THE COMPLETE
TECHNICAL PAPER PROCEEDINGS
FROM:

NCTA
Technical
Papers

A Cost and Performance Comparison Between Uncompressed Digital Video and Lightly Loaded AM Supertrunking Methods within CATV Network Upgrades and Rebuilds

Robert W. Harris
C-COR Electronics, Inc.
State College, PA

Abstract

This paper compares two types of fiber optic supertrunking methods - Lightly Loaded AM (LLAM) and uncompressed digital video - to determine the "overall" cost and performance impact on 750 MHz CATV upgrades and rebuilds employing a supertrunk. The analysis outlines channel loading, performance, fiber requirements, cost, and system compatibility issues for both AM and Digital supertrunking technologies. Included is a discussion of the advantages and disadvantages of each method as well as the associated cost and performance tradeoffs. It is found that the choice of supertrunking technology has a notable impact on the end-of-line system performance and, hence, the total cost of the CATV network upgrade or rebuild.

INTRODUCTION

The advances in AM and digital video fiber optic technology have led to dramatic improvements in CATV network designs that enhance system performance and reliability. One application for both AM and Digital fiber optic technologies has been their use in CATV supertrunking to deliver high quality signals deeper into the network.

Supertrunking is not a new concept in CATV. In the past, supertrunking was done to avoid the relatively higher costs of adding additional headends with their associated earth stations, large buildings, satellite receivers, modulators, and maintenance. Franchise extensions were accomplished through feedforward amplifiers, FM over coax, FM over fiber, and AML microwave. While signal quality at the end-of-line (EOL) improved, these technologies had their limitations with respect to transmission distance, performance, reliability, maintenance and cost.

The objective of today's CATV fiber optic supertrunks is to deliver a headend or near-headend quality signal at the receive site(s) or hub(s) relatively deep in the CATV network. The supertrunk receive site location also serves as a launch point for secondary AM fiber links which distribute services to the subscriber serving area. Future CATV supertrunk hubs may be a point of presence for telephone service transmission similar to the current telephony CO (Central Office). As such, the supertrunk receive site equipment is typically located in an environmentally controlled building with plenty of available equipment rack space.

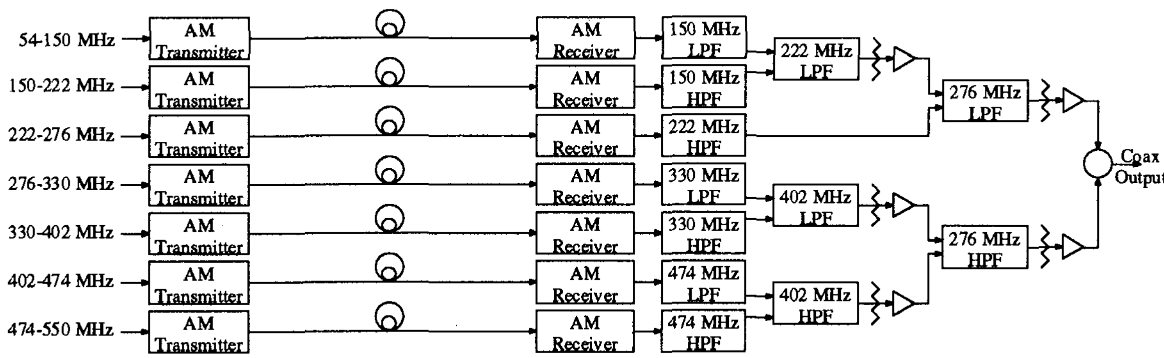
The supertrunk must also support a variety of signal formats such as VSB-AM, baseband and RF scrambled video, digital data, future digitally compressed video, and provide an output that is easily interfaced to the CATV RF distribution. It should have the capability for upgrade and expansion for redundancy and have minimal maintenance requirements.

This paper builds upon the information presented earlier in [1]. In the following pages, two types of CATV supertrunking methods are compared. The first is Lightly Loaded AM (LLAM) which transports about 10 channels per transmitter. The second is uncompressed 8-bit digital video transporting 16 channels per transmitter at 1.6 Gb/s. This analysis describes the channel loading, fiber requirements, performance, configuration, cost and compatibility issues of each supertrunk technology. This paper will also demonstrate the cost and performance impact on the secondary AM fiber link distance and RF distribution operating levels served from the supertrunk hubsite.

LIGHTLY LOADED AM SUPERTRUNK DESCRIPTION

AM supertrunking is described in [1] and [2] as a method for eliminating and/or consolidating headends. This approach utilizes a tiering method whereby groups of 9 to 13 VSB-AM signals are directly modulated onto seven DFB lasers and transported over seven fibers (four fibers using wave-division-multiplexing). A block diagram of an AM

Typical Lightly Loaded AM (LLAM) Supertrunk Configuration
Figure 1.



supertrunk is shown in figure 1. The main advantage to this approach is the frequency arrangement of each tier. Except for the first transmitter, each subsequent transmitter is channel loaded at less than one octave. By loading to octave or less, the composite second order (CSO) distortion products fall out of the band being transmitted by the respective laser.

However, the CSO products which fall out of one RF band will appear in another RF band. To reduce the effect of this problem, each frequency band is filtered such that the distortion products are attenuated before recombining the various RF bands. Pads and post-amplifiers are used to obtain proper isolation and RF output levels. Through a 26 km path, the performance of the LLAM system is approximately 57 dB CNR, -70 dBc CSO, and -70 dBc CTB.

LLAM supertrunks are usually considered when the distance to the hubsite and/or secondary headend is 15 to 35 km away from the primary headend. This means that the LLAM supertrunk will operate through about 8 to 14 dB of fiber loss, assuming 0.4 dB/km at 1310 nm. The loss budget should typically not exceed 14 dB in order to maintain an adequate CNR performance (>55 dB).

After adding in connector, splice and WDM losses, little, if any, additional budget is available for splitting the optical output power from each transmitter in order to share a bank of transmitters with two or more hubsites. Therefore, when multiple hubs are served, the AM supertrunk requires a separate and *dedicated* LLAM link (bank of transmitters and bank of receivers, filters and combinors) from the headend to each hub site.

A complete 80 channel LLAM system with RF outputs typically occupies no more than one half of a six-foot rack at the headend and about the same at the hub site. The receive equipment is usually housed in an environmentally controlled building.

DIGITAL VIDEO SUPERTRUNK DESCRIPTION

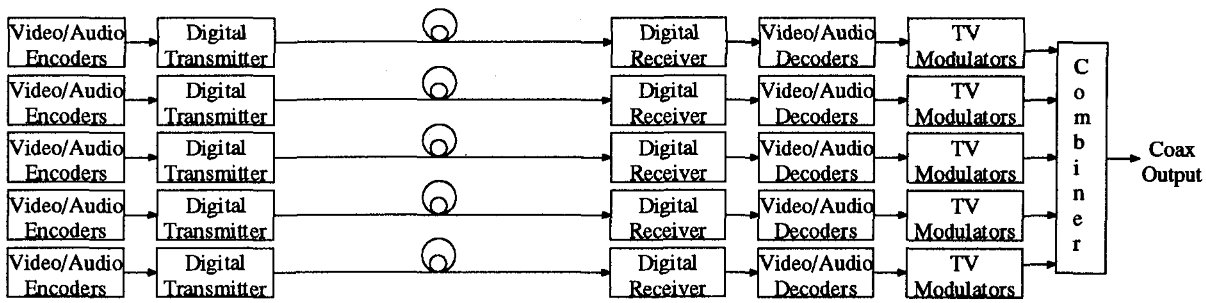
Uncompressed digital video transmission is described in [3] as another method for transporting headend quality video/audio signals to hubsites. One type of system is based on 8-bit video resolution codecs that provide RS250C medium haul performance at each hub site. A typical high speed digital transmission system has a data line rate between 1.6 and 2.4 Gb/s operating at either 1310 or 1550 nm and transports up to 16 uncompressed channels on a single wavelength through an optical loss budget of 30 dB. A block diagram of an 80 channel digital supertrunk is shown in figure 2.

In an uncompressed digital video system, each of the 16 channels are digitally encoded separate from one another then time-division-multiplexed (TDM) to create a high speed serial data stream operating at say, 1.6 Gb/s. The high speed data is then directed to the laser transmitter where it intensity modulates a Fabry-Perot or DFB laser diode.

The digital transmission system provides RS250C medium haul performance which is considered headend quality in that it delivers a 60 dB video SNR. Also, because of the use of synchronous time division multiplexing within the digital network, each channel is completely independent of one another and, therefore, *no* composite distortions are generated.

Digital transmission technology provides consistent signal performance at each hub and is not affected by channel loading, path loss variations or fiber chromatic dispersion as in AM fiber optic technology. Digital signals can be transparently

Typical Digital Fiber Optic Supertrunk
Figure 2.



dropped and repeated and new signals inserted (Add/Drop Multiplexing) at each hub. The digital network can be expanded to offer full optical component and path redundancy. Further, digital systems require very little maintenance and no operational adjustments or optimization.

An 80 channel digital system with RF outputs typically requires no more than one six-foot rack at the headend and about two six-foot racks at the hub site. The digital transmit and receive equipment requires an environmentally controlled building.

COST COMPARISON BETWEEN LLAM AND DIGITAL SUPERTRUNKING

Given below in Table 1 is a direct cost comparison between supertrunking technologies for an 80 channel system with multiple receive sites. The LLAM system cost includes all optical modules, RF filtering and post-amplification. It is assumed that each of the hub sites are fed with a dedicated LLAM system, i.e., no optical splitting of the transmitters.

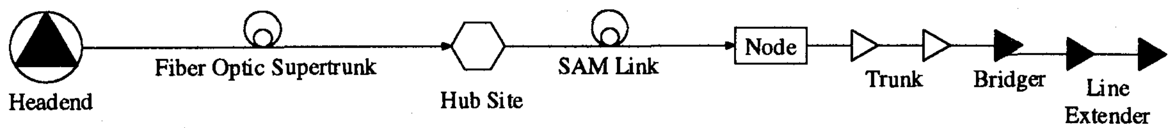
The uncompressed digital fiber network, with a 30 dB loss budget, can make use of optical splitting at the transmitter output to increase the number of receive sites served from a transmit site. Sharing the transmitters with multiple receive sites lowers the total cost of the digital network. The cost provided for the digital network includes all modulators and/or IF to RF output converters for delivering RF outputs.

Table 1 indicates that the LLAM system equipment cost is less than the digital video system equipment cost regardless of the number of receive sites. However, one must look beyond the direct cost differences and consider the *overall* impact from the supertrunk with respect to cost and performance on the entire CATV system. In other words, what effect, if any, does the slightly lower performing (near-headend quality) LLAM supertrunk have on the secondary AM fiber and RF cascade performance? And specifically, what impact do changes in that portion of the system have on the overall system cost due to the finite composite distortions (CSO and CTB) incurred through the LLAM system?

Cost Comparison Between LLAM and Digital Supertrunk
Table 1.

Number of Receive Sites	Equipment Cost for LLAM System ($)	Equipment Cost for Digital Network ($)	Cost Premium for Digital ($)
1	110,000	345,000	235,000
2	220,000	580,000	360,000
3	330,000	815,000	485,000
4	440,000	1,050,000	610,000
5	550,000	1,250,000	700,000
6	660,000	1,480,000	820,000
7	770,000	1,710,000	940,000
8	880,000	1,940,000	1,060,000

Fiber Rich Architecture Employing Fiber Optic Supertrunk, Secondary AM and RF Distribution
Figure 3.



EOL PERFORMANCE AND COST PER MILE IMPACT RESULTING FROM EACH SUPERTRUNK APPROACH IN A FIBER RICH ARCHITECTURE

Figure 3 shows the CATV network diagram that uses headend to hub supertrunking within a 750 MHz fiber rich system loaded with 80 analog carriers. The output of the hub site is a secondary AM link defined in this paper as the SAM link. The SAM link carriers the full 50 to 550 MHz AM-VSB signal format on a single fiber. Following the SAM link are two power doubling trunk amplifiers, a power doubling bridger and two power doubling line extenders. On average, the serving area covered by each secondary AM node is 10 miles of RF distribution. The EOL performance objective is 48 dB CNR, -53 dBc CTB and -58 dBc CSO.

EOL Performance from a Digital Hub

The output of a digital video fiber optic supertrunk yields a performance identical to that of a standard CATV headend. This is a best case initial starting point and allows the maximum optical path losses on the SAM link and the highest operating output levels on the RF amplifiers. The SAM link distance and the RF amplifier operating levels are determined to meet the EOL performance of the system. Table 2 shows the secondary AM path distance, RF operating levels and EOL performance.

EOL Performance from a LLAM Hub

With the hub site fed from a LLAM supertrunk, the system again meets the EOL performance by determining the secondary AM link distance and RF amplifier operating levels. The secondary AM path distance, RF operating levels and EOL performance are given in Table 3.

The LLAM supertrunk provides near-headend quality performance with a CNR of 57 dB. However, even though lightly loading the transmitters with 9 to 13 channels, a finite and measurable level of composite distortions still exist.

Table 3 shows that the LLAM performance does affect the secondary AM node optical path distance as well as the operating levels in the RF distribution. To meet the same EOL performance as achieved with a digital supertrunk, the secondary AM link distance was ***decreased by 1 dB*** and the RF operating levels of the line extenders was ***lowered by 2 dB***. Next, the total system cost impact from the SAM fiber link loss and RF level changes will be discussed.

Secondary AM Electronics Cost per Mile Using Digital Supertrunking

From Table 2, we see that the SAM link is designed for an 11 dB link loss. Typically, the SAM link uses optical splitting in order to serve two or more nodes from a single transmitter. A common transmitter-to-receiver usage ratio (ρ) is 1:2.5 or one transmitter for every 2.5 nodes. In other words, each AM link shares 40% of the cost of an AM transmitter. From this, we can determine the secondary AM cost per system mile from an ideal starting point (a headend or digital hub) using the equation,

$$SAM_{DS} = [(C_T * \rho_1) + C_N] / N_m \qquad \text{eq. 1}$$

where,

SAM_{DS} = secondary AM cost per system mile using Digital supertrunk
C_T = cost of AM transmitter
C_N = cost of AM Node
ρ_1 = AM transmitter usage ratio with Digital
N_m = number of miles of RF distribution served per node.

With C_T = \$13,000, C_N = \$2,000, ρ_1 = 0.4 and N_m= 10 miles, we find the secondary AM fiber electronic cost to be about \$720 per mile. Again, this cost is assuming that the Hub is providing headend quality signals at the input to the SAM transmitters.

Secondary AM Electronics Cost per Mile Using LLAM Supertrunking

With a LLAM fed hubsite, the secondary AM link can still use optical splitting in order to serve two or more nodes from a single transmitter. And each node still serves 10 miles of distribution. However, there is now only 10 dB of fiber link loss to work with instead of 11 dB, a 20% reduction. As

End-of-Line Performance with Digital Supertrunk (Analog carriers to 550 MHz)
Table 2.

Cascade Length	Equipment Description	Output Level 750/54 MHz	CNR (dB)	CTB (dBc)	CSO (dBc)
1	Digital Supertrunk to hub site(s)	N/A	60.0	-120	-120
1	Secondary AM fiber to Nodes, 80 channel loading, 11 dB optical link loss	37/32	51.0	-65.0	-63.0
2	29.5 dB spaced, 750 MHz Power Doubling Trunk, 80 analog channels	38.5/27.5	55.5	-72.5	-67.0
1	37 dB gain power doubling Bridger	46/35	58.5	-69.0	-70.5
2	33 dB gain power doubling Line Extender	47/36	57.5	-59.5	-62.0
		EOL Performance	48.2	-53.1	-58.5

End-of-Line Performance with LLAM Supertrunk (Analog carriers to 550 MHz)
Table 3.

Cascade Length	Equipment Description	Output Level 750/54 MHz	CNR (dB)	CTB (dBc)	CSO (dBc)
1	LLAM Supertrunk to hub site(s)	N/A	57.0	-70.0	-70.0
1	Secondary AM fiber to Nodes, 80 channel loading, **10 dB optical link loss**	37/32	**52.0**	-65.0	-63.0
2	29.5 dB spaced, 750 MHz Power Doubling Trunk, 80 analog channels	38.5/27.5	55.5	-72.5	-67.0
1	37 dB gain power doubling Bridger	46/35	58.5	-69.0	-70.5
2	33 dB gain power doubling Line Extender	**45/34**	55.5	-63.5	-64.0
		EOL Performance	48.1	-53.3	-58.8

a result, this reduces our transmitter-to-receiver usage ratio (ρ) by 20% to 1:2.0 or one transmitter per 2 nodes. In other words, each AM link now supports 50% of the cost of an AM transmitter. From this, we can determine the LLAM fed secondary AM cost per mile. Using equation 1a we find,

$$SAM_{AS} = [(C_T * \rho_2) + C_N] / N_m \qquad \text{eq. 1a}$$

where,

SAM_{AS} = secondary AM cost per system mile using LLAM supertrunk
C_T = cost of AM transmitter
C_N = cost of AM Node
ρ_2 = transmitter usage ratio with LLAM
N_m = miles of RF plant served per node.

With C_T = \$13,000, C_N = \$2,000, ρ_2 = 0.5 and N_m= 10 miles, we find the secondary AM fiber electronic cost to be about \$850 per mile. This cost is assuming that the Hub is providing a near-headend quality signal at the input to the SAM transmitters. Relative to the headend quality digitally fed hub site, this represents an additional \$130 per system mile in secondary AM fiber electronic costs.

RF Electronics Cost Impact due to LLAM Supertrunking

There is also a cost penalty for operating line extender RF amplifiers at lower levels relative to the levels when served from a headend quality hub site. Lower operating levels in the RF distribution results in an increase in the number of active and passive RF components used. Design models [4] have shown that the associated cost penalty ranges from \$175 to \$225 per 1 dB lower operating levels per one mile of CATV system with active return. An equation is given to calculate the additional cost of operating RF amplifiers at lower levels.

$$CP_{RF} = RF_{C/M} * n \qquad \text{eq. 2}$$

where,

CP_{RF} = cost penalty per n dB of RF level change in one mile of active plant
$RF_{C/M}$ = additional RF electronics cost per mile per 1 dB of level change
n = number of dB level change from an optimum headend output

Using the average additional cost of \$200 per 1 dB level change per system mile, we find that 2 dB lower RF operating levels results in an additional cost of \$400 per mile in the RF distribution.

Overall CATV Cost Impact Associated with Each Supertrunking Method

This section investigates the total system cost impact on the secondary AM and RF electronic components as a result of employing a supertrunk. We look at various system sizes and determine the cutoff point where any cost savings from the LLAM supertrunk are offset by additional expenses in the secondary AM and RF portion of the plant.

From the above discussions, we can formulate an equation (equation 3) which characterizes the total system cost penalty from using a LLAM supertrunk system. This model considers system size in miles and assumes that a certain number of hubs are required for a given system size. In this example, it is assumed that a Hub is required for every 500 miles of CATV distribution. The cost premium of the digital supertrunk is then subtracted from the cost penalty associated with the LLAM supertrunk.

$$CP_{AS} = S_m[CP_{RF} + (SAM_{AS} - SAM_{DS})] - CP_{DS} \quad \text{eq.3}$$

where,

CP_{AS} = total system cost penalty from using LLAM supertrunk
C_{RF} = RF electronics cost per mile associated with change in RF operating levels
SAM_{AS} = secondary AM cost/mile when using a LLAM supertrunk
SAM_{DS} = secondary AM cost/mile when using a digital supertrunk
S_m = total number of system miles
CP_{DS} = cost premium of digital supertrunk

The following example illustrates the point. Assume a 2,000 mile plant with four hub sites. From Table 1, the cost premium for a digital supertrunk (CP_{DS}) with four hub sites is \$610,000. From equation 3 we get,

$$CP_{AS} = 2000 * [\$400 + (\$850 - \$720] - \$610,000$$

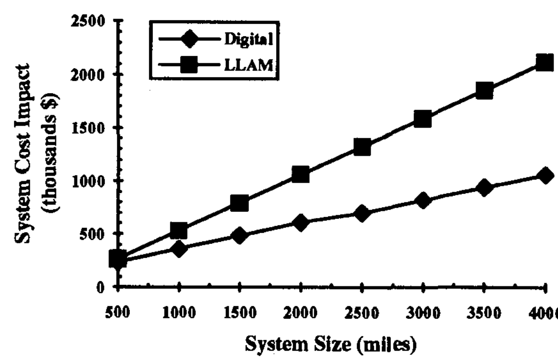
$$= 2000*(\$530) - \$610,000$$

$$= \$1,060,000 - \$610,000$$

$$= \$450,000$$

In other words, the original cost savings of \$610,000 (\$1,050,000 cost for digital minus \$440,000 cost for LLAM) from using a Lightly Loaded AM supertrunk is offset by \$1,060,000 from additional costs in the secondary AM and RF distribution portions of the network. This results in an additional \$450,000 overall system cost from using the LLAM supertrunk. Table 4 and figure 4 shows the total system cost impact from implementing either LLAM or digital supertrunk over a variety of system miles.

Cost Impact of Each Fiber Optic Supertrunk on the CATV Network
Figure 4.



Note that the digital network used in the example is a simple point-to-multipoint star configuration and it is assumed that a single bank of optical transmitters can be split to feed all the receive sites. More sophisticated digital networks with "self healing" and redundancy options and/or longer path losses requiring additional transmitter and/or repeaters will increase the cost of the digital network.

ADVANTAGES AND DISADVANTAGES OF EACH SUPERTRUNKING TECHNOLOGY

LLAM Supertrunk Advantages

There are a number of advantages when using the LLAM supertrunking approach as outlined above. Using wave-division-multiplexing (WDM) only four fibers are required to transport 80 channels. There are no signal format changes that

Total CATV System Cost Penalty When Using AM Supertrunk
Table 4.

System Size (miles)	# of Hubs (est.)	Cost Premium for using Digital ($) From Table 1	AM and RF Cost Penalty from using LLAM ($)	Total System Cost Penalty from using LLAM ($)
500	1	235,000	265,000	30,000
1,000	2	360,000	530,000	170,000
1,500	3	485,000	795,000	310,000
2,000	4	610,000	1,060,000	450,000
2,500	5	700,000	1,325,000	625,000
3,000	6	820,000	1,590,000	770,000
3,500	7	940,000	1,855,000	915,000
4,000	8	1,060,000	2,120,000	1,060,000

need to take place (AM-VSB in, AM-VSB out). Scrambled signals are transported in the same manner as non-scrambled signals. Spare equipment requirements are relatively modest and inexpensive.

For the system architecture shown here, the overall system cost is not significantly impacted when the Hub site is fed from a Lightly Loaded AM supertrunk and the RF distribution served from that Hub is less than 500 miles.

The rack space requirements are modest - about one half of a six foot rack in the headend and the same in the receive site. The receive site equipment can even be located within an outdoor pedestal that is environmentally controlled with a relatively small (< 2,000 BTU's) air conditioning unit.

LLAM Supertrunk Disadvantages

The main disadvantage of LLAM supertrunk is its cost impact on the secondary AM and RF distribution electronics in the network. This is the result of its slightly lower performance (near-headend quality) relative to uncompressed digital. For the given model above, this occurs in supertrunk Hub distribution areas larger than 500 miles.

Also, for an 80 channel LLAM supertrunk, there can be as many as twelve RF bandpass filters. The use of these filters yields seven cross-over channels. Each cross-over channel will suffer a slight degradation relative to the non-crossover channels in carrier level and CNR performance.

Additionally, each AM link will require optimization for obtaining the proper RF output and signal performance from each frequency band. If optical automatic level control (ALC) is not employed within each transmitter and receiver, path loss and transmitter output power variations over time may cause one or more of the RF frequency bands to vary in level relative to one another causing each group of frequency bands being "off" in level.

Digital Advantages

Uncompressed digital video networks provide an identical and consistent headend quality signal performance at each and every hub site regardless of path loss. Digital systems do not require optimization of each link or path. Optical loss budgets of 30 dB allows multiple splitting of the transmitter output so as to share the cost of the transmit equipment with multiple receive sites. WDM can be used to transport 80 uncompressed digital channels over three active fibers with a spare window available for future use.

Neither diplex filtering nor post-amplification is required and there are no cross-over channels in a digital system. The RF output levels and performance at each digital receive site does not vary with changes in the optical path or additional channels added to the network. At least one digital video equipment supplier can now transport all forms of IF scrambling. The technique used to accomplish IF scrambling allows transportation of digitally compressed video using 64 QAM and/or 16-level VSB carriers.

A relatively simple and inexpensive digital network (point-to-multipoint) can be installed and later upgraded to a redundant and automatic "self-healing" ring network that provides the maximum level of protection from fiber path or component failures. With an installed digital network, a platform is in place for future growth into regional networking.

Supertrunking Comparison
Table 5.

80 Channel Supertrunk Features	Digital	LLAM
Consistent Signal Performance at each Hub Site	Yes	No
Variation in performance due to different path losses or channel loading	No	Yes
RF Diplex Filters required	No	up to 12
Number of Cross-over Channels	None	7
Potential for variation in RF output level due to fiber electronics	No	Yes
Requires RF Scramblers at receive site(s)	No	No
Transports *all* forms of scrambling (Baseband and IF)	Yes	Yes
RF Scrambled channels transmitted per wavelength	16	Variable
Requires Modulators or IF/RF converters at receive Site(s)	Yes	No
Post-amplifiers required at receive rite(s)	No	Yes
CNR	60 dB	57 dB
CSO (worst case)	None	-70 dBc
CTB (worst case)	None	-70 dBc
Number of fibers required without WDM	5	8
1550 nm optical terminals available	Yes	Yes
Number of fibers required with WDM	3	4
Platform in place for interconnected hub sites and regional networking	Yes	No
Transports digitally compressed video carriers	Yes	Yes
Transparent optical repeating	Yes	No
Optical loss budget available	30 dB	< 14 dB
Maximum point-to-point distance	100 km	<35 km

Digital Disadvantages

Digital video networks do, however, require considerably more rack space and power consumption than an AM supertrunk approach. This is primarily because each channel is processed separately. Unlike the VSB-AM in/out from AM fiber systems, digital system inputs typically require baseband video and either baseband audio or 4.5 MHz audio subcarrier. Modulators or IF/RF upconverters are required at each hub site. Each digital receive site requires as many as 2 seven-foot racks and air conditioning of about 7,000 BTU's. Table 5 shows a comparison between LLAM and digital supertrunking.

CONCLUSION

Two methods of fiber optic supertrunking for CATV applications, Lightly Loaded AM and uncompressed 8-bit digital video, have been examined to determine the impact each technology has on the overall cost and performance. LLAM supertrunking provides near-headend quality performance through less than 12 dB of optical loss while digital video transmission provides headend quality through 30 dB path losses. Both technologies can employ WDM to reduce active fiber count requirements.

The slightly lower performance of the AM supertrunk relative to digital supertrunking has been shown to directly add expense in the RF distribution and secondary AM portions of the CATV system. This leads to the requirement of, 1) operate the RF distribution at lower operating levels and, 2) design lower path losses for the secondary AM links that serve the nodes.

As a result, the total system cost can actually be *less* when using a digital as opposed to a LLAM supertrunk. This is due primarily to the performance difference of the LLAM relative to the digital supertrunk which results in shorter allowable secondary AM path links and lower RF distribution levels which translate into additional AM transmitter and RF electronics costs, respectively.

This analysis has shown that there are more factors associated with the selection of a supertrunk than simply the head to head cost comparison. The CATV network design engineer must carefully consider the overall system cost impact from each method of supertrunking. Factors include transmit/receive site building size, power consumption and cooling requirements, as well as any additional RF distribution and secondary AM costs.

ACKNOWLEDGMENT

The author wishes to thank Jeff Sauter and Tim Mooney of C-COR Electronics for their suggestions and insight in preparing this paper.

REFERENCES

[1] R.W. Harris, "Digital Video versus AM Supertrunking: A CATV Cost and Performance Analysis," CED, July and Aug. 1993.

[2] D.E. Frymyer, R.E. Pidgeon, F.R. Little, "Advances in AM Supertrunking," Conference on Emerging Technologies Proceedings Manual, pp. 19-24, 1993.

[3] R.W. Harris, "Uncompressed Digital Video Fiber Optic Systems for CATV," CED, pp. 45-52, Dec. 1992.

[4] C-COR Electronics, Inc. System Design Group

A GENERALIZED FRAMEWORK FOR CATV TRANSMISSION ON FUTURE BISDN

Geng-Sheng (G.S.) Kuo

Department of Information Management, National Central University
Chung-Li, Taiwan 32054, R.O.C.
TEL: +886 3 4263086 FAX: +886 3 4262309

Abstract

The purpose of this paper is to investigate and construct a generalized framework for CATV transmission on future BISDN by using some new concepts and mechanisms of UNIX. The software developments have been made. In addition, some future works have been pointed out briefly.

INTRODUCTION

Clearly, the HDTV standard has already been decided in all-digital format, which is compatible with packet-oriented BISDN technologies. Hence, the HDTV technology can not only be used in broadcast networks, but also be used on the BISDN [8]. Furthermore, the BISDN is going to impact the CATV industry [10]. In the future, internetworking will construct one network in the world including BISDN, LANs, wireless networks, and so on. Due to frequency limitation on radio communications, BISDN will play the role of basic backbone for future communications.

The ultimate target of future communications will be the universally personal communications. Due to its established subscribers, CATV industry will keep the key role for information service provider on BISDN for business consideration. The information sources for a variety of future CATV-based multimedia information services will definitely be located on BISDN. To provide real-time-oriented CATV-based multimedia information services to large volume of users *simultaneously* is a must for the success of applications.

The purpose of this paper is to design a generalized framework for BISDN-based CATV information source (CATV server) serving the information services to a large volume of users simultaneously by using distributed client-server architecture with socket and fork mechanisms in UNIX.

A GENERALIZED FRAMEWORK

Up to now, it might be one of the best solutions that the interfaces between the CATV-based terminal system set or multimedia workstation in user's side and information sources elsewhere are based on distributed client-server architecture [3]. In [4], a new better mechanism for distributed client-server architecture on the personal communications networks than the ones currently available has been designed. This new mechanism associating with the information management system designed in [1, 2] could play the role of basic framework for *interactive* customer control of future personal communications services [4].

On existing computer networks, UNIX is a well-accepted operating system. The socket mechanism of 4.3BSD UNIX [5] [6] is a special means of application program interfaces to a variety of different communication protocols [7]. The distributed client-server architecture needs socket as the means for interconnecting many client processes to some specific server process in order to provide corresponding users with the same information service *virtually* simultaneously. However, based on the current status of 4.3BSD UNIX and our experimental experiences on Sun workstation SparcStation SS10, the total socket number available for the server process is 256 and the number of child processes generated by one parent process through fork mechanism is small too [9] [11].

In this paper, many study efforts and experimental experiences on the topic mentioned above have been made for shaping the proposed generalized framework. It is our conclusion that the multi-client multi-children-server single-parent-server architecture communicating with multiple pipes and sockets is the generalized framework for CATV transmission on future BISDN.

Parent server opens *P* pipes, generates *P* children in order to let every child own a private pipe. The value *P* is decided by `MAXFORK` defined in the following program. Then, it reads *K*-byte data from the file, sequentially sends them to its children with pipes. After the *K*-byte data have been send to all destinations, parent reads next *K*-byte data. The read-write procedure will repeated until parent reads the end of the file.

Every child opens *S* sockets, binds every socket to a unique port number. So, there is a port number range for every child, and the intersection of these ranges is null set. Then child enters a loop. Child receives data from its pipe, stores them in its buffer, and waits for a short time which is decided by `timeout` (a static `timeval` structure variable) to see if there are new client connections reaching. If there are indeed, child performs `accept()` function to establish connection with the client(s). Then, child sends the data in buffer to all connected clients with sockets. The loop procedure will not end until child can not read any data at all from its pipe.

The server, including parent part and child part, can be developed as the following program.

```
/*
 * Example of max client connection number testing
 */
#include "inet.h"

#define MAXFORK         3
#define MAXSD           3
#define BEGINPORT 10000


char *pname;
int  line[MAXLINE];

main(argc, argv)
int  argc;
char *argv[];
{
     int    childpid, pipefd[MAXFORK][2], i, n, f;
     int    sockfd[MAXFORK][MAXSD], acptflag[MAXFORK][MAXSD], newsockfd, clilen, maxfdpl;
     int    maxsd = MAXSD, readynum, cc;
     char   buffer[MAXFORK][MAXLINE];
     FILE   *fp;
     char   bufferpnt[MAXLINE];
     fd_set        writeto[MAXFORK], ready;
     static        struct timeval timeout;
     struct        sockaddr_in peer;
     int    peerlen = sizeof(peer);
     struct sockaddr_in  serv_addr, cli_addr;
```

```
pname = argv[0];
timeout.tv_sec = 0;
timeout.tv_usec = 5;


for(i=0; i<MAXFORK; i++) {
        if(pipe(pipefd[i]) <0)
                err_sys("can't creat pipe%d (No.%d)", i, pipefd[i]);
}

for(f=0; f<MAXFORK; f++) {

        if ((childpid = fork()) < 0 )
                err_sys("can't fork ");

        if (childpid == 0) {/* child */

                for(i=0; i<MAXFORK; i++) {
                        close(pipefd[i][1]);
                        if(i!=f) close(pipefd[i][0]);
                }
                printf("child(%d): pipe%d (No.%d) is ready for
                  reading\n", f, f, pipefd[f][0]);


                FD_ZERO(&writeto[f]);

                for(i=0; i<MAXSD; i++) {

                        /* set the accept flags of the sockets off */
                        acptflag[f][i] = 0;

                        /*
                         * Open a TCP socket (an Internet stream
                         *  socket).
                         */

                        if ((sockfd[f][i] = socket(AF_INET, SOCK_STREAM,
                          IPPROTO_TCP)) < 0)
                                err_dump("socket() error! Max socket
                                  No.= %d", i);


                        /* Bind our local address so that the client can
                         *  send to us.
                         */

                        bzero((char *) &serv_addr, sizeof(serv_addr));
                        serv_addr.sin_family    = AF_INET;
                        serv_addr.sin_addr.s_addr = htonl (INADDR_ANY);
                          /* INADDR_ANY _long)0x00000000 */
                        serv_addr.sin_port
                                = htons(f*MAXSD +i +BEGINPORT);
```

```
        if (bind(sockfd[f][i], (struct sockaddr *)
          &serv_addr, sizeof(serv_addr)) < 0)
                err_dump("child(%d): bind() socket%d
                  (No.%d) error!", f, i, sockfd[f][i]);

        listen(sockfd[f][i], 5);
        fprintf(stdout,"child(%d): socket%d (No.%d)
          (port %d) is listening !\n",f, i,
          sockfd[f][i], (f*MAXSD +i +BEGINPORT));

}

maxfdpl = sockfd[f][maxsd-1] + 1;
printf("child(%d): maxfdpl = %d \n", f, maxfdpl);

clilen = sizeof(cli_addr);

while(1) {

        for(i=0; i<MAXSD; i++)
                FD_SET(sockfd[f][i], &writeto[f]);

        readynum = select(maxfdpl, &writeto[f],
          (fd_set *) 0, (fd_set *) 0, &timeout );
        if(readynum < 0)
                err_sys("srever: select() error");

        if((n =readn(pipefd[f][0],buffer[f],
          MAXLINE))<=0) {
                printf("child(%d): closes pipe%d
                  (No.%d)\n", f, f, pipefd[f][0]);
                close(pipefd[f][0]);
                for(i=0; i<MAXSD; i++)
                        close(sockfd[f][i]);
                exit(0);
        }

        printf("child(%d) :\n", f);
        if(writen(1, buffer[f], n)<0)
                err_sys("child error: writen error on
                  screen\n");


        for(i=0; i<MAXSD; i++) {

                if (acptflag[f][i]) {
                        if(writen(sockfd[f][i],
                          buffer[f], n)<0)
                                err_sys("child(%d):
                                 writen error on socket"
                                 , f);
                }
```

```
                                if ((FD_ISSET(sockfd[f][i],
                                 &writeto[f]))&&(!acptflag[f][i])) {

                                   printf("child(%d): port %d is
                                    selected !\n", f, (f*MAXSD +i
                                    +BEGINPORT));

                                   if((sockfd[f][i] =
                                    accept(sockfd[f][i],
                                    (struct sockaddr *) &cli_addr,
                                    &clilen)) <0)

                                       err_sys("child(%d):
                                        accept() error", f);

                                   else
                                   {
                                       acptflag[f][i] = 1;

                                       if (getpeername(sockfd[f][i],
                                        (struct sockaddr_in *)&peer,
                                        &peerlen) < 0)

                                          err_sys("child(%d):
                                           getpeername() error", f);


                                       fprintf(stderr,"child(%d):
                                        select from %s\n", f,
                                        inet_ntoa(peer.sin_addr));

                                   }

                                   if(writen(sockfd[f][i],
                                    buffer[f],n)<0)
                                       err_sys("child(%d): writen error
                                        on socket", f);

                                }

                        }


                } /* end while: read from pipe, write to sockets loop*/

        } /* end child */

} /* end fork */

                                /* parent */

for(i=0; i<MAXFORK; i++)
        close(pipefd[i][0]);

if((fp = fopen("data", "r")) == NULL) {
```

```
            printf("error: fopen() ! \n");
            exit(0);
      }
      else printf("parent: open file ok, transmitting data ....\n");

      do {

            if(fgets(bufferpnt, MAXLINE, fp) == NULL) {
                  for(i=0; i<MAXFORK; i++) close(pipefd[i][1]);
                  fclose(fp);
                  printf("parent: close descriptors ...\n");
                  exit(0);
            }

            n = strlen(bufferpnt);

            for(i=0; i<MAXFORK; i++) {
                  if(writen(pipefd[i][1], bufferpnt, n)<0)
                        err_sys("parent: writen() error on pipe%d
                         (No.%d)\n", i, pipefd[i][1]);
            }

            sleep(1);

      } while(feof(fp) != EOF);
}
```

Client acquires a port number from command line, and binds its socket to that port in order to connect to server. It receives data from the socket and shows them on screen. The following program performs it.

```
/*
 * Example of client using TCP protocol.
 */
#include    "inet.h"
#define BUFLEN 1024

main(argc, argv)
      int            argc;
      char          *argv[];
{
      int            sockfd, port, n;
      char          buffer[BUFLEN];
      struct sockaddr_in serv_addr;

      /* name = argv[0]; */

      if(argc < 2) {
            printf("Usage: c ###      , ### indicates the port number\n");
            exit();
            }
```

```
        argv++; argc--;
        port = atoi(&argv[0][0]);


        /*
         * Fill in the structure "serv_addr" with the address of the
         * server
         * that we want to connect with.
         */
        bzero((char *) &serv_addr, sizeof(serv_addr));
        serv_addr.sin_family = AF_INET;
        serv_addr.sin_addr.s_addr = inet_addr(SERV_HOST_ADDR);
        serv_addr.sin_port = htons(port);

        /*
         * Open a TCP socket (an stream socket).
         */
        if ((sockfd = socket(AF_INET, SOCK_STREAM, 0)) < 0)
                err_sys("client: can't open stream socket.");
        else printf("client: socket() ok \n");


        /*
         * Connect to the server.
         */
        if (connect(sockfd, (struct sockaddr *) &serv_addr,
                    sizeof(serv_addr)) < 0)
                err_sys("client: can't connect to server");
        else    printf("client: connect() ok, socket is connected \n");

        str_read(sockfd);

        close(sockfd);
        exit(0);
}
```

The question is that: what is the maximum number of the clients that can be served by this server? It is dominated by the following two variables:

P: the maximum number of child server processes generated by the parent server process on the same machine.

S: the maximum number of socket descriptors that a process can open.

So, it is important to know the values of these two variables because the available client connection number is dependent on them.

FUTURE WORK

Our first-cut results on this generalized framework for distributed client-server architecture on packet-oriented networks are very impressive, which might be beneficial to all CATV industry and multimedia applications by sharing the same public BISDN. Many important issues need to be studied. The comparison between high-speed protocol XTP and TCP/IP is needed for CATV-based information services. Further research efforts and experimental trials on this generalized framework are needed and have been planned in order to improve its functionality and performance.

ACKNOWLEDGEMENT

I would like to thank Mr. Tsin-Hai Chang, one of my best students, for his contributions to the implementations of software programs in this paper.

REFERENCES

[1] Kuo, G. S., "New Systems Concepts on HDTV System," in the 1991 NAB HDTV World Conference Proceedings, in Las Vegas, Nevada, on Apr. 15-18, 1991.

[2] Kuo, G. S., "Systems Architecture of Information Management System for Future HDTV System," in the 1992 NAB HDTV World Conference Proceedings, in Las Vegas, Nevada, on Apr. 13-16, 1992.

[3] Kuo, G. S., "Some Control Issues for Future HDTV System," in the Conference Proceedings on High Definition Video, International Symposium on Fiber Optic Networks and Video Communications, in Berlin, FR Germany, on Apr. 5-9, 1993.

[4] Kuo, G. S., "A New Framework for Customer Control of Future Personal Communications Networks," in the Fifth International Conference on Wireless Communications, in Calgary, Alberta, Canada, on Jul. 12-14, 1993.

[5] Stevens, W. R., UNIX Network Programming, Prentice-Hall, Englewood Cliffs, N.J., 1990.

[6] Sun Microsystems, Network Programming Guide, Revision A, Mar. 27, 1990.

[7] Ogle, D. M., Tracey, K. M., Floyd, R. A., and Bollella, G., "Dynamically Selecting Protocols for Socket Applications," IEEE Network, pp. 48-57, May 1993.

[8] Aoyama, T., Tokizawa, I., and Sato, K., "ATM VP-Based Broadband Networks for Multimedia Services," IEEE Commun. Mag., pp. 30-39, Apr. 1993.

[9] Kuo, G. S., "A Generalized Framework for HDTV Transmission on Future BISDN," in the Proceedings of International Workshop on HDTV '93, in Ottawa, Canada, on Oct. 26-28, 1993.

[10] Dukes, S. D., "Next Generation Cable Network Architecture," in the Proceedings of Technical Papers from the 42nd Annual NCTA Covention and Exposition, in San Francisco, California, on June 6 -9, 1993.

[11] Kuo, G. S. and T. H. Chang, "A Generalized Framework for HDTV-Based Multimedia Transmission on Future BISDN -(II)," in the Proceedings of NAB '94 MultiMedia World Conference, in Las Vegas, Nevada, on Mar. 20-24, 1994.

A STANDARD SOFTWARE PLATFORM FOR DIGITAL INTERACTIVE TELEVISION

Robert Thibadeau, Ph.D.
School of Computer Science
Carnegie Mellon University

James Large, B.S.E.E. and Joe Newcomer, Ph.D.
Television Computer, Inc.
Pittsburgh, Pennsylvania

Abstract

This paper describes a standard for the representation of portable application programs in the television set and set-top environment. The representation solves many practical problems in distributing digital interactive television services.

THE TELEVISION COMPUTER

The word "digital" in "digital interactive television" is a code word for "computer." Digital interactive television, in any conceivable meaning, implies "the television set contains a computer." The television computer makes it possible, in fact, outrageously important, to investigate how to apply the "hardware-software" distinction we exploit in computer science.

As we will see, the result of the investigation reveals straightforward, economically viable and stimulative, solutions to many problems in cable television and telecommunications in general, including problems of copyright ownership and interoperability among systems.

The purpose of this paper is to present the technology that enables a sharing of the "software" of digital interactive television by varieties of "hardware." This is to reduce the "box count" in the living room, while, perhaps, at the same time increasing the "invisible box" count in the house and office. This goal is kind not only to the consumer, but also to the box maker because of manufacturing volume requirements. A hardware manufacturer can make more money by contributing components of widely accepted boxes than from diverse, low volume, boxes. Software companies can make more money by contributing components as well.

We coined the term "television computer" [1][2][3] to refer to the computer in the television. This is distinct from the term "telecomputer" coined by Jim Clark [4] that extends past computer, television, and into telephony as well. There are many television computers. Most televisions sold today contain micro controllers. Every game machine and converter box contains micro controllers. To date these controllers are "hard programmed" in the sense that they are re-programmable only at the factory. This situation will change.

Open programmability is on the horizon and already available in a number of boxes, such as the Philips CD-I television computer. However, there is no way to program, or write software, that can operate reliably across the different television computer platforms. Each platform must be individually coded. This is true whether we state that the meaning of "platform" is the "raw hardware" or, even, "the software operating system on the raw hardware." In other words, there is no way to write a piece of code that is the same across software operating systems. We believe there is a fairly simple, straightforward, and well understood way to free up this bottleneck without otherwise "giving up the store."

THE G-CODE SYSTEM

Our proposal is analogous to the NTSC standard for analog video data streams into (and out of) television sets. NTSC allows any supplier of television programs to send those video programs to any body's television set. Open programmability for digital interactive television then, quite naturally, means the transmission of computer software, digital programs, into television sets with the same guarantee of interoperability. If there is technology that can provide open programmability by controlling the form of the transmission of computer software into television sets, this technology should be examined. The technological know-how does exist. There are many computer scientists aware of this. To date the proposals have been for proprietary standards.

Our proposal is distinguished from others in that it proposes a free and public technology. Our technology does not replace, but properly augments, the many valuable contributions made by the authors of the proprietary standards, other software makers, telecommunications companies, and hardware manufacturers. Television Computer, Inc., copyrights the work, but this is in order to provide a mechanism for a single authoritative source for the software. The company is also committed to cooperating with the various industry standards groups and we have endeavored to place the contribution in the context of actual or emerging industry standards.

The clean separation of software and hardware in the television computer is achieved by providing a software layer that speaks a standard language. But what should this language be? Our view is that the language should not be the top level of the operating system, but should rather be a level that allows direct programming of the television computer. This programming level should not be a high level programming language, such as "C" or "FORTRAN," but a low level programming language similar to "assembly code." It has been recognized and well known for decades that all computers share basic sets of operations at the "assembly code" level. This shared set, at minimum, can provide a universal framework for minimal programming.

What is required of the operating system software is an "interpreter" that can read the universal assembly code (commonly called "virtual machine code") and translate this into the precise hardware environment in which that operating system resides. A second, and last, minimal requirement is for a few well defined "system calls" for input and output from the computation that the television computer can be made to perform. Every operating system must support these few system calls that take the form of code call outs to the real operating system.

We have named this code "G-Code" where "G" stands for "Group." The ***group allocator*** mechanism described below provides a means by which different "networks or channels" can operate on one set-top without the possibility of interfering with one another. But, basically, G-Code was inspired by the "P-Code" or "Pseudo Code" of the old UCSD Pascal Compiler from the 1970s. In contrast to "P-Code" and others such as the GNU C intermediate, Microsoft C P-Code or even FORTH, G-Code was developed expressly for the purpose of creating a standard software platform for digital interactive television. This idea is not new to either science or practice. Another example is the IBM-Apple consortium's proprietary "Kaleida" operating system and high level language that uses proprietary (but probably similar) instruction codes to achieve interoperability between the MacIntosh and IBM PC platforms.

For the G-Code to be interesting, it must be simple and universally interpretable. The requirement of simplicity has to do with the manufacturing and materials cost of putting a G-Code interpreter into every television or set-top box. ***The incremental cost to support G-Code must approach insignificance.***

To this end, the present proposal is to specify G-Code as a possible "payload" within a SMPTE Header/Descriptor framework. MIT initiated efforts to define a Header/Decriptor framework for advanced television and digital video systems, and this prompted formation of a Society of Motion Pictures and Television Engineers (SMPTE) task force on Headers/Descriptors. The task force completed its work early in 1992 and a working group, whose work is nearing completion, was formed to take the task force report and produce a standard. The SMPTE Header/Descriptor provides for unique identification (using the ISO/ITU registration system) of payload encoding rules. Descriptors are defined to provide for payload parameterization and the insertion of application oriented information including copyright notification. Payloads can be digital video programming, such as MPEG compressed NTSC, or, as in our case, transportable computer codes. The header standard is itself embedded in an international standard for declarative syntax known as ITU/ISO ASN.1 [6]. Done in this way, if a television does not support any G-Code interpreter, it can simply reject payloads that have the G-Code header. This is done simply because a single internationally standard authorizing number does not match. If it does support G-Code because the header number matches a number in the television's capability list, it can accept the payload and interpret it. Thus, the incremental cost of G-Code, through a standard already worked out internationally, does indeed approach zero dollar cost.

The next step up in cost is when the G-Code interpreter is present in the operating system of the television computer. We provide a G-Code interpreter for as many micro controllers as possible (without discrimination but dependent on resources). Furthermore, unlike the several propriety systems, G-Code is sufficiently simple and straightforward that a vendor can freely "roll his own version," if, for one reason or another, he does not desire to use ours. This would allow any operating system vendor to include a G-Code interpreter at minimal cost. The working G-Code Draft Document is available from Television Computer, Inc. [7].

At the core of the G-Code system is the G-Code Virtual Machine — an abstract computer that does not favor any one vendor's hardware over another's. Programs written for the G-Code virtual machine are executed in either of three ways:

- An emulator directly interprets the program's instructions and system calls.

- The program's instructions are translated to native code, and the native code is executed in a software environment that emulates the system calls.

-or-

- The hardware directly implements the G-Code instruction set.

The G-Code virtual machine is designed to facilitate translation of G-Code programs to both CISC and RISC style native instruction sets. G-Code defines a rich set of instructions that allows a translator to make effective use of the rich instruction sets of typical CISC processors. For RISC type machines, it contains hints about flow of control and storage classes that simplify register and branch optimizations.

Some set-top and television manufacturers may wish to make only a minimal commitment to supporting the G-Code standard. For this reason, the G-Code virtual machine is defined as a ***core instruction and system call set*** and a set of ***standard virtual machine extensions.*** The virtual machine core specifies the minimum set of data types and operations needed to allow software to be loaded and run on any set-top or television. It provides for integer arithmetic, simplistic graphics, and the most essential system calls.

The standard virtual machine extensions are optional instructions and data types that, if either present or absent, can be handled by the program, or by the G-Code interpreter's ***group allocator.***

To take one example: A hypothetical "atlas of the planets" program might use floating point arithmetic. When run on set-top box A, that has a floating point co-processor, the G-Code floating point operations are translated to co-processor instructions, and executed directly. On another set-top box, B, there is no floating point hardware support, but the manufacturer has elected to provide software emulation for the floating point G-Code extension in native code. Finally, on a third set-top box, C, that does not support the floating point G-Code extension, the app can still run if its group allocator provides a software emulator for the extension.

In the example above, if the app's group allocator provides floating point emulation in a discardable library, and the app is run on either set-top box A, or B, then the group allocator can discard the emulator library and reclaim the associated memory, knowing that the app-provided emulator will not be needed on those set-top boxes.

If a certain set-top box contains a unique "hardware accelerator", the manufacturer is encouraged to define a non-standard G-Code extension, thereby enabling application vendors to use the accelerator. If it is meaningful for them to do so, application vendors can define software extension emulators to mimic the custom hardware and thereby enable their applications to run on other vendor's set-top boxes. (Presumably, they will not run as fast without the custom hardware). In order to handle speed of operation variations, there is a standard extension G-Code system call provided for clocking. An application can use this when it requires certain real-time constraints to operate.

A "group" in our proposal refers to a single managed collection of software programs made to run on the television computer through initial "G-Code" booting. In practice, Groups are identified by the SMPTE Header under a branch of the ISO/ITU name space. For example, a header number {iso(1) organization(3) smpte(52) 68} could be the prefix used to identify a SMPTE defined class for downloadable television G-Code. In this hypothetical case, the ID and identifiers with this prefix (for example, {1 3 52 68 1}) uniquely identify G-Code programs in the class. A G-Code allocator, in keeping with the proposed standards, could be declared in the Descriptor portion (the parameterization) of the SMPTE Header which contains the G-Code payload.

The "Time Warner Full Service Network" might have (a version) that request to use {1 3 52 68 1} as its "G-Code publisher's" number. Since this number, as per the SMPTE standard, is unique, the group will have a unique Header ID and identifier in the operating system of the television computer. Because we cannot guarantee symmetric communications in and out of the television computer, it is impractical to have the television computer generate a supposedly unique number for a group.

However, if symmetric communications is possible in a particular plant, and in any manner that symmetric communications is possible, the G-Code Group allocator mechanism allows the definition of cooperative communication processes between any two or any one-to-many configuration of tasks running on different machines.

The most significant potential problem with something like a G-Code standard for digital interactive television is that the code may not be efficient on a particular piece of hardware or for a particular computer programming language. The microcomputer manufacturers extend the instruction sets of their computers precisely in order to provide higher computing efficiencies. Furthermore, high level programming languages like "C," "Basic," and "Pascal," and scripting languages like "Lingua" and "ScriptX," also may not compile or interpret efficiently in G-Code. Indeed, the box makers and computer language people extend the system calls and sometimes the instruction sets in order to accomodate specialized hardware "accelerators." The most well known instance of this is the "sprite controller" in game boxes such as Atari, Sega, and Nintendo. We believe that our G-Code proposal effectively and correctly addresses both the hardware interoperatability and software interoperability problems. We also believe that it simultaneously addresses the scaleability problem that allows a box maker to control the cost of supporting G-Code.

Once there is a commitment to have the set-top or television set handle SMPTE Header/Descriptors, basic G-Code support is designed to have a real cost that is as little as a fraction of a dollar. At the other end, for the more aggressive among us, it facilitates efficient code deployment for high-end digital interactive television. This in no way decreases the value, or necessity, of native code in high, or low, end hardware.

Security

We understand that the very idea of allowing someone else to run a computer program in one's television computer can send shivers up the spine. What happens if the program grabs the television's display so the poor homeowner has to "power cycle" his television to get it back? What happens if the program has a "bug" and "crashes?" Curiously, the technology that can guarantee against these failures (except when there is an outright electronics failure that interacts with computer state conditions) has been well understood for many years. The basic idea is quite simple:

- Because the G-Codes are symbolic and actively interpreted by the computer,
- because they do not run directly on the hardware or in the operating system, and
- because the Group allocator mechanism allows program groups to be isolated from each other,

any G-Code task can, at worst, get itself messed up. This technology is rarely employed in PCs, but it has been common in workstations, and mainframes for decades.

We believe that it should be overtly, and standardly, employed in television, precisely to offer the interoperatibility that permits the classic interplay of independent hardware, transport, and content. A box maker, for example a Sega, might have a G-Code demonstration that runs, albeit slowly and only illustratively, on a Nintendo box, and, vice versa. Furthermore, with G-Code Groups, Sega can distribute highly optimized G-Code to varying models of Sega boxes.

