processed to generate a predicted picture. This frame is transform coded to form a motion compensated, compressed *P* frame which is stored in encoder memory. The encoder now accesses intermediate frames skipped when the *P* frame was created. The encoder calculates forward motion vectors from the locally buffered *I* frame to the *B* picture, and reverse from the *P* frame created immediately prior. Intermediate *B* frames are processed and encoded next, and the process repeats.

In the decoder, the process is synchronized by the recognition of *I* frames, and the inverse transform applied.

MPEG-2 evolved from the precepts of MPEG-1. It is in fact more a set of rules within whose context compression designers may work than it is a rigid standard. But despite that MPEG-2 is a set of rules, observing those rules affords designers and users of the standard an extraordinary range of freedom, not previously available in analog systems.

MPEG-2 proffers a layered coding structure. Such an approach allows for scaleable levels of quality. These are determined by both the programmer and the cable operator. Both encoder and decoder are capable of latitude within the layers of a protocol. Programmers are free to send full resolution, high bit rate data, while in the selection of decoders, the operator is free to choose less complex receivers, his decision governed by his equipment cost and/or the quality level he wishes to purvey to subscribers.

Error concealment, signal to noise ratio, resolution, and entropy propagation are all within the control of MPEG-2 system designers and operators. As such, the Request TV signal assumes many forms and quality levels between production and exhibition.

Contribution quality video is 4:2:2 encoded to 720 pixel resolution, utilizing an 18Mbit data rate. That same signal might reach a digital set top decoder in 4:2:0 format, decoded at 2.5Mbits with a horizontal video resolution of 360 pixels. All fall within the parameters of MPEG-2.

For these reasons, however, when a signal is described as "MPEG-2 compatible," it too often means a great deal less than its purveyors might have us think. MPEG-2 is neither a performance specification, nor an indication of quality. It is more a set of syntactical guidelines, much like those used to make language convey information properly and understandably. And, as with language, there is ample room within its syntax to allow the creation of Hamlet or Disco Duck.

THE NVOD NETWORK

Origination Architecture

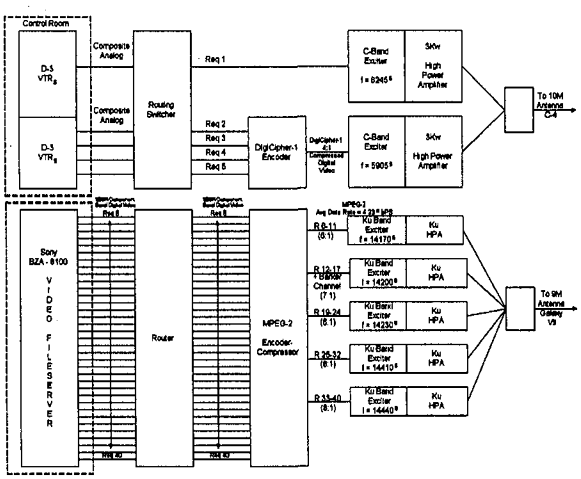
Operating from TCI's National Digital Television Center in Colorado, the Request Television multi-channel NVOD networks comprise some 40 video channels.

Occupying seven satellite transponders, operating in both the C and Ku bands, the network ensemble is designed to produce and distribute a cross-section of signals, accommodating most CATV earth station configurations.

The C-band network group comprises five video channels occupying two 36MHz transponders. Transponder 16, Satcom C-4 remains analog f.m, VideoCipher-II-Plus encoded. Transponder 2, Satcom C-4 is DigiCipher encoded, split-multiplex QPSK modulation, and compressed 4:1.

The Ku band group occupies five 27MHz transponders aboard satellite Galaxy-VII. These are MPEG-2 encoded, DigiCipher encrypted, and QPSK modulated.

Examining the origination architecture (fig-10), it is apparent that expansion to this level of origination followed an evolutionary rather than revolutionary path to full digital distribution. Request-1, like the C-band, analog earth station segment it is designed



{fig-10}

to serve, remains little changed in its 10 year history. Tape based, and functionally analog, R1 employs composite digital encoding only where that scheme might limit post-production generational degradation.

Most incoming source material originates as 24 frame per second (fps) feature length film. Averaging 118 minutes in length, programs are directly digitized and recorded in the D-3 composite digital format. This accommodates either playback to air, or encoding to the file server. If this material is destined for non-linear storage it will be compressed at 2:1 on D-Beta component digital videotape. Component signals are routed via 259M, 270Mbit serial digital interface protocol.

Following the routing of fig-10 it is seen that a common control room is utilized for channels R1 through 5. Though all are D-3 composite encoded, the channels are converted back to analog NTSC prior to input to the router, distributed NTSC throughout the plant, and are split only prior to upconversion at the exciters.

It was determined early in the development of the expanded PPV channels represented by R6 thought 40, that for simplicity of operation, and to accommodate the high degree of title duplication that characterizes NVOD programming, a tape based origination architecture would become unnecessarily complex.

NVOD would differ from conventional pay-per-view in that perhaps 40 incremental movie and event programs (*titles*) would be offered on a monthly basis. Of these, some 30 would be on-air simultaneously. Mostly feature length or longer, programs would be stagger-started. Thus, many titles would require duplication for simultaneous play to air. The initial decision was to make major, popular, and/or premiering titles available for purchase and play at 30 minute intervals. This meant that at any point in time, a feature title would be running on perhaps three networks simultaneously. Only through non-linear storage could a single movie title to be encoded once and randomly accessed for playout. The alternative, 250 video tape players, was an unattractive prospect that would undoubtedly prove both untenable and unreliable in operation.

This of course mandated the use of a file server. However, unlike the sort of server that might occupy a commercial spot playback function in a CATV headend, the type of device necessary for high-end network origination must be of a sort designed to operate within the context of a conventional broadcast automation environment, and more importantly, must be capable of the massive memory requirements needed to store perhaps 160 hours of programming, and to store them at *contribution quality* coding levels.

Contribution Quality Coding

As stated earlier, the scalability of MPEG-2 allows for variable quality levels at appropriate points in a distribution network. This becomes an indispensable attribute given the realities of network processing. In most distribution schemes, digital signals are decoded or transcoded several times along their way to the user. If quality is to be maintained, then, source material must be coded at much higher levels than video at points in the distribution chain closer to the receive sink. Video coded at low bit rates will introduce random-like square response noise to video signals. This noise will be coded, consuming precious bits to the detriment of entropy in subsequent stages if an expected coding gain is to be maintained. Lost entropy cannot be reproduced in later stages. In the case of an underdesigned network, where unrealistic coding gains are assumed, this cycle is repeated causing each subsequent stage to introduce its own artifacts. The aggregate effect is known as concatenation error. Digital artifacts, once coded, will be reproduced in the final picture when decoded at your subscribers' set tops.

Thus, video program material stored as MPEG-2 files at a headend server may be coded at a much lower level than that intended for distribution at the programmer's server. In fact, video material coded at 3.0 Mbps at a headend file server might produce perfectly adequate video quality at the set top, since such coded files need only be robust enough to overcome the normal degradations of a CATV cascade. However, the earlier in the network's hierarchy the signal is coded, the more error correction it must contain, the less entropy loss it can support, and the lower its noise floor must be if it is to produce signals with adequate margin to be coded at 3.0Mbps at the headend. Such are "contribution quality" digital video signals.

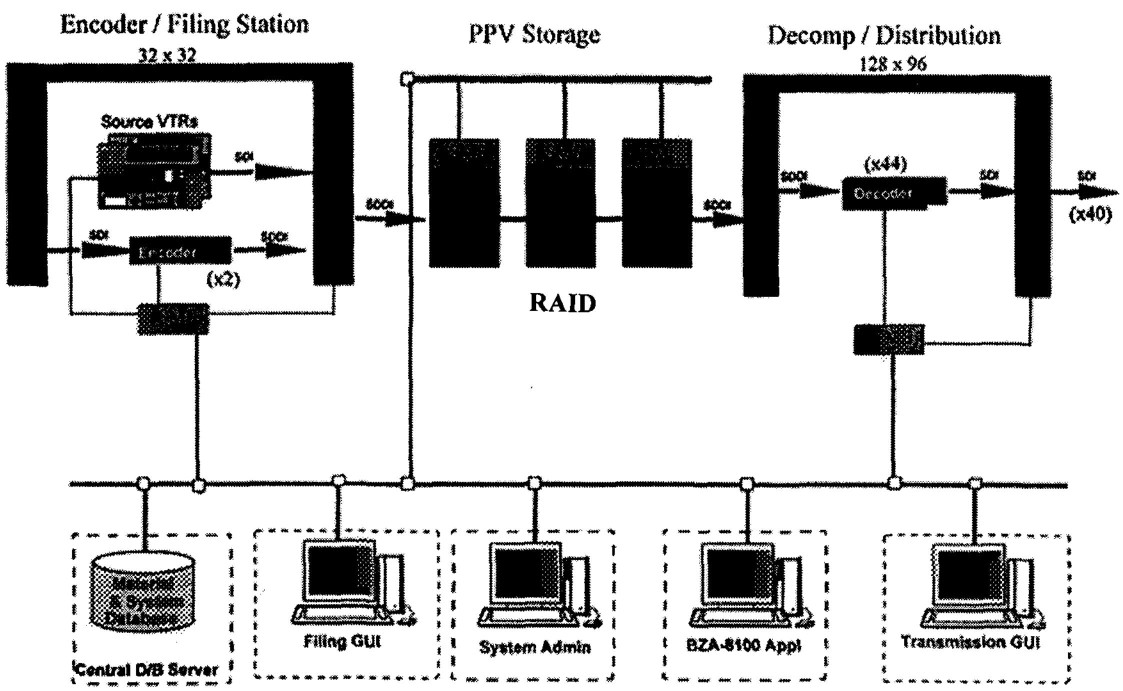
Once a decision is reached regarding the level of entropy loss that can be sustained, the various coding levels (hence memory that must be defined in hardware) that will be accommodated at each point in the distribution chain, can be determined.