In summary, the box and the operating system are secure. This security is absolute. It is impossible to create a virus using G-Codes except for a virus internal to one's own publishing group. There is no "password" mechanism outside of a group.

The word "impossible" is very strong. We should caution that, if a hardware or operating system, or software vendor, publishes a G-Code extension that opens the door on his hardware, operating system, or software, the door is opened to anyone using that extension. It is up to these people to protect their own security when they proactively generate and publish new extensions.

"First Copy" Protection

We will close this short paper with what we regard as perhaps one of the most interesting examples of the use of G-Code that we call, "first copy protection." This is the kind of protection offered by pay-per-view. However, it can be controlled on a finer grain, by the publisher. A publisher can change his encryption frequently. He uses a G-Code program to create a permission for the receipt of further programming. In receivers that can decrypt an entire program the permission can provide absolute first copy protection.

Of course, once a receiver has gained permission to play the program once, it would be possible for a hacker to gain an illegal copy in principle (even if hardware architecture may make that difficult). Nevertheless, this is copy protection that is under the control of the publisher. In effect, he downloads G-Code that hides a permission.

Interestingly, this does not disenfranchise the forms of copy protection that have been developed for pay-per-view. Technically, at least in the computer world, the method employed by box makers like General Instrument and Scientific Atlanta for "addressable converters" is known as "multicasting." First copy protection through G-Code downloading is enhanced by a multicasting architecture, because multicasting secures an *authorization,* as opposed to a *permission,* on a box-by-box and event-by-event basis. The Group Allocator mechanism provides the means of tying the permission with the authorization. A G-Code program cannot obtain *authorization* without a G-Code extension provided by the hardware manufacturer that gives the G-Code, or certain G-Code groups (e.g., the Turner or Viacom XYZ channel) such authorization.

The Information Highway

G-Code programs are designed to support the information highway, again through the adoption of international standards. Structured text is supported by the G-Code system. Structured text objects are represented by arbitrary length sequences of characters. The structuring is based on SGML or "Standard Generalized Markup Language" developed at the National Institutes of Standards and Technology and widely employed on the Internet. SGML is made to enable computer documents to be marked up for retrieval, search, and playback by specific "players" (such as MPEG players). For example, in one SGML application, World Wide Web (from CERN in Geneva, Switzerland), one can traverse the global community on Internet through hyperlinks. SGML is on the other end of the spectrum from the SMPTE Header/Descriptor standard. G-Code provides a natural method by which publishers can control the region in between the two. An SGML interpreter is possible on anybody's box, even with only core G-Code support.

SUMMARY

The G-Code system is a software system, freely distributed, that adequately and correctly describes a standard for the representation of portable application programs in the television set and set-top environment. The G-Code system has the following technical attributes:

- MODULAR SYSTEM with STANDARD EXTENSIONS requires only a minimal commitment from a set-top box maker who wishes to comply with the standard.

- EFFICIENT, HIGH-PERFORMANCE implementations are possible using many different RISC and CISC type processors.

- LOADABLE EXTENSION EMULATORS allow any application software to be run on a minimal system in which the standard extensions are not built-in.

- DESCRIPTOR BASED ADDRESSING allows each application to run in a protected address space, even when there is no specific hardware support.

- MULTITASKING allows one vendor's information services to be "on line", even when another vendor's services are "in use".

- EXTENSIBLE RESOURCE MANAGEMENT allows different applications to share not only the standard resources such as memory, audio-video display, infrared remote control, PCMCIA cards, and printers, but also APPLICATIONS

- DEFINED HARDWARE AND SOFTWARE RESOURCES.

- STANDARD, COMPACT ENCODINGS for STRUCTURED TEXT and GRAPHICS.

REFERENCES

[1] Thibadeau, R. *The Television Computer*. Pittsburgh, Pennsylvania: Television Computer, Inc., 1992.

[2] Thibadeau, R. "The television as a robot servant." Technical Report CMU-RI-TR-93-22, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, August, 1993.

[3] Thibadeau, R. "The question of standards in digital interactive television." Technical Report CMU-RI-TR-93-23, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, November 1993.

[4] Clark, J. "A TeleComputer." Unpublished manuscript, Silicon Graphics, Inc., 1992.

[5] Gerovac, B. & Soloman, R. "Protect revenues, not bits: identify your intellectual property." Unpublished manuscript, Richard Soloman, MIT DOHRS Program at the Center for Technology, Policy and Industrial Development, Cambridge, Mass. Also see June 1992 issue of *SMPTE Journal.*

[6] ISO 8824/8825 also, Rose, M.T. *The Open Book: a practical perspective on OSI.* Prentice-Hall, 1990.

[7] "The G-Code System: First Draft for Comments. " Technical Report TCTR-7, Television Computer, Inc., Pittsburgh, PA 15203. Phone 412 488 3610 or Fax 412 488 3611, $75 First Class, $90 Overnight U.S./CAN., MC/VISA/AMEX.

AC-3: The Multi-Channel Digital Audio Coding Technology

Craig C. Todd
Dolby Laboratories
San Francisco

Abstract

AC-3 is a digital audio compression technology which has been designed to provide very high quality audio at bit rates on the order of 64 kb/s per audio channel. This paper will cover the background of the development of AC-3, the details of the coding techniques which allow AC-3 to achieve very high coding gains, and the features of AC-3 which offer true solutions to a number of problems inherent with audio program delivery to a broad audience.

1. Introduction

In the United States, the High Definition Television (HDTV) standardization process formally began in 1987 with the formation, by the Federal Communications Commission (FCC), of the Advisory Committee on Advanced Television Service (ACATS). The initially proposed sound systems involved matrix-encoding a multi-channel audio source into a stereo pair, which was then digitally encoded. At the receiver, the two audio channels would optionally be decoded into four channels using a matrix decoder. This proposal basically used the 4-2-4 multi-channel matrix system as a 2-1 bit-rate reduction system, since the matrix technology reduced the bit-rate required to convey four channel audio by a factor of 1/2.

By 1990, there were suggestions by some, that the limitations of the 4-2-4 matrix technology should be avoided in HDTV, and that four discrete audio channels should be transmitted. Others felt that the small gain in multi-channel performance was not worth the required doubling of the bit-rate. It was at this point that the concept of AC-3 was born: a fundamentally multi-channel audio coder, operating at approximately the same bit-rate required by two independently coded audio channels, while offering multi-channel audio performance without the limitations of the 4-2-4 channel matrix system. Basically, it was realized that the bit-rate reduction process that was being performed by the 4-2-4 matrix system could be better performed in the coder itself, and not by an addition of the matrix technology to a two channel coder. If successful, the concept of AC-3 would allow HDTV to gain the benefit of going to discrete audio transmission without paying the price of a corresponding doubling of the required bit-rate.

While conceived in response to a need for HDTV, AC-3 was first realized in response to a similar need in the cinema. By 1990, a 4-2-4 matrix based analog sound system had been in place in the 35 mm cinema for some 14 years, and interest was growing to offer the cinema industry new digital sound technology. In 1988, a SMPTE subgroup had studied the issue of how many sound channels a new digital film sound system should offer. The conclusion was that 5.1 channels should be provided (left, center, right, left surround, right surround, subwoofer), which is identical to the 70 mm split surround format which has been in use in the cinema since 1979. In order to place digital sound data on film reliably, and yet not interfere with either the picture or analog sound area of the film, the available data rate is limited. It was determined that 320 kb/s of error corrected audio data could be reliably placed in and extracted from the film area between the sprocket hole perforations on one side of the 35 mm film. The first cinema industry demonstrations using AC-3 began in May of 1991. By Dec. of 1991, the first AC-3 coded digital film, Star Trek VI, played in three theatres. The formal roll out of the Dolby SR•D system (as it is called) was in June of 1992, with release of the film Batman Returns.

In mid-1991 the existence of the AC-3 audio coding system was publicly disclosed, and AC-3 was eagerly embraced by the HDTV audio community in the United States. The ITU-R (formerly CCIR) Task Group 10-1 met in June of 1991 and accepted the basic 5 channel audio format, making it the basis for a recommendation. In February of 1992 the U.S. Advanced Televisions Systems Committee released a document formally recommending composite coded 5.1 channel audio for the U.S. HDTV service. In Oct. of 1992, TG 10-1 accepted the 0.1 low frequency channel and modified the draft recommendation accordingly. In 1993 the AC-3 system underwent subjective testing in the United States in order to evaluate its suitability for inclusion in the HDTV system being proposed by the Grand Alliance (a consortium of the remaining U.S. HDTV proponents which has been authorized to collaborate on the U.S. HDTV broadcast system). In Oct. 1993 the Grand Alliance recommended the use of Dolby AC-3. In Nov. 1993, the full ACATS committee formally approved the use of AC-3 by the Grand Alliance HDTV system. Final system tests are expected to be completed in early 1995, final FCC approval should occur in late 1995, and some initial broadcasts may begin in 1996. While the U.S. HDTV process will take some additional time to complete, AC-3 is expected to begin a widespread roll out in cable television and direct broadcast satellite equipment in 1994.

There are a diverse set of requirements for an audio coder intended for widespread application. The intent was to make AC-3 into a universally applicable low bit rate coder by satisfying many diverse requirements. This allows AC-3 to deliver an audio signal in a form usable by an entire audience, even though different members of the audience have different needs. While the most critical members of the audience may be anticipated to have complete multi-channel reproduction systems, most of the audience may be listening in mono or stereo. Some of the audience may have matrix based multi-channel reproduction equipment without a discrete input, thus requiring a 2-channel matrix encoded (and generally not mono-compatible) output from the AC-3 decoder. Most of the audience welcomes a restricted dynamic range reproduction, while a few in the audience will wish to experience the full dynamic range of the original signal. There is a market in serving the visually and hearing impaired. All of these (and other) diverse needs were considered early in the design of the AC-3 coding technology. Techniques to satisfy these needs have been designed in from the beginning, and not added on as an afterthought.

2. Filter Bank

The choice of the type of analysis/synthesis filter bank to employ in an audio coder is a trade-off between frequency resolution, time resolution, and cost, where cost is measured in random access memory bits (RAM) and multiply/accumulate cycles (MACs). Steady state audio signals benefit from finer frequency resolution, whereas transient signals require finer time resolution. It is not possible to simultaneously achieve fine time and fine frequency resolution, so a compromise must be reached. Fine frequency resolution has a very real cost, in that longer blocks of audio must be buffered which requires larger amounts of RAM. Cost of RAM dominates in the single chip decoders which are demanded by the market. Very fine frequency resolution can allow somewhat higher coding gains to be achieved, but requires significantly more costly RAM.

AC-3 makes use of the oddly stacked time-division aliasing cancellation filter bank described by Princen and Bradley. Overlapping blocks of 512 windowed samples are transformed into 256 frequency domain points. The filter bank is critically sampled, and of low complexity, requiring only 13 MACs for computation. A proprietary 512 point Fielder window is used to achieve the best trade-off between close-in frequency selectivity and far-away rejection. Each transform block is formed from audio representing 10.66 msec (at 48 kHz sample rate), although the transforms are performed every 5.33 msec. The audio block rate is thus 187.5 Hz. During transient conditions where finer time resolution is useful, the block size is halved so that transforms occur every 2.67 msec. The low 13 MAC computation rate is maintained during transient block size switches. The frequency resolution of the filter bank is 93.75 Hz. The minimum time resolution is 2.67 msec. The full resolution of the filter bank is used; the individual filters are not combined into wider bands (or critical bands), except during a portion of the core bit allocation routine. Bit allocation can occur down to the individual transform coefficient level, with neighboring coefficients receiving different allocations.

3. Bit Allocation

The challenge of bit-rate reduction of audio signals is, of course, to remove bits from a representation of the audio signal in a manner that is inaudible (to humans). There are two broad classes of bit allocation strategy: forward adaptive; and backward adaptive. Forward adaptive (fig. 1) refers to the method where the encoder calculates the bit allocation and explicitly codes the allocation into the coded bit stream. In theory, this method allows for the most accurate allocation since the encoder has full knowledge of the input signal and (in principle) may be of unlimited complexity. The encoder may precisely calculate an optimum bit allocation within the limits of the psychoacoustic model employed. An attractive feature of forward adaptation is that since the psychoacoustic model is resident only in the encoder, the model may be changed at any time with no impact on an installed base of decoders. While forward adaptation has some attractive features, there are practical limitations to the performance which can be obtained with this technique. The performance limitation comes from the need to utilize a portion of the available bit-rate to deliver the explicit bit allocation to the decoder. For instance, the ISO MPEG1 layer II audio coder requires a data rate of nearly 4 kb/s/ch to transmit the bit allocation with time resolution of 24 msec and frequency resolution of 750 Hz. During transient conditions it is beneficial to be able to alter the bit allocation with a much finer time resolution but, of course, that would require a significantly greater allocation data rate (nearly 12 kb/s/ch for 8 msec time resolution in the case of the LII coder).

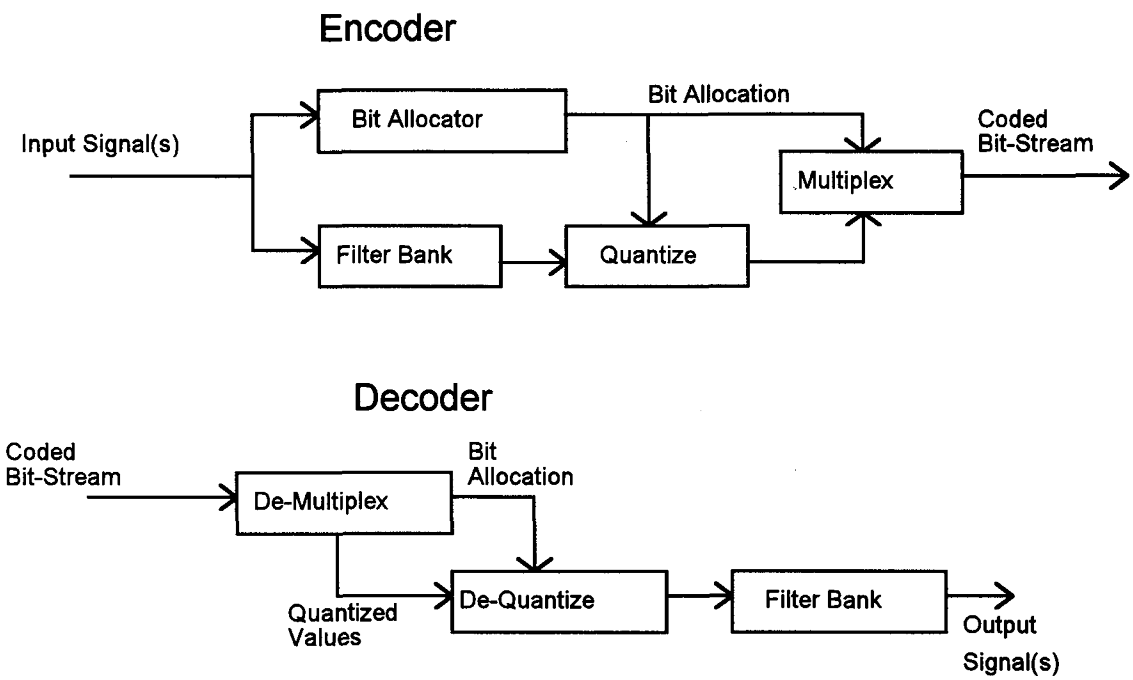


Figure 1: Forward Adaptive Bit Allocation

During steady state conditions it is useful to be able to allocate bits with much greater frequency resolution. For example, in the case of a signal with a spectrum which contains spectral lines in every 750 Hz frequency band, all bands would have to be allocated bits and very limited bit-rate reduction may be obtained before audible degradation occurs. If bits can be allocated on a finer frequency grid, then even spectrally dense signals may have bits removed at frequencies between the spectral lines, and more useful bit-rate reduction may be obtained. Like an improvement in time resolution, an improvement in the frequency resolution of the bit allocation would also require a substantial increase in the allocation data rate (an increase by a factor of 8 in order to improve the frequency resolution by a factor of 8). While attractive in theory, forward adaptive bit allocation clearly does impose significant practical limitations on performance at very low bit-rates.

Backward adaptive bit allocation (fig. 2) refers to the creation of the bit allocation from the coded audio data itself, without explicit information from the encoder. The advantage of this approach is that none of the available data rate is used to deliver the allocation information to the decoder, and thus all of the bits are available to be allocated to coding audio. The allocation may have time or frequency resolution equal to the information used to generate the allocation. Backward adaptive systems are thus more efficient in transmission, and allow the bit allocation to have superior time and frequency resolution. The disadvantages of backward adaptive allocation come from the fact that the bit allocation must be computed in the decoder from information contained in the bit stream. The bit allocation is computed from information with limited accuracy, and may contain small errors. Since the bit allocation is intended to be able to be performed in a low-cost decoder, the computation can not be overly complex, or else decoder cost would not be low. Since the bit allocation algorithm available to the encoder is fixed once decoders are deployed into the field, the psychoacoustic model may not be updated.

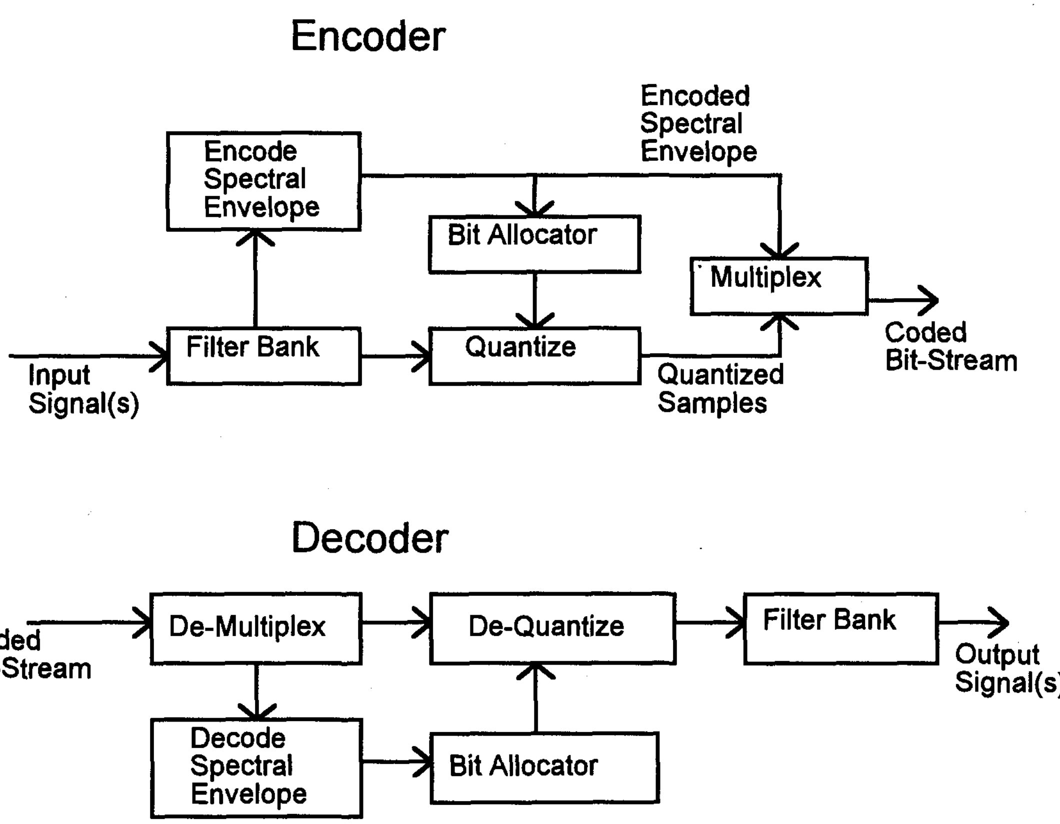


Figure 2: Backward Adaptive Bit Allocation

The AC-3 coder makes use of hybrid backward/forward adaptive bit allocation (fig. 3), in which most of the disadvantages of backward adaptation are removed. This method involves a core backward adaptive bit allocation routine which runs in both the encoder and the decoder. The core routine is relatively simple, but is based on a psychoacoustic model and, in general, is quite accurate. The input to the core routine is the spectral envelope, which is part of the encoded audio data delivered to the decoder. With AC-3, the spectral envelope has time and frequency resolution equal to that of the analysis/synthesis filter bank (5.3 msec and 94 Hz).

There are two aspects to the forward adaptation: psychoacoustic model parameter adjustment, and delta bit allocation. The core bit allocation routine employs a psychoacoustic model which has certain assumptions about the masking properties of signals. Certain parameters of the model are explicitly sent within the AC-3 bit stream. Thus, the details of the actual psychoacoustic model implemented in the decoder may be adjusted by the encoder. The encoder can perform a bit allocation based on any psychoacoustic model of any complexity, and compare the results to the bit allocation based on the core routine used by the decoder. If a better match can be made to an ideal allocation by altering some of the parameters used by the core routine, the encoder can do so. If there is a condition where it is not possible to approach the ideal allocation by means of parameter variation, the encoder can then explicitly send some allocation information. Since the allocation computed by the core routine should be very close to ideal, only small deltas to the default allocation are ever required. The AC-3 data syntax allows the encoder to explicitly send delta bit allocation information which will cause the bit allocation in small frequency regions to be increased or decreased. Since the final bit

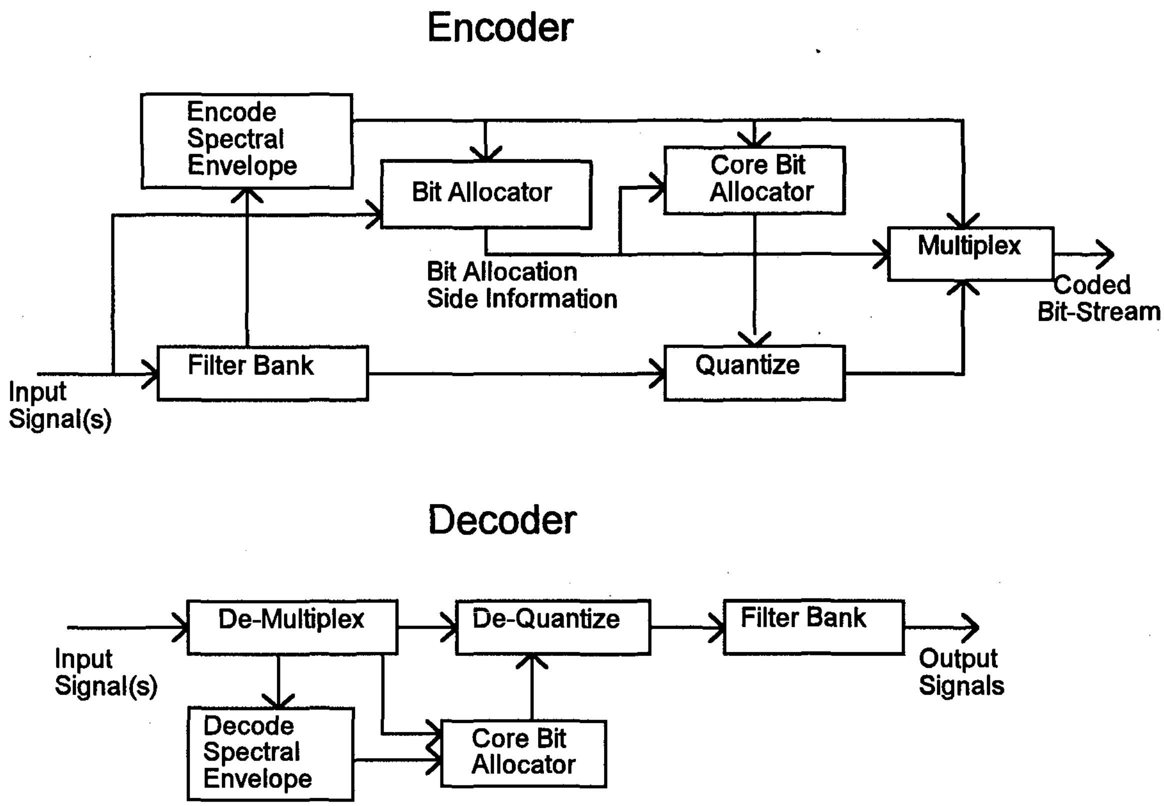


Figure 3: Hybrid Backward / Forward Adaptive Bit Allocation

allocation used by the encoder and decoder must be identical, the final allocation actually used by the encoder is the decoder core routine with whatever parameter variations and delta bit allocation are in effect.

4. Spectral Envelope Encoding

Each of the individual transform coefficients is coded into an exponent and a mantissa. The exponent allows for a wide dynamic range, while the mantissa is coded with a limited precision which results in quantizing noise. The synthesis filter bank in the decoder constrains the quantizing noise to be at nearly the same frequency as the quantized signal. The set of coded exponents forms a representation of the overall signal spectrum, and is referred to as the spectral envelope.

The filters of the transform filter bank are not perfect brick wall filters, but have gentle slopes. In general, the response falls off at approximately 12 dB per adjacent filter, or per 93.75 Hz. If the signal spectrum is analyzed with the filter bank, the level in any two adjacent filters seldom exceeds 12 dB. This behavior can be used to advantage in coding the spectral envelope. The AC-3 coder encodes the spectral envelope differentially in frequency. Since deltas of at most ±2 (a delta of 1 represents a 6 dB level change) are required, each exponent can be coded as one of 5 changes from the previous (lower in frequency) exponent: +2, +1, 0, -1, -2. The first exponent (D.C. term) is sent as an absolute, and the rest of the exponents are sent as differentials. Groups of three differentials are coded into a 7 bit word. Each exponent thus requires approximately 2.33 bits to code. This method of transmitting the exponents is referred to as D15. When employed, the D15 coding provides a very accurate spectral envelope.

Coding all exponents with 2.33 bits each for every individual audio block would be extravagant. However, fine frequency resolution is only required for relatively steady signals, and for these signals the spectral envelope will remain relatively constant over many blocks. The fine resolution D15 spectral envelope is only sent when the spectrum is relatively stable, and in that case the estimate may be sent only occasionally. In the typical case, the spectral envelope is sent once every 6 audio blocks (32 msec), in which case the data rate required is < 0.39 bits / exponent. Since each individual frequency point has an exponent, and there is one frequency sample for each time sample (the TDAC filter bank is critically sampled), the D15 high resolution spectral envelope requires < 0.39 bits per audio sample. An example of this is shown in figure 4. The individual spectra of 6 audio blocks is plotted (fine lines) as well as the coded spectral envelope (dark line).

When the spectrum of the signal is not stable, it is beneficial to send a spectral estimate more often. In order to keep the data overhead for the spectral estimate from becoming excessive, the spectral estimate may be coded with less frequency resolution. Two additional methods are available. The medium frequency resolution method is D25, where a delta is transmitted for every other frequency coefficient. This method requires half of the data rate of the D15 method (or 1.16 bits / exponent), and has a factor of two worse frequency resolution. The D25 method is typically used when the spectrum is relatively stable over 2-3 audio blocks and then undergoes a significant change. Use of the D25 method does not allow the spectral envelope to accurately follow all of the troughs in a very tonal spectrum. The coded spectral envelope is, of course, forced to cover all of the peaks.

The final method is D45, where a delta is transmitted for every 4 frequency coefficients. This method requires one fourth of the data rate of the D15 method (or 0.58 bits / exponent), and is typically used during transients for single audio blocks (5.3 msec). The transmitted spectral envelope thus has very fine frequency resolution for steady state (or slowly changing signals), and has fine time

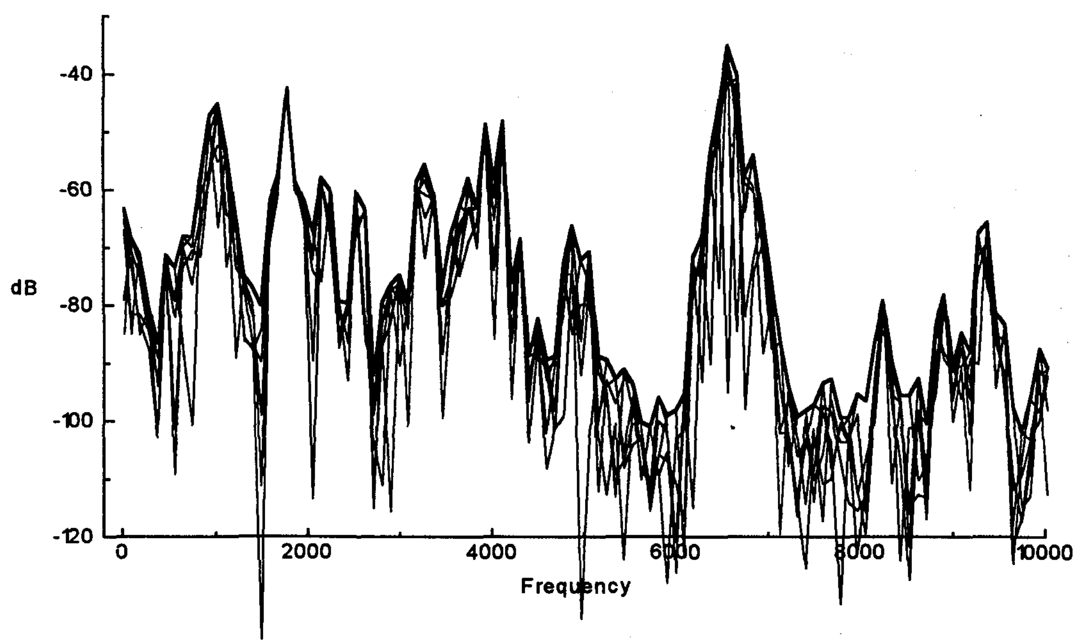


Figure 4: Spectral Envelope, D15 Method

resolution for transient signals. Typically, transient signals do not require fine frequency resolution since by their nature transients are wide band signals.

The AC-3 encoder is responsible for selecting which exponent coding method to use for any audio block. Each coded audio block contains a 2-bit field called exponent strategy. The four possible strategies are: D15, D25, D45, and REUSE. For most signal conditions, a D15 coded exponent set is sent during the first audio block in a 6 block frame, and the following 5 audio blocks reuse the same exponent set. During transient conditions, exponents are sent more often. The encoder routine responsible for choosing the optimum exponent strategy may be improved or updated at any time. Since the exponent strategy is explicitly coded into the bit stream, all decoders will track any change in the encoder.

5. Bit Allocation Details

The AC-3 core bit allocation routine begins with the decoded spectral envelope, or exponents, which are considered to be the

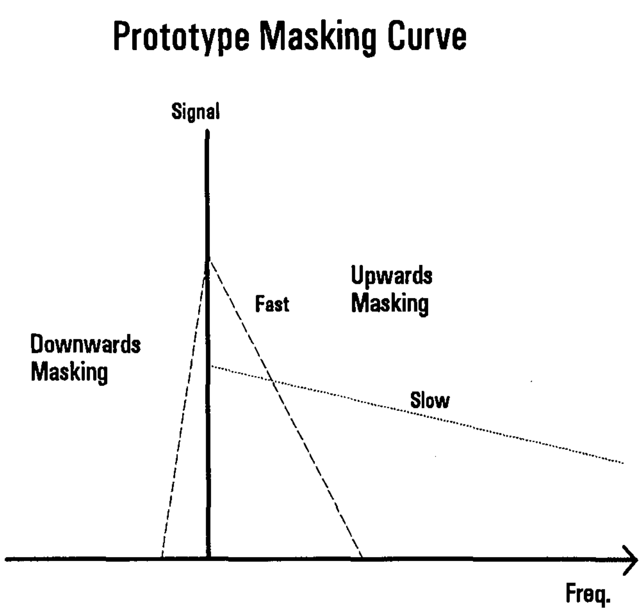


Figure 5: Prototype Masking Curve

power spectral density (psd) of the signal. There may be as many as 252 psd values (the number can vary depending on the number of exponents which have been sent, which in turn depends on the desired audio bandwidth and the audio sampling rate). The major portion of the bit allocation routine is the convolution of a spreading function matching the ear's masking curve, against the power spectral density. In order to reduce the computational load, the original psd array is converted to a smaller banded psd array. At low frequencies, the band size is 1, and at high frequencies the band size is 16. The bands increase in size proportional to the widening of the ear's critical bands. The initial (up to) 252 psd values are combined to form 64 banded psd values. A simplified technique has been developed to perform the step of convolving the spreading function against the banded psd. The spreading function of the ear is approximated by three curves: a very steeply decaying downwards masking curve; a fast decaying upwards masking curve; and a slowly decaying upwards masking curve which is offset downwards in level (fig. 5). The technique ignores the downwards masking curve for simplicity (at the expense of occasional inefficiency). A simplified convolution of the two different sloping upwards masking curves against the banded psd is performed. The calculation begins at the lowest frequency of the banded psd array and moves upwards in frequency. Two running computations are performed, one for each of the fast decaying and slowly decaying masking curves, which we refer to as the fast leak and the slow leak. The calculation is performed in the log domain, where a decay (leak) can be implemented as a simple decrement. As the calculation is moved up in frequency band by band, each new banded psd is examined. If the psd in the new band is significant with respect to the current leak values, the new psd is combined into the leak terms which are increased in value. If the new

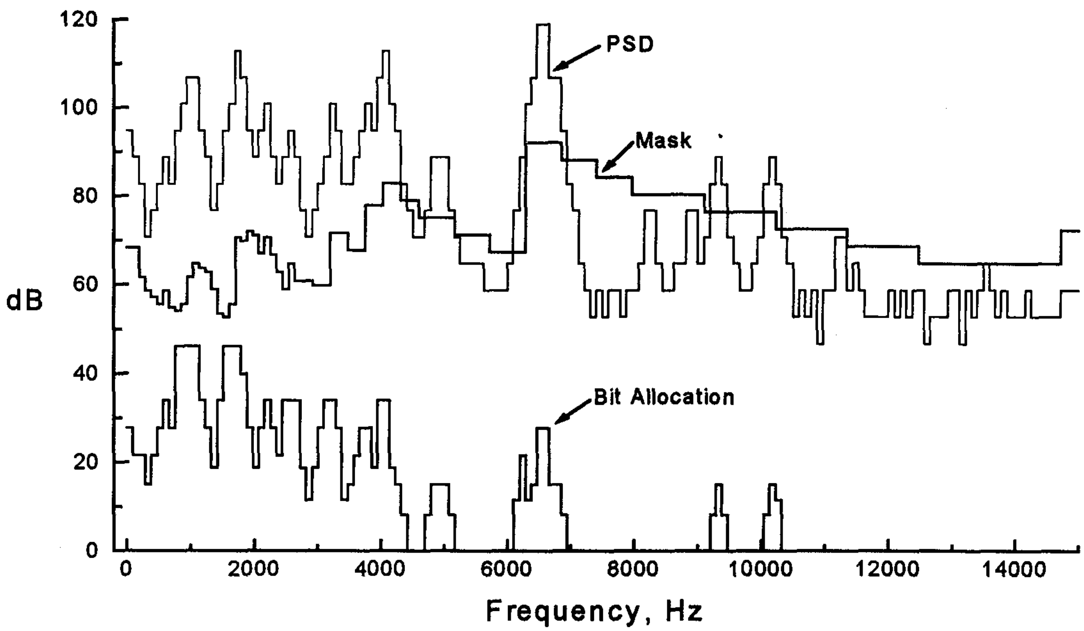


Figure 6: Bit Allocation Computation

psd is insignificant, the old leak terms are decremented, or allowed to leak down. The largest of each leak term at each frequency is held. The result of this calculation is an array indicating the predicted masking, band by band.

This curve is compared against a hearing threshold, and the larger of the two values is held. The final step is to subtract the predicted masking curve from the original (unbanded) psd array to determine the desired SNR for each individual transform coefficient. This is shown in figure 6. The array of SNR values is converted to an array of bit allocation pointers (baps). The bap array values point to the quantizers to be used for each transform coefficient mantissa. At this point the encoder does a bit count to determine if the bit allocation has used up the available number of bits. All available bits are from a common bit pool, available to all of the channels. If more bits are available, the individual mantissa SNR's may be increased, until all of the bits are used up. If too many bits have been allocated, the individual mantissa SNR's may be decreased, and/or coupling (see next section) may be invoked.

6. Coupling

Even though the coding techniques employed by AC-3 are very powerful, when the coder is operated at very low bit rates there are signal conditions under which the coder would run out of bits. When this occurs, a technique which we refer to as coupling is invoked. Coupling takes advantage of the way in which the ear determines directionality for very high frequency signals, in order to allow a reduction in the amount of data necessary to code a high quality multi-channel audio signal.

At high audio frequencies (above approximately 2 kHz) the ear is physically unable to detect individual cycles of an audio waveform, and instead responds to the envelope of the waveform. Directionality is determined by the inter-aural time delay of the signal envelope, and by the perceived frequency response which is affected by head shadowing and the ear's pinnae. Coupling takes advantage of the fact that the ear is not able to independently detect the direction of two high frequency signals which are very closely spaced in frequency.

When the AC-3 coder becomes starved for bits, channels may be selectively coupled at high frequencies. The frequency at which coupling begins is called the coupling frequency. Above the coupling frequency the coupled channels are combined into a coupling (or common) channel. Care is taken with the phase of the signals to be combined so as to avoid signal cancellations. The encoder measures the original signal power of the individual input channels in narrow frequency bands, as well as the power in the coupled channel in the same frequency bands. The encoder generates coupling coordinates for each individual channel which indicate the ratio of the original signal power within a band to the coupling channel power in the band.

The coupling channel is encoded in the same manner as the individual channels; there is a spectral envelope of coded exponents and a set of quantized mantissas. The channels which are included in coupling are independently coded up to the coupling frequency. Above the coupling frequency, only the coupling coordinates are sent for the coupled channels. The decoder multiplies the individual channel coupling coordinates by the coupling channel coefficients to regenerate the high frequency coefficients of the coupled channels.

The coupling process is audibly successful because the reproduced sound field is a close power match to the original. Within each narrow frequency band, the reproduced signal envelope out of each loudspeaker matches the original signal envelope in that loudspeaker. Additionally, the level within each narrow frequency band of each source channel is reproduced at the same overall power in the sound field, even though the power may be shared amongst several of the loudspeaker channels. The coupling coordinates are encoded with an accuracy of < 0.25 dB. This fine resolution not only allows a good power match to be obtained but, more importantly, allows small changes to be made smoothly. (This is in contrast to techniques employed by other coding technologies to combine high frequencies together, where the minimum gain step is 2 dB.)

The AC-3 encoder is responsible for determining the coupling strategy. The encoder controls which of the audio channels are to be included in coupling, and which remain completely independent. The encoder controls at what frequency coupling begins, the coupling band structure (the bandwidths of the coupled bands), and when new coupling coordinates are sent. The coupling strategy routine may be altered or improved at any time and, since the coupling strategy information is explicit in the encoded bit stream, all decoders will follow the changes.

7. Bit Stream Syntax

The fundamental time unit for AC-3 is related to the transform block size. While each transformed block is 512 samples long, the 100% overlap/add of the TDAC transform means blocks are transformed every 256 samples, or every 5.33 msec for a 48 kHz sampling rate. The transform block rate is 187.5 Hz. The AC-3 syntax groups 6 transform blocks into an AC-3 frame. The frame rate is 31.25 Hz, and each frame lasts 32 msec. Each AC-3 frame (fig. 7) begins with a 16 bit sync word. Following the sync word are 8 bits of information which indicate sampling rate and frame size. We refer to the first 3 bytes of the AC-3 frame as sync info because the information in these bytes is used to acquire and maintain synchronization with the AC-3 frames.

Following sync info is a set of data we call bit stream info, or BSI. The BSI data contains information about the number of channels which are coded, dialogue level, language code, information about associated services, etc. All of the BSI data is essentially static,

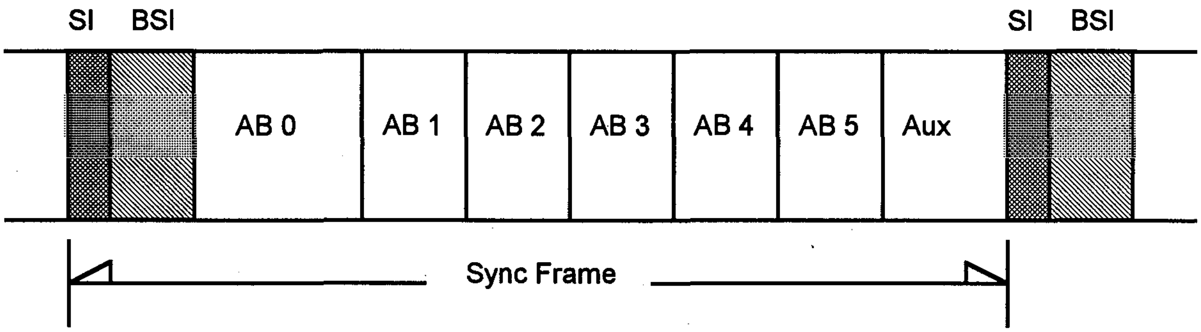


Figure 7: AC-3 Framing

and is descriptive about the data in the audio blocks which follow.

Following BSI are 6 coded audio blocks. The first block always contains a complete refresh of exponents, coupling coordinates, and all other information which is conditionally transmitted. The following 5 blocks may or may not contain information beyond quantized mantissas. Unused data at the end of the AC-3 frame may be considered aux data.

8. Features

A number of features have been designed into AC-3 in order to make it widely applicable. The goal is to have a coded audio signal which is usable by as wide an audience as possible. The potential audience may range from patrons of a commercial cinema or home theatre enthusiasts who wish to enjoy the full sound experience, to the occupant of a quiet hotel room listening to a mono TV set at a low volume who nevertheless wishes to hear all of the program content. Two of the major concerns of the AC-3 coder design were mixdown, and loudness control.

Very early in the development of AC-3, it was decided to encode the channels discretely, without the use of intermediate matrixes (as have been employed in other multi-channel coders which offer backwards compatibility with some existing two-channel decoders). It was felt that backwards compatibility with existing 2-channel digital decoders was of little value, since very few decoders were in the field to be backwards compatible with. The need to be backwards compatible with a variety of reproduction systems, such as mono, stereo, and matrix surround was seriously considered. Avoiding unnecessary backward compatibility matrixing allowed AC-3 development to concentrate on fundamental coding techniques, and not merely on techniques which only mitigate the negative effects of compatibility matrixing.

Since AC-3 delivers all channels to the decoder coded discretely, the decoder has full flexibility to mixdown the channels as appropriate to the listening situation. Differing listening situations require differing mixdown coefficients. For example, a stereo listener may wish the surround signals to be mixed out-of-phase so that they don't localize well in two loudspeaker reproduction. A second listener may wish the same mixdown because this listener has a current home theatre matrix decoder which requires a 2-channel surround matrix encoded signal (which has the surround signal out-of-phase) as an input in order to reproduce surround. A third listener may wish a mono mix-down. This third listener could not simply combine the 2-channel mix enjoyed by the first two listeners, because then the surround signal would be lost. (The currently used Dolby Surround matrix system does suffer from the problem that the surround signal is lost in monophonic reproduction. This fact is known by program mixers who consciously avoid the placement of essential program content solely in the surround channel. One of the goals for

AC-3 was to eliminate this and other limitations of the matrix surround system. It is essential that a new multi-channel coder allow all listeners to hear all of the sound from all of the channels, so there are no constraints on program production.) The mono listener thus requires a different downmix. Only by allowing the listener to select the desired type of downmix can all audiences be served. In most cases the listener will not be required to make any conscious decision about the desired downmix as products will be designed such that the AC-3 decoder will automatically make the choice depending on information such as the number of loudspeakers available.

In the United States (and perhaps elsewhere) there is a problem with the loudness of different broadcast programs. Many broadcasters highly compress the dynamic range of the audio, and fully modulate the audio channel much of the time. In this case, there is little headroom. Sometimes the entertainment program will have a more natural dynamic range with some headroom, but commercial messages (attempting to sound loud) may not. The result is that there are significant level differences between program segments on a particular broadcast channel, as well as between broadcast channels. Two of the design goals for AC-3 were: to eliminate the apparent level differences between broadcast channels; and allow broadcasters to compress the dynamic range of the programming for most listeners, while allowing other listeners (who so choose) to enjoy the full original dynamic range of the program.

Loudness uniformity is achieved by determining the subjective level of normal spoken dialogue, and explicitly coding this level into the data stream as a dialogue level control word. The AC-3 decoder may then interact with the system playback level. Differing programs may have differing dialogue levels which simply mean that they have differing amounts of headroom available for dramatic effect. When the listener adjusts the volume control, the level of reproduced normal dialogue (i.e. not shouts or whispers) will be set to the desired subjective sound pressure level. When a new program segment begins, with dialogue encoded at a different level (i.e. with a different amount of headroom), the reproduction system may use the coded dialogue level control word to make a corresponding adjustment to the system playback volume. The result is that for all programs and all channels, the reproduced level of dialogue will be uniform.

The AC-3 coder contains an integral dynamic range control system. During encoding, or at any point thereafter, dynamic range control words may be placed into the AC-3 bit stream. These control words are used by the decoder to alter the level of the decoded audio on a block basis (every 5.3 msec). The control word may indicate that the decoder gain be raised or lowered. The control words are generated by a level compression algorithm which may be resident in the AC-3 encoder, or in a subsequent bit-stream processor. The control words have a resolution of < 0.25 dB per block. The block-to-block gain variations are further smoothed by the gentle overlap add process. Gradual gain changes are free from gain stepping artifacts. For program audio levels above dialogue level, the dynamic range control words will indicate level reduction. For audio levels below dialogue level, the control words will indicate a level increase. The default for the decoder is to use the control words which will result in the reproduction of the audio program with a compressed dynamic range. The exact nature of the compression is determined by the algorithm which generates the control words, but in general the compression is such that headroom is reduced (loud sounds are brought down towards dialogue level) and quiet sounds are made more audible (brought up towards dialogue level). It is a decode option to reduce the effect of the control words of either polarity. If the control words which indicate that gain should be increased are not used, then low level sounds will retain their proper dynamics and only loud sounds will be compressed downwards. If only the control words which indicate gain reduction are ignored, the low level sounds will be brought up in level, but loud sounds will reproduce naturally. If all control words are ignored, the original signal dynamic range will be reproduced. It is also possible to partially use the control words for either polarity. Thus the listener may instruct the decoder to partially or fully remove the dynamic range compression which has been applied to either the loud (above dialogue level) or soft (below dialogue level) sounds. (Actually the audio is coded without any level alteration, and level alterations occur at the decoder. However, this is irrelevant to the listener, since the default for the listener is to reproduce the program with the compression characteristic which has been intended by the program originator. The listener must take some action to remove the compression.)

The AC-3 syntax has been defined to support the coding of one main audio service with from 1 to 5.1 channels. Additionally, associated services may be encoded into an AC-3 bit stream. Associated service types include: visually impaired (a verbal description of the visual scene), hearing impaired (dialogue with enhanced intelligibility), commentary, dialogue, and second stereo program. All services may be tagged with a code to indicate language. Single channel services may have a bit-rate as low as 32 kb/s. The overall data stream is defined for bit rates ranging from 32 kb/s up to 640 kb/s. The higher rates are intended to allow the basic audio quality to exceed audiophile expectations, as well as allow the incorporation of associated services without severely restricting the bit-rate (and thus quality) of the main audio service.

9. Conclusion

AC-3, with its high resolution spectral envelope coding and hybrid backward/forward adaptive bit allocation offers very high coding gain at modest complexity. Bit starvation is avoided during extreme signal demands by invoking the technique of coupling, which avoids any need to restrict the coded audio bandwidth. Fully discrete signal transmission avoids the need to increase coder complexity and compromise coded audio quality in an attempt to avoid compatibility matrixing artifacts. Decoder downmixing allows every listener to obtain the optimum downmix for his listening situation. The quality level of AC-3 is not limited to that obtainable with currently available encoders; future encoders are expected to achieve improved results by means of additional complexity and sophistication. All decoders in the field will automatically benefit from future encoding improvements.

AC-3 has been subjectively tested twice (Grand Alliance testing in July 1993, MPEG-2 testing in Nov. 1993) simultaneously with MPEG-2 multi-channel audio technology. In both sets of tests, AC-3 clearly outperformed the MPEG-2 technology.

The first integrated circuit designed for AC-3 (the Zoran ZR38000) became available in 1993, and several more are expected to become available in 1994. AC-3 has been selected for use in the United States HDTV system for reasons of: high audio quality; advanced state of development; and full provision of all necessary features. AC-3 is being designed into consumer electronics equipment for cable television, direct broadcast satellite, and pre-recorded media, and AC-3 encoded signals are currently being broadcast via satellite in the U.S.

An Analysis of Interoperability and Cost Factors for Regional Digital Backbone Networks

John Holobinko
Vice President, Marketing and Strategic Planning
American Lightwave Systems, Inc.

ABSTRACT

In major metropolitan areas the need exists to consolidate headends and to provide a wide variety of video, data and telephony services across the CATV network. Digital Backbone Networks which connect a number of primary transport hubs together are the primary means of establishing these metropolitan networks. Other means are broadband linear (AM) supertrunks.

Questions arise as to the degree of interoperability of these networks with the standard telephony network; the need to transport BTSC compatible video, baseband and RF scrambled video, satellite delivered MPEG-2; management of local commercial insertion, preservation of revenue generating data within the video vertical blanking interval, the need to carry telephony and data traffic, and the amount of ancillary processing equipment required at each primary hub.

This paper analyzes the need for SONET compatibility in the metropolitan network, and provides an economic and technical analysis of signal quality at the subscriber, processing equipment at the various primary headends and effects on the design and cost of the AM hybrid fiber/coax networks which are fed by the digital backbone network, based upon the technologies used.

A network design is presented which allows cost effective implementation of digital video and telephony services today, plus a graceful migration path which provides a means of network expansion for accommodating higher levels of telephony and data traffic for the future.

METROPOLITAN AREA NETWORK ATTRIBUTES

A typical metropolitan area network consists of one or two master headends (also called television operating centers, TVOC's) which are connected with primary hubs (also called remote headends). The area covered by the network may be that of a large city, or a wide region. The largest known system of this type currently is installed by Continental Cable, and covers their properties in the three states consisting of Massachusetts, New Hampshire and Maine. Whether metropolitan or regional in its area of coverage, the term metropolitan area network is used to reference these networks in this paper.

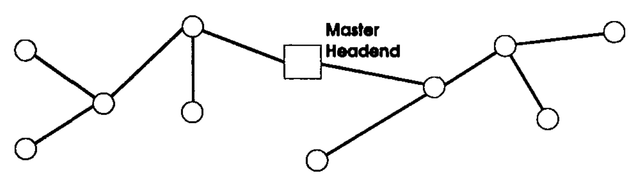


Figure 1 Metropolitan Area Network

Signals between the hubs and subscribers are not usually considered part of the metropolitan area network, but rather the distribution network. In some large CATV networks, secondary hubs are also employed. However, secondary hubs are usually considered as part of the distribution network versus the metropolitan area backbone network.

Metropolitan area networks can be designed in bus, ring and star topologies. However, some topologies are less conducive for providing high reliability and delivery of advanced services than others. This will be addressed later in this paper.

METROPOLITAN AREA NETWORK VS. PUBLIC SWITCHED NETWORK

The characteristics which define the metropolitan area network differ significantly from the characteristics of the public switched network. In the metropolitan network:

1. There is only one carrier (or a very small number of carriers) who owns and manages the network;

2. Communications in the network is highly asymmetrical;

3. For most traffic within the network, there are highly defined points of origin. Even in the case of video on demand (VOD) services, the number of origination points is small compared to the number of subscribers. Most communications enters the network via gateways (e.g. satellite receiver, digital server, etc.). Symmetrical traffic tends to leave the network via gateways;

4. Traffic types within the network are highly predictable in terms of types and numbers of each type of channel. For example, if a channel is being used to transport a service such as CNN at 5:35 PM, it is highly unlikely that this same channel will be transporting a combination of voice and data services at 5:36 PM.

SONET IN THE METROPOLITAN AREA NETWORK

To understand if SONET provides a value within the metropolitan network it is helpful to review the intended benefits of SONET, and very importantly, to whom those direct benefits are intended. SONET was designed for the ubiquitous network in which the number of channels, their size and content are not predictable. SONET is very valuable in this environment. In theory, it allows information to be transported from one carrier to another across multiple carrier boundaries without regard to content, and via universal interfaces. Note that SONET provides no direct benefits to end users. The direct benefits of SONET are to communications carriers operating in a multi-carrier environment. As exemplified above, this is not the environment of the metropolitan network.

SONET does not come without cost. For a metropolitan area network with a typical mix of cable entertainment channels, near video on demand, video on demand channels and telephony, a SONET digital backbone will cost two to three times that of an uncompressed high speed digital fiber backbone which is not fully SONET compatible. For the average metropolitan network, this amounts to millions of dollars.

One can argue that the benefits to SONET include indirect savings. But these suggested savings if any, are not comparable to the cost penalties imposed today by SONET. The video distribution network is not an unpredictable, multiple carrier environment, even when there are multiple service offerings. In the competitive world of video distribution, each dollar of capital equipment cost trickles down into the cost of providing the subscriber with specific services. This is the same reason that hybrid fiber/coax systems as opposed to traditional telephony architectures, are being designed and implemented for so many enhanced services networks.

The cost penalties of SONET for the metropolitan area network arise in four basic areas:

1. Additional cost of ancillary video signal processing equipment required at primary hubs due to the inability of compressed video codecs to support the video signal format requirements of the network;

2. Opportunity cost in terms of lost revenue, due to the inability to carry certain types of information which are additional sources of revenue to the operator;

3. Additional cost to the broadband linear

hybrid fiber/coax network for signal transport from the digital hubs to the serving areas, in order to achieve the targeted performance level at the subscriber drop;

4. Cost per Gb/s for the transport system itself including space and power requirements.

1. <u>Additional Ancillary Equipment</u>

The need for additional ancillary equipment arises out of the inability of the SONET based transport system to accommodate video signals in a format which is necessary or optimal for signal quality. One such straightforward example is an RF scrambled video channel. If a codec which employs compression is used to transport video, it cannot accommodate such a scrambled signal. Therefore, each signal to be scrambled must be sent from the master headend to the hubs in baseband format. At each hub site a separate signal scrambler is required for each channel to be scrambled. The greater the number of hubs in the network, the larger the cost penalty is for duplicate scramblers at each location. Additionally, since many channels like HBO employ three audio channels (left, right, and second language), the cost of the codec itself may be more expensive. In fact some codecs only support two channels of audio per video. Therefore, an auxiliary audio transport system may be required to accommodate the additional audio requirements of these channels.

A more subtle example but with wider ranging effects is the handling of BTSC subcarrier audio. North American video transmission standards allow for audio to be transmitted as a subcarrier to video at 4.5 MHz. Therefore, if the audio is BTSC encoded, stereo audio and SAP audio can be accommodated within the 4.5 MHz subcarrier. This can be encoded along with video as a composite signal by a majority of compression based codecs. However, this creates two significant disadvantages with regards to signal quality:

The first disadvantage is within the codec itself, and is very straightforward. Any digital encoder has a finite dynamic range. For optimum video signal to noise (SNR) performance, the video signal amplitude must be adjusted such that at maximum amplitude, it occupies the entire encoding range. If the audio subcarrier is included with the video, part of the encoding resolution must be used for the audio subcarrier.

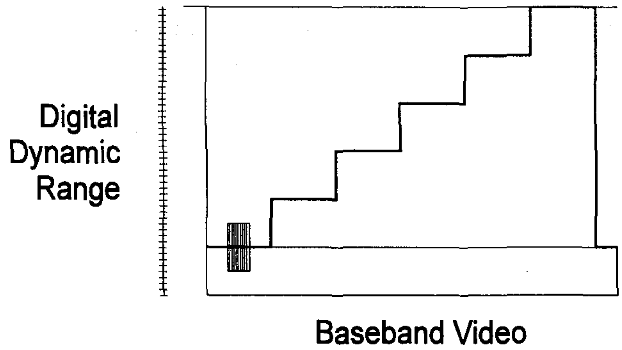


Figure 2 Baseband Video Encoding

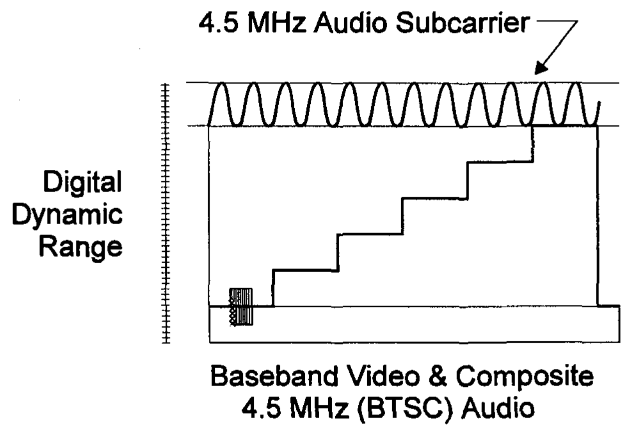


Figure 3 Composite Video Encoding
Note lower dynamic range available for video signal

Since part of the codec's dynamic range has been taken by the audio subcarrier, the absolute video SNR is less for a composite encoded signal than it will be if the audio subcarrier is encoded separately of video. A 2 dB SNR penalty is possible. This is not inconsequential. The video performance at the subscriber drop is proportional to the SNR contribution of the metropolitan network as well as the hybrid fiber/coax distribution system. This is explained in the next section of this paper. An additional problem is that encoding the two signals together produces a classic 920 KHz beat due to interaction between the 3.58 MHz color subcarrier and the 4.5 MHz audio subcarrier. This phenomenon is well known to video engineers and therefore needs no further explanation.

The second disadvantage of composite video encoding has nothing to do with the digital transport system itself, but rather with the VSB/AM modulator which follows. If a composite signal is fed to a VSB/AM modulator, it must first separate the signal into the discrete baseband video and the separate audio subcarrier so that AM modulation and sideband filtering can be performed on the video only. The highest video frequency is 4.2 MHz. If the filter required to perform this separation has too sharp a cutoff, then its group delay performance will be poor and characteristics such as video differential gain and chrominance to luminance will suffer. If the filter is made softer as is the case in normal practice, then the upper video frequencies will be cut off.

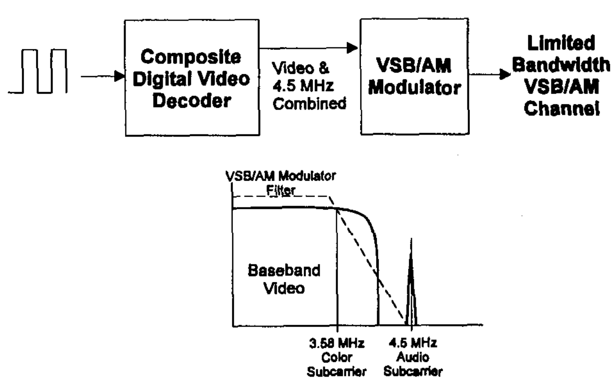


Figure 4 4.5 MHz Cutoff Filter Problems

Since video resolution is directly related to video bandwidth, the effect of this filter is to reduce video resolution and picture clarity. In today's systems, the subscribers who typically pay the largest monthly cable bills also have large screen televisions. Therefore, they will be the subscribers most deleteriously affected by the loss of resolution that results. If the digital transport system cannot carry 4.5 MHz audio as a separate discrete digitized carrier, then audio must be sent as baseband, and BTSC encoders replicated at each and every hub site. This results in significant added cost and operational penalties to the network operator.

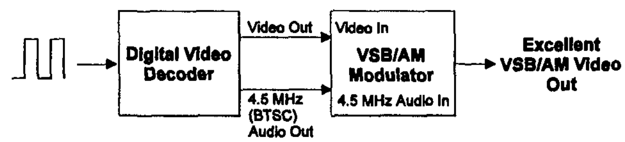


Figure 5 Proper Handling of 4.5 MHz Audio

2. <u>Lost Revenue Opportunities</u>

In the compression process, certain information is stripped out of the signal which is unnecessary to the video content. This information can be regenerated at the decode side of the system. One such area in the video waveform is the video blanking interval, referred to as the VBI or VITS area. However, in many CATV systems, the VITS area is used to carry revenue generating information services which are not related to the video itself. Examples of VITS data include: FNN stock updates, XPRESS computer service, etc. Since the location of these services may vary from video channel to channel, it is not straightforward to simply preserve a few selected VBI lines.

In contrast to the compression based codecs used in SONET systems, uncompressed video encoders do not modify or delete VITS information. Therefore, these sources of additional revenue are preserved.

<u>Advanced Digital Video Services Distribution</u>

The first uses of digital video channels which are n-QAM and n-VSB modulated will be for near video on demand applications. Initially the cost of the modulators to produce the n-M and n-VSB carriers will be quite high. Operationally, it will be desirable to digitally modulate once at the master headend and distribute channels to the hubs. An uncompressed digital system can do this. For example, the DV6000 by American Lightwave Systems, Inc. has already undergone tests with the Zenith 16 VSB encoding stream to demonstrate compatibility. Such a capability has not been developed for SONET based systems to date.

3. SNR Performance versus Distribution Network Cost

The true signal quality delivered to the subscriber is a combination of the addition of all noise contributions in the network from the master headend to the subscriber. Carrier to noise only characterizes the quality of the video carrier from the point of modulation, which in many cases will be the primary hub. Ultimately, it is SNR that is the determinant of picture quality on the subscriber's TV set. If the CNR of a distribution network is converted to an equivalent SNR contribution, then a total logarithmic addition of SNR contributions is possible. For a VSB/AM path, its SNR contribution is slightly lower than CNR. For the following illustration, VSB/AM SNR is approximated as equal to CNR for simplification. The conclusions from the following analysis are not affected by this simplification.

Distribution systems being designed for the future have total CNR contributions in the range of 47 dB to 49 dB. It is typical for CATV distribution system designers to assume that the metropolitan network feeding the primary hubs is transparent, and therefore is not accounted for in computing end of line SNR performance. But what performance level constitutes transparency?

To establish this requires examination of the addition of the noise contribution from the metropolitan backbone fiber system from the master headend to the hub and the noise contribution of the distribution system from the hub to the subscriber. If we assume the distribution system to be state of the art 49 dB CNR, then its SNR is approximately 49 dB.

A graph can be generated which shows the effect of metropolitan network SNR performance on the resulting performance at the subscriber drop. In the figure which follows metropolitan backbone SNR is shown on the X axis, and the resulting SNR performance at the subscriber drop which is the summation of the two portion of the networks together, assuming a constant 49 dB for the distribution network.

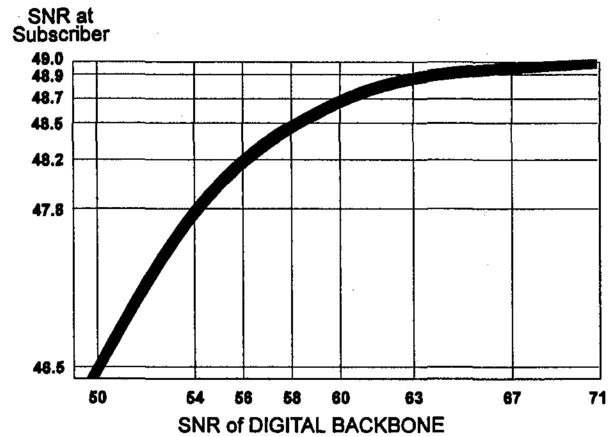


Figure 6 Subscriber SNR as a Function of Metropolitan System SNR (Distribution CNR held constant at 49 dB)

The conclusion is quite evident. At SNR levels above 59 dB SNR, the metropolitan network has a negligible effect (less than a .3 dB) on subscriber signal quality. However, when the contribution of the metropolitan network is 56 dB, almost a full 1 dB is lost at the subscriber. At 54 dB this degradation jumps to 1.2 dB, i.e. performance at the subscriber degrades to 47.8 dB. If the target at the subscriber is 49 dB, how can this be achieved with a metropolitan network whose performance is only 54 - 56 dB? (This is the CNR performance level of an unrepeatered AM supertrunk used for a star based metropolitan network versus a high performance uncompressed digital network with multiple repeaters.)

The only way this is possible is for the distribution network to be upgraded above 49 dB in CNR performance. In practice, this means either a shorter cascade or fewer optical nodes served by a single optical transmitter. Since the former is much more expensive at a performance level of 49 dB, I shall concentrate on the latter. Let's assume that the 49 dB CNR was achieved by taking a 50 dB CNR optical link in combination with a 56 dB CNR amplifier cascade. (This design assumes an average of 4 nodes served from one transmitter in a metropolitan area.) From the illustration above, the distribution system performance must be raised from 49 dB to 50.5 dB. If this is done via improvement of the AM fiber link versus shortening of the amplifier cascade, then the AM fiber link performance must go from 50 dB to 54 dB. Therefore, instead of serving four nodes per transmitter, only two nodes can be served per transmitter at this higher required performance level. Double the number of AM transmitters are required in the distribution network to achieve the original target of 49 dB at the subscriber!

The closer that the SNR of the metropolitan network is to the SNR of the distribution network, the more deleterious the effect on end subscriber picture quality. Or alternatively, the more expensive the subscriber distribution network has to be in order to achieve a desired performance level at the subscriber. As the above example shows, the expense of this can be very significant.

Given that the true SNR of compressed digital codecs used in CATV SONET links may have performance in the very low 50's dB as measured by digital SNR techniques such as ANSI proposed standard T1Q1.5/91-205R2, the cost penalty to the distribution network which follows a SONET system is very high when compared to an uncompressed digital system with SNR of 59 dB which places no cost penalty on the distribution network whatsoever.

4. System Space, Power

In addition to the higher cost of the SONET equipment itself, the cost of a SONET system for metropolitan network will be higher in terms of both space and power than a modular uncompressed digital system. A high performance uncompressed digital transmission system can fit 2.4 Gb/sec. of transmission equipment including individual channel drop/add/pass functionality, redundant optics and power supplies, plus 16 channels of encoding equipment, all into 19 inches of vertical rack space, consuming 170 watts of power.

In comparison, a SONET OC-48 system (2.48 Gb/s) will take a complete 19" rack without adding in the video codecs. Power consumption is more than double. When ancillary equipment requirements are added, the space penalty may be as high as three to one.

Cost Model

A cost model can be created to show a typical network and the cost differences between an uncompressed system and SONET based system. The ALS DV6000 was used as the uncompressed system. The model used for this paper is conservative and consists of the following:

One Master Headend
Serving Six Primary Hubs

60 Channels Basic CATV Service:

- 25 CH w/Stereo Audio (need BTSC)
- 5 CH Monaural w/2nd Language (need BTSC)
- 25 CH Monaural only
- 5 CH w/Local Commercial Insertion at Hubs (audio sent baseband)

10 Premium Channels - Scrambled, all Stereo, half with SAP
5 PPV Channel - Scrambled, all Stereo and SAP
6 DS3 Telephony Channels - Alternate Access Business
6 DS1 Channels - Service Center Consolidation

The costs of ancillary equipment were chosen to be in the middle range. The results of this analysis showed the following:

TOTAL NETWORK COST COMPARISON

	SONET SYSTEM A	DV6000 SYSTEM
MASTER HEADEND	$350,000	$300,000
6 HUBS	2,500,000	1,400,000
TOTALS:	$2,850,000	$1,700,000
ADDITIONAL AM TX'S TO ACHIEVE 48.7 dB SNR	800,000	0
TRUE COST FOR EQUIVALENT PERFORMANCE AT SUBSCRIBER	$3,646,963	$1,797,875
LOST REVENUES FROM INABILITY TO DELIVER SPECIAL SERVICES	???	0

Figure 7 Cost Summary

In summary, the cost of system and ancillary hardware for the SONET system was approximately 1.7 times the cost of the uncompressed digital solution. When the penalty for restoring 49 dB performance at the subscriber is added, the difference in costs between the systems increases to 2.15 times.

REDUNDANCY ISSUES

Today, both SONET and at least one high end uncompressed digital system offer fiber hot standby switching and other forms of redundancy. However, redundancy in video systems is based on the ability to protect and restore an individual channel. Therefore, the metropolitan system must provide protection at the hub sites so that if an individual decoder fails *or if the VSB/AM modulator which follows the decoder fails*, that active spares can be remotely switched to restore the received channel to service on the proper channel number without any manual intervention or replacement of any equipment. This is possible today in uncompressed digital systems which are modular in design and provide single channel video encoder and decoder modules, channel drop/add/pass capability exists, and interfaces have been developed so that VSB//AM modulators may also be controlled via custom software.

In contrast, SONET based systems have primarily used packaging in which multiple decoders are in the same physical box. (This is also true of some earlier vintage uncompressed digital systems.) Therefore, if one module fails, more than one channel may be lost. With this form of equipment packaging it is not possible to use 1 X N channel protection to restore service. A hard outage normally occurs, and therefore a truck roll is necessary in systems which are configured with multiple channels per module.

TELEPHONY AND DATA TRANSPORT

The metropolitan area network must provide the ability to transport telephony and data channels in addition to video channels. Today virtually 100% of these channels will range from DS0 (64 kbs) to DS3 (45 Mbs). It is obvious that higher level rates are easily accommodated by the SONET transport system. New generation uncompressed digital systems also provide DS0 - DS3 capability, and provide added flexibility for interconnecting PBX's with individual DS1 channels which avoid the use of costly M13 multiplexors. This is highly desirable for consolidating customer service centers, for example. The concept of the uncompressed digital system is to provide an economic means of providing the ability to add telephony and data capability to the network gradually, without the penalty of a second network. If telephony traffic grows to a high level (nine or more DS3 channels between sites), then at that time perhaps it makes sense to add a SONET OC-12 system to handle this specialized traffic. However, alternatively, it appears not to make sense to create a very costly SONET OC-48 network for video in order to accommodate an initial telephony and data traffic which might encompass only a limited number of DS3's. Since state of the art uncompressed digital systems provide interfaces to standard telco network monitoring systems, a single operations support system (OSS) can manage the entire network.

VIDEO TRANSPORT IN THE PUBLIC NETWORK

If the network is not SONET based, how can information make its way to and from the video distribution network? Clearly, the ability to extract information and to insert information onto the public SONET based network is imperative. Therefore, advanced video distribution networks based upon high speed uncompressed digital transmission are designed with a SONET gateway at the master headend, typically with an STS-3c interface. Channels originating in the video distribution network therefore can access the SONET network. Likewise, channels which originate in other networks can enter the video distribution network. In this way, the cost of SONET is only imposed upon those services which derive direct benefit from SONET.

NETWORK CONTROL AND STATUS MONITORING

The need exists to monitor and control all traffic into and out of the metropolitan network. For telephony and data traffic, strict monitoring of incoming and outgoing bit error rates is imperative in order to characterize system performance, provide back up switching and conduct fault isolation. Support of standards such as TCP/IP and TL1 are critical to provide universal interfaces to an OSS. Both SONET and state of the art uncompressed digital systems provide this capability. However, in the metropolitan network, information about the content of channels is also very critical. When channels must be dropped, added or inserted, it is key to be able to determine what video service is riding on each channel, and to check automatically and prevent human errors such as inadvertent replacement of one video service with another at a channel insertion site. State of the art uncompressed digital systems provide this capability.

METROPOLITAN NETWORK ARCHITECTURES

There are three major architectures which are in use for metropolitan area networks today. These consist of the Star, Bus and Ring topologies. Of these, the star is the most limiting, when considerations of high reliability and fiber miles are concerned over any metropolitan or regional area which covers significant distances.

Star Network

In the star network, virtually all signals emanate from the master headend which is the center of the star. If the distance from the center of the star to each hub is within the reach of an AM supertrunk system, then the cost of signal conversion from digital to VSB/ AM signals can be avoided. This results in considerable cost savings for the forward path.

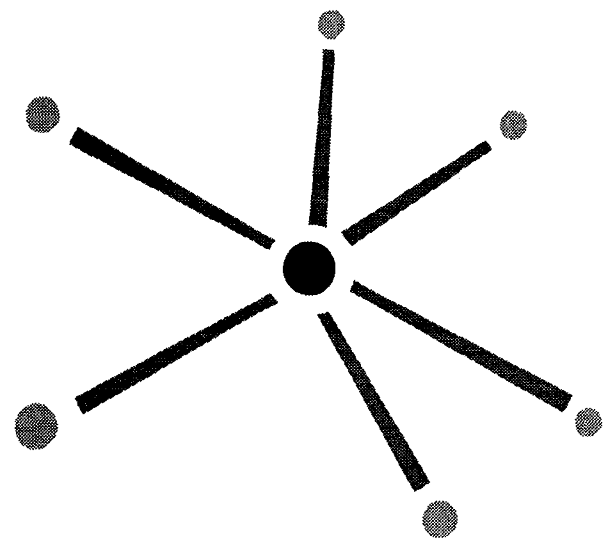


Figure 8 Star Architecture

The disadvantages of the star network are based on reliability and flexibility. Reliability of the network is proportional to the total number of fiber cable kilometers and the ability to provide automatic redundancy. As the average distance from the center of the star to the hubs gets larger and the number of hubs served by the star increases, the amount of fiber necessary to create the network grows exponentially. If a cable cut occurs, there is no simple way to provide redundancy from the master headend, and if redundancy is provided, there is a high likelihood for the necessity of repeaters with associated degradation of signal quality. Therefore, a hard loss of service will occur. Based on standard probabilities, the greater the amount of fiber, the greater the likelihood for a fiber cable cut.

Therefore, implementation of a star network calls for serious consideration of totally underground cable construction, since two-way redundant paths will be very expensive and very possibly require the use of repeaters. If the network is established using AM supertrunking instead of a digital transmission system, the use of a repeater will deleteriously affect subscriber performance, or cause the distribution system to go up significantly in cost as illustrated previously.

Bus Network

The bus network is highly flexible, especially for long distance networks in which signals can enter and exit the network at multiple hubs or gateways. Traffic can be both symmetrical and asymmetrical. Redundancy for most services can be established by providing two master headend locations, each which feeds all hub locations. In the event of a fiber cut, switch over to the alternate path can be made locally at each hub in a few milliseconds. However, some types of telephone and data communications may be more difficult to back up. This architecture is only practical with a digital system.

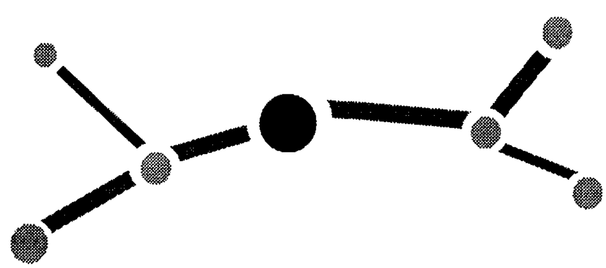


Figure 9 Bus Architecture

Ring Network

Ring networks can be designed with a number of variations. The ring may be open, closed, or closed with redundant counter rotating signals. If AM supertrunking is used, only two remote sites are possible with redundancy. If uncompressed or compressed digital systems are used, signals may be sent to many sites, dropped and added, and fully backed up with no loss of signal quality (except of course for degradations in the compressed digital system associated with the codecs themselves, as noted previously).

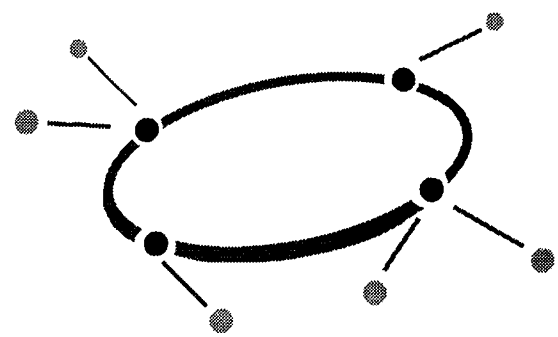


Figure 10 Ring Architecture

Star vs. Bus vs. Ring

Larger networks will preclude the use of star technology. Bus and ring topologies are both attractive for larger networks. The choice will be based both on network geography and mix of services provided.

CONCLUSIONS

It is imperative that the metropolitan network must be highly efficient and economical. This means that there is no room for added cost without direct quantifiable economic benefit. Therefore, it is not viable to extend SONET into the metropolitan network today and for the foreseeable future. If SONET did not impose such a cost and performance penalty on the network it could be argued that there is some benefit in providing compatibility deeper into the network. Someday, the cost differential may decrease between uncompressed digital backbones and SONET digital backbones. If this happens then there is a reason to implement SONET, because it will no longer impact the cost of providing subscriber services. Gateways to SONET are viable currently. However, elimination of the great cost differential between uncompressed digital transport and SONET for the video distribution system is not in sight today.

AM supertrunks are not viable for metropolitan networks due to the large distances, number of locations, channel drop/insert flexibility, and redundancy issues associated with these networks. As networks provide more demand based services, these will be more difficult to supply in a supertrunk based architecture. Supertrunks should be associated with distribution based networks.

An Efficient Digital Modulation Scheme for Multimedia Transmission on the Cable Television Network

Krista S. Jacobsen* and John M. Cioffi
Information Systems Laboratory
Stanford University
Stanford, California 94305-4055

Abstract

We present a comparison between the performances of single-carrier modulation with equalization and multicarrier modulation on simulated cable television (CATV) channels. Simulations indicate that for a given data rate, the complexity of multicarrier is significantly lower than that of a single-carrier system. Furthermore, multicarrier is able to reduce the effects of various distortions on a CATV network, including microreflections and interference from external signals such as amateur radio, with a reasonable computational complexity while achieving information bit rates on the order of 80 Mbits/s.

Introduction

The feasibility of offering high-speed interactive data services to customers on CATV networks or similar broadband coaxial networks is currently being investigated by a number of service providers. With over 92% of the homes in the United States passed by cable, the CATV network is a potential supplier of these interactive data services. However, there are several electrical transmission problems that must be overcome before these services can be supplied reliably over CATV networks.

Presently, CATV networks are generally simplex broadcast tree- or star-structured networks. However, these networks can be converted for duplex service, as some already have been, by replacing amplifier/repeater segments with so-called "diplex" filters and amplifiers that separate the downstream frequencies (50-550 MHz) from upstream frequencies (5-40 MHz) for interactive transmission. Some CATV networks can also have high-frequency attenuation reduced by the installation of equalizing diplex amplifiers that can boost the usable spectrum to 1 GHz. Even after such installation, however, the network operator finds that only some channels are available for new digital services. Consequently, the deployment of digital services is likely to be incremental and to demand high spectral efficiency in those channels allocated for the new services.

The transmission of digital signals over the CATV network is complicated by two general effects. First, the hardware in CATV networks is not ideal. Taps, amplifiers, and splitters can all cause signals to be reflected at their insertion points. In particular, splitters in subscriber homes are known to have poor isolation characteristics. The effect of reflected signals on the frequency transfer characteristic of a CATV channel is passband ripple, which is known to cause "ghosting" in received analog signals. The effects of rippling on digital signals, however, are more severe. Variations in a channel's frequency response cause successively transmitted symbols to interfere with one another, an effect known as intersymbol interference (ISI). Depending on the absolute deviations of the channel magnitude response (which are determined by the magnitudes of impedance mismatches on a CATV system) and the frequency with which those deviations occur in the frequency response (which depends on the length of coaxial cable between mismatches), a given symbol could interfere significantly with a large number of adjacent symbols. Without some scheme to combat ISI, a receiver would make detection errors. In turn, the detection errors could disable, for example, a digital video

*This work was supported by NSF contract number NCR-9203131, JSEP contract number NCR-9203131, NASA contract number NAG2-842, and an NSF Graduate Fellowship.

decoder, resulting in a loss of the signal. Consequently, a robust digital transmission technique for multimedia signals must alleviate the ISI caused by CATV channels if reliable transmission is to be achieved.

A second source of signal degradation in CATV networks is interference from other signals. For example, leaks in a CATV network caused by poorly shielded consumer devices or system hardware allow amateur radio (ham) signals to enter the CATV system. Because the frequency spectrum allocations for CATV and ham overlap, receivers on leaky CATV networks can tune interfering ham signals along with the desired signals. As a result, the received signal is degraded. Furthermore, the bandwidth of ham signals can vary from fewer than 100 Hz to several megahertz.[1] In addition, the duration, spectral location, and severity of ham interferers can vary. As a result, an effective digital transmission scheme must be able to maintain a desired performance level in the presence of interfering signals.

In the next section, we discuss two candidate techniques for multimedia transmission on the CATV network: single-carrier quadrature amplitude modulation (QAM) with equalization and multicarrier modulation. Subsequently, we present a simulated channel, and we compare the performances of the single-carrier and multicarrier systems on this channel. Our results indicate that for a fixed system throughput, the computational complexity of the multicarrier system is significantly less than that of the single-carrier system. Next, we simulate a ham interfering signal and investigate its effect on the achievable data rates and system complexities. Simulations show that a slightly more complex multicarrier system achieves bit rates and data rates only slightly lower than those achievable on the "clean" channel. In contrast, the single-carrier system requires an enormous increase in complexity to achieve a data throughput comparable to the multicarrier system's.

Background

Although numerous digital transmission techniques exist for transmitting digital signals on bandlimited channels, given a finite-complexity constraint and the desire to achieve high throughput with little latency, there are two practical options: single-carrier quadrature amplitude modulation (QAM) with equalization and multicarrier modulation. In QAM, symbols are decoded one by one in the receiver. Because practical channels cause ISI, an equalizer is used to reduce the ISI, thereby improving the performance of the system with a fixed complexity. Equalization is, however, an inherently suboptimal detection method for practical channels. Furthermore, equalizers do not perform well on channels with significant deviations in their frequency response magnitudes. As a result, we investigate multicarrier modulation for digital transmission over CATV networks.

In multicarrier modulation, a channel is divided into N equal-bandwidth subchannels, each with its own carrier, such that the frequency response is roughly constant across each subchannel. The resulting subchannels are approximately memoryless if N is large enough. *Discrete multitone modulation (DMT)*, a common form of multicarrier modulation, makes the subchannels exactly independent and memoryless in the white Gaussian noise case by using the basis vectors of the inverse fast Fourier transform (IFFT) as the subchannel carriers and adding a *cyclic prefix* to each symbol. The cyclic prefix is a block of deterministic data that is used to "clear" each subchannel after every symbol period, and it is discarded at the receiver prior to reconstructing the data stream using an FFT. The length of the cyclic prefix must be sufficient to ensure that blocks of data from one symbol period to the next do not interfere with each other. Specifically, if the length of the channel impulse response is $\nu + 1$, then the length of the cyclic prefix must be ν. Because each transmitted block must have a cyclic prefix to be transmitted over any non-ideal channel, a portion of every block is wasted by the inclusion of the cyclic prefix. Therefore, one goal in the design of a multicarrier system is to minimize the percentage of each block wasted by transmitting the cyclic prefix. For any fixed channel impulse response, the percentage of each block lost to the cyclic prefix decreases as the FFT/IFFT size (which is $2N$)[1] is increased. The penalty for the increased FFT/IFFT size is an increase in system complexity. This relationship will be demonstrated in detail in the next section.

In contrast to traditional frequency-division multiplexing (FDM) techniques, multicarrier modulation does not constrain the number of bits per subchannel

[1] A $2N$-point complex-to-real IFFT is required in the transmitter to ensure that the signal applied to the channel is real.

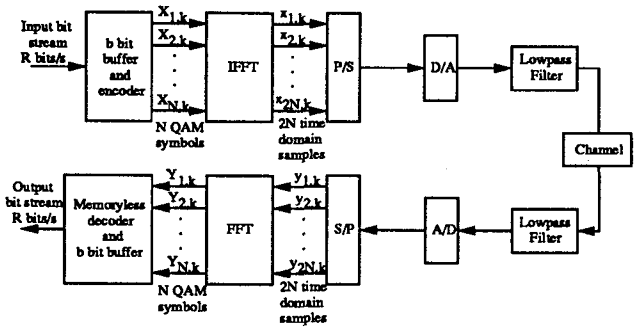


Figure 1: DMT block diagram

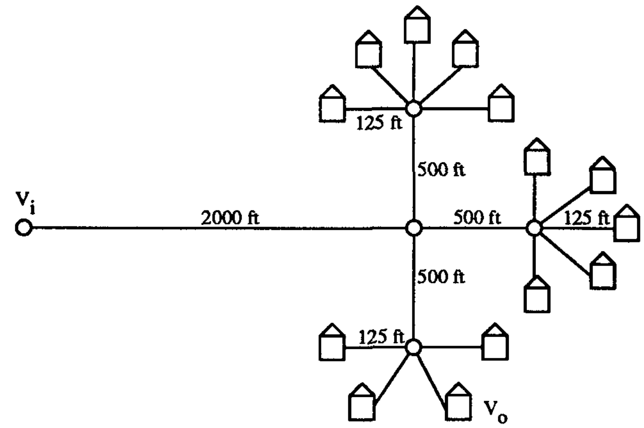


Figure 2: CATV system diagram

to be equal for all subchannels. Instead, bits are originally assigned to subchannels just after training during system initialization in direct proportion to the subchannel signal-to-noise ratios. As a result, subchannels that suffer from little attenuation and/or little noise carry the most bits, while subchannels that are severely attenuated and/or very noisy might not carry any bits. This property can be used to alleviate the problems caused by both frequency-domain ripple and interferers such as ham radio signals. Because the bit distribution is continuously updated during transmission as the receiver sends the required information to the transmitter on secure overhead channels, even severe and unpredictable interference like ham can be tolerated by a multicarrier system. If the noise on a subchannel becomes severe while the system is in use, the transmitter simply assigns fewer or no bits to that subchannel while the noise persists. Each subchannel then supports its own QAM constellation, and, because the subchannels are essentially independent and memoryless, a memoryless detector is used for each in the receiver. As a result, no equalizer is required in the multicarrier receiver. Figure 1 shows a block diagram of the DMT transmitter and receiver. For more information on multicarrier modulation and DMT, see [2], [3], [4], [5], and [6].

Simulations

In this section, we describe our simulated channel model and discuss how well our model approximates actual CATV channels. We then use our simulated channel to compare the performances, in terms of the computational complexity required to support a given throughput, of multicarrier systems and single-carrier systems with equalization. Finally, we modify our simulated channel to determine the effects of interference from amateur radio signals or other comparatively narrowband signals on the two transmission systems.

CATV Channel Model

To enable us to quantify the performances of both single-carrier and multicarrier modulation on CATV channels, we computed a "typical" channel frequency response using the system configuration shown in Figure 2.[2] Because most of the degradations to CATV signals are caused by "leaky" hardware in subscriber drops, we have only modeled the portion of the cable system that is nearest to subscriber homes. The system is modeled as 1/2" copperclad coaxial cable delivering signals through bridge taps to fourteen homes. We have assumed that both source and termination impedances are 75Ω. Since the upstream frequency allocation in most CATV systems is from 5-40 MHz, and we are trying to determine the viability of transmitting bidirectionally on the CATV network, we selected the 6-MHz slot from 30-36 MHz as our simulated channel. The frequency response magnitude of this channel is shown in Figure 3. In computing the frequency response magnitude of the channel, we have assumed the presence of a gain element that brings the maximum magnitude to 0 dB. We note that our channel meets the FCC's technical specifications for the frequency response of CATV channels.[7] Furthermore, our channel exhibits the frequency-domain rippling that is common

[2] Thanks to Professor D. G. Messerschmitt of the University of California at Berkeley for making his transmission line modeling program available to our research group.

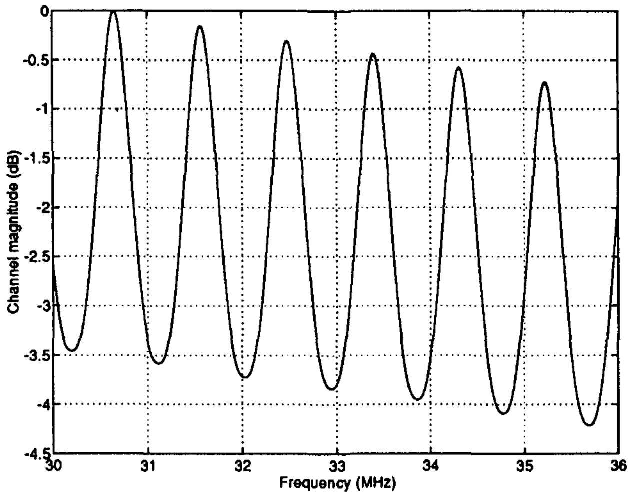


Figure 3: Frequency response magnitude of simulated channel

in actual CATV channels due to leaky consumer hardware such as splitters. Consequently, this channel will help us to illustrate the ability of multicarrier modulation to combat ISI. In addition, by modeling interfering signals as flat nulls in the channel's frequency response, we will be able to illustrate the effect of interferers on the performances of both single-carrier and multicarrier transmission systems.

System Comparison on Simulated Channel

In this section, we investigate the achievable data rates and complexities of both DMT and single-carrier systems for various signal power and channel noise levels. We assume here that no interfering signals, either from within the CATV system or from external sources, degrade the channel.

First, we simulated the DMT system described previously using $2N = 4096$ (2048 subchannels) with our simulated channel and symbol error probabilities of 10^{-7} and 10^{-9}. We assumed a flat noise power spectral density over the 6 MHz bandwidth and a flat power distribution over the subchannels. With these assumptions, we have defined the overall DMT signal-to-noise ratio (SNR) as the ratio of the signal power to the noise power for the subchannels that have 0 dB attenuation. Allocating fractional numbers of bits to subchannels was allowed in the simulation, and a 5.0 dB coding gain was assumed. Table 1 details the achievable bit rates for various DMT SNRs and the two error probabilities.[3] We see from the data in the table that, for any SNR, the throughput of the system operating with a symbol error probability of 10^{-9} is decreased by approximately 7.8 Mbits/s with respect to the $P_e = 10^{-7}$ system. For either symbol error probability, however, the throughput of the multicarrier system is adequate to support a number of multimedia services. For example, with 49 dB SNR and $P_e = 10^{-7}$, a bit rate of 83.243 Mbits/s is high enough to support 4 20-Mbit/s HDTV signals or 16 5-Mbit/s digitized NTSC signals. We note that a 6-MHz slot on the CATV network could easily be used to transmit a variety of multimedia signals with different bandwidth requirements simply by assigning an appropriate number of subchannels to each signal. Furthermore, because of the flexibility in bandwidth allocation that multicarrier affords, a given 6-MHz slot could serve several users at the same time.

Table 1: Achievable bit rates on simulated channel

	Bit rate (Mbps)	
DMT SNR (dB)	$P_e = 10^{-7}$	$P_e = 10^{-9}$
40	65.256	57.493
43	71.251	63.485
46	77.246	69.479
49	83.243	75.474
52	89.240	81.471
55	95.237	87.467
58	101.23	93.464
61	107.23	99.461

To illustrate the relationship between the channel frequency response magnitude and the number of bits supported by each DMT subchannel, Figure 4 shows the bit allocation for the 49 dB SNR case with a symbol error probability of 10^{-9}. Because both the power distribution and noise power spectral density are assumed to be flat, each subchannel supports a number of bits directly proportional to its frequency response magnitude.

Because the complexity of DMT systems is proportional to the FFT/IFFT size used in the implementation (or, equivalently, the number of subchannels into which the channel is divided), we varied the FFT size to determine the effect of a reduced number of subchannels on the achievable bit rate. With $P_e = 10^{-9}$, the results are given in Table 2. In addition to the achievable bit rates, Table 2 gives the

[3] These data were first presented in [8].

Table 2: Achievable bit rates with $P_e = 10^{-9}$ as a function of FFT size

DMT SNR (dB)	FFT size ($M = 2N$) 4096	2048	1024	512	256	128
40	57.493	57.468	57.420	57.323	57.131	56.752
43	63.485	63.457	63.402	63.294	63.078	62.653
46	69.479	69.448	69.388	69.267	69.028	68.556
49	75.474	75.441	75.374	75.243	74.980	74.461
52	81.471	81.434	81.362	81.218	80.933	80.366
55	87.467	87.428	87.350	87.195	86.885	86.272
58	93.464	93.422	93.338	93.171	92.838	92.178
61	99.464	99.416	99.326	99.147	98.791	98.084
% data	98.6	97.2	94.3	88.7	77.3	54.7
Complexity (MIPS) (assumes $2M \log_2 M$ instructions/FFT)	264	240	216	192	168	144

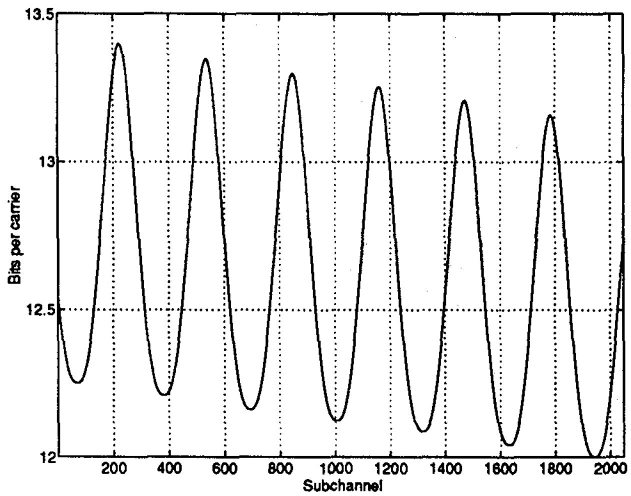


Figure 4: Bit allocation for multicarrier system with 49 dB SNR and $P_e = 10^{-9}$

percentage of each block that is actually data given that the length of our channel's impulse response is approximately $\nu + 1 = 59$ symbol periods. Finally, the complexity, in millions of instructions per second (MIPS), of the multicarrier transmitter and receiver combination is given for each FFT size. The number of instructions required to compute the M-point FFT of a real input sequence is commonly approximated as $2M \log_2 M$, where M is a power of 2. Because we must compute $\frac{2}{MT}$ FFTs per second for a passband multicarrier system operating in a bandwidth of $1/T$ Hz, we find that the complexity in MIPS for a multicarrier system is roughly equal to $\frac{4 \log_2 M}{T}$, where $M = 2N$ in our notation. We have assumed in computing the complexity values in the table that the multicarrier system uses the entire 6 MHz of allocated bandwidth.

The data in Table 2 show that for a given DMT SNR, the maximum achievable bit rate decreases only slightly as the FFT size (or, equivalently, the system complexity) is decreased. However, the percentage of each block that is data decreases more dramatically as the cyclic prefix consumes a larger percentage of each size-$2N$ block. Because we wish to maximize the amount of data transmitted, we cannot arbitrarily decrease the size of our FFT simply because the system throughput remains approximately constant.[4] Consequently, we will constrain the minimum FFT size to be 512, allowing a maximum 11.3% of each transmitted block to be lost to the cyclic prefix. In addition, because the achievable bit rates for $P_e = 10^{-9}$ with $2N = 4096$, 2048, and 1024 are nearly equal, and the data percentages per block are also similar, we have used only the FFT sizes 1024 and 512 with $P_e = 10^{-9}$ in the following analysis.

To compute the complexity of single-carrier systems yielding the same throughputs as the DMT systems, we first computed the data rates and the numbers of bits per symbol that must be supported by each single-carrier QAM system. After adjusting the throughput values according to the data percentages given in Table 2, we find that single-carrier QAM systems must support the bit rates and approximate numbers of bits given in Table 3. To make the comparison between the single-carrier and multicarrier systems more tractable for the reader, we have maintained a column of DMT

[4] In actuality, we could use the smaller FFT sizes (for example, $2N \leq 256$) by implementing a time-domain equalizer (TEQ), the purpose of which is to reduce the required length of the cyclic prefix in exchange for an increase in system complexity. A detailed discussion of the TEQ is beyond the scope of this paper, but interested readers should consult [5] or [9] for details.

Table 3: Data for single-carrier systems assuming 5.0 dB coding gain

DMT SNR (dB)	Corresponding to $N = 1024$ DMT			Corresponding to $N = 512$ DMT		
	Data (Mbps)	QAM bits	SNR_{rcv}	Data (Mbps)	QAM bits	SNR_{rcv}
40	54.170	9	32.5	50.828	8	29.5
43	59.813	10	35.5	56.123	9	32.5
46	65.461	11	38.5	61.419	10	35.5
49	71.108	12	41.5	66.718	11	38.5
52	76.757	13	44.5	72.016	12	41.5
55	82.406	14	47.5	77.316	13	44.5
58	88.055	15	50.5	82.615	14	47.5
61	93.704	16	53.5	87.914	15	50.5

SNR values. Table 3 also contains the SNRs that must be achieved by single-carrier receivers to support the required numbers of bits. In computing the required receiver SNRs, we have used the rule of thumb that, for $P_e = 10^{-9}$, the first 2 QAM bits require 16.5 dB, and each additional bit requires 3.0 dB. Additionally, we have assumed, as we did for the multicarrier system, a 5.0 dB coding gain.

Finally, we simulated a series of minimum mean-square-error linear equalizers (MMSE-LEs) to compare the single-carrier computational complexities at the various data throughput rates to the complexities of the corresponding DMT systems. Since the complexity of an MMSE-LE is proportional to the number of taps, N_f, in the equalizer, we have computed the approximate numbers of taps required for the MMSE-LEs to achieve the required SNRs given in Table 3. The results are given in Table 4. In Table 4, γ is the "complexity increase factor," and it is the ratio of MMSE-LE complexity to DMT complexity for each DMT SNR and FFT size given in the table. In computing the complexities for the MMSE-LE systems, we have assumed $T/2$ spacing and a transmit symbol rate of 6 Msymbols/s. Therefore, the bandwidths of the single-carrier and multicarrier systems are equal. In addition, the increases in complexity for both the DMT system and the single-carrier system due to the 5.0 dB code are approximately equal. Thus, we have neglected the code in our complexity calculations in both cases. The data in Table 4 show that the complexity of the MMSE-LE system increases with SNR in contrast to the DMT system, the complexity of which is constant for any SNR once the FFT size is selected. Furthermore, the complexity increase factors given in Table 4 show that, for this channel, the complexity of a single-carrier QAM system with an MMSE-LE receiver is at least 2.5 times the complexity of a DMT system operating at the same data rate.

We note that by using a decision feedback equalizer (DFE) instead of a linear equalizer, the complexity of the single-carrier system might be reduced slightly with respect to the MMSE-LE case because only the feedforward filter would need to run at the $T/2$ rate. However, our preliminary results from DFE simulations show that the complexity of a DFE receiver is still at least twice that of the multicarrier system.

System Comparison on Degraded Channel

In this section, we explore the effect of interfering signals on the performances of single-carrier and multicarrier systems. For this simulation, we have simulated a ham interferer of bandwidth 30-kHz located near the middle of our 6-MHz simulated channel. We have assumed that the ham signal causes a -20 dB null in our channel's frequency response magnitude, as shown in Figure 5.

In order to select an appropriate FFT size for the multicarrier system, we computed the channel's impulse response to determine the required cyclic prefix length. Since the impulse response length is roughly 281 samples, we chose $2N = 2048$, implying that approximately 86% of each transmitted block is data. Assuming $P_e = 10^{-9}$ and 5.0 dB coding gain, we then simulated another multicarrier system and found the achievable bit rates on this channel for the same DMT SNRs we used previously. The results are given in Table 5. Comparing the bit rates given in Table 5 with those in Table 2 for the case $2N = 2048$, we see that the effect of the interferer is a minimal decrease in the overall bit rate for each value of DMT SNR. A plot of the bit distribution, shown in Figure 6 for the 49 dB SNR case, indicates that only the subchannels overlap-

Table 4: MMSE-LE complexity data assuming $T/2$-spaced equalizer

	Corresponding to $2N = 1024$ DMT			Corresponding to $2N = 512$ DMT		
DMT SNR (dB)	N_f	Complexity (MIPS)	γ	N_f	Complexity (MIPS)	γ
40	45	540	2.50	40	480	2.50
43	45	540	2.50	45	540	2.81
46	50	600	2.75	45	540	2.81
49	50	600	2.75	45	540	2.81
52	55	660	3.00	50	600	3.13
55	60	720	3.33	50	600	3.13
58	70	840	3.88	55	660	3.44
61	75	900	4.17	65	780	4.06

Table 5: Achievable bit rates and data rates for $2N = 2048$ DMT system on channel with 30-kHz interfering signal

DMT SNR (dB)	Bit rate (Mbps)	Data rate (Mbps)
40	57.314	49.479
43	63.600	54.647
46	69.291	59.819
49	75.283	64.992
52	81.276	70.166
55	87.270	75.340
58	93.264	80.515
61	99.258	85.689

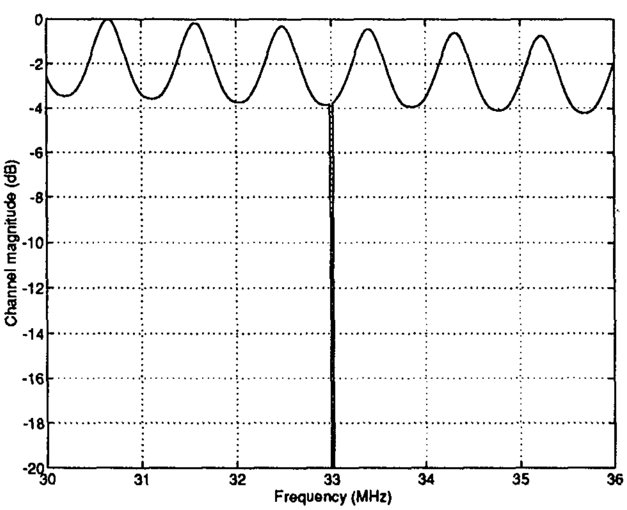


Figure 5: Frequency response magnitude of simulated channel with 30-kHz interferer

ping the interfering signal's bandwidth carry fewer bits than they did in the previous simulations. However, since the channel's impulse response length has increased significantly, the data percentage per block has decreased, yielding lower data rates. Specifically, for each DMT SNR value, the percent decrease in data throughput with respect to the data rates achievable with $2N = 2048$ on the "clean" channel of the previous section is about 11%.

Finally, we computed the required numbers of QAM bits and receiver SNRs and simulated another set of MMSE-LEs that achieve the same data rates as the DMT systems on the degraded channel. The pertinent values are given in Table 6. In computing the data in Table 6, we have again assumed a 5.0 dB coding gain and $T/2$ equalizer spacing. The values of γ in Table 6 clearly illustrate the dramatic increase in complexity required by the MMSE-LE to achieve the same data rates as the multicarrier system. We note again that the complexity of the single-carrier system increases with SNR, while that of the multicarrier system remains 240 MIPS. Furthermore, if the bandwidth of the interfering ham signal were even wider than the 30 kHz we assumed, then the complexity comparison would further favor multicarrier. Equalizers with reasonable numbers of taps are simply not effective on channels with so-called "dead-zones" caused by strong interfering signals. Thus, to achieve data rates on the order of those attainable with a multicarrier system on a channel with a dead-zone, a receiver would need an equalizer with a prohibitive number of taps.

Table 6: MMSE-LE data for degraded channel

DMT SNR (dB)	Data rate (Mbps)	QAM bits	SNR_{rcv}	N_f	Complexity (MIPS)	γ
40	49.479	8	29.5	160	1920	8.00
43	54.647	9	32.5	165	1980	8.25
46	59.819	10	35.5	170	2040	8.50
49	64.992	11	38.5	195	2340	9.75
52	70.166	12	41.5	225	2700	11.25
55	75.340	13	44.5	250	3000	12.50
58	80.515	13	44.5	250	3000	12.50
61	85.689	14	47.5	260	3120	13.00

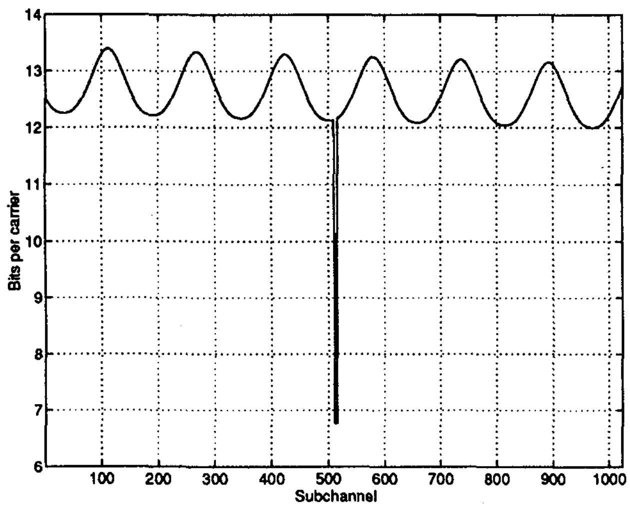


Figure 6: Bit allocation for multicarrier system on degraded channel with 49 dB SNR and $P_e = 10^{-9}$

Conclusion

The results of our simulations indicate that discrete multitone modulation is more computationally efficient than single-carrier modulation with equalization for the transmission of digital data over CATV channels with rippled passbands. Furthermore, DMT can easily adapt to a variety of channel degradations, optimizing the system bit rate for a given channel and complexity. Because system reliability is important for any consumer service, a digital transmission system must maintain a certain level of performance under a wide variety of circumstances. Given our simulation results, multicarrier modulation appears to be better equipped than single-carrier modulation to accomplish this objective.

References

[1] R. Pience. private communication. February 1994.

[2] J.M. Cioffi. "A Multicarrier Primer". In *ANSI T1E1.4 Committee Contribution*, pages 91–157, Boca Raton, FL, November 1991.

[3] J. A. C. Bingham. "Multicarrier Modulation for Data Transmission : An Idea whose Time Has Come". *IEEE Communications Magazine*, 28(5):5–14, May 1990.

[4] A. Ruiz, J.M. Cioffi, and S. Kasturia. "Discrete Multiple Tone Modulation with Coset Coding for the Spectrally Shaped Channel". *IEEE Transactions on Communications*, 40:1012–29, June 1992.

[5] J.S.Chow, J.C.Tu, and J.M.Cioffi. "A Computationally Efficient Adaptive Transceiver for High-Speed Digital Subscriber Lines". In *Proceedings 1990 International Conference on Communications*, June 1990. Boston, MA.

[6] J.C. Tu. *Theory, Design and Application of Multichannel Modulation for Digital Communications.* PhD thesis, Stanford University, June 1991.

[7] Federal Communications Commission. "Code of Federal Regulations". Title 47, Part 76.605. 1992.

[8] K.S. Jacobsen and J.M. Cioffi. "High-performance Multimedia Transmission on the Cable Television Network". In *Proceedings 1994 International Conference on Communications*, New Orleans, LA, May 1994.

[9] J.S. Chow. *Finite-length Equalization for Multicarrier Transmission Systems.* PhD thesis, Stanford University, June 1992.

AN INTEGRATED NETWORK MANAGEMENT SYSTEM FOR CABLE TELEVISION

John C. Anderson, P. Eng.
Rogers Engineering

Abstract

The Integrated Network Management System (INMS) uses computer technology to monitor and control all of the equipment in Rogers primary fiber hubs, secondary fiber hubs; and trunk amplifiers and power supplies in the coaxial cable TV system. In the event of a fiber cut, or optical hardware failure, a backup route or redundant hardware automatically switches in to restore service. The INMS system monitors all of this equipment and reports any fault conditions that occur to a central Technical Action Center (TAC). Through different security access levels, all users can monitor the network, but designated technicians have the ability to issue switching or reconfiguration commands. INMS is unique in that it provides a single graphical interface to the outside plant with a consistent representation of several different vendors' equipment laid out in a logical network presentation. Rogers has developed software that integrates the proprietary communications protocols of various vendors into one system of centralized network management. The operator now has a comprehensive view of the entire network status at any time, providing more effective problem isolation and resolution. This system is providing real gains in network operation efficiency and customer service.

Background

The past several years have witnessed dramatic changes in cable television network architectures. These changes have been driven by the need for improved signal quality, better reliability and expanded bandwidth. The local headends that formerly served a community are being replaced by regional headends that serve multiple primary and secondary fiber optic hub sites. These regional headends provide economies of scale for signal reception, processing, routing, and scrambling. However, with the use of digital or FM fiber optic backbone systems, the local fiber hubs are in essence small headends that receive the fiber signal and remodulate onto VSB-AM systems for either coaxial or further fiber distribution. The large tree-and-branch coaxial networks are also evolving rapidly into fiber-rich networks with much smaller serving areas off the fiber feeds. Some architectures are now calling for 500 to 2000 home nodes served by each fiber. Fiber optic receivers are being placed in large numbers throughout the distribution network.

With this network evolution we are also seeing a proliferation of new components in the network, such as fiber secondary hubs and optical bridgers. Future changes as part of the full service network such as DVC, video-on-demand, and high speed data services will only add to the number and complexity of the installed equipment base.