Coding some 160 hours of programming at contribution quality bit rates requires massive storage capabilities in an origination server. TCI and Request TV decided on the Sony BZA-8100 video server to fill their storage and quality specifications.

Signals are encoded and stored at main level, main profile, 4:2:2 MPEG-2 coded and distributed in memory for 1:1 redundancy. Coding levels are scaleable at between 5 and 18 megabits depending on application and equipment configuration. Initially, primary video files will be coded and stored at 10 megabits while backup files will be encoded at 5Mbps. The requirement for the nearly 2 terrabits ($1x10^{12}$) of memory needed to support these coding levels was met by the Sony BZA-8100 (fig-11).

and router frames, eliminating single point source(s) of failure.

Encoding into the server represents the first data reduction operation of the chain. Each encoding station will accept analog NTSC, analog component, S-VHS or SDI. For purposes of this discussion, it is here, where SMPTE 259M serial digital video, incoming at 270Mbit is real time encoded



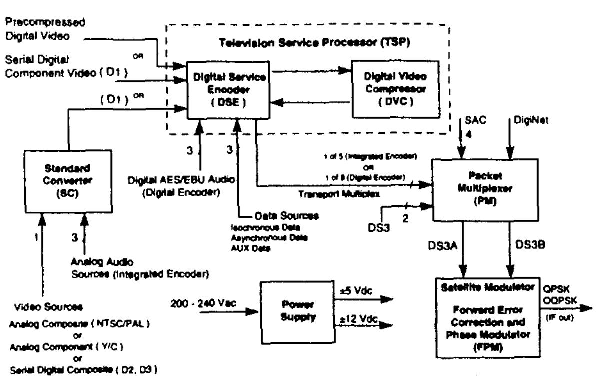
{fig 11}

As configured, the distributed, scaleable architecture of the BZA-8100 server will store program material as MPEG files across some twenty RAIDs (redundant array of independent drives) each configured at 80 gigabits (80^9bps). While initially intended as a spot storage system, the scaleable nature of the device allows both short and long form material to be stored into RAID devices with the long form files distributed automatically across multiple disk arrays

and compressed to 18 megabit, 4:2:2, "studio" MPEG-2 files. Studio MPEG is the highest "contribution" quality compressed signal we will encounter. In early 1997, Request signals will be encoded at 10 megabits, with redundant (backup) files stored at half that level (5Mbit/4:2:0).

It is anticipated that in the backup mode, entropy loss and the inevitable encoding of quantizing square-noise will propagate degraded signals to subsequent network stages. It is believed that video encoded at 5Mbits so early in the chain will not provide adequate margin to cantatenation errors and result in subscriber signals exhibiting objectionable artifacts. Later in 1997, Request signals will be encoded at full studio MPEG-2 (18^6 bps, 4:2:2 sampled). Tests are being conducted to determine the precise contribution level coding that best suits PPV origination. It is projected that a coding level of 9Mbits, 4:2:2 will provide the best compromise, allowing both adequate storage efficiencies to accommodate contribution quality signals, and storage efficiencies appropriate to both primary and backup video files.

The server is capable of 60 discreet outputs. Recalled video files are routed to MPEG-2 decoders and decompressed. Each video output from the decompressors will again be 270Mbit SDI (SMPTE 259M compliant) component digital video. Audio is AES digital encoded at 48Khz and embedded in the serial digital bit stream. This group of signals is routed to the standards converter of the DigiCipher-II encoder (fig-12).