Fiber optic systems have brought new vendors into the cable TV marketplace and traditional vendors have increased the range of their products. However, it is extremely rare to find a cable TV system built with a single vendor's hardware. In most cases, a network consists of hardware from numerous vendors and of varying vintage. Unlike the telecom environment, many cable TV manufacturers and suppliers have either not addressed network management issues at all, or the solutions that they offer are proprietary and only apply to their particular hardware. These stand-alone systems do not communicate with one another to rationalize alarms, create trouble tickets and outage reports, or interface to other operations support systems. Network management protocol standards and interoperability are virtually nonexistent in the cable TV industry. The result is a mix of various computer systems from different vendors each interfacing with a particular type of hardware. These single-vendor systems focus on technology control, rather than the end-to-end process of delivering signals reliably to the customer. Consequently, the system operator must deal with an array of independent computer systems and an astute operator must interpret the cause of alarm conditions.

With all of the activity in fiber optic and full service network deployment, the ongoing operational implications have received very little attention. Part of Rogers Cable TV fiber optic and customer service strategies was the recognized need for an automated network management system to aid in the operation and maintenance of this growing base of network equipment. Without some form of automation, there would be a severe strain on staff skill levels and overall manpower requirements.

Network Management System Requirements

A Network Management System is a tool to co-ordinate, monitor and control the distributed hardware resources throughout a network. To be fully effective, the network management system requires cohesive interaction between the human operators, the software applications and the network hardware elements.

Rogers had a number of key requirements for an integrated network management system:

1. The system must integrate the various stand-alone systems into one so that an operations technician could view the network on one computer screen as a system, rather than a series of disjointed devices. The graphical presentation of multiple devices at multiple locations to multiple end-user workstations must be in a consistent format.

2. The system must be able to handle various device protocols, no matter how proprietary or standardized, complex or simple they might be.

3. The system must be flexible in configuration to handle multiple users logged on simultaneously, both from a central location and distributed throughout the operating divisions.

4. The system must be very user-friendly, providing automatic updates to operations technicians and allowing them to concentrate on solving problems rather than operating the network management system.

5. The system must be justified financially in operating efficiency improvements and better customer service.

6. The system must use off-the-shelf components, with low cost points-of-presence and have a linear growth cost.

7. The system must have several levels of security that could be partitioned either geographically or by equipment type, to prevent unauthorized access to the system.

Rogers Integrated Network Management System

A network management project team from Rogers spent over a year researching various systems that were in the marketplace. Almost all of these proved inadequate for the requirements identified. Most operated on high-end expensive workstations, and, with software, a total solution was in the order of several hundred thousand dollars. Most of these solutions only supported telecom and LAN/WAN protocol standards. Very few systems could handle simple monitoring such as voltages or contact closures, and many had limits as to the number of devices they could handle. A number of systems were simply "managers of managers" that brought proprietary vendors' systems together on one screen but didn't really integrate the alarm and database functions.

Consequently, Rogers assembled a team of software developers and began to develop an Integrated Network Management System. "Off-the-shelf" components were sourced which included both hardware and software tool kits. All workstations and hub site communication servers (c-servers) were to be 386 or 486 type PCs with standard interface cards installed. Hewlett-Packard's Openview software adapted best to Rogers' applications. Although originally designed for WAN/LAN monitoring, this software tool kit allowed for user customization. The various workstations and c-servers would communicate via a Novell LAN interconnection, since this technology was already in place for Rogers office LAN systems.

Technical Overview

INMS is a highly distributed system; it uses 386 and 486 type PCs interconnected using any standard Local Area Network (LAN) technology for both hub site servers and local workstations. The distribution of processes allows for system modularity, and flexibility while the adherence to existing standards offers a solid base for future growth.

A number of INMS system software modules, or processes, cooperate to manage the wide variety of information packets present in the system at any one time. These processes include message routing, database access, security, status management, and system testing. Security is active on all users, network devices, c-servers, message packets and LAN activity. INMS information is not accessible to a given user without approval from the security process. The status management process is responsible for maintaining a current snap-shot of the network that rapidly issues updated alarms to work stations using INMS. Without this process, it would take an inordinately long time to log onto the network and get updated on the status of all devices. The distribution of processing provides INMS with a linear growth path, with no practical upper limit concerning number of users, locations or devices.

Each hub site has a communication server (c-server) connected to the INMS backbone using a variety of async, sync, X.25 or T1 bridge hardware. An INMS workstation can be collocated with the c-server at a hub site for local monitoring and control of devices by field staff. Each c-server typically supports up to thirty-two physical ports, with each port supporting an individual vendor's protocol (both proprietary and non-proprietary). The c-server acts like a network management multiplexer and translator, combining a number of proprietary sources into a single INMS communication port. Under typical operation, the c-server either polls devices cyclically for a change of state, or is interrupted by one of the managed devices when a change of state occurs. This initiates an event within the c-server that takes the proprietary source and converts it to a standardized INMS message. Once converted, the c-server packages the event message within an envelope and issues a datagram to the status manager for alarm filtering. The status manager compares the datagram to see if its contents result in a change-of-state of the current network image. If a change of state has occurred, the status manager issues an event to all workstations registered (via security) to access this information.

The INMS user workstation uses both graphics and text to represent the current state of the managed network or devices within its access. The system design ensures that monitoring and control procedures are similar for a wide range of devices, even though the underlying protocols or technologies are vastly different. This ability allows even junior TAC staff to function as seasoned professionals. The end user workstations of INMS are typically 80486 based computers, equipped with super-VGA, high resolution monitors, a mouse and 4 MB of RAM. A company LAN connection is the only requirement for a workstation location. INMS is accessible (with security limitations) to everyone within the company from technicians, to supervisors, to executives. The workstation presents the user with a simple graphical (Windows-based) representation of all networks and facilities, plus access to the textual database, trouble ticket, e-mail and login screens.

INMS uses a series of "layered" screen displays to graphically illustrate the network topology and status. Each layer presents a set of icons that provide access to different successive layers containing more detail. The color of the icon indicates the status of the underlying devices: red indicates an alarm, yellow a warning, green is OK, blue indicates the device is out of service, purple indicates a disabled alarm. Using a mouse, a user can "drill down" through several layers of detail to isolate the cause of an alarm or to check the status of a particular device. After the appropriate log-in procedure, the opening screen is that shown in Figure 1, the Ontario Inter-city Fiber Optic Network. A mouse click on Toronto, for example, takes the user to the local Toronto fiber network, Figure 2, which shows the primary fiber hubs. A mouse click on the Sheppard hub takes the user to the display of the fiber detail screen, Figure 3. The user can then drill-down further to the equipment detail on each fiber. Figure 4 shows the optical equipment on fiber block A. At the equipment level screen, the user can interrogate each device and retrieve real-time status of all measurable parameters. For example, Figure 5 shows the secondary optical receiver status for Block A. Remote video monitoring via a tunable demodulator and signal restoration controls are available using the screen in Figure 6. Similar navigational steps take the user down into secondary hubs and out into the coaxial trunk network, Figure 7. Figure 8 illustrates the interrogation of individual trunk amplifier parameters. Standby power supplies are monitored and controlled in a similar fashion. The entire layered display system is analogous to an inverted tree that allows the user to navigate across and into the network through thousands of possible paths easily and logically.

Every physical component of INMS is off-the-shelf, multi-vendor, multi-source equipment. Manufacturer independence allows the implementation of INMS at the lowest competitive price, and provides multiple sources for backup or emergency purchase.

On high priority links, INMS typically operates as an "out-of-band" management system. Out-of-band means that the network management information data traffic is transported external to the network being monitored. On links with lower priority, the carriage of INMS data traffic is on the network, with dial-up redundancy available in case of total network failure. Small fiber repeater sites, that do not warrant a full c-server installation, communicate via automatic dial-in and dial-out access.

Operational Results

Rogers has implemented INMS in all of its operating divisions interconnected with fiber. INMS has few limitations on what devices can be monitored or controlled. It is being used for Rogers Cablesystems fiber distribution network including digital, AM, and FM fiber links. In each primary fiber hub INMS provides control of agile modulators and demodulators; and video and RF switches. Using INMS, redundancy equipment and procedures are in place to deal with entire fiber cable cuts, loss of individual fibers or associated equipment, or the loss of an individual output channel. In any of these events, signal restoration occurs in seconds rather than hours. INMS is expanding to secondary fiber hubs as they are installed, to provide monitoring of optical transmitter and receiver status, and control of redundant path switching. Approximately 4000, or one third, of the company's trunk amplifiers are now on-line with INMS, with the remainder scheduled for the balance of the year.

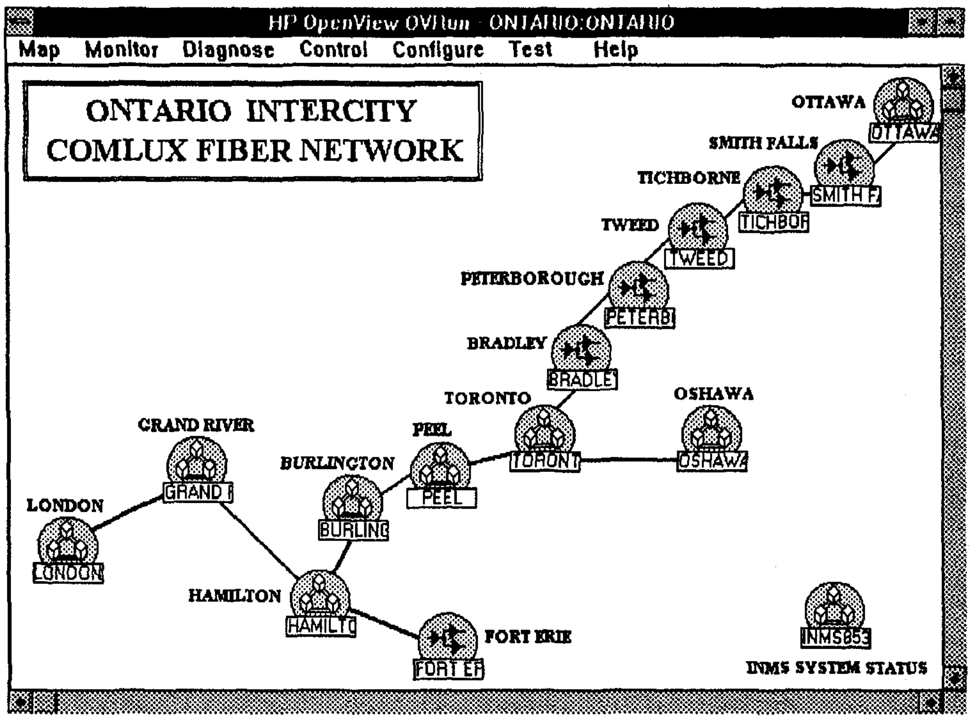


Figure 1
Ontario Inter-city Fiber Optic Network

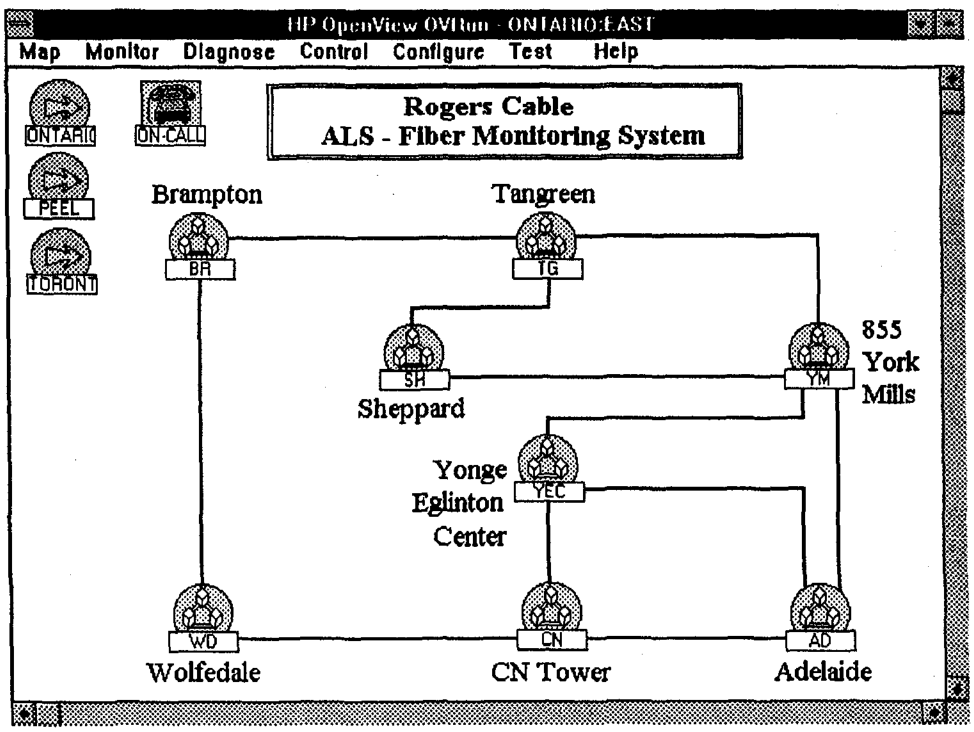


Figure 2
Toronto Primary Hubs

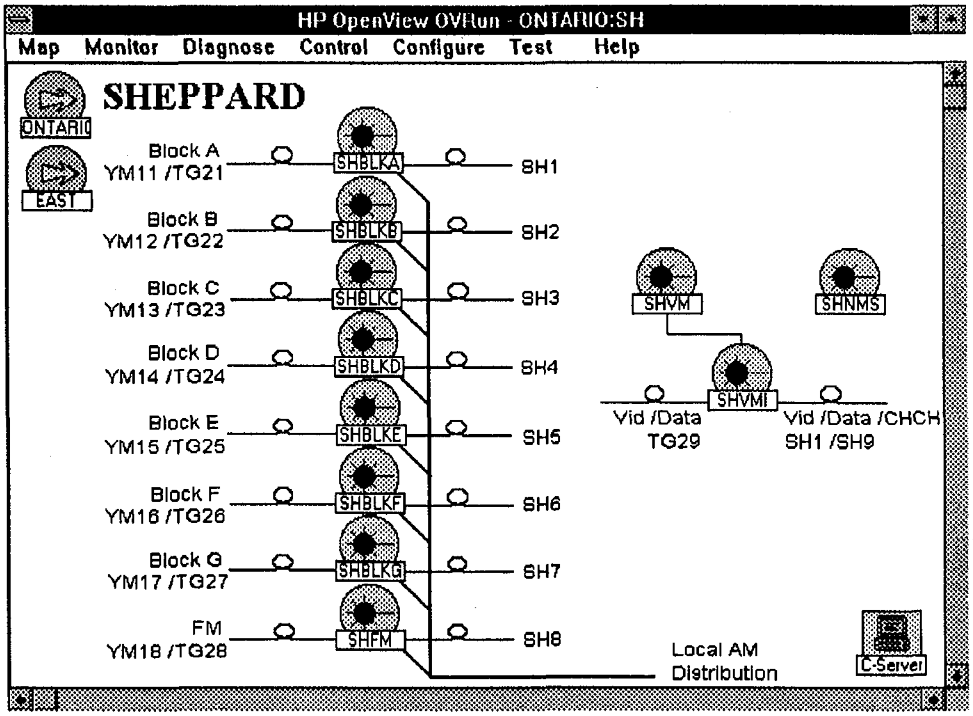


Figure 3
Sheppard Hub Fiber Detail

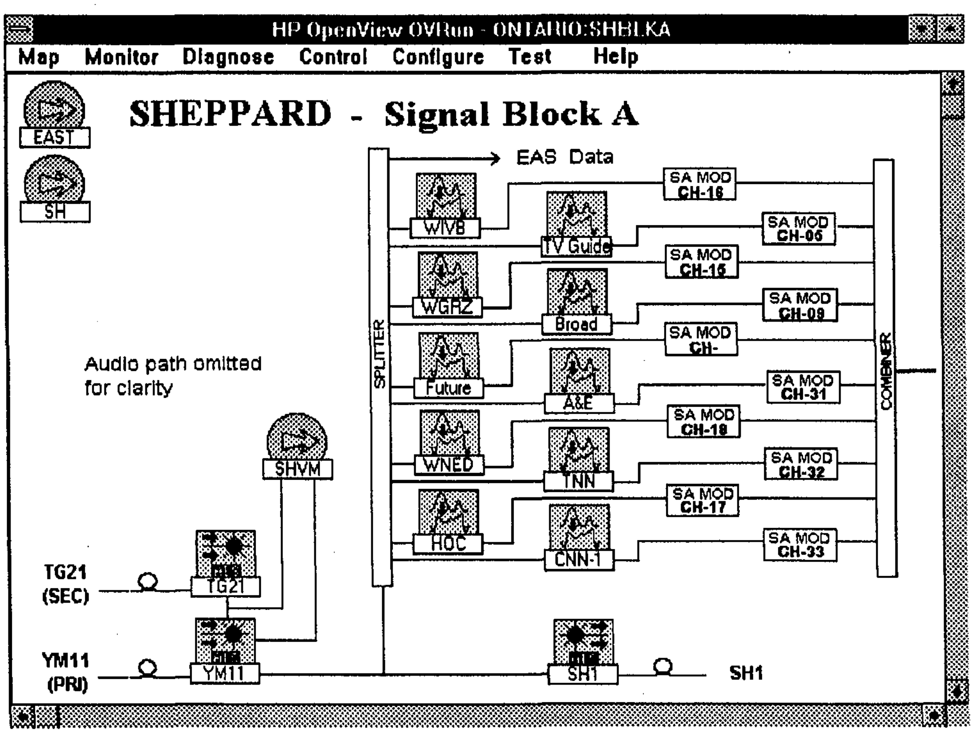


Figure 4
Fiber Block Equipment Detail

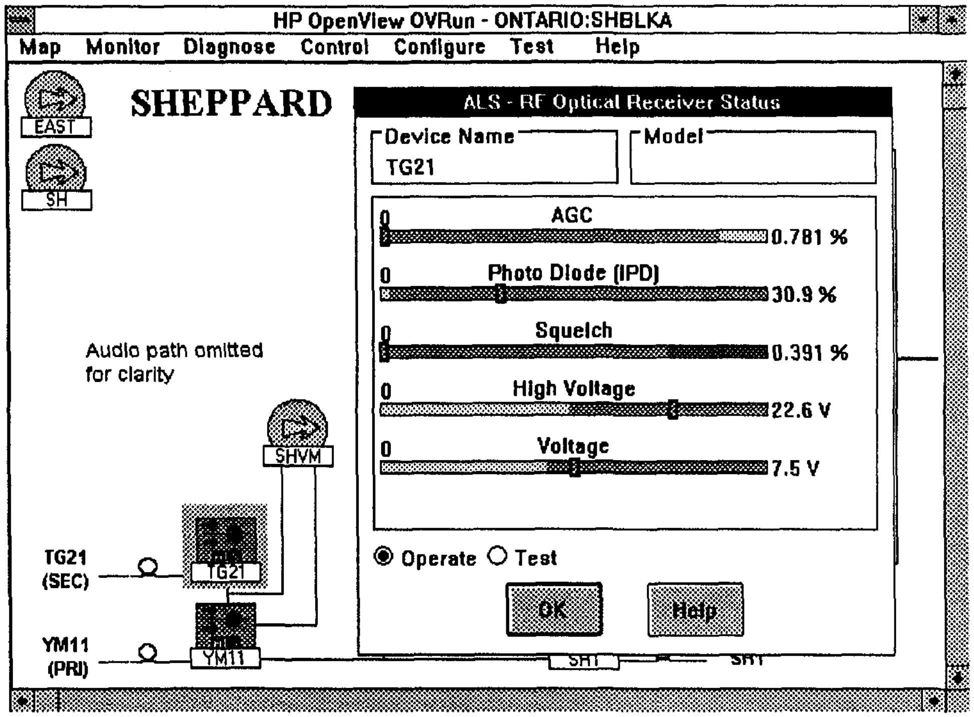


Figure 5
Optical Receiver Status

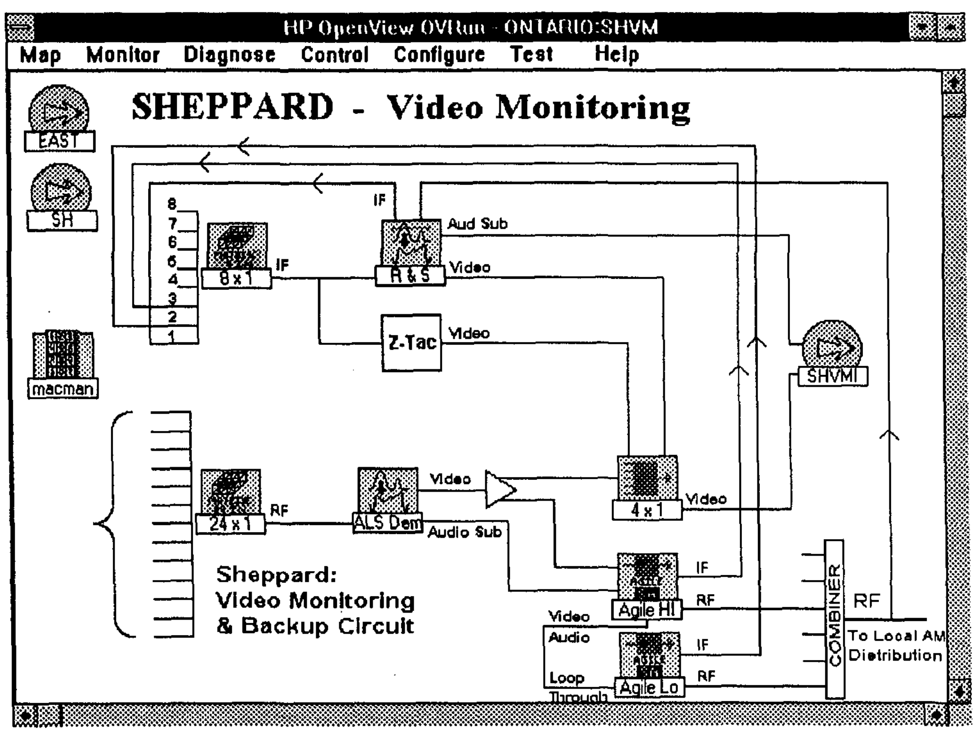


Figure 6
Video Monitoring and Backup Switching

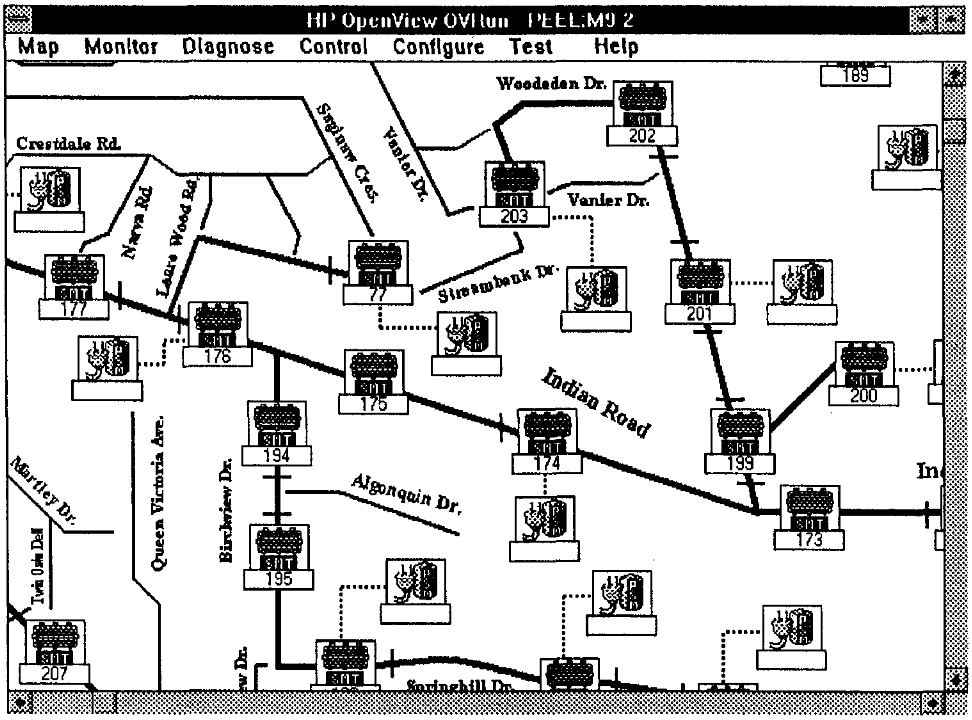


Figure 7
Typical Coaxial Plant Screen

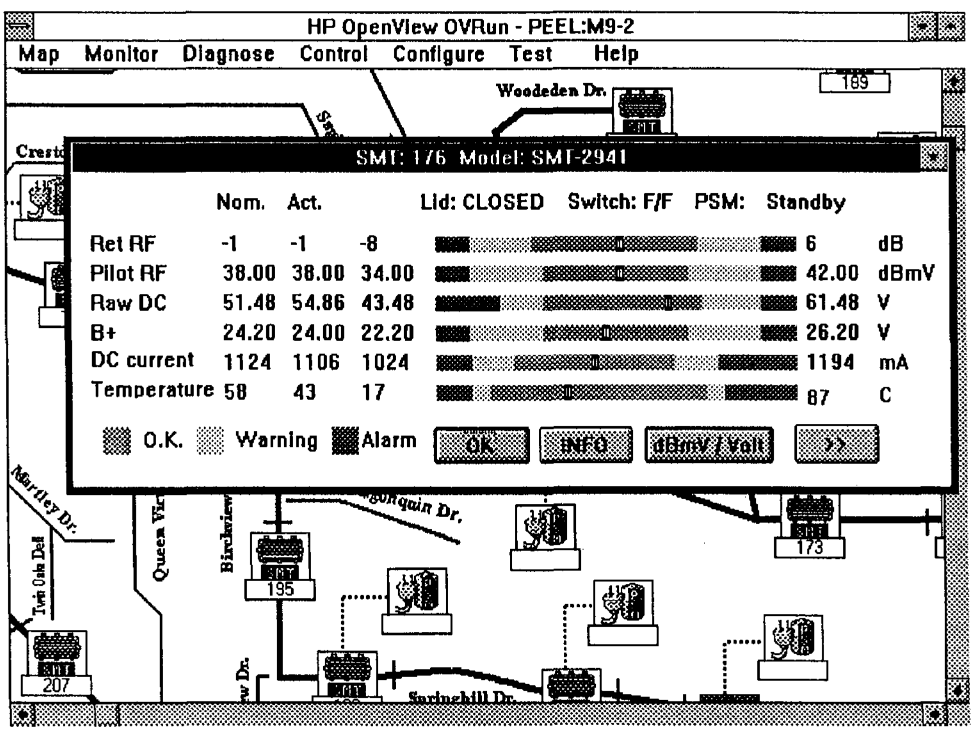


Figure 8
Amplifier Status Window

Using INMS, a technician, manager, or executive can graphically review the status of the network, its components and relationships. The network status can be viewed either globally, by city, by hub site or by equipment type. Real-time alarm conditions can be reacted to remotely without the sometimes unnecessary dispatch of technical staff. Should the Technical Access Center (TAC) operator feel that a problem requires hands-on attention, a technician can be dispatched using the INMS Trouble Ticket System. Color-coded icons represent all components in the overall network within a geographical display. All control functions use a consistent style across all vendor's equipment. Full auditing and security levels ensure that only authorized users are able to control devices specifically assigned to them.

On-call technicians have access to INMS from home using dial-up facilities and a laptop PC. In many situations, they can take corrective action even before leaving home. During the day, certain field technicians can access INMS from their vehicles using cellular dial-up facilities.

INMS monitors a number of hub site environment conditions such as temperature, primary AC power, standby generator power, fire and smoke alarms, etc. This provides an early warning of potentially major problems.

Each hub site has a current "on-call" technician list icon attached to it listing various phone numbers . The TAC center operator can quickly establish who to call and how best to reach them depending upon the nature of the problem.

Each alarm is time stamped and logged in the INMS database. Report writer software facilitates ad-hoc and regular reports that sort and provide statistics on alarms by site, by type of equipment, by date, etc. This functionality provides management with the tools to analyze the effectiveness of network operations.

Future Directions

Rogers Cable TV currently operates with three main operations support systems: the customer database and billing system (Supersystem), the Integrated Network Management System, and the automated mapping and facilities management (AM/FM) system. These three "islands" of technology currently cannot be linked electronically but yet they support three independent databases with overlapping functionality. The long-range goal of Rogers Cable TV is to provide an integrated customer service system that would link these three systems together into a comprehensive operations support system. For example, it would be desirable to relate an equipment alarm on INMS to the affected addresses on the AM/FM system, and then relate these addresses to subscriber accounts on Supersystem. A technical service representative (or perhaps even a voice response unit (VRU)) would have up-to-the-minute information on the technical status of each subscriber when they phone in. The municipal base map on the AM/FM system would contain information such as demographics of subscribers, property information, zoning information and building types. The AM /FM system would be able to sort and correlate information that is now either not available or which must be assembled manually, such as address lists of subscribers fed from specific hubs or equipment, inventories and repair histories of installed equipment, subscriber bandwidths, support structure rental detail for billing reconciliation, etc. After linkage, these three core systems would then feed off into computer-aided dispatch systems, GPS systems for vehicle tracking and radiation patrol, and automated work management systems.

Recommendations

Cable television operators are facing numerous challenges with the deployment of new technologies and network architectures, the introduction of new services, new and vigorous competition in the video marketplace, along with pressures from both customers and regulators for better service and reduced costs. It is imperative that operating efficiencies be applied in order to meet these challenges. Network management is one of these efficiencies.

However, the CATV industry is also facing a number of obstacles to the implementation of integrated network management. A lack of industry standards and a proliferation of proprietary vendor solutions has made the task of integrating network management solutions very difficult. The CATV industry must come together and agree on standard protocols or application programming interfaces in order to make integrated network management easier to implement. Vendors seem to be reluctant to offer software solutions either because they add cost to their product, or they are not willing to invest in software development.

Much of the CATV plant hardware in place today is not equipped for network management. The industry needs to start addressing the need for more intelligent devices in the network. Higher levels of functionality are feasible through the use of imbedded microprocessor technology. Some estimates claim that less than 10% of the CATV coaxial plant in the U.S. is active two-way. Most recently, coaxial rebuilds and fiber optic plant are being built two-way active. These are prime areas for immediate deployment of network management technology. New video-on-demand and high speed data services, and network technologies such as SONET fiber rings and ATM will change the single "physical" network connection to the customer to a "logical" connection with "virtual" channels. Network management will be an essential tool in managing these more complex networks in the future.

Acknowledgments

I would like to pay tribute to the talented software developers at Rogers Engineering who have done the "impossible" with INMS.

Windows is a registered trademark of Microsoft Corporation
Openview is a registered trademark of Hewlett-Packard Company

AN OPTIMAL APPROACH TO A FULL-SERVICE BROADBAND COMMUNICATIONS NETWORK

Reem Safadi
General Instrument Corporation
GI Communications Division
2200 Byberry Road
Hatboro, PA 19040

Abstract

There has been a number of past R&D efforts to provide extended-service networks addressing data communications, voice communications, and interactive applications utilizing the cable distribution plant. This paper provides a brief survey of such efforts in terms of special network considerations with respect to a fully integrated system of the aforementioned services. A summary of considerations for an efficient and cost effective approach is provided along with the benefits that such an approach offers to the system operator.

Relevant Work

R&D efforts addressing implementation of non-traditional services on cable systems date as early as late 1970's. Some of these efforts did not come to fruition and others were implemented but lacked consumer demand.

Many such efforts served to be more than mere academic exercises. The larger subset focused on optimization of multiple access protocols (medium access control, MAC) for a cable based metropolitan area network, the other subset described the design of the overall system as in [3] and [7], for interactive video-text based applications. While optimization of medium access control is an important element to enabling non-traditional services, it is not the only area of consideration. Nonetheless a brief account is provided below for illustration. (A summary of the various media access schemes proposed for cable based distribution networks, and referenced herein, is provided in [6] and repeated in [2].)

A common goal to many of these efforts was to achieve performance optimization based on well identified and objective evaluation criteria. The two main figures of merit are network efficiency - amount of network capacity devoted to data (video, audio, and text) transmission, and access efficiency - the time between the queuing and the transmission of a packet. A careful examination of each approach, however, pointed to the following often common drawbacks:

1) Definition and modeling in many cases relied on invalid assumptions. In [4] the approach was optimized specifically for a file transfer application (support of interactive applications was secondary). It pointed out however the practical invalidity of approaches based on memoryless systems.

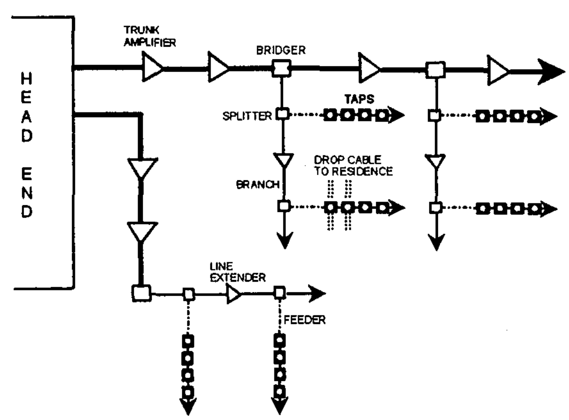


Figure 1. Placing substantial intelligence throughout the distribution plant as suggested by [5], [8], and [9] (at bridgers or splitters) is considered objectionable by Cable Operators.

In [5] the spread spectrum access scheme assumed persistent (next slot) retransmission of unsuccessful transmissions in order to be able to model system performance. However that assumption caused the system to move to a saturation point for throughput values considerably lower than 1. The effects of a random scheduling time were not provided. An entire system was described in [3] and [7] but, again, a common assumption was that traffic was generated by interactive applications only. None of the mentioned references thus far accounted for support of isochronous traffic (for video and voice based applications). An exception to that is found in [8]. The assumption there was that data packets were shorter in length than voice packets. While that facilitated non-blocking access for voice packets, the potential indefinite delay of data packets was not addressed.

2) Imposed certain requirements on the distribution plant that were not favorable from a cable operator's standpoint. In the system described in [8] two tuned notch filters and frequency converters were required at each regional subgroup boundary (neighborhood of less than a mile in diameter). This was an attempt to increase bandwidth reuse on pure coaxial tree & branch distribution networks. A control node was required in [5] consisting of a CPU with memory for buffering at each bridge to aggregate the upstream load from a given neighborhood and forward it to the headend. Similarly [9] proposed placing store and forward switches on trunks (and bridgers as an alternate approach) that performed routing, flow control, and error correction and detection.

3) Each approach grew increasingly complex when an attempt was made to extend it to support and address practical assumptions and requirements. In [5] buffering capacity at the control nodes was assumed to be infinite. Beyond a given threshold (saturation region, i.e., a given number of active users) the system became unstable in terms of successful transmissions. Moreover, spread spectrum processing gain, choice of FEC (imposed to guard against collided transmissions) and packet lengths posed conflicting optimization criteria with respect to one another.

A practical consideration such as the effect of round-trip propagation delay on feedback based multiple access protocols often required further refinement (and additional relative complexity) of the proposed schemes as in [2] and [11], where once the distance (from the headend to the most distant communicating node) was changed from the one mile assumption to a more typical value (35 miles), the protocol had to be augmented in order to maintain the same level of desired performance. In this example, impact of round-trip propagation delay is not the only practical factor leading to additional modification of the proposed scheme[11]. The impact of a non-ideal transmission channel (not a major issue in LANs since such networks are environmentally controlled, but is for a cable plant), is yet to be fully characterized. In addition, there has been a general tendency to depend on large upstream transmission rates (in excess of 5-10 Mbps) for performance improvement. Suffice it to state that a signal with such a rate has a considerable wider bandwidth that puts it at a disadvantage when taking into account upstream narrow-band interference.

Optimal-Design Requirements & Considerations

As discussed above, it is imperative to establish a full set of applicable requirements as a first step to defining the overall system. This in turn will have an impact on subsequent definition of system sub components. The system level design of a full service network has to address the following requirements:

1. Enabling a successful deployment of a given service by optimizing system performance, operation (including reliability) / administration, and cost.

2. Enabling incremental investments towards expanded or additional services.

3. Providing an end-to-end solution that is not optimized strictly towards a specific service or application. Such optimization is bound to incur penalties when considering the addition of other services. This is also applicable when considering monitoring and control (operations, administration and maintenance, OA&M) of the overall system.

While the three criteria are not mutually independent, the following addresses each individually.

1. Enabling a successful deployment: The success of a given service is determined by its appeal to the consumer and cost effectiveness of its offering (each governing generated profits). For many services, the appeal to the consumer remains an unknown, thereby potentially invalidating any assumptions technical or otherwise that may be taken into account in the overall design. Consider the previous example of multiple access protocol optimization. In many instances traffic modeling and usage pattern assumptions have been made or even mandated to justify the validity of a given approach. Relying on such assumptions may only serve as a theoretical justification but not a practical one. As illustrated, Medium Access Control is one of many areas that are impacted by the definition of the overall system.

The initial cost effectiveness of a design is mandatory. Relying on economies of scale to justify a more costly initial implementation is undoubtedly risky especially when consumer demand for a given service/application may not be fully characterized. Conducting trials is beneficial and in some instances is required when a given system design mandates a substantial initial investment. The latter is a characteristic of a revolutionary rather than an evolutionary approach (more below). An evolutionary approach allows the system operator to invest incrementally in expanding or adding a given service. As such the initial design has to have a self contained migratory path towards a more capable system. This is applicable to regional and local source equipment design, distribution plant design, and set-top design, etc.

2. Enabling incremental investment: As stated above, the initial design of an evolutionary system has to have a self contained migratory path towards a more capable system. For example, with respect to the distribution network, a gradual increase of return path capacity can be attained initially through extending fiber to a smaller number of homes/node, as in the Hybrid Fiber-Coax (HFC) architecture, illustrated in [12], followed by frequency block up conversion for sub-split systems when it becomes necessary to do so (i.e., when maximum capacity threshold has been reached). With respect to the set-top design, additional capabilities to support enhanced applications can be realized through add-on modules (e.g., increased processing power, memory and graphics capabilities). At a later date secondary modules may be upgraded or integrated into the base unit (when the use of such modules prevail). As for the design of the source equipment, it has to be modular and scalable in order not to dictate an unwanted initial investment while allowing incremental investment as economically deemed desirable. (It is assumed that the necessary switching equipment already meets this criterion). Figure 2 illustrates.

Enabling multiple services and ensuring proper allocation of resources to such services is a key consideration that the system operator has to address. Depending on how the system is designed, system operators may be faced with misplaced choices and options (or lack of). For example, systems that allow dynamic bandwidth allocation per service (and per user) have the flexibility and efficiency of BW reuse that is not present in BW dedicated systems. Additionally, systems that do not allow a transparent transport of information associated with a given service are likely to unnecessarily increase the complexity of the system. This is particularly true for existing services that evolved utilizing different connectivity infrastructures, e.g. telephony (POTS), PCS, data communications, etc. However this approach (transparent transport) if taken to an extreme may also pose unwanted disadvantages: if a given service is to be ported intact onto a Hybrid Fiber-Coax (HFC) infrastructure, such porting while suitable for the associated service may be entirely unsuitable for another, leaving the system operator with significant resource and operational issues. The previous example of multiple access protocols can be extended towards the currently defined MAC schemes such as Ethernet (CSMA/CD), Token Bus, or Token Ring (IEEE802.3, .4, .5 respectively), as an additional illustration. While there are applicable features in some (as is the case for a CDMA scheme), none were intended for a system utilizing a HFC infrastructure with given operational end-node characteristics. The traditional MAC schemes are also unsuitable for applications requiring guaranteed bandwidth. Even when considering DQDB, IEEE802.6, one finds the operational requirements widely varying from those governing the system at hand. All of this however does not imply that isochronous and asynchronous data can not be efficiently transported over the HFC distribution plant. What is important to realize is that there are boundaries that need to be established with respect to transparent information delivery. In the case of telephony, it would not be economically nor technically advantageous to redesign the Host Digital Terminal and switches to recognize new protocols nor reinvent the interfaces, where applicable (consider the effort invested by Bellcore to produce technical requirements specifications for telephony systems). Similarly for data communications, it would not be desirable to impact the design of IP routers, X.25 or ATM switches as an example. Establishing such boundaries becomes a relatively simple task when communication models are applied that separate the various functional layers. Which layers to preserve and which can be replaced in order to optimize overall system design becomes rather self-evident.

Architecturally, there is a common denominator to enabling additional services on the cable infrastructure. PCS, POTS and video telephony serve as an example. Each may have a suitable interface from an origination/destination point and a corresponding interface to the controlling (switching) point. As long as the interfaces to equipment residing at the demarcation points is kept intact, a single architecture can be supported.. Preservation of a given service platform outside the demarcation points is desirable since it provides the system operator the freedom to select individual platforms with the most competitive offerings. The same applies architecturally to interactive or Video on Demand applications. The later, however, is an example application where end-to-end system definition (of all applicable functional layers) has to be provided since it is a new application.

3. Providing an end-to-end solution: A fully integrated system implies a fully integrated Operations, Maintenance, Administration, and Provisioning subsystem (OAM&P) covering: billing, network, resource and service management, access control, remote monitoring and control, etc. for the different types of services: broadcast video, NVOD, VOD, POTS, video telephony, interactive applications, telecomputing, etc. This however is a long term objective especially when integrating in an overlay manner two or more existing systems of distinct service types (to achieve near term objectives).

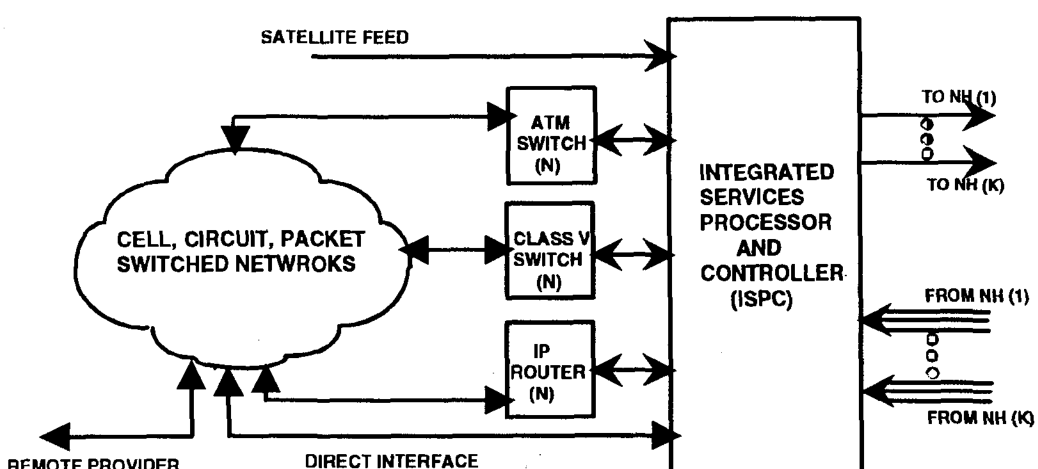


Figure 2a. Example of Scalable Network Stages External to the ISPC. (Some of the elements in the intermediate stage, to the left of ISPC, may be combined via ATM).

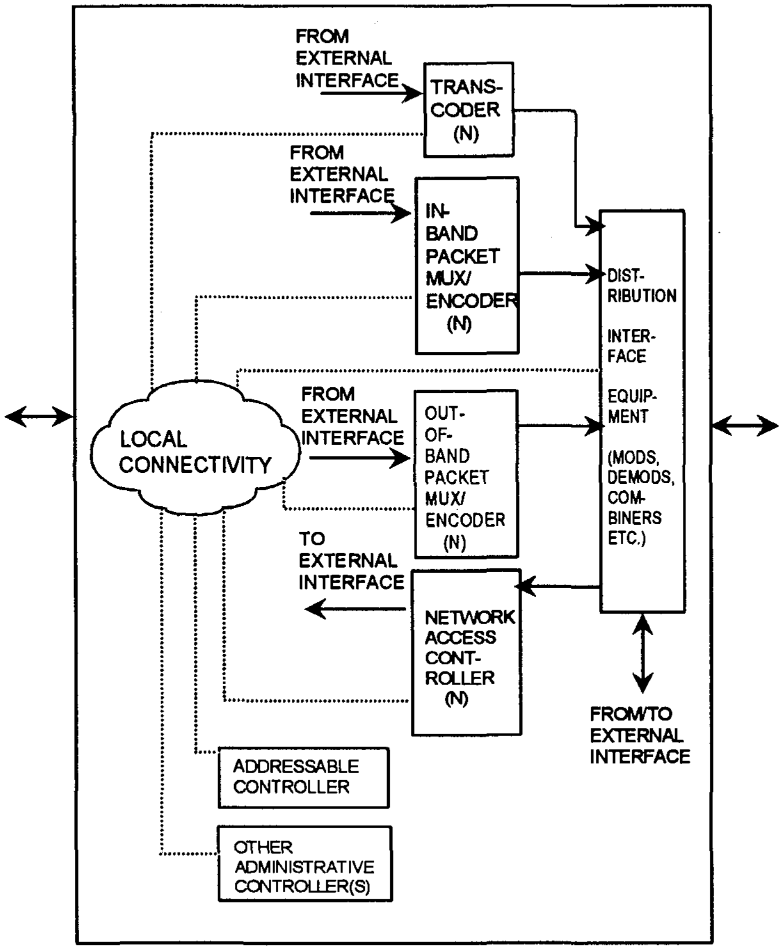


Figure 2b. Example of ISPC Functional Blocks.
Notes:
1. Some functional blocks may be remotely located.
2. Future (conditional) ubiquitous use of ATM may facilitate combining certain functional elements.
3. See [12] for example illustration of subscriber side. (There ISPC is ISHDT.)

A homogenous operational system offers advantages associated with a single set of skills required to administer the system as well as understanding the underlying OAM&P processes. However fully-developed enabling technologies for a homogeneous system (not just the OAM&P subsystem) are a few years away and thus it is best to rely on proven approaches rather than risk offering a system that can be considered experimental at best. What is learned from experimental systems however can be used in further advancing existing systems and as such should not be disregarded.

Acknowledgment

The writer would like to thank the countless individuals especially those who prefered anonymity for their strong interest and ongoing support of this effort and all other related activities.

References

[1] IEEE 802.6-1990," Local and Metropolitan Area Networks, Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network," IEEE Inc., July 1991.

[2] G. Campbell, R. Khasawneh, "DQRAP: A Proposed Standard For the 'Last Mile' of An IEEE802.6 MAN, " Dept. of Comp. Science, Illinois Institute of Technology, Submittal to IEEE802.6 Committee (IEEE P802.6-93/13).

[3] M.L. Ellis et al., "INDAX: An Operational Interactive Cabletext System," IEEE J. on Selected Areas in Communications, vol. SAC-1, no. 2, February 1983.

[4] D.C. Feldmier, "An Efficient Access Scheme for A CATV-Based High-Speed Packet-Switching Metropolitan Area Network", Lab. for Computer Science, Massachusetts Institute of Technology, IEEE INFOCOM '88, March 1988, New Orleans, LA.

[5] H.M.Hafez et al, "Spread Spectrum Access in Two Hop CATV Data Networks," IEEE J. on Selected Areas in Communications, vol. SAC-3, no.2, March 1985.

[6] A.I. Krashmer, J. N. Thomas, "Computer Networking on Cable TV Plants," IEEE Network, November 1992.

[7] R. McNamara et al., "CATV Communication System", General Instrument, U.S. patent 4533948, August 1985.

[8] N.F. Maxemchuck, A. Netravali, "Voice and Data on a CATV Network," IEEE J. on Slected Areas in Communications, vol. SAC-3, no. 2, March 1985.

[9] T.N. Saadawi, M. Schwartz, "Distributed Switching for Data Transmission Over Two-Way CATV," IEEE J. on Selected Areas in Communications, vol. SAC-3, no. 2, March 1985.

[10] A. S. Taylor, "Characterization of Cable TV Networks as the Transmission Media for Data," IEEE J. on Selected Areas in Communications, vol. SAC-3, March 1985.

[11] W. Xu, G. Campbell, "A Distributed Queuing Random Access Protocol for a Broadcast Channel," Dept. of Comp. Science, Illinois Institute of Technology, Submittal IEEE 802.6 Committee (IEEE P802.6-93/13).

[12] C. Eldering et al, "Engineering Requirements for Hybrid Fiber-Coax Telephony Systems," Proceedings of the 1994 NCTA Convention & Exposition, (New Orleans, May 22-25, 1994).

Analysis of Cable System Digital Transmission Characteristics

RICHARD S. PRODAN, PH.D.—VICE PRESIDENT, ENGINEERING
MAJID CHELEHMAL, PH.D.—DIGITAL SYSTEMS ENGINEER
THOMAS H. WILLIAMS—TRANSMISSION ENGINEERING SPECIALIST
CRAIG M. CHAMBERLAIN—ELECTRICAL ENGINEER

CABLE TELEVISION LABORATORIES, INC.

ABSTRACT

CableLabs has performed an evaluation of cable channel characteristics to statistically quantify the cable environment for high speed, band-limited transmission of digital information in a study conducted in cooperation with several of our member cable system operators. Approximately three hundred subscriber home sites in twenty cable systems were measured and analyzed over a wide range of channel frequencies.

The range of channel characteristics, degree of impairment, and relative frequency of occurrence in a statistical distribution of both the cable plant and the subscriber home wiring is presented. These results provide the purveyors of digital cable modem equipment with valuable design information which can be used to determine the receiver interference mitigation techniques required, the complexity and performance characteristics of a demodulator design, and the relative percentage of cable subscribers who can satisfactorily receive a digital transmission utilizing a specific demodulator implementation.

Various measurements of relevant channel characteristics provide useful information needed for both cost and performance optimization of the digital demodulator, as well as the cable system transmission equipment.

INTRODUCTION

The range of channel characteristics, degree of impairment, and relative frequency of occurrence in a statistical distribution of both the cable plant and the subscriber home wiring is needed to determine the receiver interference mitigation techniques required, the complexity and performance characteristics of a demodulator design, and the relative percentage of cable subscribers who can satisfactorily receive a digital transmission utilizing a specific demodulator implementation. Thermal noise level, carrier power level, channel frequency response (or alternatively impulse response) and interference due to ingress characterized by a statistical distribution of stationary disturbances in the cable environment provide useful information needed for both cost and performance optimization of the digital demodulator, as well as the cable system transmission equipment.

CableLabs has completed a digital channel characterization project that quantifies these stationary impairments on a multiplicity of cable systems across the U.S. and Canada. A number of geographically separated cable systems of varying size, age, and technology were investigated. Within each cable system, multiple distribution tap and home locations on as diverse a set of system branches as possible were measured. This data is available to provide the purveyors of digital cable modem equipment with valuable design information.

METHODOLOGY

The Digital Transmission Characterization project conducted field measurements of both cable television transmission channel and subscriber home distribution characteristics. The field measurements were performed using computer automated instrumentation installed in a vehicle integrated into a self-powered mobile laboratory testing environment. The data acquired at each location within each cable system was compiled into a database for generation of a statistical distribution of cable plant and subscriber home wiring transmission characteristics including frequency response, shielding effectiveness, ingress, thermal noise level, carrier power level, time domain reflection response, and composite triple beat intermodulation distortion level. This measured data can be linked to a description of relevant cable system design parameters obtained and catalogued for each cable system prior to measurement.

At each cable system subscriber location tested, a series of ten automated measurements were performed for assessing these cable system performance characteristics. All measurements were performed in the upper frequency region of each cable system where spectrum was available but not utilized for conventional analog signals (usually in a recently upgraded system prior to deploying expanded channels). These measurements are as follows:

1) Headend to Home Outlet Amplitude Flatness and Group Delay
2) Headend to Tap Amplitude Flatness and Group Delay
3) Cable System Signal Spectrum at the Tap
4) Cable System Noise Spectrum at the Tap
5) Cable System Ingress Spectrum at the Tap
6) Composite Triple Beat Level at the Tap
7) Local FM Field Strength
8) Home Wiring Shielding Effectiveness
9) Home Wiring Amplitude Flatness and Group Delay
10) Home Wiring Time Domain Reflection Response

An easy to use menu driven software program for instrument control and data acquisition had been developed to control various instruments and catalogue measurement data via a GPIB board connected to a DOS computer. All ten measurements are available in the measure menu. In addition, instrument states and calibration data are saved and restored to the instrument.

Instrument control parameters were provided in input script files that were easily edited on site for local cable system parameters. Because of the diversity of each cable plant, this script file scheme allows changing a given measurement by editing the input script files at each measurement site. This greatly simplifies data collection in the field which is much less controllable and predictable than laboratory measurement. In addition, no changes to the control program are necessary from one site to another.

After starting a measurement, instrument settings can be readjusted and data re-acquired. This process can be repeated until valid measurement data is obtained. The software has great flexibility by allowing complete control over the measurement process. Measurement data is stored into output data files along with instrument settings in the header information in the computer database for later analysis.

RESULTS

A major cause of digital signal degradation at the receiver results from corruption by intersymbol interference caused by a deviation from the ideal (i.e., Nyquist) channel response required for high speed data transmission over band-limited channels. The intersymbol interference introduces time dispersion, causing each received data symbol pulse response to overlap with many adjacent symbols in a destructive manner. The deviation of the channel characteristics from the ideal are caused by signal reflections due to impedance discontinuities in the transmission path(s).

Examples of amplitude flatness and group delay responses at one subscriber test location are shown in Figures 1 and 2 respectively. The degradation due to intersymbol interference is most easily assessed in the impulse response as reduced amplitude, delayed versions of the main impulse. The frequency response data is inverse Fourier transformed from the previous measured data into the impulse response of Figure 3. An echo approximately 24 dB down delayed 100 nanoseconds is apparent in the figure.

As a very large number of responses were measured, a statistical representation of the impulse responses is calculated using a two-dimensional histogram depicting the relative frequency (or probability) of a given echo amplitude attenuation at a given delay time. This result for the cascaded cable plant plus inside home wiring is shown in Figure 4. The echoes of more significant amplitude (20 dB attenuation or less) occur at delays of less than 500 nanoseconds. Longer delay echoes are more attenuated. More densely packed modulation formats with a larger number of states are more susceptible to echoes; hence even very weak echoes may be significant as data capacity is increased.

By projecting the data onto either the amplitude or delay axis of Figure 4, a relative frequency of each echo parameter is obtained independently. The distribution of echo amplitudes is shown in Figure 5. The distribution of echo delay times is shown in Figure 7. A useful characterization of the measured echo distributions would depict the total percentage of sites with either echo amplitudes or delays larger than a given value. This is available from the cumulative distributions obtained by integration of the relative frequency distributions. The cumulative relative frequency of echo amplitudes is shown in Figure 6. The cumulative relative frequency of echo delay times is shown in Figure 8. One may ascertain by inspection of the cumulative distributions shown in Figures 6 and 8 that 90 percent of the home sites measured had echo amplitudes less than -27 dB below the desired signal and delay times less than 570 nanoseconds.

The shielding effectiveness of the home wiring was measured in the FM radio frequency band. The strength of the local FM field was measured with a dipole antenna at the test site. The home wiring was disconnected from the cable plant and terminated. The terminated home wiring replaced the antenna and the FM field strength measured again. The shielding effectiveness is calculated as the ratio of these two measurements averaged over all the local FM stations received at that location. The relative and cumulative distributions of home wiring shielding effectiveness are shown in Figures 9 and 10 respectively. Although the average as well as the median shielding over all sites is greater than 58 dB, a small but significant number of homes (about 5 percent) provided less than 36 dB shielding. These homes are much more susceptible to ingress from external RF signals.

The carrier-to-noise ratio relating the received signal energy to the thermal (additive white Gaussian) noise level is a fundamental performance parameter in determining the received bit error rate. The carrier power relative and cumulative distributions are shown in Figures 11 and 12 respectively. The noise power relative and cumulative distributions are shown in Figures 13 and 14 respectively. The carrier-to-noise ratio relative and cumulative distributions are shown in Figures 15 and 16 respectively.

The spurious power measured indicates the total non-thermal peak noise power due to intermodulation components and the ingress of external RF signals into the cable system. The spurious power relative and cumulative distributions are shown in Figures 17 and 18 respectively.

The specific spurious component generated as the composite triple beat (CTB) located 12 MHz above the last analog channel was measured separately. The CTB relative and cumulative distributions are shown in Figures 19 and 20 respectively.

A summary derived from the cumulative distributions of all the previously described impairments tabulates the maximum impairment values for the 50 (median), 90, 95, and 99 percent relative number of test sites measured. The results for microreflection impairments are shown in Table 1. The results for noise and interference impairments are shown in Table 2.

MicroReflection Impairments	**50%**	**90%**	**95%**	**99%**
Delay (nano sec.)	230	570	730	1280
Amplitude (dB)	-37	-27	-24	-19
System Delay (nsec.)	230	640	860	1520
System Amplitude (dB)	-37	-29	-26	-19
Home Wiring Delay (nano sec.) 50 - 200 MHz.	140	350	430	640
Home Wiring Amplitude (dB) 50 - 200 MHz.	-37	-26	-22	-16
Home Wiring Delay (nano sec.) 200 - 560 MHz.	130	330	440	1250
Home Wiring Amplitude (dB) 200 - 560 MHz.	-37	-28	-25	-19

Table 1. MicroReflection Delay and Amplitude Impairments Summary.

Noise / Interference Impairments	**Ave.**	**50%**	**90%**	**95%**	**99%**
Carrier / Noise (dB)	48	43	35	33	25
Carrier Power (dBm)	-32	-37	-48	-53	-60
Noise Power (dBm) in a 6 MHz. Bandwidth	-72	-82	-72	-68	-60
Spurious Power (dBm) in a 6 MHz. Bandwidth	-68	-78	-66	-62	-56
CTB Power (dBm) in a 30 kHz. Bandwidth	-93	-107	-90	-86	-80
Home Wiring Shielding (dB) in FM Band	67	58	42	36	27

Table 2. Noise / Interference Impairments Summary.

CONCLUSION

A statistical characterization of cable system impairments to digital transmission has been presented. Three hundred home sites in twenty diverse cable systems have been measured and analyzed. The relative percentage of the total measured population exceeding a given level of impairment has been derived from a statistically significant number of cable subscriber homes.

The range of cable channel characteristics, degree of impairment, and relative frequency of occurrence in a statistical distribution of both the cable plant and the subscriber home wiring presented may be used to determine the receiver interference mitigation techniques required, the complexity and performance characteristics of a demodulator design, and the relative percentage of cable subscribers who can satisfactorily receive a digital transmission utilizing a specific demodulator implementation. This data provides the manufacturers of digital cable modem equipment with valuable design information as well as insight into the existing cable environment.

ACKNOWLEDGEMENT

The authors wish to thank the following individuals without whose help this project would not have been successful:

Stuart Gibson of GTS Enterprises, test engineer and road warrior extraordinaire, whose diligent field test operation and data collection enhanced the quality and value of this premiere study.

David Eng, Frank Wimler, and Moisey Goldshteyn at CableLabs for their continuous efforts in the preparation and support of the field test van construction and operation.

Jonathan Patrizio and Ewan Shepherd of Hewlett-Packard Queensferry Microwave Division, for providing hardware and software expertise and support for the key instruments utilized in the field measurements.

Finally, we wish to extend our appreciation to the CableLabs member companies and their personnel who participated in the digital transmission characterization, providing their cable systems and their support for this valuable study.

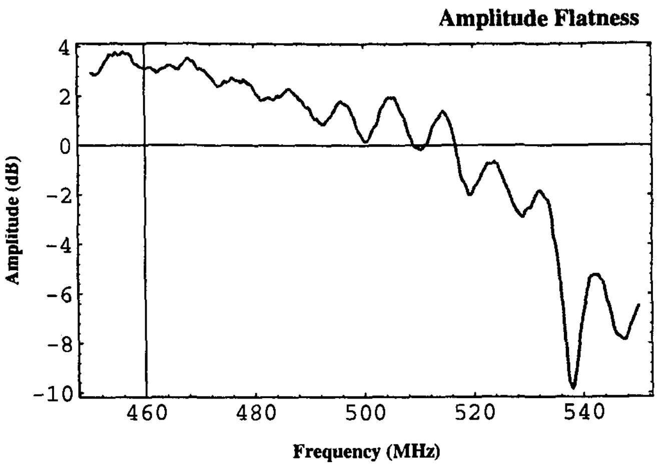


Figure 1

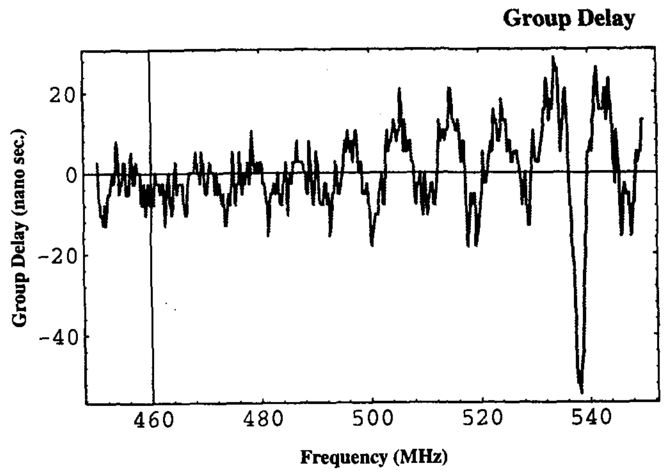


Figure 2

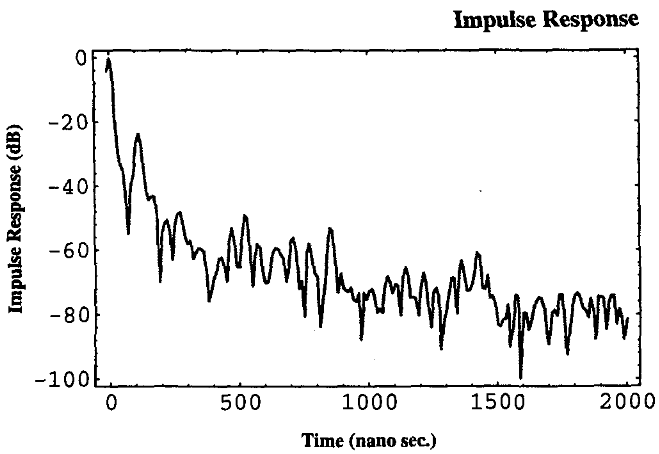


Figure 3

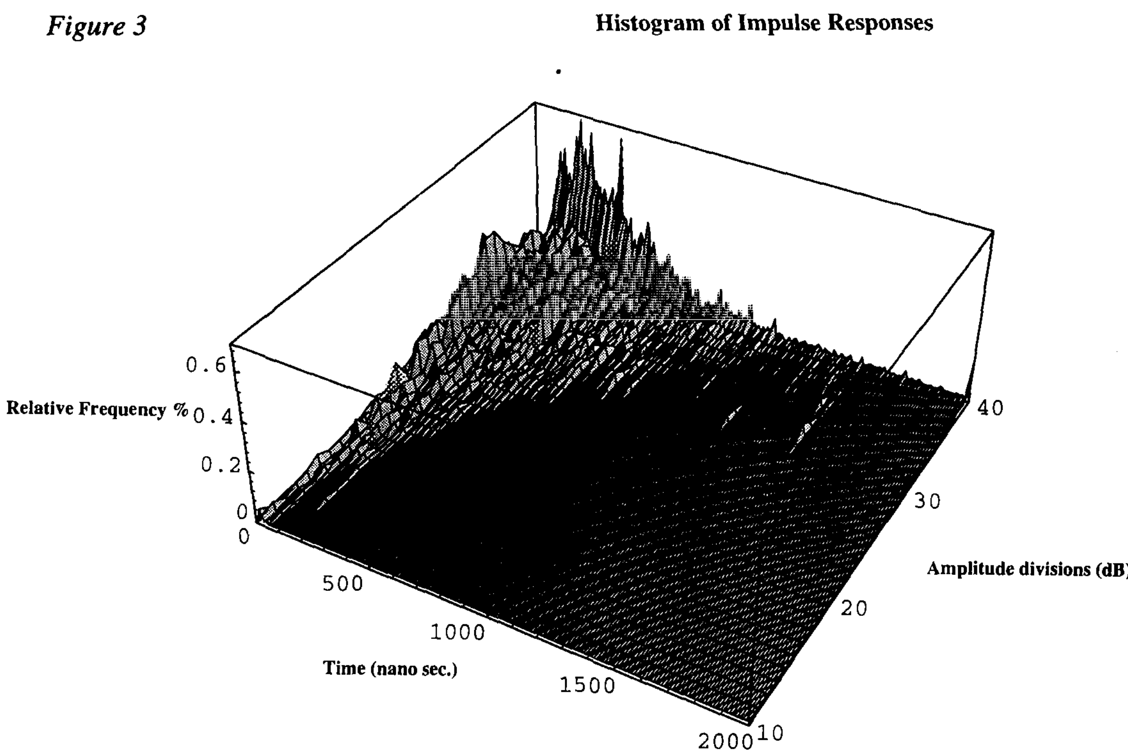


Figure 4

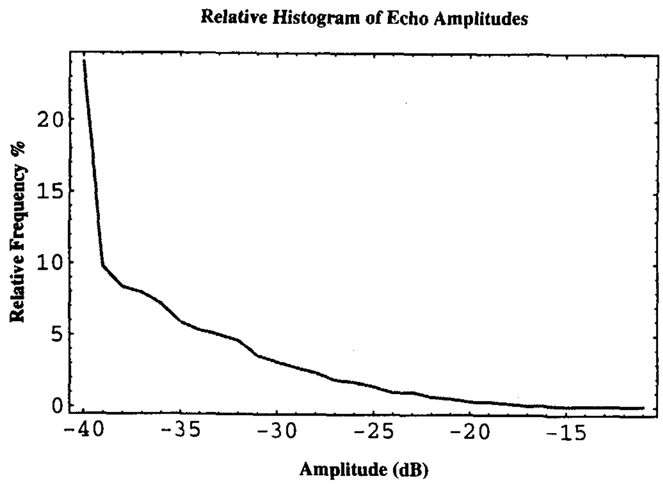


Figure 5

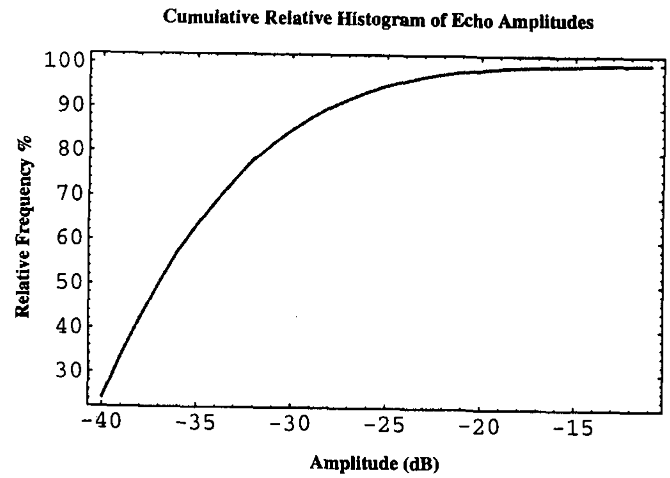


Figure 6

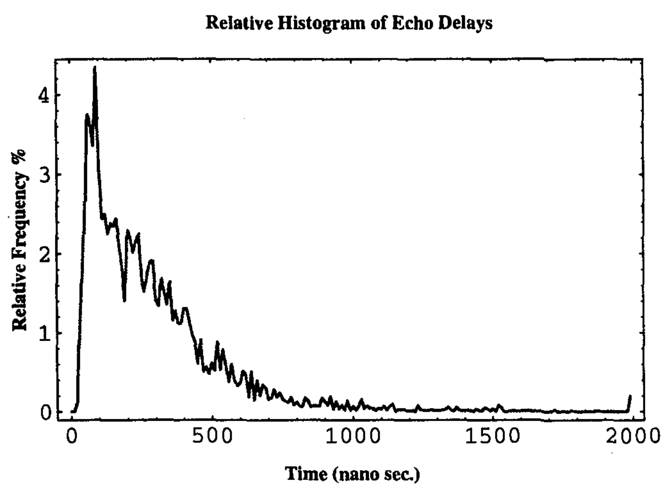


Figure 7

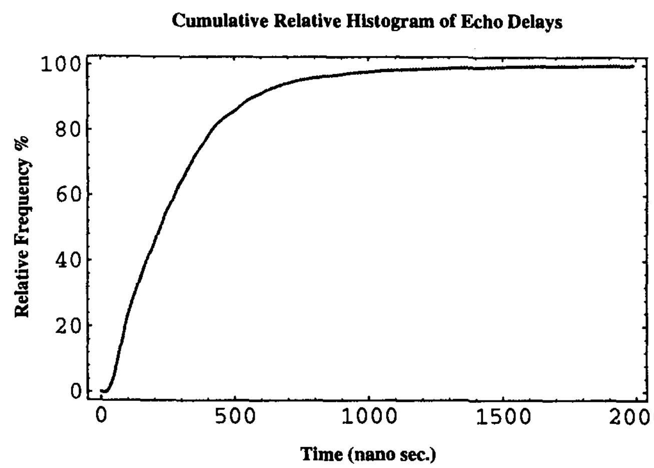


Figure 8

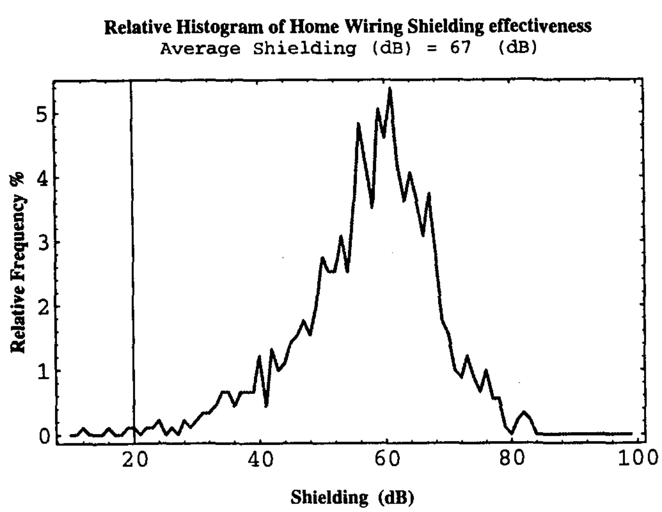


Figure 9

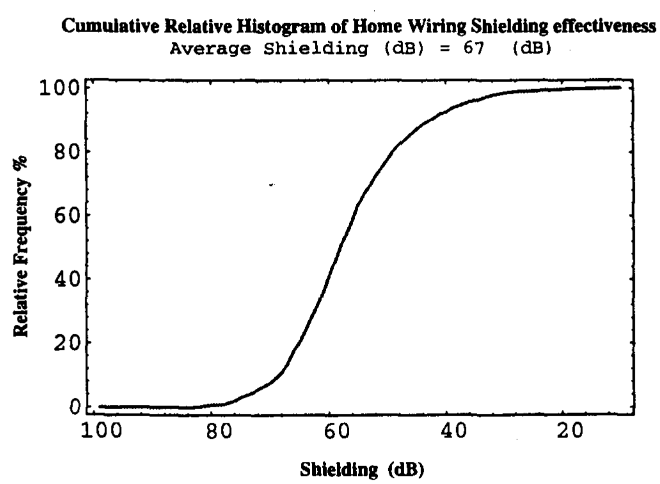


Figure 10

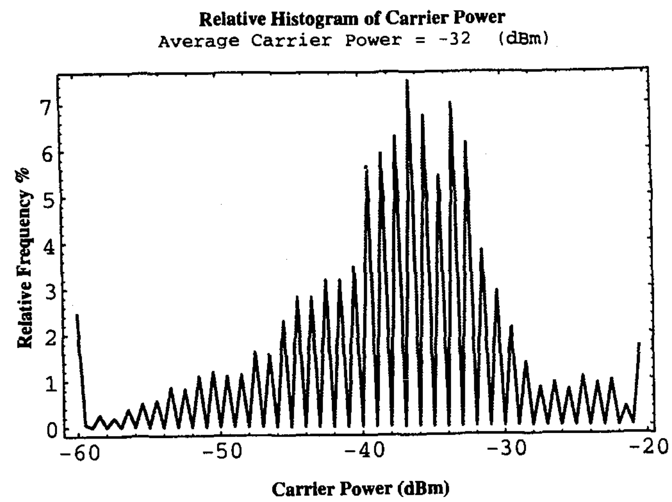


Figure 11

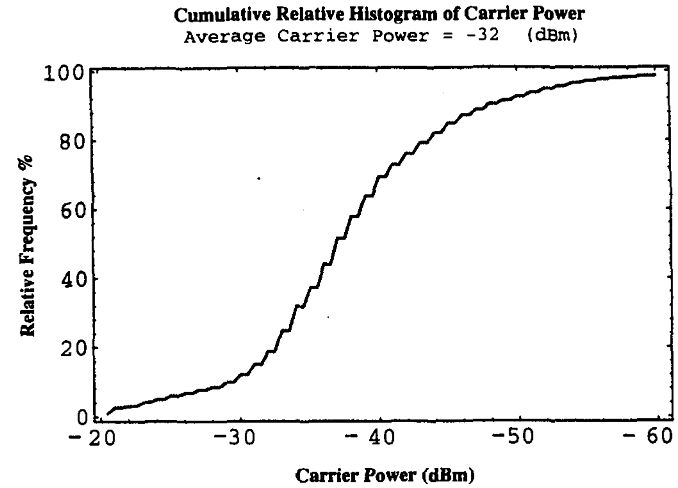


Figure 12

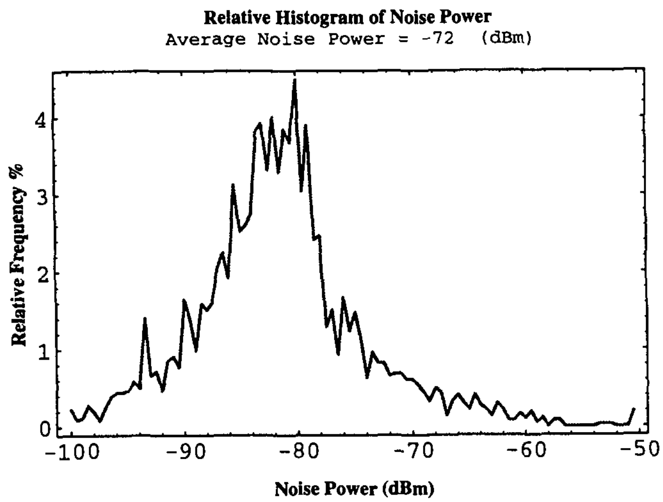


Figure 13

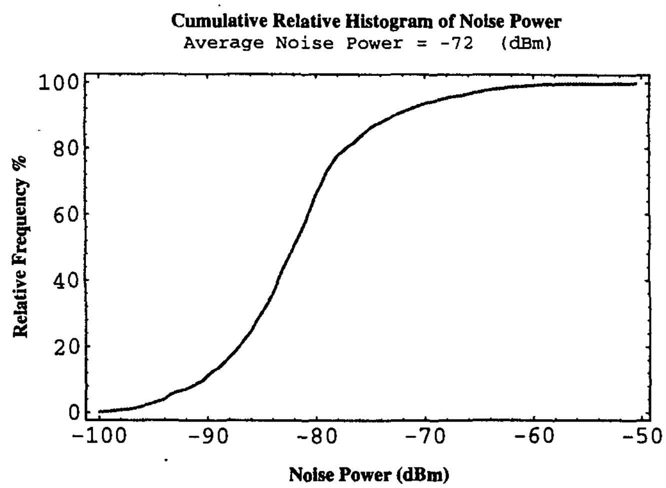


Figure 14

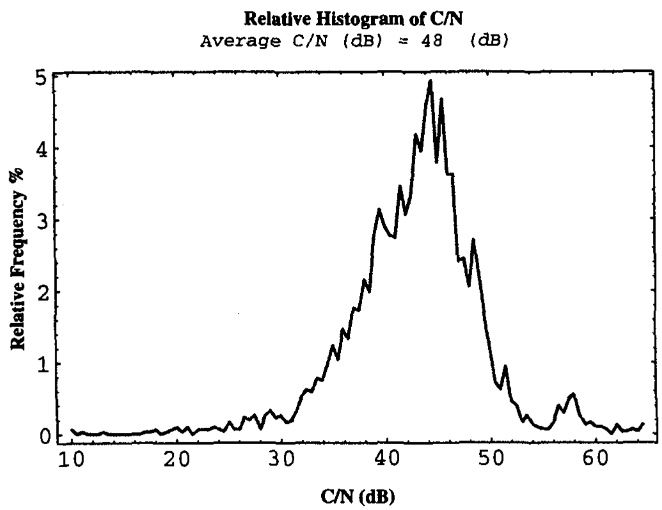


Figure 15

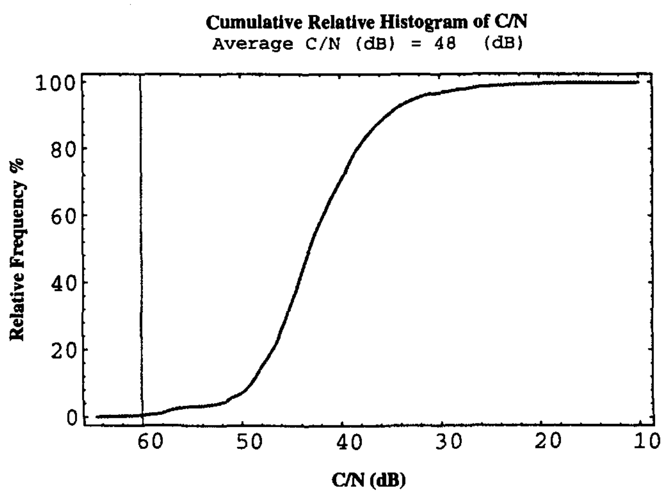


Figure 16

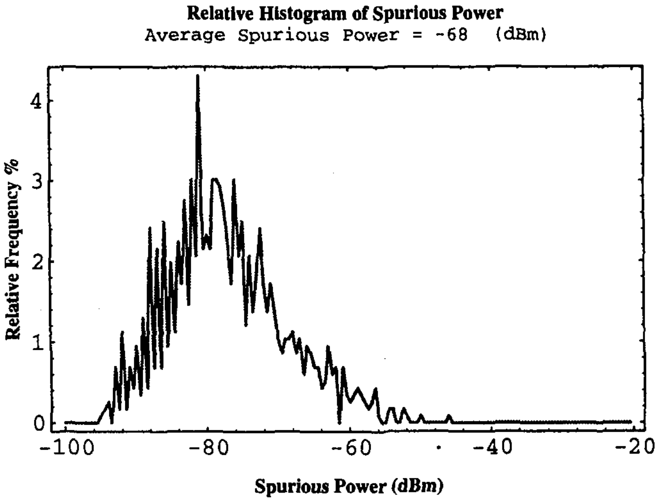


Figure 17

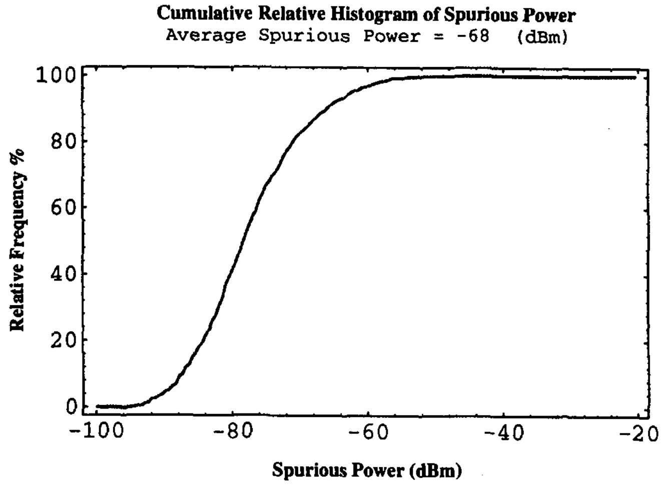


Figure 18

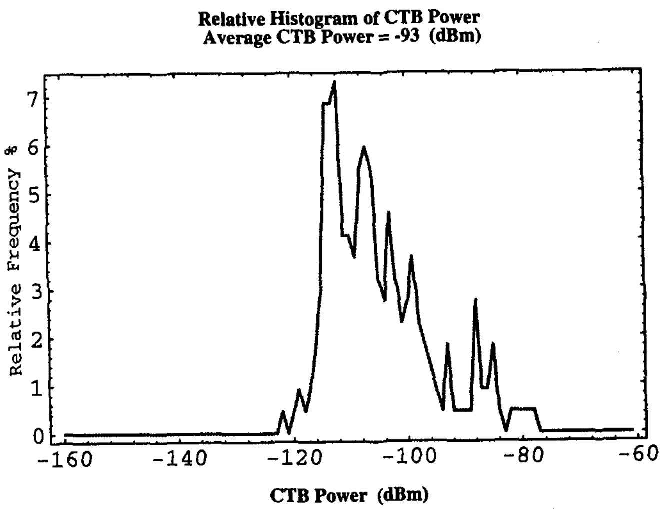


Figure 19

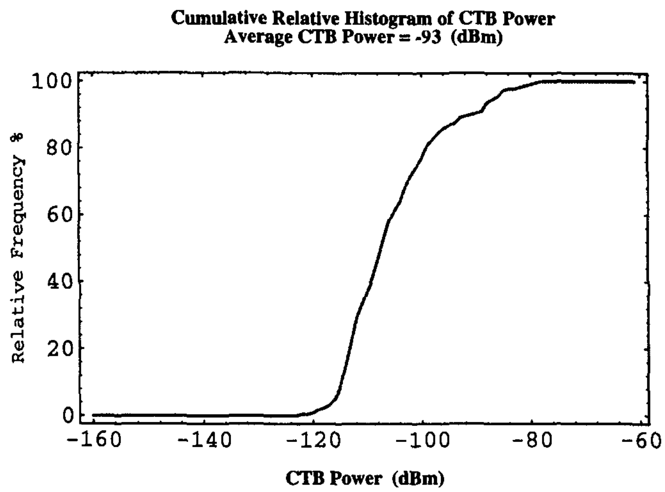


Figure 20

Architectures for Video Servers

Manu Thapar and Bill Koerner
Hewlett Packard
1501 Page Mill Road, Palo Alto, CA 94304
thapar@hpl.hp.com

Introduction

In the past few years a number of important activities and technology advancements that will influence the future of the industry have emerged. Cable companies have been laying out fiber cable providing enormous amounts of bandwidth capacity to residential neighborhoods. Memory storage is quickly becoming cheaper. Compression and decompression techniques are becoming more well understood and standards such as MPEG are gaining significance. The above factors now enable the possibility of various new services that can be provided to residences. One such service is video on demand (VOD). VOD will allow customers to choose the movies they want to view. Instead of driving to a video store to rent a tape, they will be able to choose a movie that is delivered to their home over a high speed digital network. VOD has the potential of being a multi-billion dollar business in a few years.

The essence of video on demand is that a number of digitized, compressed movies are stored in a video server and transmitted over a distribution network to viewers in their homes. The overall architecture is shown in figure 1. There are a number of possibilities for the distribution network [1]. The bandwidth into the residence is limited by the final interface at the home. Due to cost constraints, the equipment in the home is assumed to be minimal, comprising little more than decompression, decoding, demodulation, and display on a TV. The movie is played back in real time over the network and is not downloaded into storage in the home. The system cost is dominated by the cost of storage, and it is less expensive to concentrate the storage in the server where only enough storage for the number of simultaneous viewers is needed, rather than having idle storage in the homes of all non-viewing subscribers.

This paper describes the key issues in the design of video servers. We provide an overview of some of the video server architectures currently being explored. We also briefly overview the potential services, service requirements, features, and communication facilities required for the whole infrastructure. We highlight the design issues that to us are the interesting research problems and outline some of our views on their solutions.

Video Server Requirements

There are a number of services in the area of entertainment and information/education that can be conceived. These include video on demand, news services, shopping services and educational information, to name a few. Of all these services motivating the development of a video server, the most immediate is the delivery of movies to home viewers. Cable companies want to expand their revenues by providing new services as well as providing telecommunication services. Already there are cable companies that are utilizing the large bandwidth of cable to provide internet services to the home. Not to be left out, and experiencing a threat to their revenues, the telephone companies are planning new integrated services. The telephone companies, cable companies, and third parties see an opportunity simply to displace a major part of the revenue collected by video stores for movie rentals. There are a number of other service opportunities. Hotels will offer movies on demand to their guests plus on-demand videos of hotel facilities, tourist attractions, and other information. Companies large and small will offer training programs and company news

Architectures for Video Servers

Manu Thapar and Bill Koerner
Hewlett Packard
1501 Page Mill Road, Palo Alto, CA 94304
thapar@hpl.hp.com

Introduction

In the past few years a number of important activities and technology advancements that will influence the future of the industry have emerged. Cable companies have been laying out fiber cable providing enormous amounts of bandwidth capacity to residential neighborhoods. Memory storage is quickly becoming cheaper. Compression and decompression techniques are becoming more well understood and standards such as MPEG are gaining significance. The above factors now enable the possibility of various new services that can be provided to residences. One such service is video on demand (VOD). VOD will allow customers to choose the movies they want to view. Instead of driving to a video store to rent a tape, they will be able to choose a movie that is delivered to their home over a high speed digital network. VOD has the potential of being a multi-billion dollar business in a few years.

The essence of video on demand is that a number of digitized, compressed movies are stored in a video server and transmitted over a distribution network to viewers in their homes. The overall architecture is shown in figure 1. There are a number of possibilities for the distribution network [1]. The bandwidth into the residence is limited by the final interface at the home. Due to cost constraints, the equipment in the home is assumed to be minimal, comprising little more than decompression, decoding, demodulation, and display on a TV. The movie is played back in real time over the network and is not downloaded into storage in the home. The system cost is dominated by the cost of storage, and it is less expensive to concentrate the storage in the server where only enough storage for the number of simultaneous viewers is needed, rather than having idle storage in the homes of all non-viewing subscribers.

This paper describes the key issues in the design of video servers. We provide an overview of some of the video server architectures currently being explored. We also briefly overview the potential services, service requirements, features, and communication facilities required for the whole infrastructure. We highlight the design issues that to us are the interesting research problems and outline some of our views on their solutions.

Video Server Requirements

There are a number of services in the area of entertainment and information/education that can be conceived. These include video on demand, news services, shopping services and educational information, to name a few. Of all these services motivating the development of a video server, the most immediate is the delivery of movies to home viewers. Cable companies want to expand their revenues by providing new services as well as providing telecommunication services. Already there are cable companies that are utilizing the large bandwidth of cable to provide internet services to the home. Not to be left out, and experiencing a threat to their revenues, the telephone companies are planning new integrated services. The telephone companies, cable companies, and third parties see an opportunity simply to displace a major part of the revenue collected by video stores for movie rentals. There are a number of other service opportunities. Hotels will offer movies on demand to their guests plus on-demand videos of hotel facilities, tourist attractions, and other information. Companies large and small will offer training programs and company news to employees at their individual convenience. Real estate agents will allow customers to view potential acquisitions and retail merchants will be able to provide information about products more effectively.

The specific requirements for a video server are not totally defined, since commercial trials are only about to begin. We expect to see small trial systems with 50-100 titles serving 10-100 subscribers. Pilot roll outs will probably have 100-1000 titles and 200-3000 subscribers. And real services, when fully configured, will likely require 10000 titles per server (with subscriber access to a total of over 100000 titles on a distributed array of servers) and 10000 subscribers per server. These numbers are also likely to vary with the area served and the service provider. Because of the wide range of these numbers, we feel it important to make the server scalable in at least the dimensions of number of viewers and number of titles, so that a common design and set of components can be the basis of a whole family of servers.

Another factor that influences the design of the server is the set of features that it must provide. Reliability, quality and privacy (security) are some of the key features. Since the video server provides a service that is can be accessed by thousands of viewers and each call last around 2 hours, reliability of the server is of key importance. We use redundancy to increase the reliability of the architecture. The quality of the video stream, which translates into jitter free bitrate, is another important feature. We assume that movies are encoded digitally using MPEG (the standard from the Moving Picture Experts Group) at rates that begin at around 1.5 Mb/s but could also include rates such as 3, 4.5, 6 or 8 Mb/s. Ideally, a server would allow subscribers to select their quality level, so it would have to play movies back at a variety of bitrates. In some cases, the distribution network bandwidth and settop specifications will determine an upper bound to the bitrate that can be delivered. The storage requirement for a movie is proportional to the bitrate of its encoding; a minimal bitrate of 1.5 Mb/s results in a 2-hour movie taking up 1.3 giga bytes of storage. Enough buffering is included in the system to allow a continuous stream of data to be delivered to the network. Privacy and security are provided by encryption of the digital data stream. The encryption key is changed periodically to eliminate piracy of the data, so that no one views something they are not paying for. Another feature is the degree of interactivity. To explain this, we describe two varieties of services.

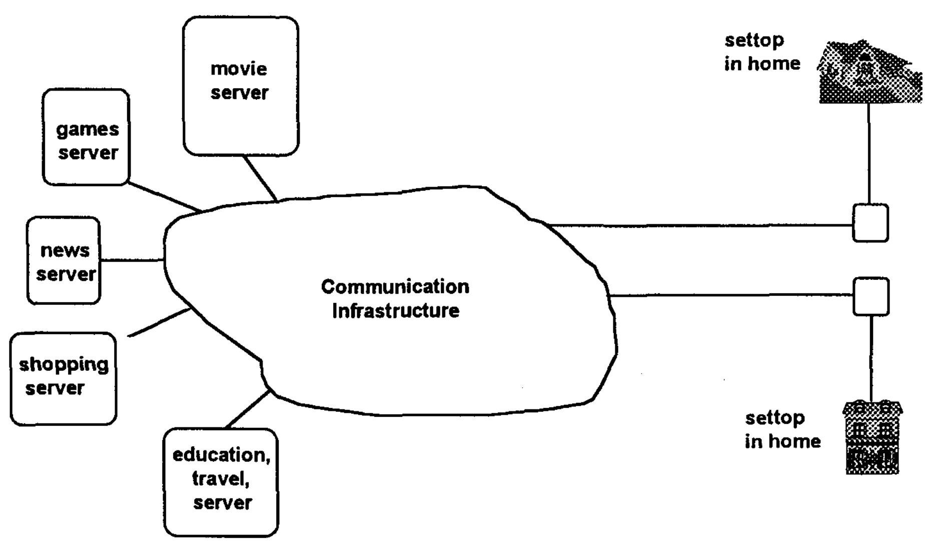


Figure 1: VOD Infrastructure

One potential service is near video on demand (NVOD). In NVOD, a popular movie is repeated at intervals of, say, 5 minutes, and a viewer can simply choose to watch continuously or jump between streams. The movie is broadcast to all the viewers, hence a large number of subscribers can be served by the same stream. This delivery mechanism is good for very popular movies (hot movies). The server needs to transmit only a single stream to satisfy multiple requests. For example, a two hour movie transmitted at an interval of 5 minutes requires 24 streams to serve an unlimited amount of viewers. At the high end, we have fully interactive video on demand (IVOD), sometimes called true video on demand, in which the viewer has complete virtual VCR functions of pause, fast forward, and rewind. Whereas NVOD allows a considerable degree of broadcasting and therefore a very large number of viewers per copy, IVOD is restricted to one viewer per stream. It is therefore the most expensive in terms of storage and network bandwidth per viewer.

Network Delivery of Video

There are a number of delivery mechanisms possible. We provide a brief overview here. Further details are available in [1]. Currently there are two main physical communication channels entering homes. These are the twisted pair wires used for telephones and the coax cable used for the cable services. Asymmetric digital subscriber line (ADSL) can be used to transmit digital data over twisted pair. ADSL I has an upper bound of 1.5 Mb/s for transmission of data. ADSL II improves the upper bound to 6.4 Mb/s. Delivery of information over twisted pair is thus limited by the data rate and cannot support very large data rates. Coax cable has a bandwidth of around 1GHz. It is split up into 6MHz analog channels. Using 64 QAM (quadrature amplitude modulation), 25Mb/s data rates can be transmitted in each of the 6MHz channels. Typically VOD related information will be carried in the frequency spectrum above 450 MHz, with most of the lower frequency spectrum being used for traditional cable delivery that is being done today. A limited number of channels are reserved to carry reverse, or upstream, information to the server. However, upstream amplifiers do not exist in all locations and need to be included. A hybrid combination, using the twisted pair for upstream communication, and the cable for downstream data delivery is also possible.

We assume that many servers will reside in telco central offices or cable company head ends. Customers will directly access movies on their local server. If the movie they wish to view is not stored there, their server will establish a link to other local servers or regional libraries. We believe that by the time video on demand is widely deployed this communication infrastructure will be based on SONET/SDH and ATM (asynchronous transfer mode). Already the long-distance trunks are predominantly fiber and SONET based.

Video Server Architectures

The main physical components of a video on demand service are a video server, data delivery network and set-top converter box. Figure 1 shows these main components. The video server consists of the storage and control required to store movies in compressed format and play them back on request. It differs from a traditional database server in various ways. It has to perform a number of functions such as admission control, request handling, data retrieval, guaranteed stream transmission, stream encryption and support of VCR functions. Admission control is done for each request by determining if the request can be serviced by the available resources in the system. Because the transmission of video data is stream oriented, it needs to be delivered to the end viewer without any glitches. The system can service the request only if continuous delivery of the video stream can be guaranteed, once the stream has been started. Because of the non-deterministic nature of disk accesses, intermediate buffer memory is used to transform the bursty disk accesses into a continuous stream that is guaranteed to be glitch free. The communication paradigm is also different from that used in a conventional system that is part of an enterprise network. In the case of VOD, the data is continuously sent from the server to the settop. There is no feedback from the settop asking for more data when the settop buffer is about to underflow. In an enterprise network the settop acts as a client and is continuously pulling the data from the server. In the VOD case, the data is "pushed" to the settop instead of being "pulled" by the settop. The buffer requirements at the settop need to be sufficient to prevent any underflow or overflow to occur.

The digital data is taken in compressed format from the disks and delivered to the network interface, where it is modulated for transmission over the distribution network to the home receivers, or set-tops. The set-top box, next to the TV, decompresses the data in real time and displays it on the TV screen. Decompression chips conforming to the MPEG standard are available from a number of vendors. The set-top box will provide the user interface that will allow the user to choose movies from a menu and also have a VCR-like interface for play, pause, rewind, and fast forward. The set-top box has to be very inexpensive in order to compete with existing CATV converters and VCRs. Therefore it should contain only enough storage to compensate for network jitter, not enough to download the whole movie.

There are a number of important issues that need careful consideration in the design of a video server. Current computer systems are mainly designed for good computations. For the VOD application, the compute power of the server is irrelevant. Good input/output bandwidth, or efficient data movement, is significantly more important. Multiprocessors may be used as video servers, but that would not be a very cost effective choice since a lot of hardware is not utilized.

One possible architecture for a video server is to use an existing shared memory multiprocessor. Figure 2 shows a shared memory multiprocessor. The data is taken from the storage subsystem, transferred over the bus to main memory, and then transferred again over the bus to the network interface. The caches and processors are not utilized and the bus is also not effectively utilized due to the dual transmission of the data over it.

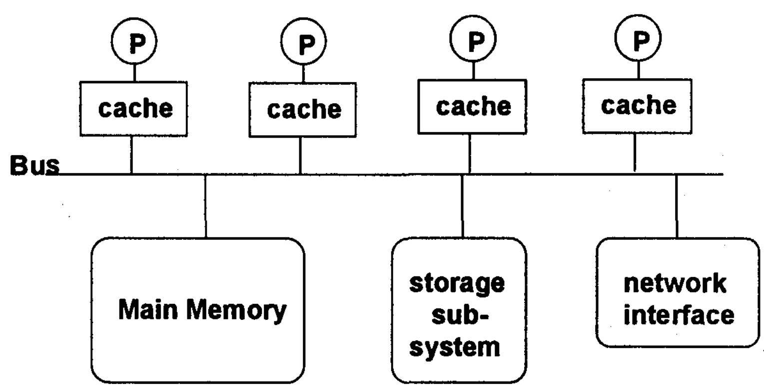


Figure 2: Shared Memory Multiprocessor

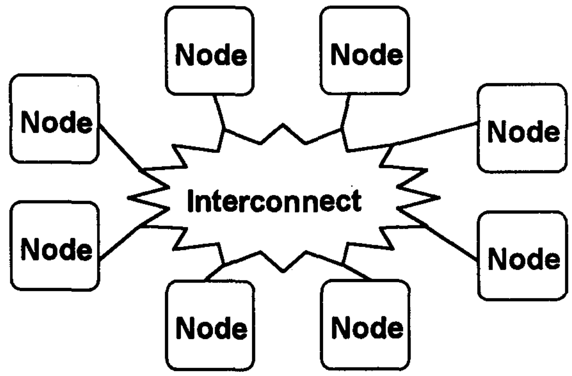


Figure 3: Distributed Memory Multiprocessor

Another possible video server architecture is a distributed memory multiprocessor. Figure 3 shows a distributed memory multiprocessor. Each node consists of a processor, cache, memory and interface to the interconnect. These architectures are designed for large scale computations using the message passing programming paradigm. Each node can be adapted to include a network and or storage subsystem interface. The computation related resources again are underutilized.

Scalable Architecture

As had been mentioned earlier, it is important for the design to be scalable, both in terms of the number of titles and in terms of the number of viewers. The VOD application involves large amounts of data transfers and comparatively small computations. We now describe an architecture that can be used effectively for the large data transfers that are typical for video servers. It is tailored for the *data movement* problem instead of a general purpose *data computation* computer. Figure 4 shows a generic architecture for a scalable video server, consisting of a main controller and a number of data servers. Each data server contains its own storage devices. Scalability in the number of titles is achieved by increasing the amount of storage per data server. Scalability in the number of viewers is achieved by increasing the number of data servers. This makes the design modular in structure and scalable to larger systems. We follow an open systems design approach by using standard interfaces and protocols. These interfaces range from SCSI for the storage devices to ATM (asynchronous transfer mode) packets that are transported over a SONET interface over the distribution network.

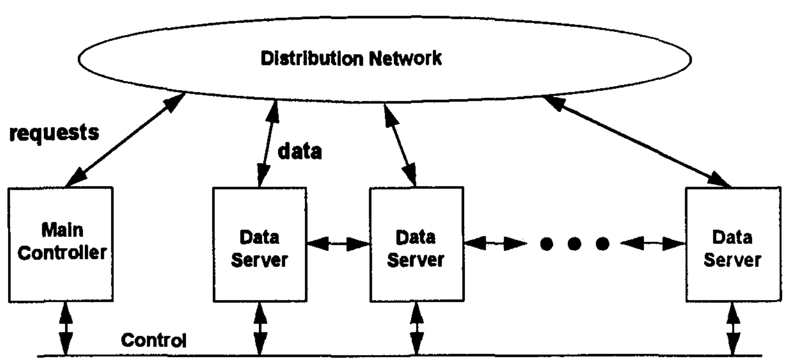


Figure 4: Scalable Video Server Architecture

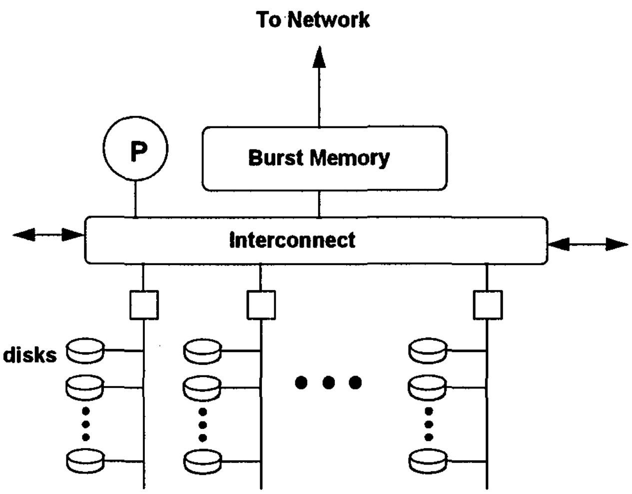


Figure 5: Data Server

The main controller can be a general purpose workstation, or an embedded device. It receives requests from the network and performs the admission control functions. It has the mapping of movies on I/O devices and controls the data servers. Control information consists of commands such as "play stream A", or "rewind/fastforward stream A". A data server is responsible for a particular video stream. Once it receives a command from the name server, it determines locally how to manage the available resources most efficiently. Each data server is connected to the storage devices.. A data server is the replication unit and additional data servers may be added to the system to scale the bandwidth of the system. The main controller can run Unix. We follow the open systems approach throughout the architecture. Besides controlling the data servers, general services such as billing are also performed by the main controller. Viewer requests are received at the main controller via the back channel. The main controller has the necessary software and database to perform admission control and control the data servers. Each data server is connected to SCSI subsystems and to an ATM switch. For small configurations requiring only one data server, the ATM switch is not required, resulting in lower overall cost.

Figure 5 shows the structure of a data server. Each data server has information about the movies that are stored on the subsystems it controls. The movie can be striped across multiple disks. By striping the data fully across all the disks in a data server, load balancing can be achieved more effectively and hot spots in the system can be prevented. A data server performs its own resource management for the movies that are stored locally. It can operate as a NVOD server or a IVOD server. A standard control protocol between the main controller and the data servers allows more than one implementation of the data server to be used in the same overall architecture. The data server controls the transfer of data from the disks to the burst memory. The burst memory is used to buffer the disk blocks and remove the non-determinism in disk access times. Data is taken from the burst memory and transmitted in a continuous stream over the distribution network. One likely network interface is a SONET OC3 interface that is capable of transmitting data at a rate of 155Mbps. Other interface conforming to the open systems approach are also possible. Each data server module has the potential of serving a few hundred viewers.

Storage

Traditionally, database servers have been designed for a large number of small transactions. The throughput for these transactions is optimized (transactions per second), but there is no notion of guaranteed, or real time delivery. In a video server, once a transaction (movie) is started, the complete movie file needs to delivered in a guaranteed, stream oriented manner. A video server also needs to access a large amount of data of the order of a terabyte. This makes the storage cost a very important design criteria. The data layout on the disks needs to be optimized in order to obtain maximum utilization of the system.

Compression is required to reduce the storage requirements for the video database. A 2 hour NTSC format movie requires approximately 100 GBytes of storage. At $0.80/MByte for disk storage, this translates to $800,000 per movie, which would make the system uneconomical. Compression reduces the storage requirement to approximately 1.35 GBytes per 2 hour movie using a 1.5 Mb/s MPEG compression scheme. From the experiments in our lab, this compression rate produces good visual results on the monitor. But this has not been tested commercially, and if 1.5 Mb/s yields inadequate quality, 3 Mb/s or more will be needed, resulting in close to 3 GBytes per movie. To cop with the vast amounts of storage needed even for compressed movies, we expect to use a combination of magnetic tape and disk storage to reduce the total system cost, which is dominated by the storage cost.

Figure 6 shows the various technologies in the storage hierarchy. RAM is the most expensive and has the highest bandwidth. At the low end we have jukeboxes that can be used to store large amounts of data, but they have large access times. A proper intermix of the various technologies can result in a cost-effective storage system for the video server.

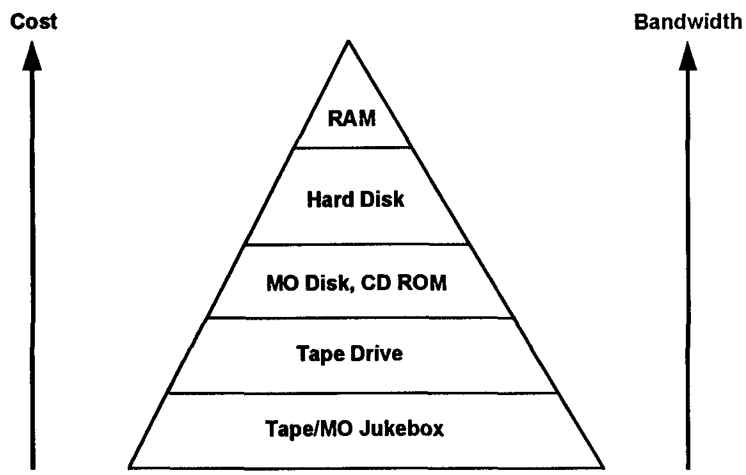


Figure 6: Storage Hierarchy

Storage Technology	Bandwidth (MBytes/sec)	Cost ($/MByte)
RAM	> 80	60
Disk	4	0.80
Tape	2	.025
CD-ROM	.3	.05

Table 1: Storage bandwidth and cost

To have a cost effective design, the storage hierarchy described above needs to be exploited by using forms of caching. We expect SCSI devices to be used widely, with fiber channel devices to follow. The small computer system interface (SCSI) is a parallel, multimaster I/O bus that provides a standard interface between computers and peripheral devices. SCSI is a widely used protocol for connecting disks, tape drives, CD-ROMs and magneto-optic (MO) drives. Fast and wide SCSI systems can provide a peak bandwidth of 20 MBytes per second.

Disk densities are increasing at a rate of 60% each year, and prices are falling at a rate of 12% per quarter. The video server is a disk based cache of data that is retrieved either locally or remotely from archival storage. Corruption of data on disk is therefore relatively unimportant as it can easily be retrieved from tape or other media. However the failure of an entire disk is important as it will disrupt the service of paying viewers. High availability is therefore more important than data integrity. Specifically the storage system must have excess bandwidth available to cope with component, usually disk, failures.

To provide the bandwidth necessary to supply many streams from the same material the movies will be striped across a number of disks. Redundant Arrays of Inexpensive Disks (RAID) techniques that add additional disks can be used. There are a number of raid levels that have been defined ranging from RAID 0 to RAID 1. For VOD, data retrieval needs to satisfy some hard real time constraints. The array needs to be able to supply the data at the correct rate even when one of the disks malfunctioning. Due to this hard real time constraint, the RAID levels that are most useful for video are RAID 1 and RAID 3. In RAID 1 data is mirrored, or replicated.. The cost is thus double than that of a simple array of disks. In RAID 3 one additional disk is required for each array to store parity. If a disk fails, the parity data is used to generate the correct data. However the sophisticated rebuild algorithms required to reconstruct the data on a replacement drive while maintaining bandwidth to the application may not be necessary as the movie can potentially be replaced from tape.

If a movie is not available at the server, the data can be obtained from an archival storage. It is likely that this transfer will be non real-time initially, with real time capability in the future.

Summary and Conclusions

We have described the various issues related with the design of video servers and described some architectures that can be used for this VOD application. The scalable architecture that we have described allows a very cost-effective solution for the VOD application and other multimedia applications. It follows our commitment to open standards and reflects our belief that the server, while computer controlled, is not a computer. The VOD application is primarily a data movement applications as compared with a data computation applications, hence neither a scaled down mainframe nor a scaled up workstation provides an optimal cost effective solution.

Acknowledgments

We are thankful to our colleagues at Hewlett Packard Laboratories and the Video Communications Division for the many stimulating discussions. The discussions with Dan Pitt, Vivian Shen, Nalini Venkatasubramanian, Al Kovalick, John Youden, and Paolo Siccardo were very useful.

References

[1] Yee-Hsiang Chang, David Coggins, Daniel Pitt, David Skellern, Manu Thapar, and Chandra Venkatraman, An Open-Systems Approach to Video on Demand, *IEEE Communications Magazine*, May 1994.

ASYNCHRONOUS TRANSFER MODE (ATM)

Its Development and Adoption: Migrating towards cable

by Roger D. Pience

ABSTRACT

Asynchronous Transfer Mode (ATM) has taken the telecommunications and cable television world by storm. It is widely recognized to be the only technology that bridges each element of the emerging National Information Infrastructure (NII); *i.e.,* interoperability between telephone, data and computer networks and cable television and multimedia networks. For the management executive, this paper will explore the various business applications and benefits of ATM along with a contrasting historical perspective of other technologies. The engineer will be presented with the technical aspects of the medium, from network applications and network architecture to basic details of the ATM protocol layers. In summary the paper will examine where the applications of the technology for cable television industry (*e.g.,* video-on-demand or multimedia), and ATM integrate various network architecture's.

INTRODUCTION

The emergence of a large number of new telecommunications services has influenced the growth of new communications protocols that will provide ever increasing speed and network flexibility. ATM is an enabling technology that process data, video, voice and image information simultaneously. It is one of a general class of packet technologies that relay data via an address contained within the packet, which provides this speed and network flexibility. First and foremost ATM is a high bandwidth, implying high data rate, switching and multiplexing <u>technology</u> [1]. ATM is not a <u>service</u> that can be compared to interactive multimedia, for example. It is a technology which will deploy most all of the advanced services, such as video-on-demand, HDTV, videophones and high speed data transfer for example, within a full service network. It is an outgrowth of B-ISDN[2] and is intended to be carried on a synchronous fiber or coaxial network.