Encoder Functional Elements

{fig 12}

Transport Quality Coding

At the DigiCipher-II encoder multiple channels of serial digital video, each representing one Request video service, are input to the standards converter. MPEG-2 encoded and compressed at the television service processor, these signals now exist at data rates that vary in proportion to the operating bandwidth of the satellite transponder they'll occupy, divided by the number of services multiplexed. For a galaxy-VII Ku band, vertically polarized transponder, signaling data rates and coding budgets would break down as follows:

CODING BUDGET

8:1 Video Compression:	
Satellite Data Rate..................	39.02^6 bps
FEC Data Rate	12.05^6
Information Rate	26.97^6
Messaging & Timing..............	1.62^6
Number of Video Svcs.	8
Data Rate per Svc....................	3.16^6
Audio Data Rate per Svc.	0.40^6
Average Video Data Rate......	**2.77^6**
Required E_b/N_0 Performance.	**3.5dB**
7:1 Video Compression:	
Data Rate per Service	3.62^6 bps
Audio Data Rate.....................	0.40^6
Average Video Data Rate......	**3.22^6**
6:1 Video Compression:	
Data Rate per Service	4.225^6 bps
Audio data Rate	0.40^6
Average Video Data Rate	**3.83^6**
5:1 Video Compression:	
Data Rate per Service..............	5.07^6 bps
Audio data Rate	0.4^6
Average Video Data Rate.......	**4.67^6**

Satellite Distribution:

A digital receive signal's required E_b/N_0 of 3.5 compares very favorably to the 9.5dB C/N ratios we're accustomed to dealing with in the analog domain. But what we gain in earth station performance through digital coding and modulation, we lose in required margin above receiver threshold. Shorter wavelengths make Ku frequencies susceptible to atmospheric attenuations (rain fade). Our 50 watt Ku downlinks will vary in receive EIRP as much as -13dB in the presence of heavy rains. This is far greater than anything we've experienced on the CATV C-band earth station network. While it's not common practice, a C-band receive station can be designed with margins above rf threshold of perhaps 2 or 3dB and never experience a discernible atmospheric fade.

Economics and good operating practice dictate that robust digital coding, shorter, Ku, wavelengths and higher EIRP levels should accrue to the advantage of receive station antenna aperture.

Antenna gain varies as the log of frequency. A parabolic receive antenna of given aperture will exhibit substantially higher gain at Ku than a comparable device at C-band. The following equation shows the relationship:

$$G_a = (20_{\log 10} f)+(20_{\log 10} d)+7.5\text{dB}$$

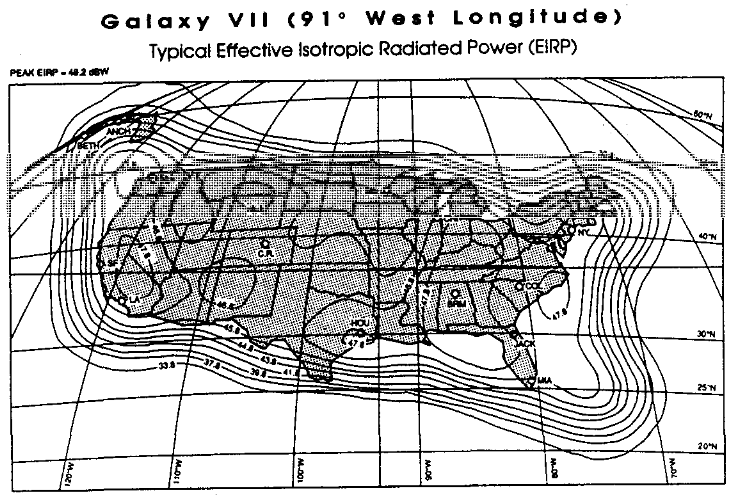
where: Ga = Antenna Gain

f = Frequency in Ghz

d = Antenna Diameter in Feet

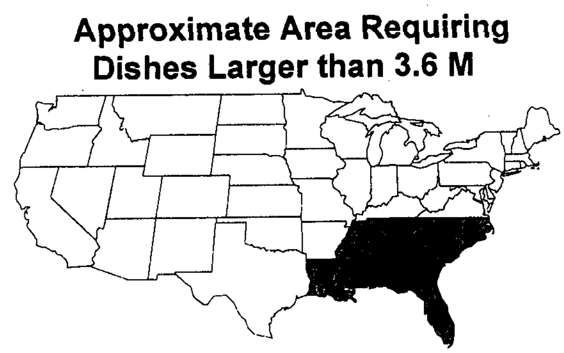
Efficiencies of 55% are assumed

Medium power Ku signals such as those typical of Galaxy-VII will allow the use of antennas as small as 3.6m in most parts of the United States yielding 'four-nines" availability (52 minutes per year).