Future market expectation can clearly delineate two categories of customers for high speed communications networks using ATM: the home and the office or business. Each has its own separate and distinct network service requirements. The home is mostly interested in entertainment type services and the office, especially health-care and education, is concerned with increased efficiency and productivity. However, as increased network capacity becomes available the home is ever more so interested in data services as well.

ATM is foreseen by the cable industry as a technology that may allow cable to fulfill its potential of performing to the standards of a broadband-switched network[3]. Indeed, many communications companies are exploring the inherent advantage of ATM over existing technologies as an efficient means of provisioning advanced data and video services.

HISTORY

Asynchronous Transfer Mode is really not a new technology, but an outgrowth of Asynchronous Time Division Multiplex (ATDM) networks which have been extensively studied at Bell Labs since 1969 [4]. The imminent confluence of computing and communications resulted in a need to interconnect several computers and peripherals together at widely separated physical locations with a high-speed data network. Packet switched networks were just beginning to emerge in 1969 and time sharing was in its infancy. The need to connect computers called for a high-speed network which would strive for efficiency, economy, and full instantaneous benefit of the bursty nature of communications required. Synchronous communications platforms, of the day envisioned for future digital telephone services did not fulfill the requisite technical needs. By 1975, research progressed to the point where the ATDM communications requirements had out-grown the present synchronous switches. There are three inherent problems with synchronous network switch designs: lack of flexibility, unnecessarily high switch speed, and synchronization of the sending, receiving computer and any intermediate switch node. Further research developed the ATM concept and in December of 1983, AT&T introduced the first commercial ATM product: Datakit VCS. From its start in 1985, B-ISDN has encompassed circuit and packet switching technologies. ATM is a derivative B-ISDN since its cell structure enables it to work in partnership with circuit switching in the same network[5]. 1984 saw the first ATM product and service availability and since that time has grown into the technology we know today and which in 1992, exploded from virtual obscurity to general awareness. This recent volcanic awareness is driven by the response of ATM meeting the users changing needs rather than being network standards driven.

ATM OVERVIEW

It is important to understand the reasons why the technology has so rapidly come from obscurity to communications hottest commodity. ATM is:

- ***bandwidth efficient***
  Low header overhead and variable length data packets form the ATM cell. This ratio provides high bandwidth utilization.
- ***scalable***
  ATM is capable of simultaneous carriage of data streams of varying rates. Video, voice and computer data require vastly different rates of data transmission and each are equally accommodated within ATM.
- ***transparent***
  Some networks require guarantee of timely delivery of information at the receiving point. Real time services such as telephony and video demand must have minimal end-to-end transmission delay. Packet switching and frame relay cannot guarantee such timely delivery[6].
- ***network flexible***
  The ATM network need not be modified or suffer efficiency losses in order to accept new and different

service characteristics. It is also capable of multicast functionality; *i.e.*, broadcast or point-to-point.

- *protocol optimized*

  Each ATM cell is fixed at 53 octets (bytes): 5-octet header for identification and 48 octets for information. The allocated information octets are commonly called the payload. This particular equation of header to payload ratio provides particularly jitter free transmission characteristics which is especially important for voice networks.

Now that it is known what ATM is and what it can do it is also important to understand what ATM is not. The comparison with other transmission systems such as PACKET, FRAME RELAY, and SONET will provide a little insight to these differences.

PACKET

Packet systems such as X.25 and other closely related X.? packet protocols were first developed for use over noisy error prone analog transmission channels such as radio or telephone lines. The protocol for these noisy channels was necessarily robust and bandwidth inefficient. Hence it is slow. DS-1 performance is about all that can be expected of packet systems. Another difficulty with packet is its time insensitivity. X.25, in particular was designed to allow data packets to arrive at a receiving location out-of-sequence and then be reassembled in proper order at the destination. This is not suitable for data such as video and voice because of the inherent delay[7].

FRAME RELAY

A fairly recent development, frame relay, is more efficient than packet in carrying data traffic because it was developed to operate over fiber optic data channels that are virtually error free. It provides higher throughput, higher bandwidth and more cost effective transport than does standard X.? packet technologies. However, this strength is its downfall. Frame relay is too inflexible to cope with changing traffic rates such as video and voice which in the future will be mixed into today's purely data networks. Frame relay can be approached in three separate defined implementations: an interface, a network signaling protocol and as a network-provided service[8]. Some of the benefits of using frame relay as an interface are:

- true international network interface standard
- multiple users per physical interface access line
- high speed of access due to low packet overhead
- higher throughput to high-speed applications

The benefits of using frame relay as a protocol are:

- various size frame transport
- increased performance over older packet technologies
- reduced overhead for backbone networks
- maximum link efficiency
- performs multiplexing functions
- reduces nodal latency
- improved bandwidth utilization

The benefits of using frame relay as a service are:

- cost effective because of network efficiencies
- transport speeds of up to T3
- utilizes all available bandwidth
- provides first true bandwidth on
- high speed - low delay
- fills gaps between X.25 and broad-band services
- "any-to-any" connectivity

Frame relay is sometimes confused with the term "fast packet." Fast packet is a generic term used for many high-speed packet technologies such as frame relay and cell relay. Frame relay and frame switching are synonymous only with CCIT switching implementation of Type II frame relay where it is defined as an end-user service under ISDN service standards.

SONET

SONET, Synchronous Optical Network, is a Bellcore term for the Synchronous Digital Hierarchy standardized by the CCITT in Europe and Asia[9]. It was conceived as a method of providing a high-speed international fiber optic transmission standard interface between disparate standards of various countries. SONET is a transport interface and method of transmission only and is not a network unto itself. SONET is planned to eliminate the different transmission schemes between countries.

SONET uses both synchronous and asynchronous transfer modes through he use of a fixed data transfer frame format. Using a unique framing format, SONET relies upon timing as the most critical element. Frame payloads of SONET are not synchronized by a common clock even though it is a synchronous technology. The actual data payload of SONET can take many forms such as DS3, FDDI or SMDS. Some of the benefits of SONET are:

- true fractional DS3 demand service specifications
- aggregation of low-speed data transport channels into common high-speed backbone trunk trans-port
- increased bandwidth management
- reduced overall network transport delay
- supports B-ISDN, SMDS, and other high bandwidth services
- add/drop channels without the need to demultiplex or remultiplex

As a communication process between two end systems, ATM like frame relay, relies only on a two layer protocol stack (header and payload). At every point in the network, therefore, processing is done only on each ATM cell individually without consideration of any other cell. The network is not concerned with the arrival of groups of cells , sequencing, or acknowledging, as is X.? packet transmissions, therefore forwarding the cell is much faster[10]. No processing is of data is done in the information field of the cell. Unlike packet, ATM assumes one and only one defined path exists through the network for the transport of cells.

ARCHITECTURE STANDARDS

Standards for ATM developed as part of the overall B-ISDN evolution. It is based upon a layered architecture similar in concept to that used by the International Standards Organization (ISO). The seven layer open systems interconnection (OSI) model is explained in many reference texts and will not be discussed in detail here. Basically, such models divide any communications process into subprocesses, called layers, arranged in a hierarchical stack, *Figure 1.* Each layer provides services to the layers above to aid in communications between the top layer processes. Another important concept behind layering is the ability to revise or change a layer without impacting the protocol layers above and below. For example the physical layer for ATM may be changed from SONET to DS3 without impact on the ATM layer above or the services provided to higher layers.

For each plane, a layered approach as in OSI is used with independence between layers[11]. According to CCITT, the layers can be further divided into a physical layer which mainly transports information (cells), the ATM layer which performs mainly switching/routing and the AAL (ATM Adaption Layer) which is mainly responsible for adapting service information into sublayers. The ATM layer is fully independent of the physical medium used to transport the ATM cells. The cell, as mentioned earlier fixed in length at 53 bytes total with a 48 byte payload and a 5 byte header. The six different functions:

- Virtual path ID
- Virtual channel ID

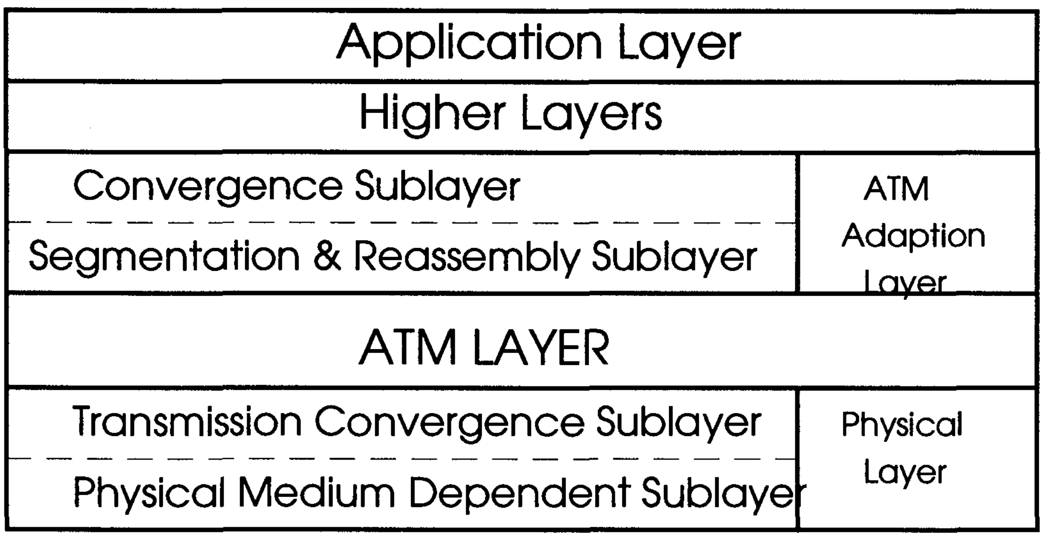


Figure 1

The ATM protocol model can be divided into three phases:

- a user plane to transport user information
- a control plane composed of signaling information
- a management plane used to maintain the network

- Payload type
- Reserved
- Cell loss priority
- Header error control

One of the most important functions of ATM is its ability to integrate and transport data services of varying rates: voice at 64 kbps to multiplexed video channels at 30+ Mbps to data information transfer at 100+ Mbps.

VALUE TO NETWORKS

Businesses today have been building private ATM networks with great success. These users are interested in high speed data transfer and are increasingly demanding desktop video telephony. Residential customers are mostly interested in plain old telephone service and TV distribution. The proposed architecture will offer increasingly high speed connectivity for all customers. It is foreseen that the terminal equipment installed at both the residential and business customers can be modularly upgraded as new services are introduced.

ATM is not without its detractions. ATM is not yet a fully accepted standard. There are various portions of the ATM OSI model that have not totally evolved out of the standards committees. ATM is not ubiquitous in that there are very few public networks with ATM capable switching apparatus. ATM does not have global acceptance. Although the CCITT sponsors further international standards development of ATM not all participating nations agree on the finer details of deployment.

HDTV transport layer work by the Grand Alliance (GA) has ensured interoperability with two of the most important alternative transport systems, namely, MPEG-2 and ATM. The GA transport packet size has been selected to ease transferring ATV transport packets in a link layer that supports ATM. Three techniques have been developed that are fully interoperable[12]. Since several of the CATV and DBS systems being designed are considering using a variant of the MPEG-2 transport layer the degree of interoperability will remain high for some future period of time.

REFERENCES

[1] Bowick, C., "From the Headend", *CED,* pp. 22, April 1993.

[2] Fraser, A.G., "Early Experiments with Asynchronous Time Division Networks", *IEEE NETWORK, vol 7, No. 1,* pp. 12-26, January 1993.

[3] Green, R. Dr., "Keynote Speech to the ATM Forum Meeting" San Francisco, CA, August, 1993.

[4] McGrath, C. J., "Digital Transport for Video and Communications Systems", AT&T Bell Labs, November, 1992.

[5] Fraser, A.G., "Early Experiments with Asynchronous Time Division Networks", *IEEE NETWORK, vol 7, No. 1,* pp. 12-26, January 1993.

[6] Anon., "ATM", *Sprint Technical Report,* April, 1993.

[7] [8] Spohn, D. L., "Data Network Design", *McGraw-Hill, New York, NY,* 1993

[9] Knapp, E., "SONET, An Emerging Transport Technology", *NYNEX Science and Technology,* July, 1988

[10] Lane, J., et al, "ASYNCHRONOUS TRANSFER MODE: Bandwidth for the Future", *Telco Systems Inc.,* 1992

[11] de Prycker, M., "ASYNCHRONOUS TRANSFER MODE, Solution for Broadband ISDN", *Harwood,* UK, 1993

[12] "TRANSPORT LAYER FUNCTIONAL DESCRIPTION FOR THE GRAND ALLIANCE ATV SYSTEM" a report, October, 1993.

Cable System Transient Impairment Characterization

RICHARD S. PRODAN, PH.D.—VICE PRESIDENT, ENGINEERING
MAJID CHELEHMAL, PH.D.—DIGITAL SYSTEMS ENGINEER
THOMAS H. WILLIAMS—TRANSMISSION ENGINEERING SPECIALIST
CRAIG M. CHAMBERLAIN—ELECTRICAL ENGINEER

CABLE TELEVISION LABORATORIES, INC.

ABSTRACT

CableLabs is performing a field study of transient disturbances occurring in cable systems to evaluate the impact of nonstationary impairments on high speed, band-limited transmission of digital information. This paper describes the design and operation of a system which has been built to determine the nature of non-stationary impairments that interfere with digital signal transmission.

The system inserts a CW (carrier wave) signal at the headend in the center of a vacant channel. At the receive site, the carrier is quadrature demodulated to baseband. Dual triggers are set for the in-phase and quadrature channels. When triggered by a disturbance, a digital oscilloscope captures data from both channels for later time and frequency domain analysis.

The type of transient disturbance, duration, and inter-arrival statistics are derived from the accumulation of events over a significant time period (weeks to months). This system can also be used for reverse system characterization. Samples of test data captured and analyzed are presented.

At this time, the system has been built and some preliminary testing on a cable system in Boulder, Colorado has been done.

INTRODUCTION

Interference is a ubiquitous property of the environment in which a multiplicity of communication systems and man-made noise sources coexist with the cable transmission medium. The ingress of these (possibly spurious) sources of interference introduce impairments to digital transmission over the cable plant. Also, non-ideal mechanisms present in the equipment used in the communication process can introduce transient interference and distortion.

Interference of a transient or impulsive nature can occur via two distinct mechanisms: ingress of external interference energy from cochannel signals or nonstationary (man-made) noise sources, and dynamic impedance or return loss changes due to time-varying terminations or signal path components.

The characterization of these types of impairments does not lend itself to a simple, elegant model. Thus a time-domain transient representation of such observed phenomena with duration and inter-arrival statistics would seem to be the best approach.

IMPAIRMENT CHARACTERIZATION METHODOLOGY

Figure 1 is a constellation diagram of a 16-QAM modem used for carrying data over an RF

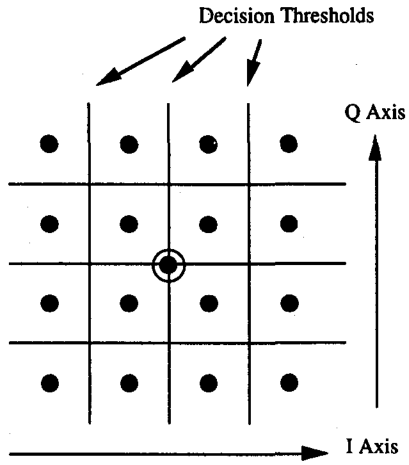


Figure 1:
Modem Testing

channel. Each of the 16 constellation points is identified as the modulated phase and amplitude that a carrier voltage may be at the correct sampling instant. Also shown in Figure 1 are decision thresholds. When the modem is carrying random data, the modem switches between states at the symbol rate. Any particular state out of 16 possible states can be identified from the voltage of the demodulated carrier on the I (in phase) axis, along with the corresponding voltage on the Q (quadrature) axis. At RF frequencies, the channel is occupied with uniformly distributed energy from the random carrier modulation.

If an impairment, such as Gaussian noise, is added to this channel, each constellation point will be moved from its nominal position. If the impairment is strong enough to move a received constellation point across the decision threshold, errors will occur. Testing with this energy present in a channel is very difficult because the energy masks the impairment. About the only type of testing that can be done is bit-error measurement, which requires that the channel be taken out of service for the test. Bit error testing is useful, but gives results that are a mixture of channel performance and modem performance. The cause of the bit errors is not apparent, only their presence and (possibly) duration.

This diagram can be contrasted with the constellation in Figure 2, which consists of only a single point. This single constellation point is generated by a CW signal. The distance from the point to the origin is the magnitude of the carrier. A rectangular threshold region is established around the point. If the level of impairment is sufficiently weak so that the point is continuously within this region, then the impairment is sufficiently benign so that an equivalent modem would operate error-free. If an impairment drives the constellation point outside of the threshold region, the event is detected and recorded for later examination. Impairments that can be identified by this technique are shown in Table 1 along with the characteristic appearance. In addition, Fourier analysis of the captured time waveforms could identify the nature of the impairment.

The block diagram of the test instrumentation is given in Figure 3. The down converter mixes the CW carrier down to an IF frequency of 44MHz with a very low phase noise local oscillator, which may be either an inexpensive crystal oscillator, or an expensive agile signal generator. For testing the reverse band, this block will be an up converter. The bandwidth of the SAW (surface acoustic wave) filter determines the width of the channel, and has been chosen to be

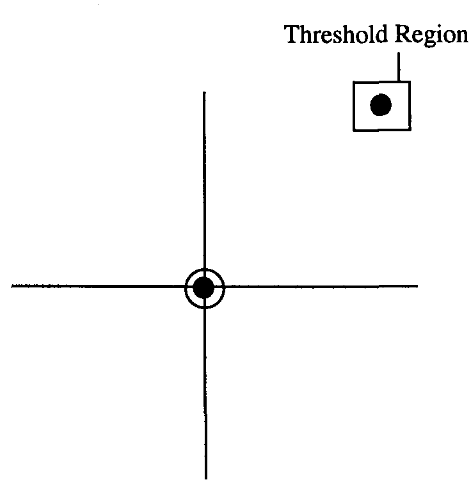


Figure 2:
Carrier Wave Testing

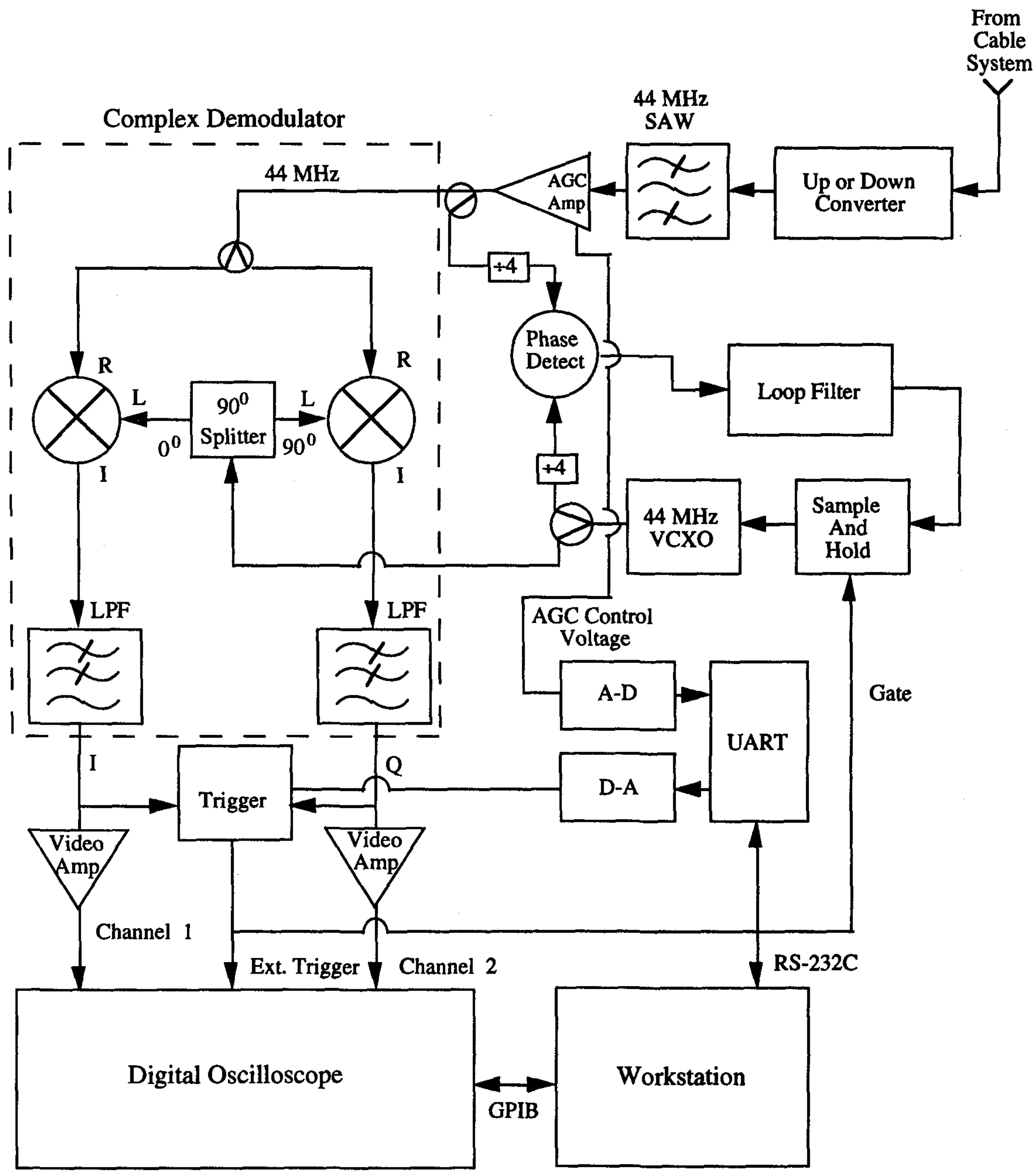


Figure 3: CW Tester Block Diagram

about 5MHz wide, so that the test may be carried out in a single 6MHz channel. The AGC amplifier serves to hold the constellation point at a constant distance from the origin as the strength of the received carrier varies slowly, due to the cable plant experiencing temperature variations. The AGC does not need to have a large dynamic range, but it does have a slow time constant and has been designed to have low phase shift with attenuation changes. A phase locked loop, based on a MC4044 type IC phase detector operating at 11MHz, keeps the voltage controlled crystal oscillator (VCXO) in lock with the incoming carrier.

The carrier is demodulated by a conventional complex demodulator that yields both I and Q components. The low pass filters in the demodulator have a 5MHz cutoff frequency. This frequency was chosen to be wide enough to pass the energy through the SAW filter (+/- 2.5MHz), but narrow enough to reject spurious components.

The trigger circuit consists of 4-AC coupled high speed comparators. Comparators establish the upper and lower trigger points for both the I channel and the Q channel. The OR'ed trigger outputs from the CW tester is supplied as an external trigger input to the digitizing oscilloscope. After triggering, a one-shot circuit opens the phase locked loop long enough to allow the data to be captured without responding to the phase error. This is accomplished through the control line labeled "gate."

Video amplifiers boost the signal level for input into the digital oscilloscope. A Unix workstation controls the threshold levels for the trigger circuit via a UART (universal asynchronous receiver/transmitter) interface to a digital-to-analog converter. The threshold levels are operator selected for a test. The UART also gathers the control voltage of the AGC circuit through an analog-to-digital converter. This information can be translated into the incoming carrier strength. The UART circuit communicates with the workstation via the RS-232C port.

The digital oscilloscope (TEK TDS 420) samples both the I and Q channels at 10M samples per second for up to a maximum of 60,000 samples on each channel. This sample rate is above the Nyquist limit for the channel. After an event is captured, the data is extracted from the oscilloscope via a GPIB (general purpose instrumentation bus) interface for storage on the work station's hard disk drive.

Effective shielding is necessary to avoid capturing interference that is actually originating locally.

A workstation was chosen to allow the apparatus to operate over an extended period of time in an unattended mode. The controlling software stores data in three types of files. The first is an AGC file that periodically monitors and logs, once each minute, the current AGC voltage. The second is an event log file containing the event time of day, the trigger voltage threshold, the event number, and the current AGC voltage. The third file contains the number of points and the actual demodulated I and Q time waveforms triggered by an event. The system is designed to prevent overflowing the hard drive should pernicious interference be experienced. This can occur when sweeping a system, causing an event to be logged every few seconds. Although in this case, the data from every event would not be stored, all events would be logged with time information, thereby capturing the event total duration with samples of the interference.

Impairment	Characteristic Appearance
Gaussian noise	fuzzy ball on I-Q plot
Phase noise	arc shape on I-Q plot
Signal drop	point drops to origin on I-Q plot
Sweep interference(Wavetek)	chirp on I-T and Q-T plots
Sweep interference(Calan)	burst on I-T and Q-T plots
Sudden change in echo	a change in point position on I-Q plot
CW interference	sine wave on I-T, and a circle on I-Q plot
AM hum	a line heading towards origin on I-Q plot
Impulse noise	"flower petals" on I-Q plot
CTB or CSO	interference with large peak excursions

Table 1

EARLY RESULTS

Preliminary testing was done with an downstream CW signal at 445.25MHz using a signal carried at video level originating at a cable system headend. Data was captured at CableLabs, which was one fiber link and one line away, and one home site that was one fiber link and 18 amplifiers away. Figure 4 is a sweep-like interference that was captured. It was probably a piece of agile headend equipment that was turned on and passed through the test channel on

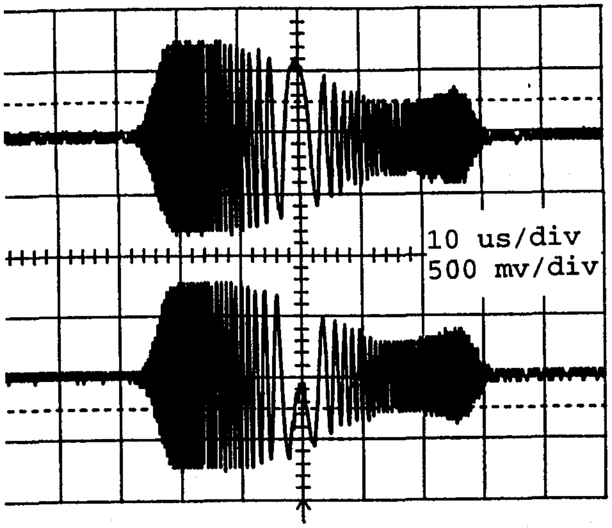


Figure 4

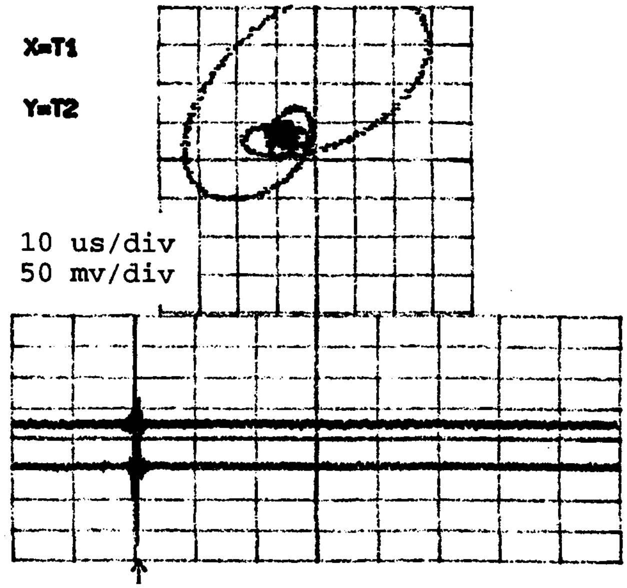


Figure 5

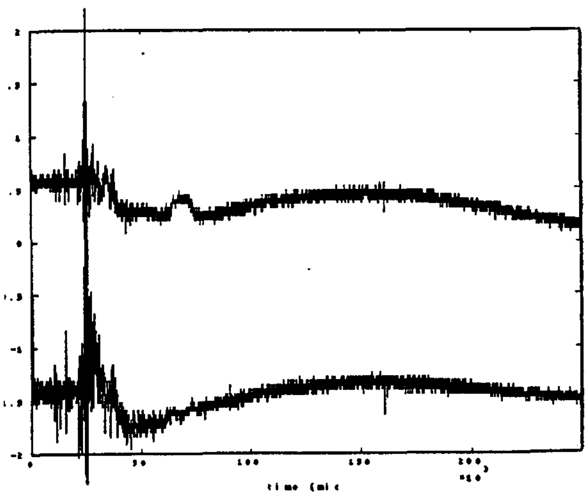


Figure 6

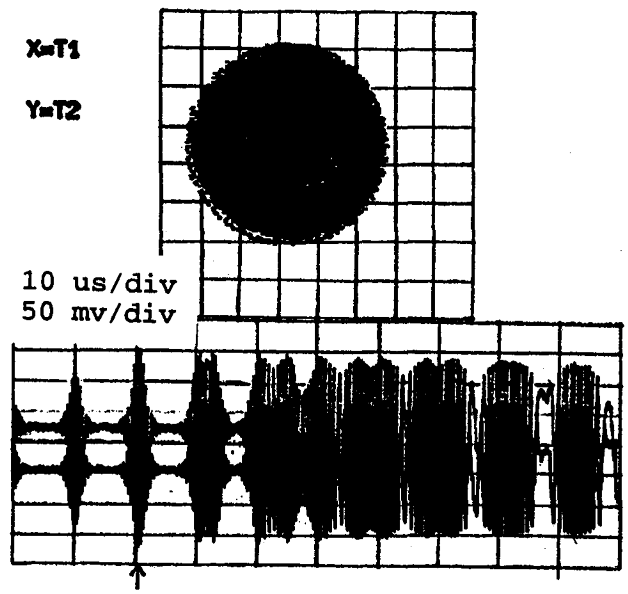


Figure 7

its way to its final frequency. Figure 5 is an impulse. Figure 6 is a nearby lightning strike that got into the phone system on the test home as well as the cable system. This particular lightning strike destroyed a home computer that was connected to the phone line through a fax-modem board in the test home. The cable equipment was not harmed. Figure 7 is interference of unknown origin. This data indicates that signals may not get out of headends unscathed unless special care is taken.

FUTURE WORK

Two impairments not characterized by this test method are (possibly dynamic) channel frequency response and group delay distortion. This can be remedied by sending an in-phase only reference signal from the headend intermittently, perhaps once an hour. One good reference signal to use would be the standardized ghost cancellation reference (GCR) waveform currently used on line 19 of many NTSC broadcast signals. This signal possesses high energy, compact duration, and a flat frequency spectrum. This signal would trigger the CW receiver, and be automatically captured by this equipment for later analysis.

The channel availability lost due to errors in fixed time intervals is a useful parameter for estimating transmission reliability. An approximation of this number can be found by modifying the CW tester circuit in the following fashion.

Remove the one-shot trigger circuit that breaks the phase locked loop. The loop will now remain locked. Add a gate circuit that is open whenever the constellation point is outside the threshold region. This gate will pass a clock oscillator to an event counter. Set the threshold region to be the same as the distance between adjacent constellation points on the simulated modem. Over a time interval, the counter will contain a number of pulses from which error duration can be estimated. For example, if the clock is a 10MHz square wave (100ns between pulses) and the counter contains 10,000 pulses, then the duration of the disturbance is one millisecond. Very long disturbance durations could be quantified this way, along with short time samples of the actual interference, as previously described.

The data will later be analyzed to determine the frequency and duration of transient impairments that would affect a given modulation and/or forward error correction scheme. A short time Fourier analysis can determine if the interference originated from a transient impedance change (changing reflection), ingress of a narrow-band interferer, or a wide-band impulsive noise spike.

CONCLUSION

A method for characterization of transient impairments on cable systems has been proposed. A system developed at CableLabs for evaluating the nature and effect of these impairments on digital transmission using this methodology has been described. Analysis techniques on the captured impairment data were outlined.

Many types of impairments may be recognized by analyzing the actual disturbances captured by this system. Also, a characterization of the duration and inter-arrival statistics of the transient disturbances are available. Channel availability and related statistics such as errored seconds may be derived. These test results can be used to design burst error mitigation schemes such as error-correction coding with interleaving for any desired type of single carrier modulation for digital data transmission over cable systems.

CAN VIDEO NOISE REDUCTION SAVE MONEY?

Rabab K. Ward[1], Donald Monteith[2], John C. Madden[3]

Abstract

In this paper we discuss the results of research aimed at making the reduction of signal impairments such as thermal noise, composite triple beats (CTB) and composite second order beats (CSO) inexpensive enough for inclusion in DVC set-top converters and other consumer equipment. Simple algorithms which reduce thermal noise, CTB, and CSO by approximately 7 dBs have been developed and demonstrated. Subjective visual improvement is frequently better than actual SNR measurements would imply. Several potential implications of this development on the operations and economics of cable systems are briefly discussed.

Introduction

Inspired by support from Rogers Canadian Cable Labs Fund, the group at U.B.C. which is headed by one of us (Ward) has been researching and developing techniques to detect, measure, and reduce a number of common visual impairments to television signals since 1990. Some of the earlier theoretical and experimental work has been reported elsewhere.[4, 5, 6]

In the past three and a half years we have employed digital signal processing techniques to detect, measure and reduce thermal noise, composite triple beats (CTB), composite second order beats (CSO), impulse noise, co-channel interference and single frequency interference.

Our previous publications have focussed on the theory behind the processes and have reported on our ability to detect the presence of the impairments automatically. Our work during the past year has been directed primarily towards noise reduction. This is because both we and our sponsors, Rogers Cablesystems, sensed that with the advent of Digital Video (de)Compression (DVC) set-top converters, the rather complex digital electronics which these devices embody might be put to good use improving ordinary TV pictures when subscribers were tuned to conventional analog broadcast stations, or were watching programming played back on conventional VCR's.

In this paper we discuss some results of this work, and finish with a brief consideration of some of the implications for cable television service providers and their subscribers.

Background

A variety of techniques for reducing Gaussian (or thermal) noise in television and motion picture signals have been developed and applied over the years. For the most part, as with the digital compression process, these algorithms have not been applied in real time. Rather the process has taken place off line, with the results recorded for later play-back in real time. Notwithstanding, some real time studio and head end equipment targetted at some specific signal impairments, such as impulse and thermal noise, has recently appeared on the market.

Our research group had, in the past, been more interested in developing accurate video signal impairment detection and measurement algorithms, a process which does not involve real time processing of sampled video data. The question we asked was whether the algorithms we had developed for these tasks could be simplified enough to work in real time using a low cost chip or chip set affordable by cable subscribers.

Procedure

Working with calibrated Betacam tapes produced with the assistance of Brian James at the Cable Television Laboratories Inc. test facility in Alexandria, Virginia, we developed efficient techniques which reduce the visual appearance of the major video impairments encountered in cable distribution plants. We have studied each impairment and found appropriate algorithms which cancel or significantly reduce the visual appearance of each of these impairments or a combination of them. Our universal algorithm also slightly enhances the picture's visual appearance. None of our algorithms introduces any visible changes in the picture. Our techniques involve processing the frames in the spatial and time dimensions. Each frame is first processed spatially, and then consecutive frames are processed together. The inter-frame processing involves detection of motion in the sequences. The algorithm specifically developed to cancel composite second order beats when incrementally related carriers (IRC) or harmonically related carriers (HRC) transmission systems are used requires finding the frequency spectrum of sequential lines of the video signal.

The rationale behind cancelling the CTB impairment depends on the idea that over a small region of any line of the picture, the CTB may be approximated as a constant luminance signal. Thus the effect of the CTB impairment over such a small segment of a line in the image may be modelled as a change in the DC luminance by an (unknown) constant over that signal. To filter out that constant, the digital signal is divided into vertical stripes or sub-pictures. The average intensity of the signal in each line of the sub-picture is determined. These average intensities are then filtered by a sliding averaging window of length equal to 3 or 5 pixels. The filtered value of each average intensity is then assumed to be the corrected DC level of the corresponding line in the sub-picture. In order to further improve the noise reduction, one can apply a multiframe CTB removal scheme. The average of the DC level in a given line in a given sub-picture is based on the average of the same line in the consecutive frames of the sequence. This method, and a slightly modified version of it, are also found to be able to measure the carrier to noise ratio of CTB-impaired pictures within ± 2 dB's. For measurement purposes, the advantage of this technique is that it is non-intrusive.

To reduce thermal noise, inter-frame averaging is most efficient. One of the practical constraints is the number of frame buffers used. In order to achieve 10 dBs reduction in the signal-to-noise ratio using direct averaging, ten frames are required (provided there is no motion in the scene). We have modified the direct averaging process so that similar performance can be obtained using only four video frames.

The algorithm for reducing CSO also uses spatial and time directions filtering. However for IRC or HRC transmission systems, much better results are obtained using a notch filter in the frequency domain. When notch filtering is used, algorithm parameters have to be chosen so that the continuous Fourier transform and the discrete Fourier transform yield exactly the same results within a scaling constant.

To reduce the computational complexity in order to meet demanding cost constraints on the hardware if it is to be implementable in DVC set-top converters, the different algorithms for reducing CTB, CSO, and thermal noise have been modified, significantly simplified, and combined into one algorithm. The resulting universal algorithm achieves different values of noise reduction depending on the allowed cost of the hardware, the SNR, and the type of signal impairment, but a good working performance improvement estimate would be 7 dBs. Video clip samples of the results obtained will be presented during the paper presentation. Those reading the paper will have to rely on the less satisfactory evidence of the still photographs of frames included in this paper.

Results

The photographs shown on the next four pages of this paper illustrate the results of the noise reduction process. All the pictures presented herein are modified by the same universal algorithm. Our process works equally well with colour and black and white sequences. Regretably, the static nature of the photographs, combined with noise introduced by the page reproduction process, make it difficult for the viewer to gain a true appreciation of the degree of noise reduction from the photographs. Those present at the paper presentation will be able to view complete video sequences from which these results are drawn.

It should be noted that employment of these noise reduction techniques frequently results in the addition of "noise" elsewhere in the signal, usually in a fashion which is not visually objectionable, and which may even be picture enhancing. Indeed, the image enhancement process is, by definition, the addition of some carefully constructed "noise", in the sense that it is not a part of the signal. Thus, when considering the performance of noise cancellation algorithms, improvements in SNR are not usually good metrics of performance. It is essential to view the actual results.

Discussion

The signal processing techniques described above require the following:

- an A/D converter;
- a microprocessor,
- up to 2 Mbytes of memory,
- special purpose custom circuitry,
- a D/A converter

MPEG-2 DVC set-top converters already contain a microprocessor of adequate power, a D/A converter, and 2 Mbytes (or 16 Mbits) of memory if they comply with the Main Profile, Main Level of the standard (i.e. are B-frame compatible). These are likely the most expensive components if a stand-alone analog video signal noise reduction system were to be developed.

Our experimental results leave us confident that custom silicon circuitry comprising 30,000 to 50,000 gates would be adequate to provide well over 6 dBs of noise reduction for thermal noise, CTB, CSO, and potentially some other impairments. Using current technology, this can easily be accommodated on a single chip, or even a part of a chip. Chip clock speeds of about 40 MHz. are believed to be adequate for the task. Appropriate A/D converters are inexpensive. It is thus feasible to think in terms of a $20 incremental selling cost for MPEG-2 set-top converters fitted with analog video noise cancellation for these impairments.

When IRC or HRC transmission systems are used, CSO noise reduction (almost to the point of total elimination) is possible, but the process presents a greater technical challenge. Although CSO noise reduction of 10 dB's or more can be obtained, today at least, the algorithm requires the use of high speed FFT chips which are costly even in mass production.

The ability to reduce video noise at the set-top converter (or further upstream in the cable system) has some interesting economic and service implications. Among them are:

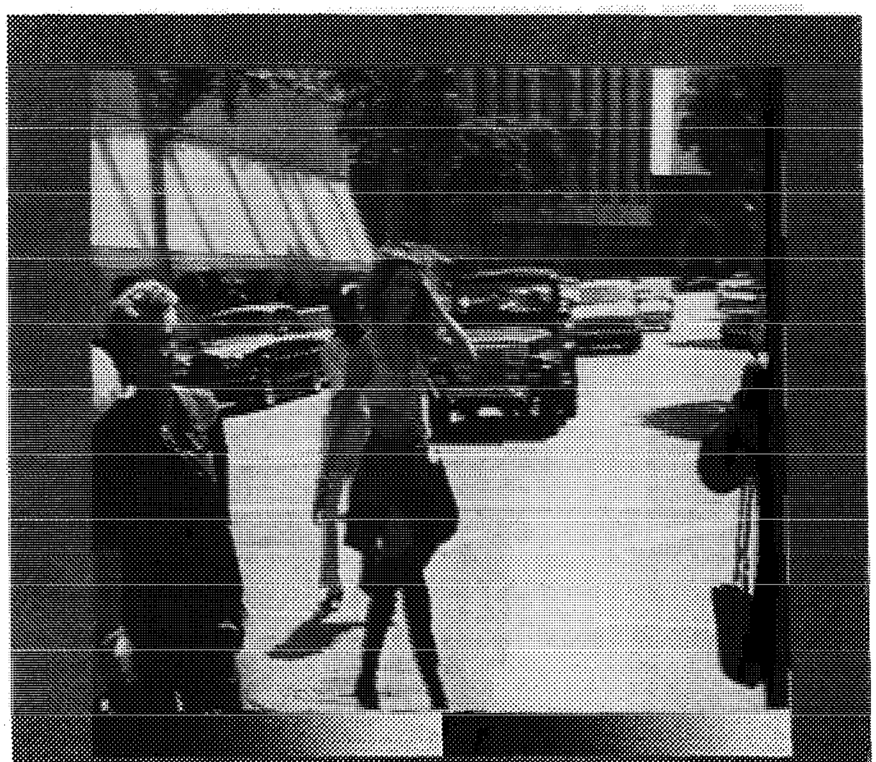
Picture with Thermal Noise (C/N = 35 dbs)

Cleaned Picture

Picture with CTB and Thermal Noise (C/N = 35 dbs)

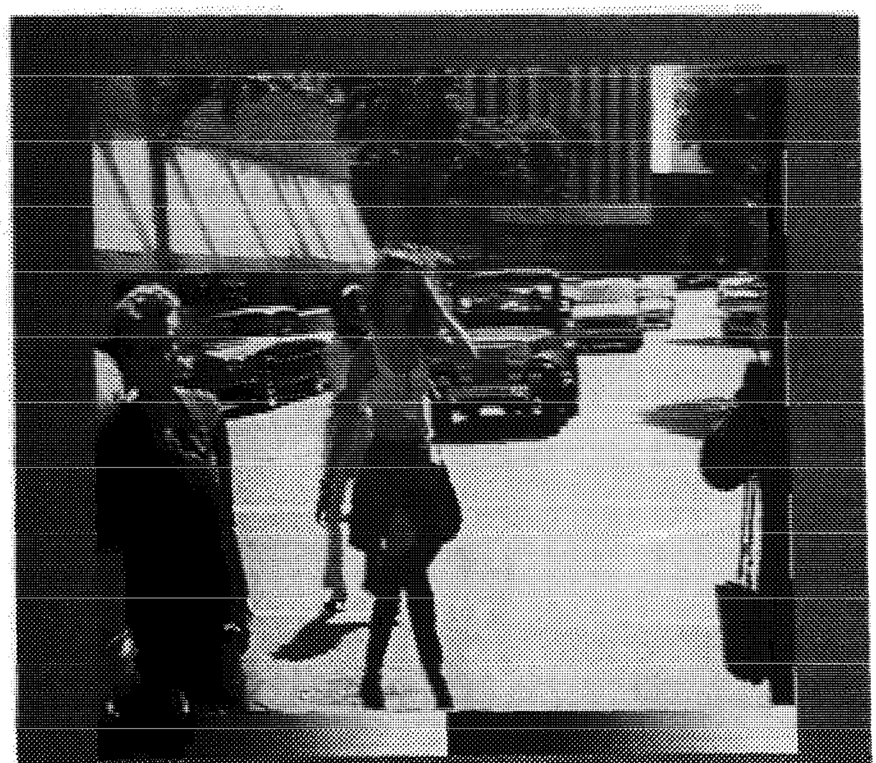
Cleaned Picture

Picture with CSO and Thermal Noise (C/N = 35 dbs)



Cleaned Picture

Picture with CSO, CTB and Thermal Noise (C/N = 35 dbs)



Cleaned Picture

1. DVC Converter Penetration Rate Enhancement

Initial penetration of DVC converters will be small, but there are strong financial incentives for cable service providers to move rapidly towards higher penetrations. There are some who predict that subscribers who have access to good quality digital satellite video signals will become dissatisfied with the technical quality of analog video signals currently delivered by cable. We believe that the techniques described in this paper could provide the cable industry with a practical response to this criticism, while at the same time giving subscribers yet another reason to obtain a DVC set-top converter.

Paradoxically, another potential incentive for boosting penetration may be to offer subscribers a feature which reduces the noise on a competitor's product, i.e. VCR cassettes obtained from video rental stores! Viewers of these cassettes typically experience thermal-like carrier to noise ratios of 30 to 40 dB, depending on the quality of the cassette and of their playback machine. By providing for a VCR input to the converter (and using the same noise reduction algorithm), subscribers could have access to cleaner rental videos. The potential effect of such an incentive on the penetration rate of DVC converters, and on the sale of DVC pay TV products is highly speculative, but, in our view, is likely to be significant, especially once viewers become accustomed to virtually noise-free digital TV programming.

2. Noise Cancellers at Nodes

The installation of increasing amounts of fiber trunking has led to a corresponding increase in concern about CSO, especially in the mid-band frequencies. A bank of noise reduction filters located at critical nodes offers a potentially attractive means of meeting noise specifications at a low cost per subscriber, while simultaneously obtaining a noise measurement and performance monitoring capability. However, while technically feasible, the additional cost of down conversion and subsequent up conversion of channels selected for noise reduction could render this approach economically infeasible in practice.

3. Noise Reduction through In-Home Noise Cancellation

As the penetration of DVC converters increases, it becomes possible to factor the noise reduction from set top converters into plant engineering calculations. For example, it may be possible to engineer capacity upgrades (e.g. from 330 MHz to 450 MHz) while conserving existing amplifier spacings (for cost savings of from $15-$25 per home passed or roughly $24-$40 per subscriber[7]). This could be done through the provision of DVC converters to the relatively small percentage of customers in high noise segments of the system, thus postponing expensive plant rebuilds to a more propitious time from a business point of view. System operators could, in such circumstances, de-activate the digital pay per view function of the DVC converters for non-subscribers to the service.

Conclusion

It would appear that the utilization of analog video noise reduction systems could have a substantial favourable impact on cable plant financial returns and investment, as well as on subscriber satisfaction.

We are of the opinion that there is a good parallel to be drawn between the audio noise reduction circuitry which is now widespread in the audio industry and the kind of video noise reduction circuitry which is discussed in this paper. In time, such circuitry will be incorporated into television sets and VCR's. Of course eventually, and perhaps as early as 2004, analog TV signals will be replaced by much more noise-free digital TV.

But at least until 2004, and perhaps well beyond that time, there will be noisy analog video signals which will be improved by noise reduction.

We believe that it is best for the industry if noise reduction know-how be made generally available in a fashion which enhances the speed of its development and promotes a commonality of approach. To that end, a new company (Ward Laboratories Inc.) has been established which has negotiated a licensing agreement with the University of B.C. for the purpose of accelerating the development of video noise reduction techniques in collaboration with interested commercial entities.

Acknowledgments

The authors wish to acknowledge the contributions of Dr. Qin Zhang, Dr. Pingnan Shi, Mr. Julong Du and Dr. Qiaobing Xie of U.B.C. to the technical developments described herein, as well the contributions of Mr. Rob Balsdon, Mr. George Hart and Mr. Nick Hamilton-Piercy of Rogers Cable, and Mr. Dan Pike of Prime Cable, to the development of ideas about the engineering applications of the developments. The work has been funded by the Rogers Canadian Cable Labs Fund and by the (Canadian) National Science and Engineering Research Council.

Footnotes and References

[1] Dr. Rabab Ward is a Professor of Electrical Engineering at U.B.C. (2356 Main Mall, Vancouver, B.C. V6T 1Z4). She is also Chief Scientist of Ward Laboratories Inc.

[2] Donald Monteith is Vice-President, Technology, Rogers Engineering, 853 York Mills Rd., Don Mills, Ontario, M3B 1Y2.

[3] John Madden is the Executive Director of Rogers Canadian Cable Laboratories as well as being the President of Ward Laboratories Inc. His address is: 16th Floor, 4710 Kingsway, Burnaby, B.C. V5H 4M5.

[4] R. K. Ward and Q. Zhang, "Automatic Identification of Impairments Caused by Intermodulation Distortion in Cable Television Pictures", *IEEE Trans. on Broadcasting, vol. 38, no. 1,* March, 1992, pp. 60-68.

[5] Q. Zhang and R. K. Ward, "A Non-Intrusive Scheme for Measuring the Signal-to-Noise Ratio of Television Pictures", *Proc. of the Can. Soc. of Elect. & Comp. Engineers Conf.,* Vancouver, Sept., 1993, pp. 656-659.

[6] P. Shi and R. K. Ward, "Automatic Recognition of Intermodulation Beat Products in Cable TV", *IEEE Trans. on Broadcasting, vol. 39, no. 1,* Sept., 1993, pp. 318-326.

[7] A penetration rate of 63% has been assumed.

CATV NETWORK PLANNING, A SYSTEMATIC APPROACH

A.C.Martelleto, H.P.Pinheiro, M.CarneiroS., P.B.França, P.D.Silveira

Inter Net Comunicações
Av. José de Souza Campos , 304
Campinas-SP BRAZIL 13.092-020

Abstract

This paper describes a proposed methodology for CATV distribution network planning, its mathematical model, and a discussion of some results obtained in this application.
The methodology has been applied in two operations: NET São Carlos and NET Franca (cities in the state of São Paulo, BRAZIL) where there were no CATV services and the networks are in process of construction.

INTRODUCTION

Computational tools used in support of decision-making have enjoyed greater and greater usage in several productive sectors of the economy. Several factors are contributing to the expansion of these computational systems. Among them is an endless increase in computer processing power coupled with a corresponding reduction in cost, the improved applicability of mathematical modeling, and the evolution of solution techniques. The main factor, unquestionably above all else, is the increasing necessity to work with optimized solutions which minimize implementation costs adjusted to specific quality levels.

The search for optimized solutions within systemic approaches has become increasingly difficult as the magnitude of the systems grow and more control variables are included. In these cases, most of the time, it is appropriate to recur to other types of resources. The construction of mathematical models and the development of optimization tools and artificial intelligence for model resolution have become one of the most studied technique around the world.

Networks are among the systems which are most frequently studied. The great number of applications in strategic areas such as electrical power, telecommunications and transportation, associated with the easy and applicable systems representation through graphs and powerful problem resolution algorithms, turn them into excellent objects to be treated by support tools in decision making.

In the CATV area, the outside plant is an important system component representing sizable investments in initial cost and operation. In this context, it is not enough just to identify potential subscribers, install the network and pray that everything works out.

This paper presents a mathematical model and its resolution, using optimization algorithms and knowledge-based heuristics, creating an expert system for the problem of CATV planning.

The methodology is designed to:

(1) allow a systemic approach to the problem;

(2) consider present requirements of cost minimization and quality improvement ;

(3) visualize long term aspects focusing on the market and technology (new services, opticalization, digitalization, etc).

MATHEMATICAL MODEL

The problem is modeled mathematically, through a graph, composed by a set N= { 1,2,... n} nodes and a set M= { 1,2,..., m} branches. Each node represents a utility pole of the area under study and associated to it is a demand composed by the number of potential homes that receive signal from that pole. The branches represent the possible connections between the poles and have as a characteristic their length. They also form the set of possible ways for the routing of trunks and feeders.

Several problems have to be solved through the use of this graph. They include the location of hubs, the routing of fiber optics, the routing of trunk lines, feeders, the control of the signal, and others. If you deal with the problem in a general way, an extremely complex model is created. The problem becomes so complex that one can antecipate major difficulties in determining its solution. Eventually we realize that it is impossible to use techniques of mathematical programming (MP). To soften the diversity aspect of the problem and its dimension, the general problem was subdivided into parts that ended up producing a graph composed of two levels (Figure 1).

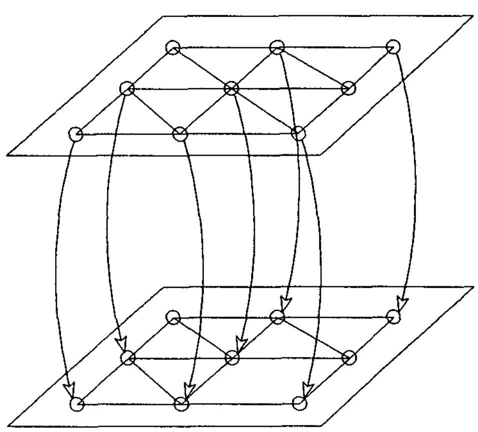


FIGURE 1 Representation of the mathematical model

On the first level, the problems of location of hubs and of routing of fibers are modeled. On the second level, the problem of the design of the network of a cell, which is delimited by the area of influence of a hub positioned on the above level. Since these hubs are the signal feeding points of the cell, connecting arcs are created to link these nodes to their corresponding ones on the bottom level. This will "allow" the entrance of the signal to each of the cells. It is possible to define which nodes will shelter a hub and which will not. For the latter the link by connecting arcs between the top and the bottom levels are not made. In Figure 1, for example, two of the nodes will most probably not shelter a hub. In the final solution, only the arcs that represent chosen hubs will be present.

The great number of variables, along with concavous cost functions (installation cost), make it difficult to approach the problem through MP. A resolution to the problem is accomplished in a hierarchical form. First, the top level problems are solved (location of hubs and routes of fibers), and then the design of each cell is accomplished. In order to relate the solution obtained on the top level with the one of the bottom level, producing a complete solution, a heuristic methodology is proposed. This methodology also approaches the problem of the number of hubs.

First, taking a look at the top level, the problem is modeled as an incapacitated location problem that defines the location of hubs and their areas of influence. The fiber optics routes that will be used to feed the hubs are modeled through a minimal spanning tree problem.

On the bottom level, the design of each of the cells is accomplished. At first, the design is done with a topological view that seeks to "discipline" the network and to generate an initial solution. It can, however, be unachievable or present poor results with regard to RF signal level.. To improve this resolution, heuristic changes based on engineering rules (expert system) that propose to allocate equipment and redefine the "microcells", taking into account the signal level, are being developed.

The topological solution consists of subdividing each of the influence regions of a hub into microcells, making the routing of feeder lines in each of them and defining the feeding points of each microcell through the trunk line. The number of generated microcells follows the criteria of number of homes and maximum distance of servicing.

The division into microcells is modeled in a similar way to the problem of hub locations. The routing of the secondary feeders is done through a shortest path problem, while the connection of the primary network to the microcells is modeled as a Steiner problem.

RESOLUTION METHODOLOGY

The resolution methodology is composed of several heuristic procedures and consists of increasing the number of hubs by 1 and producing a thorough solution. It is represented in Figure 2.

The following operation provides an estimate of the initial number of hubs:

$\lceil ST/SC \rceil = Hi$

where
ST : means the total number of homes
SC : the maximum number of the homes for one cell
Hi : initial number of hubs
and $\lceil . \rceil$ represents the minor integer higher than the quotient of the division.

The initial number of hubs is disposed in the graph so as to minimize the criterion demand x distance. This way, the aim is to place the hubs at the baricenter of the "charge". As the process of location the hubs occurs, the area of influence of each one of them is defined. Some cells, however, are likely not to respect the maximum number of subscribers. That being the case, one rule is created which allows the process to continue or to increase the number of hubs by 1, providing it with a new division. Due to the criterion used at this part of the program, the cells with many homes tend to suffer new divisions. The generated rule consits basically in checking if the "excess" of homes is diluted through the cells that burst the limit or is concentrated in some specific cells. If the problem is that of dilution, the process is allowed to continue. Otherwise, the number of hubs is increased. It is helpful to observe that the decision maker itself is able to tell the program to continue or to increase the number of hubs. After the hubs are located and the areas of influence defined, the process continues. The hubs are fed from the headend using fiber optics. At this point, the criterion is to minimize the length of fiber while making the direct connection between the points. This process shows a sequence of connections (the fiber´s route) and defines the number of pairs of fiber in each segment. To make this system reliable, it is required that the head-end degree be $\geq$ 3 at the generated tree. In other words, at least 3 branches must leave the tree at the head-end. In the future, features of disjoint paths and rings are likely to be included to feed each separate hub, which would make it significatively more reliable.

Having reached that solution, it became possible to define which connecting arcs are active, that is to say, which ones emit signal to a specific area. From this point on, the network design takes place.

The topological solution for the design at one of the cells consists of subdividing the cells into microcells, all with a similar number of customers, and all customers within a pre-established distance limit. As in the case of hub locations, these two conditioning aspects represent a decisive factor in the number of microcells to be generated. However, the network installation cost, based on the amount of cables used and equipment cost, becomes an important factor.

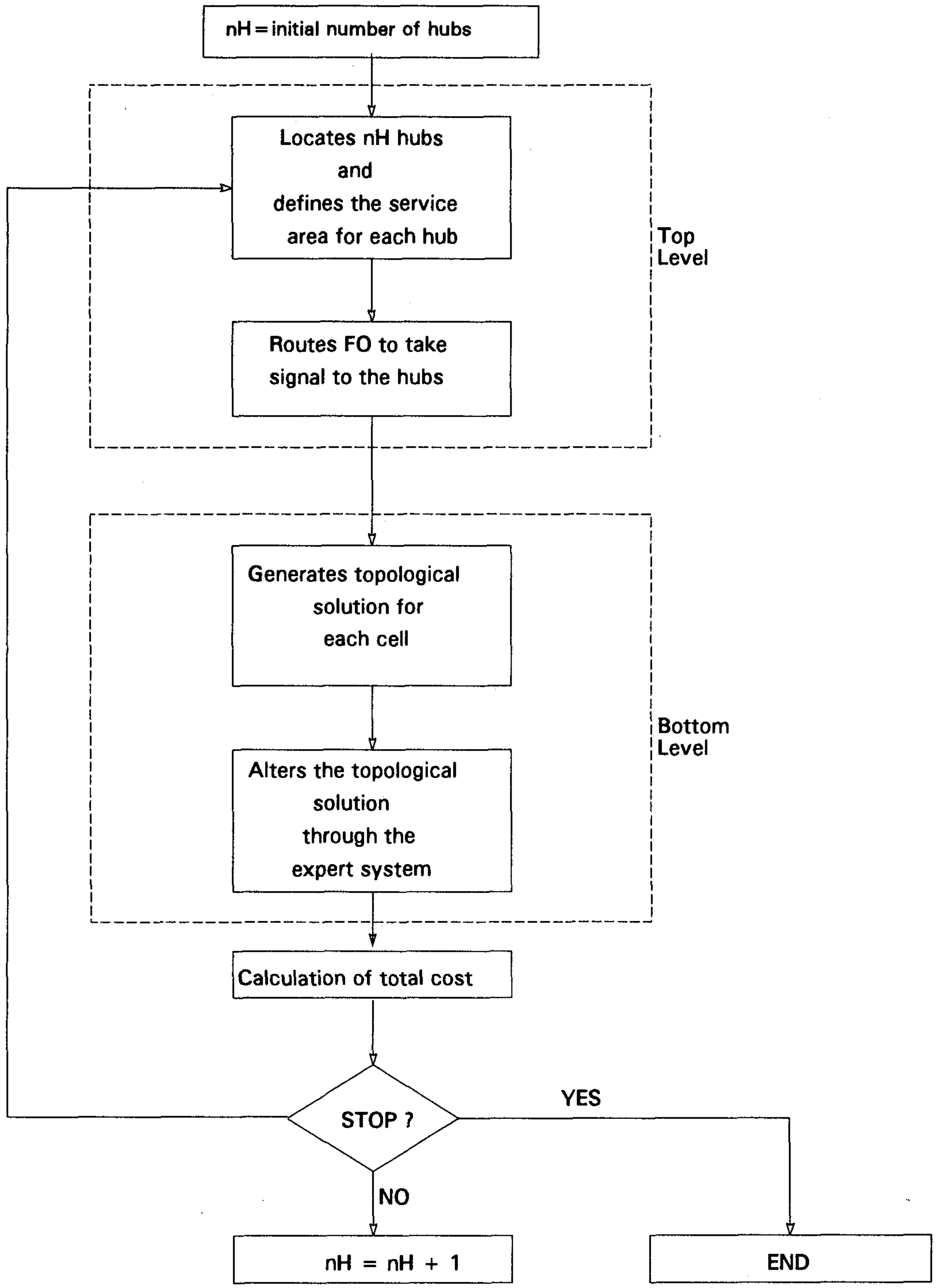


FIGURE 2 Methodology represantation

One might expect that the greater the number of microcells the less one should spend on distribution and the more one would spend on trunk. From the commitment to these figures and the obedience to other rules -- the number of subscribers, for instance -- the solution is reached. In fact, this cost will only be known after the second step of the process, which consists of changing the solution to assure adequate signal level to all potential consumers and the wise use of installation equipment, departing from the topological solution by the means of engineering rules and heuristics of change.

It is as if in the topological solution the network were totally passive having at each microcell a point with a signal level equal to 1. The distribution takes place from that point on. The point where the signal is equal to 1 is fed by a trunk. It is likely that there is a need of facilities to guarantee the signal level. Such an equipment allocation is handled by an expert system that tries to change the microcells frontiers for better equipment operation without departing too far from the topological solution which has the lowest cable cost, theoretically.

Having accomplished the design, the final cost of the best solutions (which depend on the number of microcells) are presented to the decision maker.

When all the designs are accomplished, a global cost is estimated and that will be the test for methodology stop.

RESULTS OBTAINED

The described methodology has been applied by Inter Net, MSO in Brazil. It is not yet consolidated, but its development is in progress. It was applied to two cable operations, in which the FSA network topology was adopted.

For a proper evaluation we followed two lines of application of the methodology: one group applied a conventional approach to the network project and the other the tool herein described. The immediate results of a comparative evaluation show that :

- investment: to adopt the methodology it is imperative the outside plant data availability be in digital form, which today is an imposition. In relation to the computational environment, we have adopted PC standards, available at any operator office.

- performance: the tool allows a simulation of several settings and presents immediate results whereas in the conventional approach it is difficult to change the settings and the obtention of a complete design may take days depending on the availability.

- network costs: the most prominent gain from systemization was in the reduction in distribution cable needed. In some cases a reduction of ten percent in the amount of cables occurred however, it did not result in a reduction in the equipment aspect. There was also an increase to the trunk line and in terms of equipment the amount used was not affected. At the adopted approach, we believe that the knowledge based system with intrinsic rules to the equipment installation is the point that we should focus from now on. As it has already been mentioned, the expert system incorporates the rules defined by its designers and the resolution capacity is related to the endurance of the adopted rules..

- modularity: the systemic approach presented a rather uniform topological solution, which was translated to very similar microcells and with a better servicing in terms of signal to all homes. This feature is very interesting in terms of a future network partitioning with a larger use of fibers, so to speak.

At this moment many hypothesis about the future of the networks are being formulated. Within this context, the engineering teams try to foresee which necessities should be contemplated. However, this difficult task is frequently accomplished based on suppositions. Technologies that some months ago were considered for the future, today are called technologies of the present. As the global economy relies more and more on the movement of information, the decisions have to be made faster and with greater reliability.

CUSTOMER PREMISES EQUIPMENT PERFORMANCE AND COMPATIBILITY

Claude Baggett
Cable Television Laboratories, Inc.
Boulder, Colorado

Abstract

In part 17 of the Cable Act of 1992, the congress made provisions for the specification of those features of television sets, VCRs, and cable converters which would determine whether they could truly be considered "cableready" or not. The primary performance shortfall experienced by most consumer hardware which made them unsuitable as a cableready unit involved their susceptibility to direct pickup interference (DPU). However, in addition to DPU, there were a number of other tuner performance factors which were important in this definition. CableLabs performed certain work to size the DPU problem, to measure the actual performance of a number of types of consumer premises equipment related thereto, and to determine a threshold level for the perceptibility of such interference. In addition, CableLabs devised test procedures for measuring both DPU and the other tuner performance factors, and tested an additional number of consumer and cable units to develop a state of the industry in these performance areas. This report subsums these efforts and their results.

1.0 GENERAL SUMMARY

1.1 Background

To facilitate the viewing of any television programming over a cable system, a long, multi-industrial chain of processes must take place in a fortuitous manner. A television segment begins with the creative talent and proceeds through the production, distribution, financial, and legal engines, subsequently to be delivered to the over 10,000 cable television headends in the United States. These various operational elements perform their efforts in relative harmony with well-defined and cooperative interfaces, considering the complexity of the system. But once the signals pass through the cable headend and are transported to the customer home over the cable system, there is yet one additional interface which must be bridged, that being between the cable delivery system and the consumer electronics display devices which are the last parts of the electronic system before the visual and aural signals are available to the human receptor, the object of the whole system. This interface has been problematic since the very early days of cable, about 50 years ago, is still so today, and provides the rationale for the research and measurements documented herein.

There has been considerable effort over the last decade to find a solution to the disharmony at the consumer interface through joint negotiations between the consumer electronics and cable industries. While some progress has been made, there are still many issues yet to be resolved. In fact, Section 17 of the Cable Television Consumer Protection and Competition Act of 1992 ("Cable Act of '92") specifically addresses many of these unresolved problems.

The efforts described in this document are a direct result of these long-standing interface problems and the guidance given in the Cable Act of '92 as to their solution. The EIA/NCTA Joint Engineering Committee ("JEC") has within its purview, responsibility for specifying the performance characteristics of a tuner which would be able to receive cable signals without the degrading side-effects exhibited by the broadcast-oriented tuners found in all consumer products today. In this committee, the absence of two primary data items has blocked real agreement between the main parties, that being the sizing of the problem as to the consumer population impacted by direct pickup interference (DPU), for instance, and the development of a necessary and sufficient specification for the performance of the tuners themselves. Related to both items is the development of acceptable baseline test procedures for each of the factors considered.

Recognizing the difficulty of developing the above extensive materials by such volunteer committees as the JEC, Cable Television Laboratories, Inc. ("CableLabs"), as directed by its member companies and in cooperation with the NCTA and the EIA, moved to fund the necessary projects to develop the data required to free the JEC to act.

In January, 1992, CableLabs asked Stern Telecommunications Corporation ("STC") of New York City, NY to undertake a study to answer the first question, that relating to the extent and degree to which television receivers in the United States are subjected to various levels of co-channel interference in the cable delivered picture through the mechanism of direct pickup interference. STC was selected because of its interest in and understanding of the problem, and its long-standing and even-handed relations with the broadcast, cable, and consumer industries. As a check of the STC results, the EIA hired Jules Cohen, a respected consultant in the Washington DC area, to perform a similar analysis. The Cohen results were virtually identical to the STC answers, except as to the percentage of CPE which resides in 1 volt/meter fields or greater. The STC analysis indicates about 6% of CPE reside in such field intensities, whereas the Cohen study shows just less than 2%. Even in this, however, there is no significant difference.

In December, 1992, CableLabs issued open solicitations addressing two of the other issues. The first solicitation called for the development of baseline test procedures for determining the susceptibility of customer premises equipment, including television receivers, videocassette recorders, and cable converter units, to the effects of radiated and conducted direct pickup interference, and the testing of a sample of such CPE to determine the general state-of-art. The second solicitation covered the same tasks, but as related to the deteriorating effects of other receiver performance characteristics, such as the re-radiation of cable signals, local oscillator leakage and backfeed, A/B switch isolation, DPU backfeed, VCR through-loss, adjacent channel rejection, image rejection, and tuner overload performance. After evaluation, both contracts were awarded to the Carl T. Jones Corporation ("CTJ") of Springfield, VA. Further discussion re the irradiated and conducted DPU, and the other listed performance factors can be found in sections 1.3 and 1.4 of this Report.

Finally, in August, 1993, CableLabs asked Dr. Bronwyn L. Jones, a much respected researcher relative to the psychophysical effects on the viewer of perceptible degradation in the television picture, to relate the objective measurements of DPU susceptibility made at CTJ to the actual impact as to perceptibility in the television viewer.

1.2 Statistical Model

While there was a great deal of anecdotal data regarding the cost of cable's response to DPU complaints by subscribers, mainly based on service calls where DPU was the cause, and/or the number of non-descrambling converters deployed specifically to address this problem, there was no definitive data set which allowed the consumer manufacturers to analytically assess the magnitude of the impact of this problem. Frequently requested by the consumer caucus of the JEC was a chart which would apportion television households in the United States as a function of the field strength of the UHF and VHF off-air signals in which they resided. This data is required, so that when combined with that developed in the other efforts contained herein, the DPU problem could be properly sized and an appropriate response planned. This is the thrust of the modeling and verification accomplished by Stern Telecommunications Corp. ("STC") under contract to CableLabs.

The procedure used by STC was to combine literature searches, computer modeling, laboratory measurements, and some specific field verifications to develop the data represented in the summary histogram in figure 1.2.1 below.

The geographic areas evaluated in this study consisted of the top ten television ADIs, which represents approximately 30% of total US households. The varied demographics and physical attributes of these ten ADIs permitted the data to be extrapolated to represent all urban and suburban television homes in the US.

Iso-contours representing varying VHF and UHF field strengths were calculated by the model and the numbers of television homes within each level were developed. Other factors, such as the shielding effectiveness of the CPE, the location within the dwelling, the orientation of the unit to the interfering transmitter, the effects of building shielding, and urban-suburban clutter are part of the CTJ study and not contained in this effort.

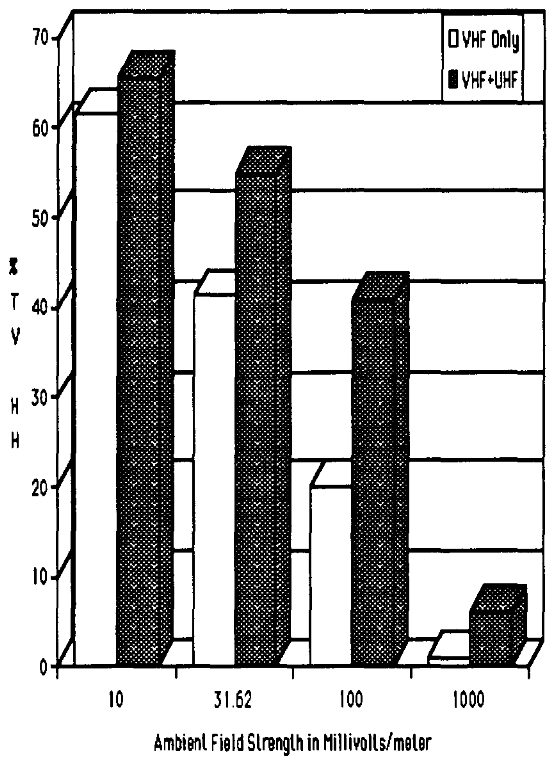


Figure 1.2.1 Allocation of TV households in the STC study as a function of the ambient field strength in which they reside (all US TV HH).

To sum up this chart, it represents that some 65.4% of all US TV households reside in fields of 10 mv/m or greater, that 54.8% of all US TV households reside in fields of 31.62 mv/m (90 dBu) or greater, that 40.8% of all US TV households reside in fields of 100 mv/m or greater, and that 6% of all US TV households reside in fields of 1 volt/meter or greater.

1.3 Direct Pickup Interference Determination

Direct pickup interference, or DPU, exists when external co-channel signals ingress into the tuner in CPE and interfere with and degrade the signal delivered by cable. This ingress may be in the form of an electromagnetic field irradiating the CPE, or in the form of signal currents entering the unit on the power cord or the braid of the coaxial cable connected to the CPE. A need was felt to determine the current state-of-art of consumer grade television receivers, videocassette recorders, and converters relative to their susceptability to direct pickup interference. It was noted in this testing that the tuners in consumer electronics hardware seemed to be sensitive to the differential voltage between the braid and the center conductor on the drop cable. This would be consistent with a tuner designed specifically for broadcast signal reception. The tuners in the cable converters typically have a single-ended input from the center conductor of the coax. The results of testing performed which is listed below consider both irradiated and conducted signals.

Technically, any kind of signal not delivered on the cable which is of the proper frequency and amplitude could interfere with and degrade the video. There are two general types of interfering signals, coherent and non-coherent. Coherent signals are those which originate from the same original source, such as a television transmitter. Non-coherent signals are those which come from other sources, such as business radio transmitters, and leakage from other household appliances. The chart in figure 1.2.1 depicts the field strength levels ambient to typical households in the US resulting only from VHF and UHF broadcast television transmitters, thus being coherent signals.

The CableLab's effort at Carl T. Jones Corporation, a nationally recognized and honored testing laboratory which has been conducting programs such as this since 1936, existed in part to measure the susceptibility of television receivers, videocassette recorders, and cable converter boxes to both irradiated and conducted DPU. The units tested included the numbers shown in figure 1.3.1 as received from a sampling of manufacturers ranging from the highest to the lowest in marketshare.

Television Receivers	**37**
Videocassette Recorders	**11**
Cable Converters	**14**

Figure 1.3.1 Sample universe of CPE tested.

In CableLab's arrangements with the EIA regarding the DPU test program, it was agreed that we would not make public the performance of any particular brand or model tested. Rather, the testing results from DPU, as are shown in figure 1.3.2 below indicate only the best performing unit tested, the poorest performing unit tested, and the median, or that set, if it existed, where half the units performed better and half worse.

The criteria used in the CTJ tests recognizes the generally accepted point of perceptibility for interference from coherent signals; notably when the interfering signal reaches an amplitude 55 dB below the sync tip of the desired video signal. This point is further discussed below. Thus, the results shown in figures 1.3.2 and 1.3.3 below represent that point when an interfering signal will be just perceptible to the average viewer observing a television receiver or as displayed from a VCR with that level of interference, or as processed by a cable converter with that level of interference.

Further, the numbers below, as with the STC chart above, do not take into consideration any mitigating factors, such as placement in the home, the degree of building attenuation applicable, urban/suburban clutter, unusual height above ground, ducting possibilities, the orientation of the set relative to the transmitter of the interfering signal, or the type of programming being displayed. These factors will be discussed in our summary conclusions which are found in section 1.6 below and relate to the performance specifications of CPE as described in part 17 of the Cable Act of '92, and in IS-23, which is being developed in the JEC.

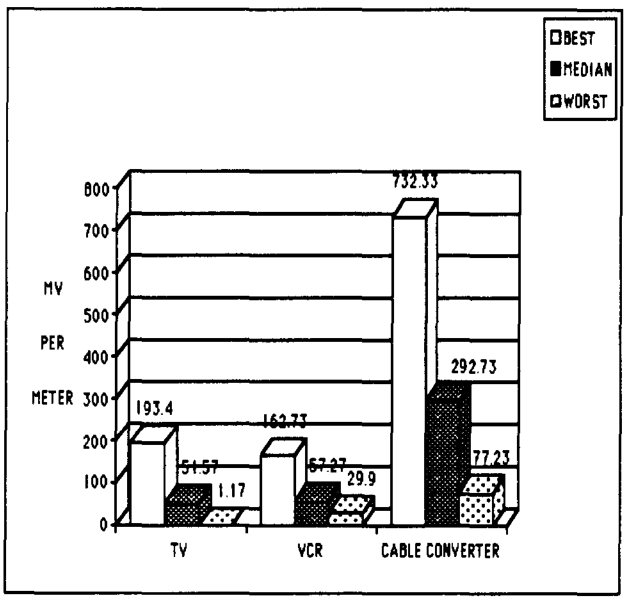


Figure 1.3.2 Ambient field strength in millivolts/meter to achieve the -55dB point of perceptibility averaged over channels 6, 12, and 78.

There are several interesting points which are contained in this histogram. First it must be noted that all CPE exhibit a considerable spread between the best and worst units, television sets being about 165 times, VCRs about 5.4 times, and cable converters being about 9.5 times. This spread must be considered when setting minimum specifications for DPU performance in CPE. Secondly, since it appears that cable converters perform well enough on average to cure the DPU problem, then improving the TV and VCR performance numbers to match those of the converters would be adequate, but only if the performance span is considered in setting the acceptable range. One of the differences here is that cable technicians often work with several converters until one is found whose performance is equal to the problem, a capability not really available or practical for the consumer electronics customer trying to purchase a new TV or VCR. The conclusion that might be reached is that TVs and VCRs need to improve their worst case performance to a level comparable with the media converter.

In Figures 1.3.3 and 1.3.4 below, the performance of the CPE in the presence of conducted signals on both the coaxial signal cable and the power cord were measured. These tests were performed utilizing signals on channels 6 and 12, and measured at the -55 dB point of perceptibility. Among the test items there did not seem to be a consistent susceptibility to a particular mode of conducted DPU.

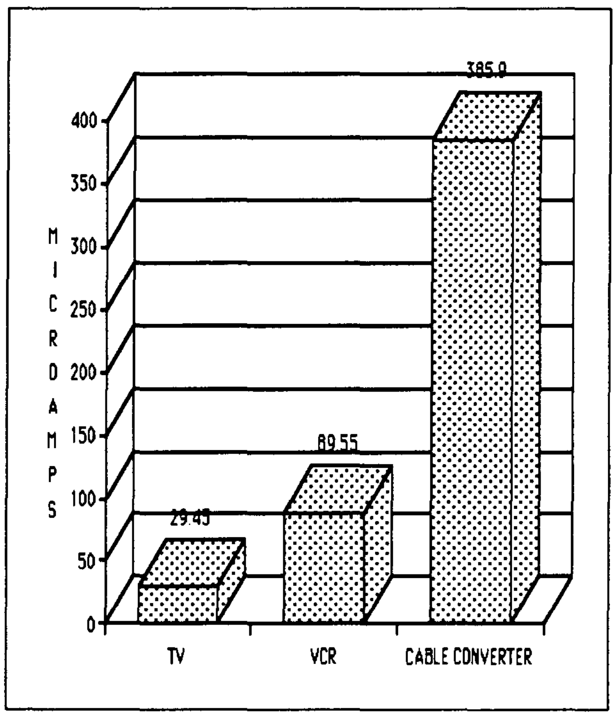


Figure 1.3.3 Susceptibility of CPE to conducted coherent direct pickup interference through current on the coaxial cable braid at the -55dB level of perceptibility, channels 6 and 12 averaged.

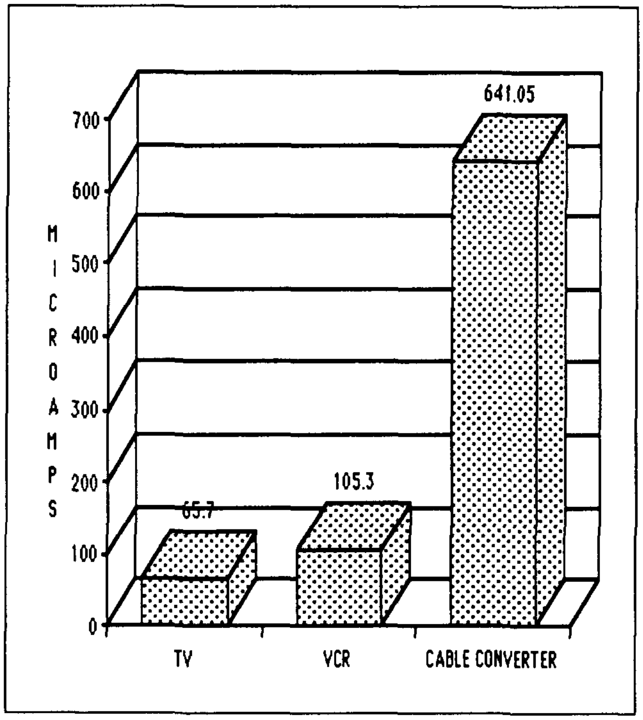


Figure 1.3.4 Susceptibility of CPE to conducted coherent DPU on the power cord at the -55dB level of perceptibility, channels 6 and 12 averaged.

In considering this data, one must remember that it is generally accepted that all consumer electronics hardware sold today have excellent broadcast or off-air tuners, whereas the cable converter has a tuner which is designed and oriented toward cable delivery and not toward the reception of off-air signals. The broadcast tuner is designed to be exceptionally sensitive to signals derived from all sources including those which are conducted, thus making it a perfect conduit for DPU when the unit is connected to a cable television delivery system. Therefore, if a piece of consumer gear is intended, labeled, promoted, and sold to be used only with broadcast signals, or only with a cable converter placed before it in the signal stream, the current tuners in the television receivers are sufficient.

However, if the consumer gear is intended to be usable with cable systems either without a cable converter or with a decoder interface, then an accommodation must be made when connected to such systems to remove the interfering DPU signals regardless of their source of ingress, whether through irradiated or conducted modes, to a level which will not be perceptible to the customer.

1.4 Receiver Performance

In the past, the FCC has been reticent to impose any performance standards on consumer grade hardware, preferring to leave that as a marketplace issue. Since cable has become the marketshare leader in the delivery of video in the United States, certain factors at the retail point-of-sale has caused the consumer to become confused as to the expected performance of the intended purchase. In all but a very small percentage of high-end consumer electronics stores, the television receivers are all displayed with the exact same material delivered primarily from a nearby laserdisk player or other such source, and/or broadcast signals. In this most common case, the consumer cannot ascertain the consumer electronics' compatibility with cable (which is much different than the other sources), at the time of purchase, because the stores do not generally provide access to the local cable system. Therefore, the selection of consumer hardware based on their compatibility with cable delivery systems is not currently resolvable at the point of sale. For this reason, there are several other tuner performance issues which need to be specified to avoid confusing and misleading the consumer.

In devising the following tests procedures, eight parameters of CPE performance were considered. These parameters are taken from those listed in the IS-23 performance document which is currently being crafted in the EIA/NCTA Joint Engineering Committee (JEC) through a bi-industry subcommittee. These test procedures were selected after consideration by CTJ of existing practices for performing such tests among consumer and cable manufacturers both domestic and foreign, and as discussed in extensive technical documents. Considerable time was spent in proving the test procedures by varying the facility elements and finally in retesting many of the same products used in the DPU considerations as a sort of "test of the test." While the procedures must be represented as unapproved as yet by the consumer caucus of the JEC, the exhaustive certifications performed at CTJ has convinced CableLabs and the cable caucus that any substantive and supported changes suggested will vary the results in only a very minor fashion, and not the scale or proportion of the results in any way.

The goal of this work at CTJ was two fold. First, it was anticipated that the test procedures developed for each of these performance parameters will be canonized, or modified and canonized, or equivalent alternatives procedures suggested, these becoming the benchmark techniques for measuring each of the eight parameters. Secondly, it was anticipated that the actual performance numbers developed will give a good measure of the state-of-art currently existing among the consumer electronics and cable equipment manufacturers relative to these factors. Note that the work at CTJ was not designed and is not meant to show preference to any particular consumer product, or between the consumer manufacturing or cable industry view of the interface, but rather to build a foundation of commonly accepted procedures, technology, and information from which meaningful remedies for problems at the interface can be developed.

1.4.1 Re-radiation of Cable Signals

The cable industry is required under the Cumulative Leakage Index (CLI) performance standard set by the FCC in Part 76.605 (a)(12) of their Rules and Regulations to limit the leakage or the egress of RF energy out of the delivery system back into the environment. As cable has complied with this directive, concern has been expressed as to determining what contribution, if any, CPE might make toward the CLI through the leakage and reradiation of cable signals. Note that this does not imply that CPE is currently required to meet the Part 76 standards. Nonetheless, this federal standard limits the radiation from a cable delivery system to 15 microvolts/meter at a distance of 30 meters for frequencies less than or equal to 54 MHz or greater than 216 MHz. Similarly, limits are placed at 20 microvolts/meter at 3 meters between 54 and 216 MHz. The test measurement system, shown in Figure 1.4.1.1 below, was designed and calibrated to have sufficient sensitivity to measure signals in light of the FCC requirements.

Only two of the 56 CPE tested, both television receivers, presented any re-radiation of cable signals above the FCC limits. All units tended to re-radiate higher levels of signals as frequency increased, but all were well within limits except these two units, both of which would have failed the Part 76.605 (a)(12) cable CLI emission limits on channels 37 and 53, if it were to be applied against CPE.

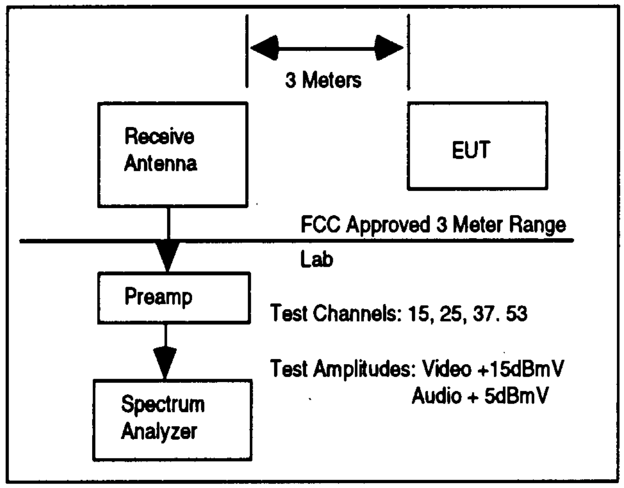


Figure 1.4.1.1 Test Configuration for Re-Radiation of Cable Signals.

1.4.2 Local Oscillator Leakage and Backfeed

In the single conversion tuners found in most consumer electronics, there is a single local oscillator (LO) whose signal is used to heterodyne the incoming RF television transmission down to an intermediate frequency which is in the 40 MHz region. In converters built for the cable industry, a double conversion process is utilized, requiring two such oscillators, each being isolated from the other and from the balance of the circuitry. In the single conversion tuners in most consumer gear, the local oscillator frequency falls within the cable delivery band, while the frequencies used in the cable converters typically fall outside the band. It is recognized that in the future, consumer manufacturers may exercise some changes to the frequency of these signals, and this will require further investigation. The test configuration is shown in Figure 1.4.2.1, below.

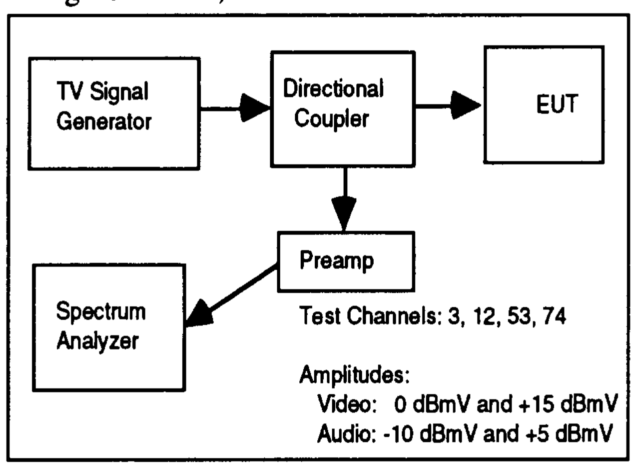


Figure 1.4.2.1 Test Configuration for Local Oscillator Leakage and Backfeed.

In the in-band case, if the emanations from these oscillators are conducted out of the tuner package on the interconnecting wiring, they serve as a source of degradation to other CPE devices connected nearby, and if sufficiently strong, to traverse the cable drop and ingress into the cable system as noise, or through the cable tap to the drop of a neighboring subscriber as interference. The degree to which these local oscillators are isolated to within the tuner package is a performance factor of interest at the interface, and is part of the IS-23 deliberations. In this test, any spurious emissions which exceed -35 dBmV within the spectrum up to 600 MHz were recorded. The worst case situation is where the LO in the TV being generated while viewing one channel degrades the reception of another channel being recorded on the co-located VCR, or vice-versa.

This phenomena, like many others in these investigations, seems to be directly proportional to the frequency of the channel tuned. Television receivers showed no emanations above the recommended level on channel 3, but numerous instances of problem levels on channels 12, 53, and 74. Specifically, 17% of the test items had problems on channel 12, 31% on channel 53, and 60% on channel 74.

Videocassette recorders were quite similar, showing 13% having problems on channels 3 and 12, 26% on channel 53, and 50% on channel 74.

Cable converters showed no emissions above the recommended level on any channel, but did follow the precedence set by the other CPE in scoring somewhat worse on the higher frequencies.

1.4.3 A/B Switch Isolation

All VCRs, some TVs, and some cable converters are equipped with an A/B switch to provide for the selection of more than one input source to the CPE for use. In the case of the VCR, the choice is between the tape playback unit in the VCR or passing the incoming cable signal on to the television receiver.

If insufficient isolation exists between the two external ports, or between any unselected external port and the common port, then a condition of signal crossfeed will exist, which may result in degradation to the viewed or recorded signal. The FCC requires in Part 15.606 of their Rules and Regulations that A/B switches present a minimum of 80 dB of isolation for frequencies between 54 and 216 MHz, and 60 dB from 216 to 550 MHz. The external mechanical A/B switches currently in use by the cable industry all generally exceed 90 dB of isolation under any of the above operating modes at any frequency, and often exceed 100 dB at any frequency.

It was deemed important to know what the state-of-art is for A/B switches among currently manufactured CPE. This did not seem to be a great problem among the television sets tested, having only 3 instances where the results did not meet Part 15.606 on the 4 channels tested. In fact, 75% of the receivers had no problem on any frequency. The test configuration is shown in Figure 1.4.3.1 below.

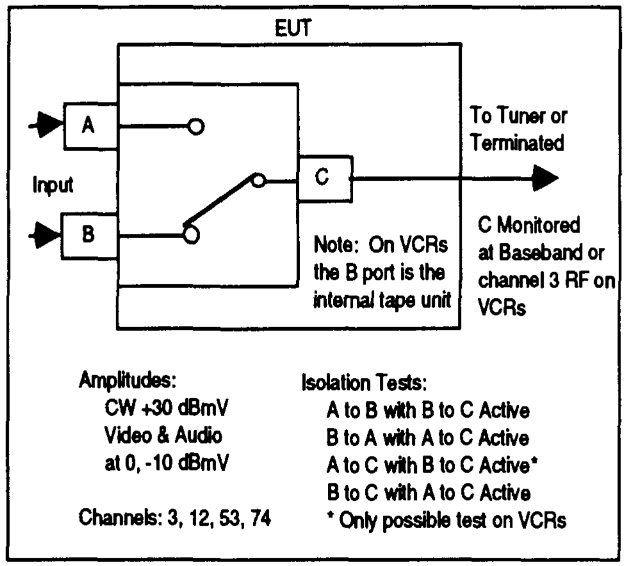


Figure 1.4.3.1 A/B Switch Isolation Test Configuration.

1.4.4 DPU Backfeed

This represents the condition wherein CPE which is susceptible to DPU not only causes its own picture to be degraded, as explained in Section 1.3 above, but also allows these ingressive signals to be fed back up the coaxial cable, thus serving as a degrading factor to other hardware in the home, or customers on the cable plant. The values measured represent the voltage present at the input port of the CPE while it is being irradiated with a 100 millivolt/meter field at the unit's point of maximum susceptibility, as explained in the DPU section. Tests showed that the DPU backfeed is a function of the impinging radiation, and not related to the channel tuned on the EUT. The FCC requires a minimum of 18 dB of isolation between customers, but standard practice shows a practical isolation figure to be about 22 dB. If the -55 dB D/U ratio is the point of perceptibility, then a -33 dB level out of the RF input of the EUT is sufficient to reach the threshold of perceptibility. The test configuration is shown in Figure 1.4.4.1 below.

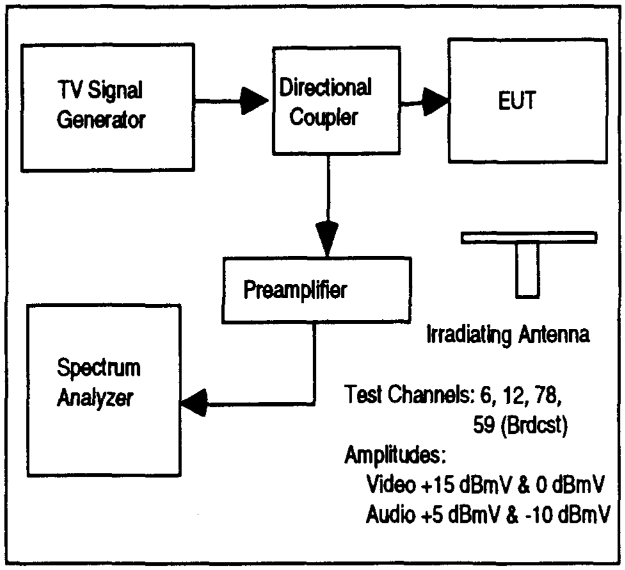


Figure 1.4.4.1 Test Configuration for DPU Backfeed.

The results show that 37% of television receivers had DPU backfeed above the point of perceptibility on channel 6, followed by 34% on channel 12, 26% on channel 78, and 17% on channel 59. In general, there was a high correlation between those television receivers which tested poorly on DPU, and those which did poorly in DPU backfeed. Interestingly, no videocassette recorders or cable converters displayed DPU backfeed above the level of perceptibility.

Further tests of spurious backfeed were conducted to look for any emission from the EUT present at the cable fitting in the 54 MHz to 550 MHz range which might cause interference in connected CPE. Any signals greater than a -35 dBmV were recorded. The emissions found in television receivers all seemed to be related to the 40 MHz IF frequency, but in no case were they above the -35 dBmV level. No such emanations were found in videocassette recorders. All converters except one, a commercial aftermarket tuner-only product, showed no significant emanations. The failed converter showed a very strong -1.7 dBmV backfeed at 144 MHz when tuned to channel 12, and -2.8 dBmV backfeed at 336 MHz when tuned to channel 53. It was below the test facility sensitivity on the two other channels.

1.4.5 VCR Through-Loss

The usual and accepted configuration for introducing cable to an arrangement in the home having both a TV and VCR is to route the cable drop to the cable converter, thence out of the converter to the VCR, and thence to the television receiver. In cable systems where a converter is not required, the usual practice is for the drop to be connected to the VCR input port, with the signal path thence traveling to the TV. If there is excessive signal loss in passing through the VCR, a noisy picture may be presented to the viewer. This probably is not an issue at the FCC prescribed input levels if the division of the input signal energy available on the cable is performed in a symmetrical fashion, with half going to the VCR and half to the balance of the system. However, it has been noted that some VCRs perform asymmetrical splitting of the signal, diverting a disproportionate amount of the signal energy internally to the recording unit, thus obviating the cost of some amplification within the VCR. The test configuration is shown in Figure 1.4.5.1 below.

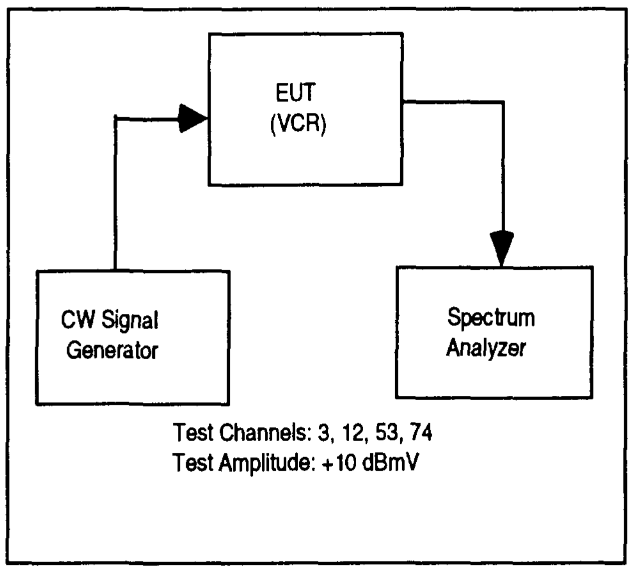


Figure 1.4.5.1 VCR Through-Loss Test Configuration.

It was decided to perform a VCR through-loss measurement on our sample set to determine the state of art in the industry today. Of the eight VCRs tested, all were contained in the range of 3.4 to 5.0 dB through-loss, as tested over channels 3, 12, 53, and 74. The average through-loss for the eight samples over the four channels was 4.05 dB.

1.4.6 Adjacent Channel Rejection

The ability of CPE to reject the in-band noise contribution from channels adjacent to the tuned or desired channel is an important factor in the performance of receivers. It is primarily the lower adjacent aural and color subcarriers, and the upper adjacent video carrier which create the problem. This issue has been addressed in the broadcast industry by leaving adjacent channels vacant, where possible. However, in the cable industry, where virtually all channels have upper and lower adjacents, and at relatively high levels, as specified by the FCC, adjacent channel interference becomes a very important performance factor having great impact on picture quality. The test configuration is shown in Figure 1.4.6.1 below.

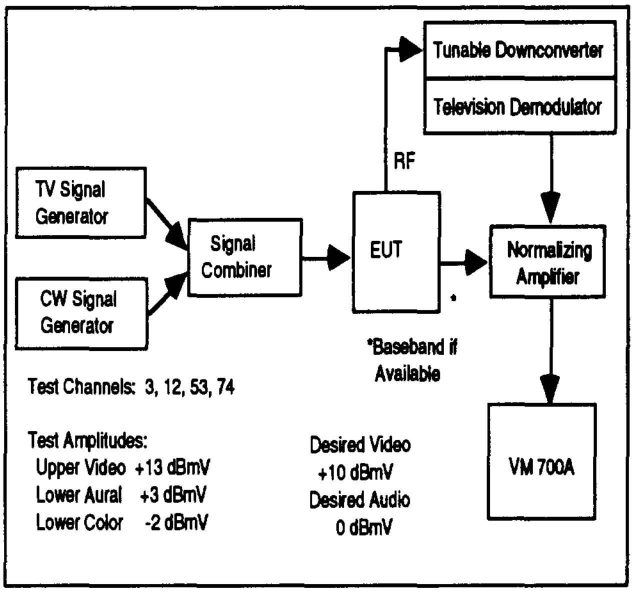


Figure 1.4.6.1 Test Configuration for Adjacent Channel Rejection.

The measurements, were divided into three parts, that for the lower adjacent color subcarrier, that for the lower adjacent aural carrier, and that for the upper adjacent channel video carrier. Again, the -55 dB D/U ratio was used as the threshold of perceptibility for degradation in the viewed picture. The desired video carrier was set at +10 dBmV. The upper adjacent video carrier was set at +10 dBmV, the lower adjacent color subcarrier at -2 dBmV, and the lower adjacent aural carrier at 0 dBmV.

Television Receivers - Lower Adjacent Color Carrier Rejection. Three of the 35 television receivers showed a failure in lower adjacent color subcarrier rejection, all on a single channel only, one unit on channel 12, the other two on channel 53. All of the other numbers for the television receivers were quite respectable, resulting in averages for all receivers at least 10 dB in margin better than the point of perceptibility.

Videocassette Recorders - Lower Adjacent Color Carrier Rejection. One VCR failed the lower adjacent color subcarrier rejection test on all channels, averaging a -49.9 dB D/U level. Two of the eight test units failed on channel 74 by very small margins. All other numbers were acceptable, and the averages for all test items were at least 5 dB better than the level of perceptibility.

Cable Converters - Lower Adjacent Color Carrier Rejection. No cable converter exhibited insufficient rejection of the lower adjacent color subcarrier to the point of perceptibility on any channel.

Television Receivers - Lower Adjacent Aural Carrier Rejection. Two of the 35 units had insufficient lower aural rejection on all channels while two other sets were perceptible only on channels 53, and 74. The average D/U for all 35 receivers tested on all channels was at least 9 dB better than the level of perceptibility.

Videocassette Recorders - Lower Adjacent Aural Carrier Rejection. Three of the eight VCRs tested showed no perceptible interference on any of the four channels tested. Three of the test items failed on all four channels. One VCR failed on channels 12, 53, and 74, while the final unit failed only on channel 74. However, there was little or no margin in the averages over the four channels above the point of perceptibility, the highest being only 1.8 dB above the 55 dB ratio.

Cable Converters - Lower Adjacent Aural Carrier Rejection. None of the cable converters exhibited insufficient rejection of the lower adjacent aural carrier.

Television Receivers - Upper Adjacent Video Carrier Rejection. No television receivers exhibited insufficient rejection of the upper adjacent video carrier.

Videocassette Recorders - Upper Adjacent Video Carrier Rejection. No VCRs exhibited insufficient rejection of the upper adjacent video carrier.

Cable Converters - Upper Adjacent Video Carrier Rejection. No converters exhibited insufficient rejection of the upper adjacent video carrier.

1.4.7 Image Rejection

The effects of insufficient image rejection is a noise contribution to the desired channel similar in nature to adjacent channel interference. However, it is the signals occupying the upper portion of the channel which is 14 channels above the desired, and the lower portion of the channel which is 15 channels above the desired channel which contribute the noise energy.

For NTSC, it is the aural carrier 14 channels above, and the visual channel 15 channels above which combine with the desired signal to generate the distortions. The FCC requires on cable that visual carriers measured at the CPE vary no more than 10 dB, plus 1 dB for each 100 MHz of bandwidth above 300 MHz. Therefore, in our test 550 MHz spectrum, there can be a variation as high as 13 dB. In the tests, the desired video carrier was set at 0 dBmV and the image carrier at +13 dBmV, according to the above. The test configuration is shown in Figure 1.4.7.1 below.

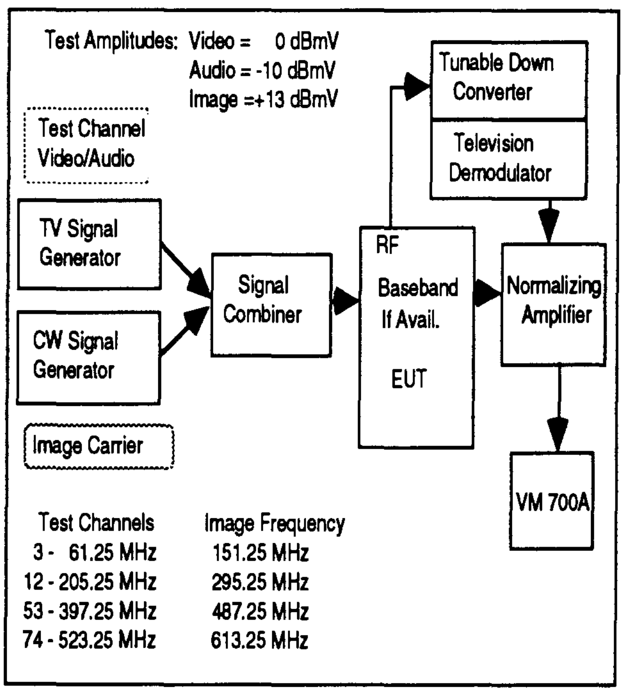


Figure 1.4.7.1 Test Configuration for Image Rejection

A review of the results shows that the image rejection performance of television receivers varies significantly according to the selected channel. On channels 3 and 12, the average rejection was found to be 66 dB, while on channels 53 and 74, the average image rejection was 58.5 dB. None of the television sets exhibited perceptible or greater interference on channel 3, and only four were above the threshold on channel 12. However, 50% of the receivers had perceptible interference on channel 53 and 67% were above threshold on channel 74.

1.4.8 Tuner Overload Distortions

Tuner overload performance has become increasingly important for CPE connected to cable as the number of channels carried thereon has continued to increase. The inability of the tuner to accept the aggregation of these high-level signals without generating excessive distortion products through the non-linear characteristics of the receiver's RF input circuitry results in the issue. The two most destructive products generated in this process are referred to as Composite Triple Beat (CTB) and Composite Second Order (CSO). The CTB product falls on the video carrier of the desired channel, and the CSO is either at $\pm$ 0.75 MHz or $\pm$ 1.25 MHz from that carrier, according to the channelization plan of the system, standard or IRC. The test configuration for Tuner Overload measurements is shown in Figure 1.4.8.1 below.

A total of 66% of the television receivers tested exhibited degradation above the "just perceptible" level on at least one channel and in at least one category. The television receivers showed perceptible degradation on all four of the channels tested.

A total of 50% of the VCRs tested showed perceptible degradation on at least one channel and in at least one category. However, the degradation was perceptible only on channels 53 and 74, and not on channels 3 and 12.

A total of 36% of cable converters showed perceptible degradation on at least one channel and in at least one category, with the visible degradation present on each of the four channels.

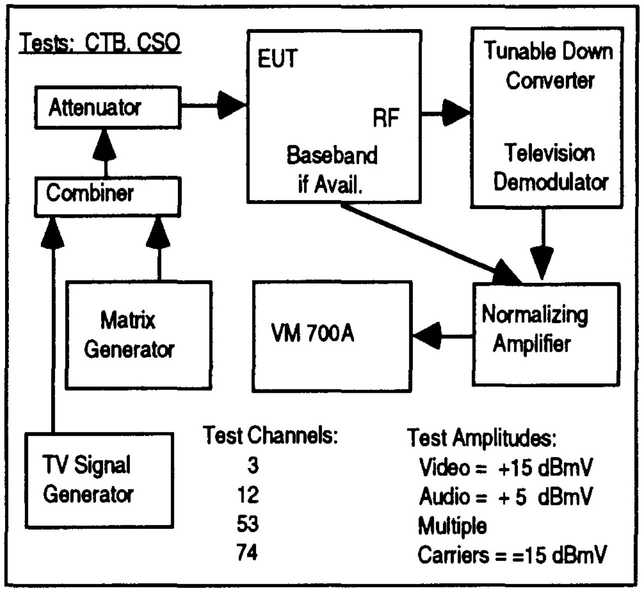


Figure 1.4.8.1 Test Configuration for Tuner Overload.

1.5 Viewer Perceptibility and the Selection of the -55dB Benchmark

There is always a great paradox in testing factors related to viewing television. One would like to devise objective techniques for testing receiver performance, since subjective tests often lack repeatability and precision. Subjective tests which are designed to have precision and reliability, often require such extensive procedures as to make them impracticable for widespread use. On the other hand, it is the subjective viewing of the television receiver, regardless of its signal source, which determines the acceptability of the delivered video to the viewer. Extensive testing has been accomplished in times past, attempting to bridge the gap, to enjoy the advantages of objective testing, while yet relating the results to some measure of subjectivity, such as the threshold of viewer perceptibility or the threshold of viewer annoyance.

These past studies have always been indexed to some particular form of interference, such as NTSC co-channel signals interfering with NTSC signals, or NTSC signals being degraded by non-coherent, non-video signals, such as in-band business radio emissions. Then, based on the type of interference, some measure could be made which would state that statistically the threshold of perceptibility occurs when the interfering signal is *N* dB, below the sync tip of the desired video signal.

The point of Dr. Bronwen Jones' effort was to perform a series of subjective perceptibility tests using exactly the same facility CTJ had used for performing their objective DPU measurements. We had assumed at the outset of the project that the -55dB desired to undesired ratio (D/U) was appropriate for the configuration used at CTJ to test for DPU susceptibility. In fact, the DPU measurements are linear, so the results contained herein can be adjusted based on the consensus ratio without redoing the tests.

The result Dr. Jones developed in her research shows, however, that the -55 dB point is indeed the threshold of perceptibility among the viewing panels assembled and supports the supposition made in establishing the DPU test procedures at CTJ. Please refer to Section 5. below for a full treatment of the findings.

1.6 Conclusions and Recommendations

In this section an attempt will be made to summarize the results of the above testing in a narrative format, drawing such conclusions as are supported by the data.

We have learned from the Stern study:

- That 65.4% of all US TV households reside in fields of 10 millivolts/meter (mv/m) or greater,
- That 54.8% of all US TV households reside in fields of 31.62 mv/m (90 dBu) or greater,
- That 40.8% of all US TV households reside in fields of 100 mv/m or greater,
- That 6% of all US TV households reside in fields of 1 volt/meter or greater.

In each of the 60 million cable homes there are an average of 2.3 television receivers, 1.8 videocassette recorders, and 0.33 cable converters. This yields about 138 million TVs, 108 million VCRs, and about 20 million cable converters in cable TV households in the US. Using the Stern findings, this would mean that:

- 90.25 million TVs, 70.6 million VCRs, and 13.1 million cable converters *which are in cable households* reside in fields of <u>10 millivolts or greater.</u>

- 55.2 million TVs, 43.2 million VCRs, and 11 million cable converters *which are in cable households* reside in fields of 100 millivolts or greater.
- 8.28 million TVs, 6.48 million VCRs, and 1.2 million cable converters *which are in cable households* reside in fields of 1 Volt/meter or greater.

Consider the CTJ results for irradiated DPU, which shows that for an average of three channels and with the CPE in an average orientation for susceptibility:

- The worst case TV reaches the point of perceptibility at 1.17 mv/m
The worst case VCR at 29.9 mv/m, and
The worst case cable converter at 77.23 mv/m.
- The median TV is perceptible at 51.57 mv/m,
The median VCR at 57.27 mv/m, and
The median cable converter at 292.73 mv/m.
- The very best TV is perceptible at 193.4 mv/m,
The best VCR at 162.73 mv/m, and
The best cable converter at 723.33 mv/m

Superimpose these two lists and the size of the DPU problem begins to emerge. Note that these numbers consider only irradiated co-channel, and not that conducted on the cable braid or on the power cord. The two modes are interrelated, and we have listed the conducted results in the body of the report for reference, but do not list them in the conclusions because the conducted test procedures developed at CTJ have not yet been accepted by all involved consumer manufacturers and the EIA.