{fig 13}

Other areas, such as those depicted in fig 14 will require larger R.O.s.



{fig 14}

CONCLUSION

Examining these system architectures and the various data rates employed, it becomes apparent that their determination should be as much a product of economics as of engineering. As was stated at this paper's beginning: bandwidth is precious. But let me add now, that it's precious only to the extent that it can be marketed. As subscribers become sensitized to higher and higher levels of video quality, largely through their exposure to high resolution, progressively-scanned, noise-free computer displays, they make sub-cognitive comparisons to their television picture.

Further, as subscribers migrate toward ever-larger-screen television displays, they grow accustomed to the levels of resolution delivered through well maintained, broadband analog systems. Reducing resolution in a digital system to increase effective bandwidth will yield predictable results, results that are understood intuitively, and are not unlike what can be expected in an analog system. Digital pictures, however, are more complex than that. With MPEG coding we've been handed more latitude to change our product than has ever before been ceded to the technical sector. And our choices are rife with pitfalls that the system designer must understand cognitively as well as intuitively. Nowhere is a lack of such understanding more apparent than in the misguided pursuit of coding gain.

A review of the data rates and levels of coding gain we have discussed in this paper, should leave us all with a sense of control. The feeling, however, is too often one of discomfort as well. For the first time, we, in the CATV community, have at our disposal the wherewithal to exchange bandwidth for content, and do so without apparent loss of quality. We can stuff more and more channels into a given space by forcing higher coding gains, greater entropy losses, and more frequent artifacts to become part of our picture. The too-frequent comparison is made to VHS tape and its consumer popularity. The comparison is grossly invalid.

VHS simply exchanges bandwidth for entropy, rolling off gently and remaining that way.

In the digital domain we not only sacrifice entropy when we sub-sample, we introduce artifacts: arbitrary images not part of the original picture. How many and how severe they appear, is critical.

As stated earlier, the human eye is easily fooled. But only to a point. Response to temporal frequencies, the amount of detail passing the eye per unit time, falls off at roughly the same rate as it does to rapid light changes. Thus, motion artifacts existing at temporal frequencies above 50Hz are easily ignored. However, the eye compensates for this - effectively raising its temporal frequency response - by scanning. The eye scans by simply moving along with a passing image. To illustrate this point, try a simple exercise. Place your open hand eight inches from your eyes. <u>Without moving your eyes</u>, move your hand in a moderately slow waving motion. This creates a high temporal frequency, and the fingers are a blur. Now do the same thing, but <u>follow the movement with your eyes</u>. This lowers the temporal frequency. You see the detail of your hand.

This relates directly to how we perceive motion artifacts in a reproduced digital video picture. When we view a picture on a large screen, and from a moderate viewing distance, the spatial expanse of the image is great enough for our eyes to scan it. Thus, artifacts that might have slipped by on a small studio or control room monitor, will be visible to our large-screen subscribers. Artifacts just below the noticeable level are even more troubling. Artifacts we don't cognitively notice, cause our eyes to fatigue as they try and try to scan, and discern them, especially when experienced over the course of a feature length movie. Eye fatigue causes viewer discomfort. The subscriber recalls the viewing experience as an unpleasant one. Without knowing why, he hesitates before buying again. No marketing survey ever created can identify this potentially devastating circumstance.

We have attempted, therefor, to anticipate what would be perceived as excellent quality signals, while maintaining an aggressive approach to channel loading. The final assessment will rest with our subscribers.

REFERENCES

It is said that copying another's work is plagiarism, while copying the work of several others is called *research.*

The following reference works have been indispensable to the *research* I've committed in developing this paper. I recommend them.

Issues in Advanced Television Technology
by: S. Merrill Weiss

Compression In Video and Audio
by: John Watkinson

Broadband Telecommunications
by: William Grant

Interoperability On The Information Superhighway
by H. Allen Ecker
1994 Technical Papers
NCTA

Television Image Quality
Society of Motion Picture and Television Engineers

Reference Manual for Telecommunications Engineering
by: Roger L. Freeman

NAB Engineering Handbook,
Eighth Edition, 1996
National Association of Broadcasters