In the past, because of the cost of extra outlets, customers have not typically reported all TVs and VCRs in the home as connected to cable. Because of the Cable Act of '92 ruling making extra outlets free of monthly costs, the cable industry has already noted a surge of extra outlets either being connected or admitted to for the first time. It is expected that this trend will continue until virtually all TVs and VCRs found in cable households are connected to cable.

The subject of mitigating factors was mentioned in several earlier paragraphs. These are factors which tend to modify the actual interfering signal level or the exposure of the CPE to such signals. These mitigating factors consists of:

- Ducting - An unusual phenomena in which the transmitted signals are channeled through a building in a non-standard fashion as a function of the building's physical structure.
- Building Attenuation - Considers the shielding properties of the building structure. This is a statistical function which can range from zero to significant attenuation, based on the placement of the CPE in the home and the type of building construction encountered.
- CPE Orientation - As demonstrated in the DPU testing, CPE is usually most susceptible to interference if the radian passing through the tuner of the unit is directed toward the interfering signal source. Other orientations offer some attenuation to the signal.
- Height Above Ground - There is a nominal 20 Log increase in the strength of the interfering signal relative to the height of the CPE above ground level.
- Urban/Suburban Clutter - A factor of the relative juxtaposition of the interfering signal source and the CPE, and the topology of the earth and structures between the two.

The difficulty in applying mitigating factors is that one cannot depend upon achieving a high level of protection over a broad range of installations, or over a long period of time in any one installation, with the results being highly stochastic. The experience of the cable industry is that service calls which are shown to result from DPU most often occur with the introduction of new CPE in the home, or when existing units have been relocated therein. While the activity level of the consumer electronics retailer is pretty much isolated to the time of sale, the cable operator's responsibility for satisfactory service continues from month to month. From the cable standpoint, it matters little whether the DPU problem became apparent with the purchase of new hardware, or at a point years down the road, it still requires a truck roll, the services of a technician, and the installation of a cable converter to fix the problem.

What has become clear during these studies is that adding mitigating factors to the other parameters in an attempt to develop a single performance number for CPE has the effect of dilating the window of acceptability of the test items. As a result of the Cable Act of '92, the FCC has called upon the cable and consumer industries to negotiate a single number which shall represent a limit below which no TV, VCR, or cable converter shall show perceptible DPU interference. This implies that a statistical process, which represents a range of performances based on both controlled (such as CPE design and construction) and uncontrolled (such as mitigating factors) issues, must be reduced to a single number, below which the performance range cannot extend. Considering the Stern data, it would appear that a 95% solution to the problem will fall into the 1.0 volt/meter range.

Unlike the DPU findings above, we published the results of the receiver performance tests more as a benchmark to determine the current state-of-art for both consumer and cable grade hardware, and primarily to assist in the completion of the engineering labor supporting the IS-23 effort in the JEC. As mentioned earlier, the test facility drawings and procedures used in these tests have all been forwarded to the membership of the JEC for analysis and comment. There will no doubt be changes suggested for these tests, all of which will be included where appropriate. Considerable non-trivial variations on the individual test plans, procedures, and facilities were tried at CTJ with the differences in results varying little more than 2.0 dB. This leads us to have confidence in these procedures as to their accuracy in portraying the performance parameter tested. However, since others may derive procedures which are either easier or cheaper to implement, or for any other legitimate reason may show increased promise, all constructive comments will be considered, tested, and implemented or rejected based on merit in the eyes of the JEC.

What is clear from these performance measurements, is that all CPE could bear some improvements in several of the eight areas considered. Among TVs and VCRs the problems measured here result almost exclusively from the use of designs which are optimized for off-air reception and are not appropriate for the different delivery conditions encountered when connected to a cable system. Since cable cannot regress to simulate the conditions found in the broadcast environment without losing its advantages and its very reason for existence, improvements can only be found in adapting the terminal devices, the CPE, to the delivery system. This is precisely what has occurred in all other areas of telecommunications, including wire and RF telephony, LANs, MANs, business satellite systems, microwave delivery systems, and in home DBS terminal equipment.

As we now stand on the threshold of the era of digital television transmission, tests performed at the Advanced Television Test Center, by CableLabs, and at other facilities underlines the fact that all receivers must be improved in all eight of the above factors, plus several more, including phase noise, group delay, and residual FM, if the transitional hybrid CPE currently being planned by the consumer manufacturers is to be usable through its expected lifetime.

Database Analysis of CATV Network Interconnectivity

by
Robert V. Moel
Time Warner Cable, Houston Division
William Spies
Time Warner Cable, Cincinnati Division

Abstract

Current outage detection systems ignore the interconnectivity between transmission elements. A failed element is analyzed in isolation without considering its impact on devices downstream from it. A simple database algorithm that describes this interconnectivity would improve outage detection speed and accuracy. This paper will describe such a program written in a common fourth generation language (4GL) called Clipper dBase.

Introduction

Cable television systems can be modeled as "weakest" link systems. Any device failure causes subsequent failure of all devices downstream from it. The probability of a point in the system failing is related to the probability of all previous devices failing[1]. Being able to know which other devices have failed would aid in customer communication about outages and would minimize falsely detecting another outage that is part of the first (and thereby not wasting manpower on chasing phantom outages).

Elements of an Interconnectivity Database Program

The questions that need to be answered are: (1) What relationship exists between devices in the network that uniquely define their relationship; in effect how do you encode which device is dependent on which. (2) The database and program must be able to model the weakest link attribute of the network. (3) The program that manipulates the network database must be fast and it must be real-time. There is no use for a program that tells you an outage occurred when the day is done. Finally, the program must be recursive. It must be able to function the same at any point in the system and recursively determine all other failed devices.

One of the key items in creating the database is determining what fields create links between database records that are unique. Consider figure 1. The three amplifiers are named 1, 2 and 3 (any convenient naming scheme will work). Amplifier 1 feeds amplifiers 2 and 3 and has them as successors. Similarly,

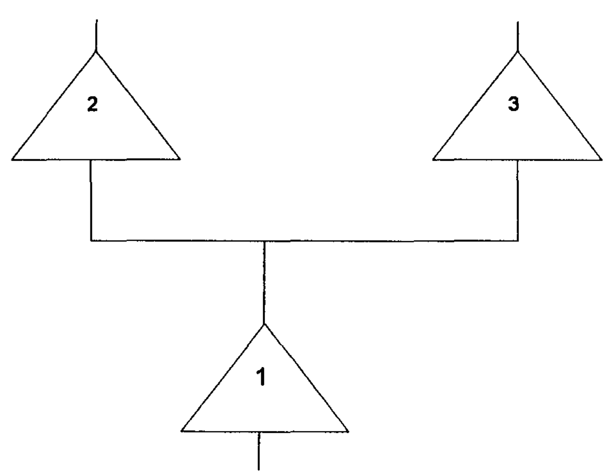
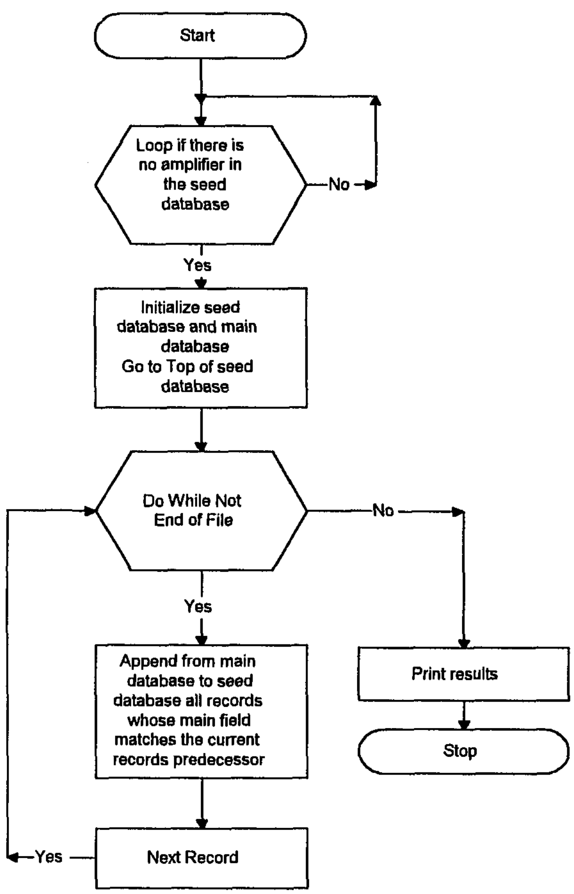


Figure 1

Amplifier 1 is fed by no other amplifier. Amplifier 1 has a null predecessor. Amplifiers 2 and 3 are fed by Amplifier 1 and hence Amplifier 2 has Amplifier 1 as its immediate predecessor and Amplifier 3 has Amplifier 1 as its immediate predecessor. Amplifier 1 can have either 2 or 3 as its immediate successor. Using the successor is, therefore, not a unique way to relate amplifiers in a database network. However, Amplifier 2 has only one predecessor, that of Amplifier 1. Similarly, Amplifier 2 has only one predecessor, again Amplifier 1. Devices in cascade as described above can be determined uniquely if the predecessor is specified as the linking field in the database record.

By stringing database records together by using the predecessor as the linking field, a database would be created that models the weakest link behavior of a real cable television system.

Figure 2



The code that was written to determine the effect that the weak link has on the rest of the cable system operates on a PC. The program is compiled to execute in real-time. As the flow chart will demonstrate, amplifiers are entered in real time and the list of amplifiers that have failed are produced instantly. Table 1 contains the program listing. Figure 2 is the flow-chart of this program.

The program works by starting with a seed amplifier. A pointer is set initially to this first amplifier. Then the program appends all the amplifiers whose predecessor fields match the amplifier currently pointed to. After all the amplifiers are copied to the end of the database, the pointer moves to the next record and the process starts again. When the pointer moves past the last record, the program stops looking for more amplifiers, prints the results and terminates.

Amplifiers are not the only devices that can be added to the database. Any device can be a part of the system as long as one knows how the devices interconnect with each other from a reliability sense. For example, while a main power supply is installed approximately in the middle of a group of amplifiers, from a reliability view point it is at the beginning of a cascade.

Because the program uses matching and copying, it is very fast. A database with 1,000 amplifier records and an average trunk run of twenty amplifiers will take approximately two seconds to respond to an inquiry when run on a "486" machine. These characteristics would make it feasible for this program to be integrated with a billing host's outage detection system either as a part of the host's program or off line a as separate processor with which the host would interconnect via a wire link. Any single amplifier outage would be used to determine what other amplifiers were also out. The additional amplifiers determined to be out would be up loaded to the host for dissemination to customer service agents.

Outage Detection In Cincinnati

Our system in Cincinnati has been using a combination of two-way monitoring of the plant and the database program to improve outage detection and correction, determine the underlying outage problem, develop a plan to improve the performance, and finally, measure the results.

The first step was to use the two-way response information from converters in the field to identify the location of outages. Part of the difficulty in using these converters is that there is always a group of converters that do not answer back. These "no-answer" converters may not be experiencing an outage but may have some other localized problem such as being on a switched outlet. This no-answer rate determines the "noise" threshold. Converters must be non-responding in excess of this rate to indicate an outage.

Once the exact location of the outage is determined, the next step is to feed the information into the database and get a list of all other amplifiers that were affected. We now have a true location of all the subsequent amplifiers affected and a better understanding of how a particular outage disrupts the system.

These disruptions were further categorized by outage fix code. In some instances the fix codes were changed to provide better information of cause. For example, the repair code labeled "intermittent" was not really a repair code because it did not tell us anything about what was done.

Once an analysis of the fix codes and outages was completed, two areas emerged as being key to resolving the problem. The first deals with the placement of power supplies. The analysis changed the way we placed supplies. Before we used to place supplies in anticipation of future growth. These under loaded supplies added to the number of supplies in cascade and increased the probability that an area would be out. Sometimes these areas were affluent and influential. Currently we minimize supplies and maximize loading. If an area grows, we will return to it and redesign the powering if needed.

The second deals with repeat outage offenders. What we discovered by using the database was that we were experiencing outages from devices that had failed and had recently been repaired by our in-house facility. The facility, it turned out, was ill equiped to perform the repair. They were merely bandaging the problem instead of thoroughly testing and repairing the unit. The few dollars we were saving were costing us dearly.

We arranged to have the original manufacturer repair the equipment by establishing an "extended warranty" agreement with them. This approach gave us a cost effective alternative to in-house repair that guaranteed the problem would not return to soon. Ultimately, this change gave the technicians more confidence in the refurbished equipment and, instead of "box" swapping, they are more likely to look for fundamental problems with the plant.

Conclusion

A database and program can be constructed to simulate a cable television system. Using this system one can detect outages faster, find and repair fundamental plant reliability problems, and better assess powering needs and power supply location.

Table 1 -- Program Listing

```
* Program -- OUTDET.PRG
* Writer -- Robert V. Moel
* All rights reserved
* This resets all files and initializes several criteria
close all
set safety off
set talk off
set confirm on
set bell off
* Clears screen
@ 0,0 clear
@ 23,0 say "OUTDET written by Robert V. Moel, all rights reserved"
@ 24,0 say "Press Enter to continue"
read
@ 0,0 clear
* Opens amplifier database as the source database
use amp alias source
sele 2
* Opens seed database
use out alias sink
* Sets up do while loop
repeat=.t.
do while repeat
 sele sink
* Clears seed database
 zap
 mampnum=space(8)
* enter the amplifier that failed
 @ 24,0 clear
 @ 24,0 say "Enter first failing amp" get mampnum
 read
 append blank
 mampnum=ltrim(rtrim(mampnum))
 replace amplifier with mampnum
 go top
* this is the iterative do loop that appends from the main database to
* the seed database each amplifier that points to another amplifier
 do while .not. eof()
* Set the pointer
  point=recno()
* Look for all amplifiers for which the current amplifier pointed to
* is a predecessor
  temp_amp=amplifier
  append from amp for pred_amp=temp_amp
* because appending moves the pointer, reset the pointer and
* then skip to the next record
  goto point
  skip
 enddo
 answ="N"
 @ 24,0 clear
 @ 24,0 say "Would you like to print affected amplifiers?" get answ
 read
 if upper(answ)="N"
  return
 else
* print or display amplifiers that were affected by the outage
  if upper(answ)="Y"
   answ1="S"
   @ 24,0 clear
   @ 24,0 say "(S)creen or (P)rint " get answ1
   read
   if upper(answ1)="S"
    display all
    @ 24,0 clear
    @ 24,0 say "Press Enter to Continue"
    read
   else
    if upper(answ1)="P"
     list all to print
    else
     quit
    endif
   endif
  endif
 endif
* do another outage?
 answ2="Y"
 @ 24,0 clear
 @ 24,0 say "Another Outage (Y/N)" get answ2
 read
 if upper(answ2)="N"
* if no other outages, leave the loop
  exit
 else
  @ 0,0 clear
  loop
 endif
enddo
* reset all and quit
close all
return
```

1. Moel, Robert and Spies, William Reliability Modeling of Cable TV Systems, 11-4 to II-5, CableLabs, September 1, 1992.

DEMAND SIDE MANAGEMENT CONSIDERATIONS IN ADVANCED NETWORK DEPLOYMENT

Dean Ericson, President
Dan Carter, Vice President
Media Management Services, inc.

Synopsis

As cable operators contemplate upgrading their cable TV systems to full service networks, one services that may be included in the list is Demand Side Management (DSM), either as a joint venture with an electric utility, or by providing DSM as a contracted service. In this paper, we will examine why an electric utility wants DSM, and how a cable operator may be successful in providing DSM services.

Information presented here is based on MMS's experience in assisting cable operators develop and implement full service networks, and in assisting electric utilities evaluate their alternatives for the installation of DSM networks and services.

❒ 1.0 Defining Demand Side Management

Demand side management is the general term applied to an application that permits an electric utility to communicate with its customers. A demand side management system utilizes a *smart* two-way interactive network linking the power company's command and control center with interface devices at customer locations. These devices monitor and sometimes control power usage. This communications network can also be used by the power company to provide: real time tracking of power usage and outages; meter reading; service connects and disconnects; and, detection of unauthorized service use and meter tampering.

The DSM network can also allow utility customers to make real time *lifestyles* driven decisions regarding their energy utilization levels. Consumers gain control and the ability to make energy consumption choices. The electric utility achieves higher levels of operating performance through the collective decision making of customers who use DSM technology. The result is energy cost savings and higher levels of customer satisfaction, and better financial results for utility company stockholders.

❒ 2.0 Why Two-Way Networks and Demand Side Management Applications Are Important to Utilities

A number of regulatory, competitive, technological, customer, and economic factors are causing fundamental changes in the electric utility industry. Two-way network deployment to end-user locations and demand side management applications are one important way to utilize technology to help deal with these forces and changes.

2.1 Regulatory Factors

A number of important regulatory trends are shaping power company thinking. First is the opening-up of the "grid". This opened-up grid permits large capacity users to purchase their electrical power out of the local power company's service area, and the local power company must provide transport to the user. With this open grid, utilities are brought into direct competition.

Second, is the relaxation of restrictions on end-users from themselves getting into the power generation business (both to serve their own needs and to feed power into the grid). Third are the strict requirements with respect to environmental impacts and public safety which affect both the cost and time to design and build new power generation capacity. Finally are public sector driven considerations with respect to the efficiency of utilization of non-renewable resources and the cost to the consumer for energy use.

➢ A two-way network to end-user locations and demand side management applications enable a utility to operate more efficiently in this increasingly competitive environment. In addition, a DSM system can help reduce and/or delay the commitment of substantial capital for new power generation plants. Further, a DSM system can demonstrate to regulators that the utility is responding to calls for higher efficiency.

2.2 Competitive Factors

Increased competition among power generators, wholesalers, distributors, and large retail customers is a reality for utilities. The result is that utilities recognize that they must strive to lower costs, improve efficiency, and/or provide additional services in order to achieve a sustainable competitive position.

➢ A two-way network for demand side management applications is one of a number of means that a utility can use to strengthen its competitive position.

2.3 Technological Factors

Electric utilities have extensively deployed fiber/digital/two-way networks in order to perform "command and control" functions in the electrical distribution grid. These functions include system monitoring and power generation/distribution adjustment among power stations and various relay and substation points on the distribution network. They also include communications among utilities regarding the purchase and sale of power on the *opened-up* national grid, and the full range of plant management and maintenance functions.

➢ Demand side management represents a natural extension of this "private network" to end-user locations and thus greater utilization of a telecommunications asset already in place.

2.4 Customer Factors

Overall demand for power is increasing at a modest rate. At the same time, the "consumer power demand profile" exhibits substantial seasonal and daily peaks which, to some degree, are caused by factors like weather which are not within the control of the utility.

Residential customers are highly sensitive to price and, while generally satisfied with power company "service", do not exhibit a great deal of "brand loyalty" to what is perceived as a "commodity" service provider.

Commercial customers are demanding reliability in delivery, and utility responsiveness to problems. Further, large commercial customers have two alternatives with respect to their source of power. The first is construction of their own facilities to serve their needs. Second is purchase of power from an *out-of-region* supplier who, under the more open regulatory framework, can profitably deliver power over the local power company's grid.

➢ A demand side management system helps a utility build "brand loyalty" by giving customers some control over management of their power consumption costs and direct feedback on how their decisions save them money.

2.5 Economic Factors

Electric utilities face complex and challenging capital and operating issues.

The source of a power company's product is obviously a power generation plant. A power company faces two types of challenges in terms of capital deployment to increase its generation capacity.

First is the utility's response to *long term* increases in the demand for power. That is, bringing a new power plant on line represents a very costly investment (estimated at $1 to $3 billion dollars). The planning, design, and construction process for a new power generation "base unit" is very lengthy -- up to ten years. The regulatory requirements that must be met in terms of safety, the environment, and so forth are one of the factors that add to cost, timeframe, and risk with respect to new plant construction.

Second is the utility's response to the *seasonal and daily* variations in the demand for power. Meeting peak demands requires that the power company build "peaking units" which are brought on line on an as-needed basis. Construction and operation of these peaking units results in under utilization of capital resources and higher cost per kilowatt-hour as these units are not always operated at levels of peak efficiency.

➢ A demand side management system can help an electric utility operate more efficiently in the short term while having a positive impact on long term capital requirements for new plant construction.

❒ 3.0 DSM Network Technical and Operational Requirements

3.1 DSM Network Applications

A Power Company's applications for a demand side management system are based on a general hierarchy of internal requirements. These priorities will vary somewhat from utility to utility, but in the broadest perspective most power companies have the following priorities:

Primary Uses

1. Demand-side management / load control
2. Real time price information
3. Distribution automation capabilities for the electric power network
4. Power outage detection and notification

Secondary Uses

5. Automated reading of electric meters
6. Automated billing
7. Automated service connects and disconnects
8. Detection of unauthorized service use and meter tampering

Ancillary Uses

If a power company installs an advanced DSM network they could also offer:

9. Contracted services to other utilities (water, gas)
10. Security services
11. Information services
12. Cable television services
13. Telephone services

Before cable operators become too concerned that electric utilities are going to build extensive DSM networks and then get into the cable TV and telephone business, it should be pointed out that PUCs, at this point, typically stipulate that utility-customer information networks have to be justified based on savings and revenues from utility related services. Any revenues from non-

utility related businesses cannot exceed 10% of the total revenues from the communications network. More on this later.

3.2 DSM Network Components

How can the identified functions of a DSM network be realized? To address this question, we must first define the components of a DSM system. Broadly categorized, the components are:

- The customer premises communications system
- The electric meter interface
- The communications link between customer premises and the power company.
- The command and control computer hardware and software for desired applications

All of these components contribute to the overall cost of the network with the major cost being in the communications link between the customer premises and the power company.

Many of the peripheral devices such as electronic kilowatt-hour meters and remotely operated disconnects have been developed and some are readily available for purchase. The primary missing link between the power company and its customer is a two-way communication system which has been designed and configured for DSM. Also the hardware and software required to communicate within the customer's premise, and to interface with the electric meter needs more development, but this component will probably evolve into to a workable system before the DSM networks are in place.

3.3 Available Communications Technologies

Parts of the infrastructure which could provide communications between the power company and its customers currently exist:

- Power Companies: First on this list are power companies themselves which have a considerable amount of fiber in place, however, the fiber is generally low count and not deployed in locations where it provides ready access to businesses and residences.

- Phone Companies: The public switched telephone network is deployed throughout power company territory. This network consists largely of twisted-pair copper wire which has very limited bandwidth. An additional line to each DSM location would probably be required for DSM services.

- Cable TV Companies: Cable TV companies have fiber optic and coaxial cable deployed in most metro areas, yet the cable, on average, accesses only 60% of the residences. Cable systems provide mostly one way communications toward the customer. Return bandwidth, if available, is limited.

- Mobile Radio Companies: Mobile Radio companies provide wireless radio frequency (RF) communications, however, deployment is very limited.

- Satellite Networks: Satellite Networks could provide one-way communications toward the customer, but this limits its usefulness.

3.4 DSM Network Installation Considerations

The functional requirements of a two-way DSM system, as with other systems and technologies, will determine the type of technology implemented and its cost.

The following discussion covers some power company considerations for the installation of a

DSM Network as they relate to network requirement and costs.

Bandwidth: The most significant factor in selecting the type of communications system to be used for DSM services is network bandwidth, or capacity. The amount of data to be transferred and, how quickly the transfer is required, will be determining factors in the components chosen for the network. For example, an automated meter reading system requires little bandwidth, and the data is not particularity time sensitive. On the other hand, demand side management "load control" applications require large amounts of data which must be delivered in real-time.

Cost: Network Installation Cost: Network installation is by far the largest single cost for a power company to build their own DSM network. A very generalized number that gets used is that power companies would like to install a network for between $200 and $400 per residence. Roughly three-fourths of this cost is allocated to providing the electric meter interface and the communications connections back to the utility. (A *smart* meter alone can cost over $100.) The rest is for the hardware and software for data management and processing. It can be stated that cost is almost always directly proportional to bandwidth.

Reliability: A power company's uses for a DSM system, because of their application, will require that the network be highly reliable. While reliability requirements may ease somewhat as we move toward the bottom of their hierarchy of potential uses, network reliability for the primary uses will be a leading consideration in network planning. If a power company is planning to offer service to other utilities, or non-utility services to other entities, reliability requirements will be even higher.

If the utility is managing large power user loads based on being able to quickly reduce demand from these users as a peak load approaches, a failure in the communications network could have disastrous results. Because load management is a dynamic interactive process, broadband, two-way communications becomes a necessity. The central processor must be able to quickly confirm if the load reduction has taken place and if so by how much.

For a cable operator considering partnerring with a power company to install a DSM network, reliability (network availability) requirements could be more stringent than those for competitive access provider (CAP) services. For the highest capacity CAP services, the network availability requirements typically stipulate that the service be in operation 99.997% of the time. That is approximately 16 minutes per year of allowable outage, which may be too much time for primary (or high capacity) DSM services.

Ownership: With network reliability being a primary concern for power companies, we can easily surmise that power companies would prefer to not have another entity between them and their customer. In fact, it is safe to say that if cost of the network installation and operation was not a factor, power companies would not even consider outsourcing these services, particularity to large capacity users.

Security: Security of the network is another power company consideration. Is the communications link to the electric customer secure from unauthorized intrusion?

Geographic Coverage: How can ubiquitous coverage within the entire electric utility service area be provided?

Standards and Compatibility: At this point, there are no clear standards for DSM services, networks, or network interfaces. Even though some development is underway, there is no clearly superior technology, and low cost hardware and software is not available.

❐ 4.0 Future DSM Network Planning Considerations

Because of the extensive technical and operational requirements for a DSM network, power companies logically would like to install and operate their own networks even with the high installation cost.

In addition, a power company's construction of its own network would ensure that the utility's capacity requirements would be met , and would in fact enable the utility to use excess network capacity to generate incremental revenues from utility related and non-utility services.

However, electric utilities tend to perceive non-utility services (information services, cable TV, telephone, etc.) as very speculative, high-risk businesses and probably would not use non-utility revenue potential as justification for installing their own network.

Where this all seems to be leading us is that even though power companies would prefer to install and operate their own DSM networks, they generally expect that a fully deployed DSM system will use all of the technologies discussed previously, and will consist of facilities both leased from commercial communications providers and those owned and operated by the power company. The type of technology deployed will be a function of availability, pricing, and company needs at the time.

So -- for cable TV operators planning to install full service networks, including DSM in the strategy is probably a reasonable expectation, provided that the network meets the compatibility, reliability, and security expectations and requirements of the power company.

❐ 5.0 Electric Utility Industry Demand Side Management Experience to Date

Demand side applications of two-way interactive networks are *high on the agenda of many electric power utilities.*

A number of published reports indicate that "utilities are investing in demand side management technologies". For instance, Entergy took a minority position in First Pacific Networks in 1991. The Southern Company announced that it would invest in First Pacific in 1993. Boston Edison has asked for regulatory approval to invest in an unregulated subsidiary that would invest in demand side management technologies.

There are also indications that the venture capital community is taking an interest in demand side management investment opportunities.

A large number of mature and start-up suppliers are working on demand side management technologies. These technologies are being developed to operate in a number of network environments, including broadband coaxial/fiber, twisted pair, wireless/RF, and satellite. Supplier companies include BellSouth, First Pacific Networks, RAM Mobile Data, Ericsson/GE, CableBus Corporation, AT&T, and TranstexT.

Industry-wide standards are under development, via the involvement and leadership of the Electric Power Research Institute (EPRI) and member utility companies.

Utility monitoring services are in operation in a large number of gas customer homes.

Electric power utilities have undertaken field trials of DSM technology in a limited number of homes. The objectives of these tests have been to gain experience with the technology, to gain in-market understanding of the features and functions that are most attractive to the consumer, to determine the nature of customer use of and satisfaction, and to develop an information base that will help with packaging, pricing, and marketing decisions.

The general sense of those involved in developing the DSM business is that the power utility industry is on a "fast track" with respect to development and deployment of the technology. As we pointed out earlier in this paper, there are strong strategic and economic reasons for electric utility interest in DSM. The enabling technologies are being rather quickly developed, with functionality and price bringing product into line with utility economic requirements and customer expectations. There are still many business unknowns in terms of marketing, consumer take rate, pricing, and so forth. However, it is likely that these unknowns will be addressed in the very near term.

Finally, of all the issues that must be addressed by electric power utilities in order to make demand side management a success, perhaps the most important is the design and construction or the two-way interactive network that will link the utility with customer locations.

❒ 6.0 Collaborative Opportunities for Cable Operators and Utilities

The cable industry's advanced network deployment thrust and the electric power industry's desire to deploy demand side management services suggest that there may be substantial collaborative opportunities for the two industries, including:

- Cable company provision of the two way network to the customer.
- Cable company utilization of existing electric power utility fiber networks as part of a cable "regional networking" strategy.
- Design of in-home hardware to include functionality that meets the needs of cable companies and utilities.
- Joint marketing and promotion of services.

DEPLOYING TELEPHONY SERVICES OVER CATV SYSTEMS: SYSTEMS AND ARCHITECTURAL CONSIDERATIONS

Gaylord A. Hart, XEL Communications
17600 East Exposition Drive,
Aurora, CO 80017
(303) 369-7000

Richard Lyford, ANTEC
8101 East Prentice Avenue, Suite 210
Englewood, CO 80111
1-800-FIBER-ME

ABSTRACT

Today's cable systems are in a unique position to deliver telephony services via the two-way transmission of digital signals. Several methods exist for combining telephony and entertainment video signals within the distribution plant and for separating those signals at the subscriber end of the system. Each of these methods has unique technical and economic advantages and disadvantages.

This paper discusses the architectural and design considerations when developing a network capable of delivering telephony services. Architectural options are outlined, and the tradeoffs associated with design choices are discussed. Node sizing, optimization, and evolution are considered, as well as fiber counts, bandwidth utilization, reliability and redundancy, network management, and subscriber terminal deployment.

INTRODUCTION

The ability to deliver a multitude of new services to the home (both analog and digital; both interactive and non-interactive) via cable systems has become feasible in the last five years due to the rapid deployment of fiber optic technologies. Indeed, most experts now agree that hybrid fiber/coax networks provide the most economical means of delivering these new services. One such service is POTS, or Plain Old Telephone Service.

Deploying telephony over cable systems is only a short time away. Many manufacturers have announced cable telephony products or are already in production. Effective deployment of these systems requires a good understanding of spectrum utilization, network architectures, drop architectures, system evolution, and the tradeoffs surrounding decisions made today for tomorrow's services.

SPECTRUM UTILIZATION

CATV Spectrum Allocation

New technologies and services are bound to place new demands on future allocations of CATV spectrum, requiring much more bandwidth than has been available in the past, including a much larger or more densely utilized return path allocation. Services such as telephony, HDTV, data communications, video-on-demand (VOD), and multi-channel compressed NTSC video delivery are already staking out portions of the CATV spectrum, and as it usually turns out, services expand to consume the maximum amount of bandwidth available. Of course, new digital compression technologies will help operators efficiently utilize this spectrum.

Figure 1 shows a typical frequency allocation plan for a 1 GHz cable system. Below 550 MHz, the system looks exactly like current systems, with analog TV channels occupying 50-550 MHz and the return path occupying 5-30 MHz.

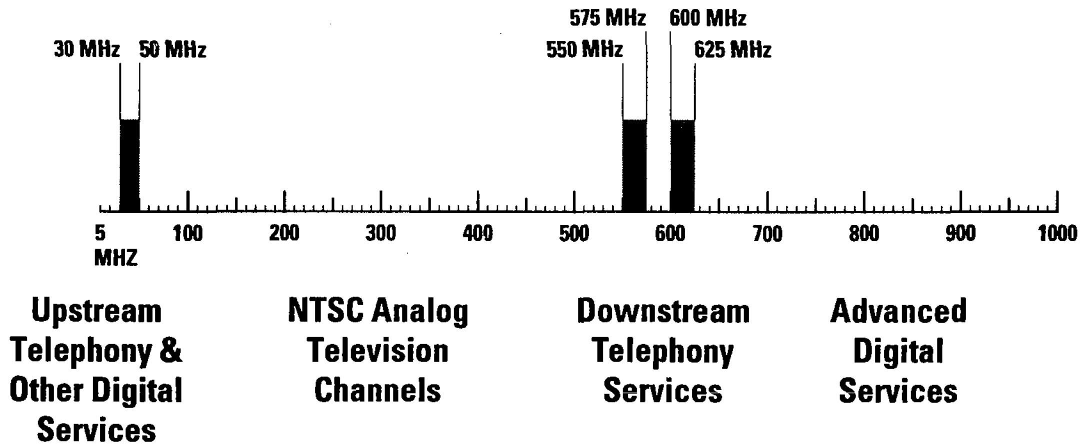


Figure 1. CATV Spectrum Allocation with Telephony and Advanced Digital Services

It is in the area above 550 MHz that the greatest changes are taking place. In this example, 25 MHz of bandwidth has been allocated for telephony in the reverse path, but in practice this may not be possible (or at least not without forward error correction or other means of maintaining a low bit error rate). System ingress originating in the subscriber's home wiring is not uncommon, and this ingress may come from several sources. It is also in the 5-30 MHz band where many amateur radio and CB transmitters are active in urban neighborhoods. In this example, another 25 MHz has been allocated for telephony in the forward path (575-600 MHz, with guard bands shown on either side; depending on system hardware, these guard bands may not be required). It is important to point out that the telephony signals in the forward path could be placed anywhere (even below 550 MHz), and that the choice of frequency location is left to the operator to maximize the particular application.

Typically, the spectrum above the telephony channels would be used for delivering advanced services such as compressed digital NTSC or VOD. As the deployment of interactive services and digital communications becomes ubiquitous, it is likely that the 5-30 MHz return path used today will not be adequate to support these new services.

Additional return spectrum can be gained by converting a system to a mid-split return (e.g., 5-112 MHz return, 150-1 GHz forward), but this is unlikely due to the analog spectrum currently in use for TV services and the vast number of consumer products supporting this spectrum. It is far more likely that additional return spectrum will be carved out of the higher end of the spectrum, likely from 850-1000 MHz.

This raises important questions about the 5-30 MHz spectrum. If additional return spectrum is provided at the top of the CATV spectrum, the 5-30 MHz spectrum may not be needed for return, thus eliminating the need for two sets of diplex filters for the two return paths. In this case, this spectrum could be allocated for forward use. Or perhaps more interesting, the 5-30 MHz spectrum could be utilized as a fully bi-directional path operating completely passive, without any intervening amplifiers. The cable losses at these frequencies are very low, and if the node sizes are small enough reliable transmission should be achievable. Such a passive path would be highly reliable since transmission could be accommodated even if power failed in the amplifiers. This has positive implications for services such as telephony and network management which need to be highly reliable.

Telephony Spectrum Allocation

Fortunately, only a relatively small allocation of bandwidth is required to deliver telephony services over cable systems. This is because not all frequencies discernible by the human ear (typically >15 kHz) need be transmitted for intelligible, natural conversation. In fact, the actual bandwidth occupied by a voice signal is limited to 4 kHz in telephony to conserve spectrum. Of course, many telephony channels have to be provided to maintain reliable service, and collectively these can occupy significant bandwidth.

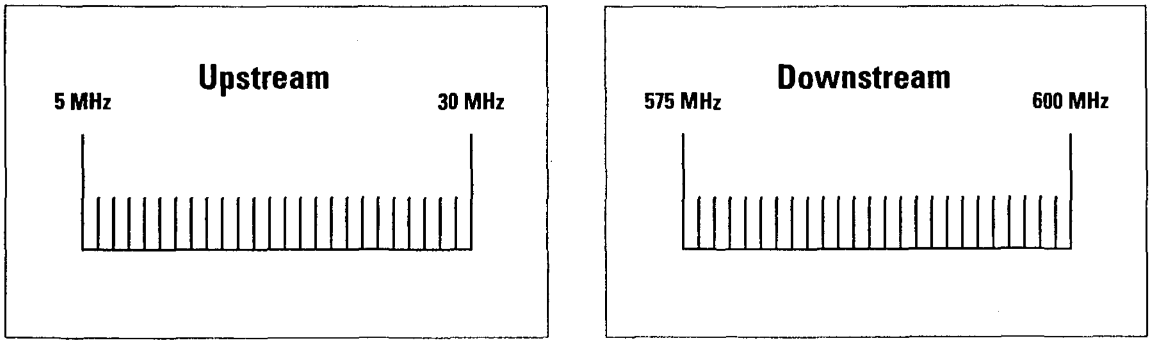
Fortunately again, the coming of age of fiber technology in cable networks provides all the bandwidth necessary for adding telephony services to existing cable systems. This is accomplished through fiber division multiplexing several voice channels on several fibers. That is, for a given number of voice circuits, telephony bandwidth on an individual fiber is conserved by spreading out the voice circuits over multiple fibers, each fiber serving a different geographical area. The fiber counts and bandwidths required for economical telephony delivery on cable systems coincide very well with bandwidth requirements and fiber counts required for existing services, and in fact provide additional capabilities and leave room for contemplated new services.

Universally, digitized voice signals are created as 64 kb/s digital channels (commonly referred to as a DS0). These individual channels may then be multiplexed together in a number of ways and at varying bit rates. Higher rates, of course, support more voice channels. In North America, the most commonly used next order of multiplexing is the T1, which consists of 24 digitized voice channels (24 DS0's, for a total of 1.544 Mb/s, including framing overhead). Although many different transmission rates could be used for delivering telephony over cable systems, this is a logical rate supported by much existing hardware and allowing relatively flexible use of existing CATV spectrum and smooth migration as service penetration increases.

Channel overhead is also required in the bit stream for control purposes within the cable telephony delivery system itself (e.g., for remote provisioning of the Subscriber Terminal Unit). This overhead may come at the expense of capacity taken from the T1, or by using the T1's embedded extended data facility (Extended Super Frame format only), or by providing additional channel overhead by transmitting at a higher data rate. For the purpose of discussion, this paper assumes T1 transmission rates will be used.

Given a T1 transmission rate for the bit stream, many choices exist for modulating the T1 onto an RF carrier: FSK, BPSK, QPSK, QAM, and others. The choice is typically a tradeoff between bit rate vs. bandwidth used and circuit complexity. Thus the modulation scheme chosen directly effects the overall system traffic capacity. QPSK presents a good balance here (providing a transmission efficiency of two bits per Hertz with relatively inexpensive circuitry and good noise performance), and many proposed cable telephony systems employ QPSK. For the purposes of discussion, this paper assumes QPSK data transmission in both the forward and return transmission paths.

Given the same data rates and modulation methods, the transmission bandwidth required for the system will be the same for upstream and downstream. The upstream path is the limiting factor, providing 25 MHz of usable bandwidth from 5-30 MHz (there is currently active interest in extending this to 40 MHz, thus providing 35 MHz of usable bandwidth). Referring to Figure 2, 25 MHz of bandwidth allows 24 QPSK signals to be transmitted, each carrying one T1 (1.544 Mb/s divided by 2 b/Hz = .772 MHz bandwidth; allowing 30 percent more for adjacent channel guard bands gives 1 MHz of bandwidth required for actual transmission). Given that each T1 supports 24 voice channels, 576 voice circuits are available in 25 MHz.



24 QPSK Signals
= 576 Voice Channels (DSOs)

Each QPSK Signal
= 1 TI
= 24 Voice Channels
= 1 MHz Bandwidth

24 QPSK Signals
= 576 Voice Channels (DSOs)

Each QPSK Signal
= 1 TI
= 24 Voice Channels
= 1 MHz Bandwidth

Figure 2. Typical Telephony Spectrum Allocation

One great advantage of transmitting telephony over several different RF data channels is the great flexibility provided. RF channels need only be added as telephony service penetrations increase (capital expenditures are tied directly to revenue generating services). RF channels may be flexibly assigned frequencies around existing or contemplated services (particularly important in the return path where service is often shared with network management carriers and PPV authorization channels utilized by addressable converters). Finally, frequency agility of the return and forward paths allows efficient use of CATV spectrum and the ability to shift spectrum usage in the future as new services and technologies are introduced.

Contention and Service Penetration

A discussion of spectrum utilization would not be adequate without some consideration of contention and penetration. Two basic approaches exist for handling telephony traffic over cable systems: dedicated circuits and contention based channel assignment. A dedicated voice circuit may be assigned to a subscriber (actually a virtual circuit, consisting of an RF channel assignment and a time slot within the T1 carried on that channel). Whenever that subscriber uses his phone, that carrier and time slot are accessed. When that subscriber is not using the phone, those resources are idle and cannot be used by other subscribers. A voice channel must be dedicated for each telephony subscriber, thus potentially requiring a large number of dedicated voice channels. For spectrum efficient telephony systems, this is not a problem. Dedicated channel assignment has the advantage that a subscriber always has guaranteed access to his voice channel and the channel assignment process is simplified.

Contention based channel assignment leaves idle voice circuits free in a pool. When service is requested either by an incoming call or by a subscriber trying to place an outgoing call, a free channel is taken from the pool and assigned to a subscriber for the duration of the call. When the call is finished, the channel is returned to the pool for use by other subscribers. Since it is highly unlikely that all subscribers would attempt to make (or would receive) calls simultaneously, this approach allows a finite number of voice circuits to be used by a much larger number of subscribers, thus increasing hardware utilization and system efficiency.

When a subscriber attempts to make a call and no voice circuits are available, no call connection can be made, and this system state is defined as blocking. An effective statistical model (the Erlang B formula for lost-calls-cleared) exists for calculating channel capacity vs. traffic vs. probability of blocking. Any use of contention requires careful analysis to guarantee a minimum grade of telephony service. Contention allows a smaller portion of RF spectrum in the cable system to be used to serve an equivalent number of subscribers as with dedicated channel assignment, but the tradeoff is in potential blocking and customer dissatisfaction during peak traffic periods.

Contention based channel assignment also has another great advantage: it allows dynamic bandwidth allocation on a per subscriber basis. Services such as videophone and high speed data transmission require bandwidth in excess of that provided in the basic voice channel (4 kHz or 64 kb/s). Contention allows assignment of additional capacity to a subscriber on a demand basis, typically in multiples of the basic data rate of 64 kb/s. A subscriber desiring to use one of these advanced services would be given channel capacity for that service only for the duration of the connection, thus saving the subscriber money while making more efficient use of the system's total traffic capacity.

Networks must be designed to meet today's needs and those contemplated in the future. However, it is not always appropriate to fully build out a network today to provide services which may not actually be used until quite some time in the future. It is likely that the number of cable subscribers may be quite high compared to the initial number of telephony subscribers on a cable system. Networks should be designed and built with this in mind, and a network evolution plan should be devised to allow natural growth in traffic capacity as penetration increases.

DISTRIBUTION ARCHITECTURE

Current FSA Network Topologies

Current cable systems utilize a fiber to the serving area architecture (FSA), with each optical receiver typically serving a node of 500-2000 subscribers. This architecture is shown in Figure 3.

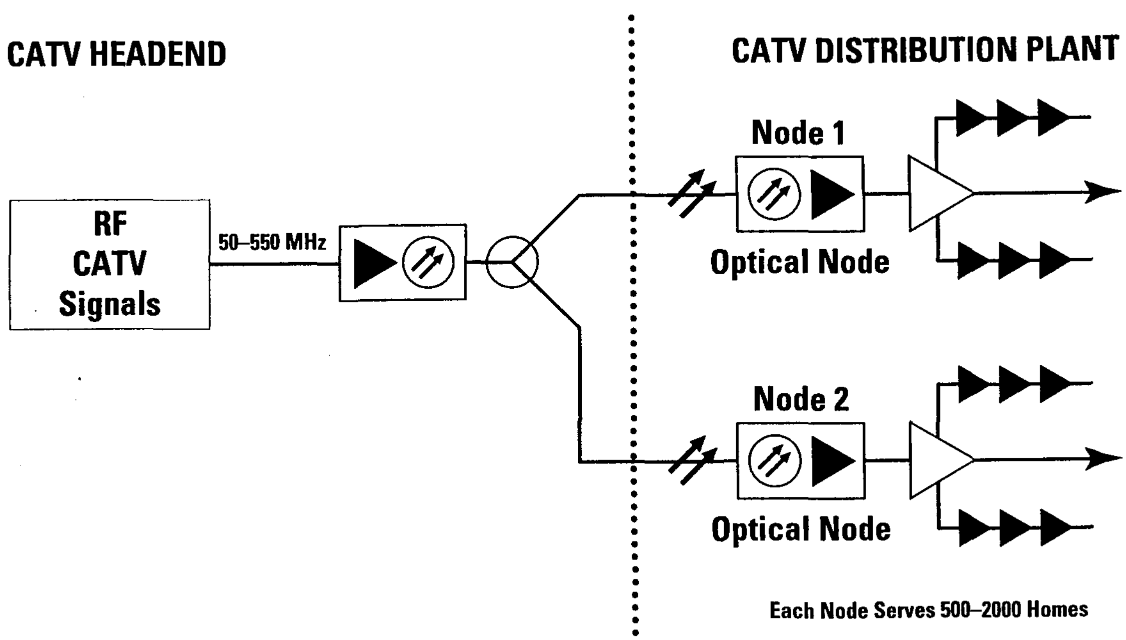


Figure 3. Current FSA Architecture

One laser in the headend may be optically split to serve two or more nodes since the television services provided are the same for each node. This is a cost-effective approach since relatively high output power DFB lasers are available, and only one laser need be used for two or more nodes. This approach is also effective in reducing power consumption, improving distortion and noise performance, increasing reliability, and reducing signal ingress and leakage. The coaxial trunk amplifiers have been eliminated, and the size of the failure group has been greatly reduced. This architecture is very flexible for overlaying additional services and allowing cost-effective system evolution.

FSA Network Topology with Telephony Overlay

Figure 4 shows the same FSA architecture with a fiber telephony overlay. In this case, the two receive nodes are treated as one logical node for telephony services. Note that this approach uses the existing downstream CATV service laser to provide downstream telephony services as well.

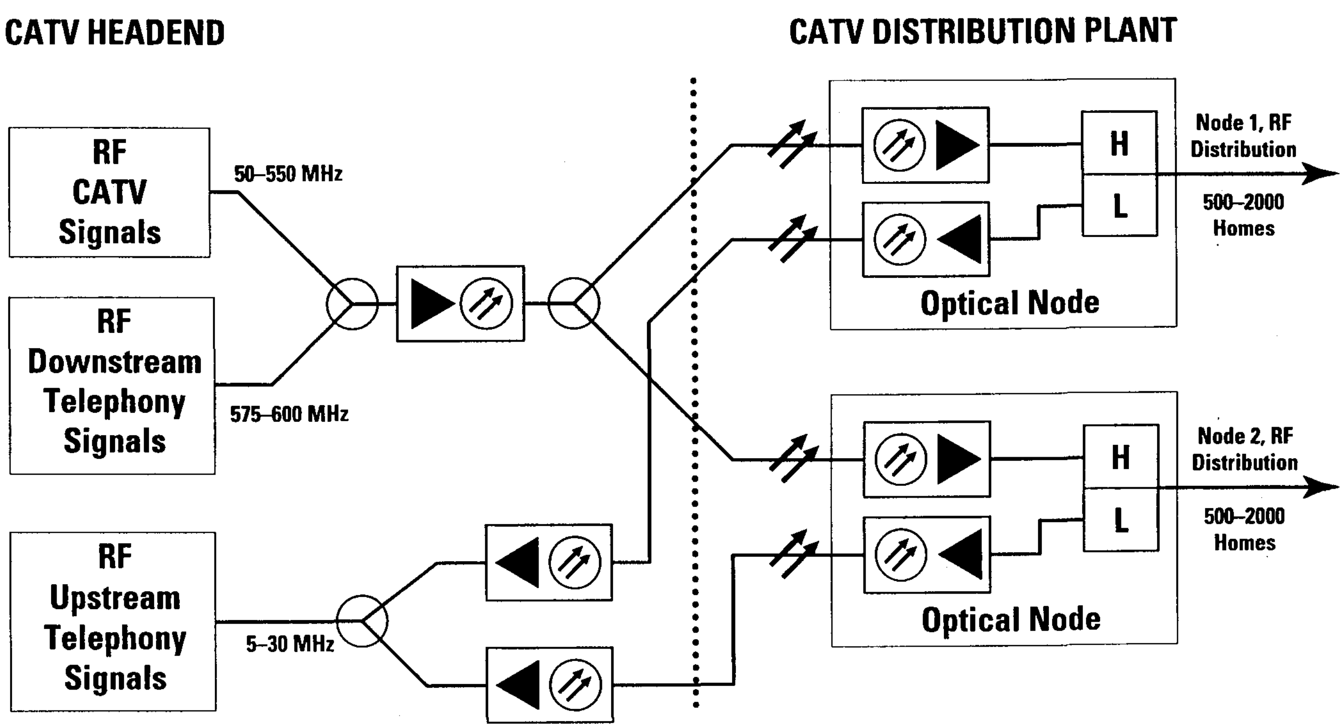


Figure 4. Current FSA Architecture with Telephony Overlay

Providing enough bandwidth is available, the downstream architecture need not be modified at all for this upgrade. The upgrade is facilitated even more since no new fiber need be pulled: dark fibers probably already exist in the cables already feeding the remote receiver nodes. If you are currently in the process of adding fiber to any of your systems, it is wise to plan now for future applications by providing additional fibers in the sheath.

The return path upgrade is readily accomplished in the distribution plant by swapping the original optical node receivers with receivers with return path lasers for the 5-30 MHz band. Many receivers on the market or already installed already provide plug-ins for retro-fitting lasers in receivers already in the field. The only additional equipment required for signal transport consists of two relatively inexpensive optical receivers in the headend for the return path.

Since this topology treats two optical nodes as one telephony node, migration to this configuration is easily and economically accomplished. However, each optical node is now served by only half as many voice channels as are available on from the telephony node at the headend. Using the spectral model shown in Figure 2, 576 voice channels would be available for the two nodes (1,000 to 4,000 homes passed). For initial service offering, telephony penetration may be low, and this architecture can be a cost-effective means of matching capital expenditures with revenue.

When service penetration increases, or if high penetration is anticipated quickly, the two CATV service nodes may be treated as two independent telephony nodes by separating the telephony services for each on different fibers, thus doubling the traffic capacity on each node. Of course, additional telephony hardware must be provided in the headend to support this increased capacity. Using the spectral model shown in Figure 2, 576 voice channels would now be available for each node (500 to 2,000 homes passed).

Figure 5 shows this architecture, which allows yet another cost-effective step in the natural evolution of the network as additional channel capacity is required.

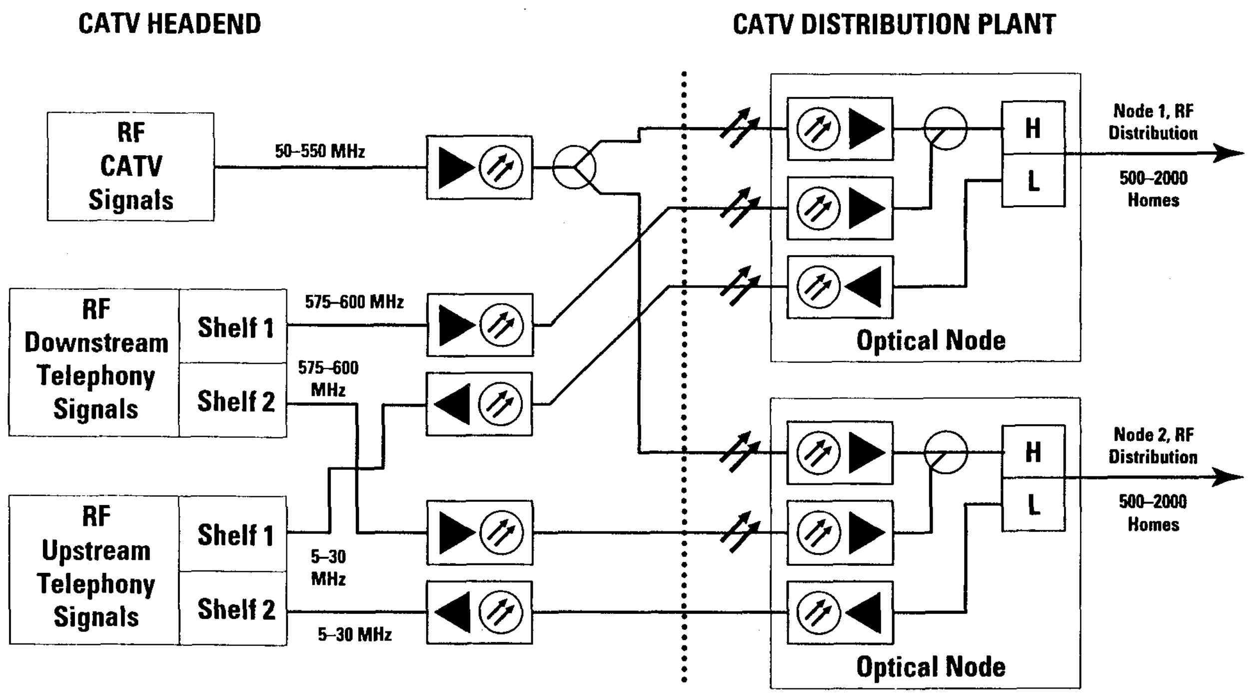


Figure 5. Current FSA Architecture with High Penetration Telephony Overlay

Note that for economic reasons only one DFB laser is used to feed television services to the two nodes. This does necessitate, however, the use of separate downstream and upstream telephony fibers for each node. It should be possible for this application, once again, to use dark fibers already in place. This topology also requires one relatively low-cost laser per node at the headend for downstream telephony and an additional optical receiver at the node for combining the downstream television and telephony services for distribution over the coax. Node receivers such as shown in Figure 5 are already available on the market.

Advanced Services FSA Network Topology

Figure 6 shows an advanced FSA architecture capable of delivering analog television, telephony, and advanced digital services to the node. The bandwidth here need not necessarily be 1 GHz if the advanced digital services are offered over a smaller spectrum or are not included at all: this architecture will work just as well at 600 MHz or 750 MHz (750 MHz amplifier hybrids are currently available, though in short supply; 1 GHz hybrids with adequate performance are still not commercially available).

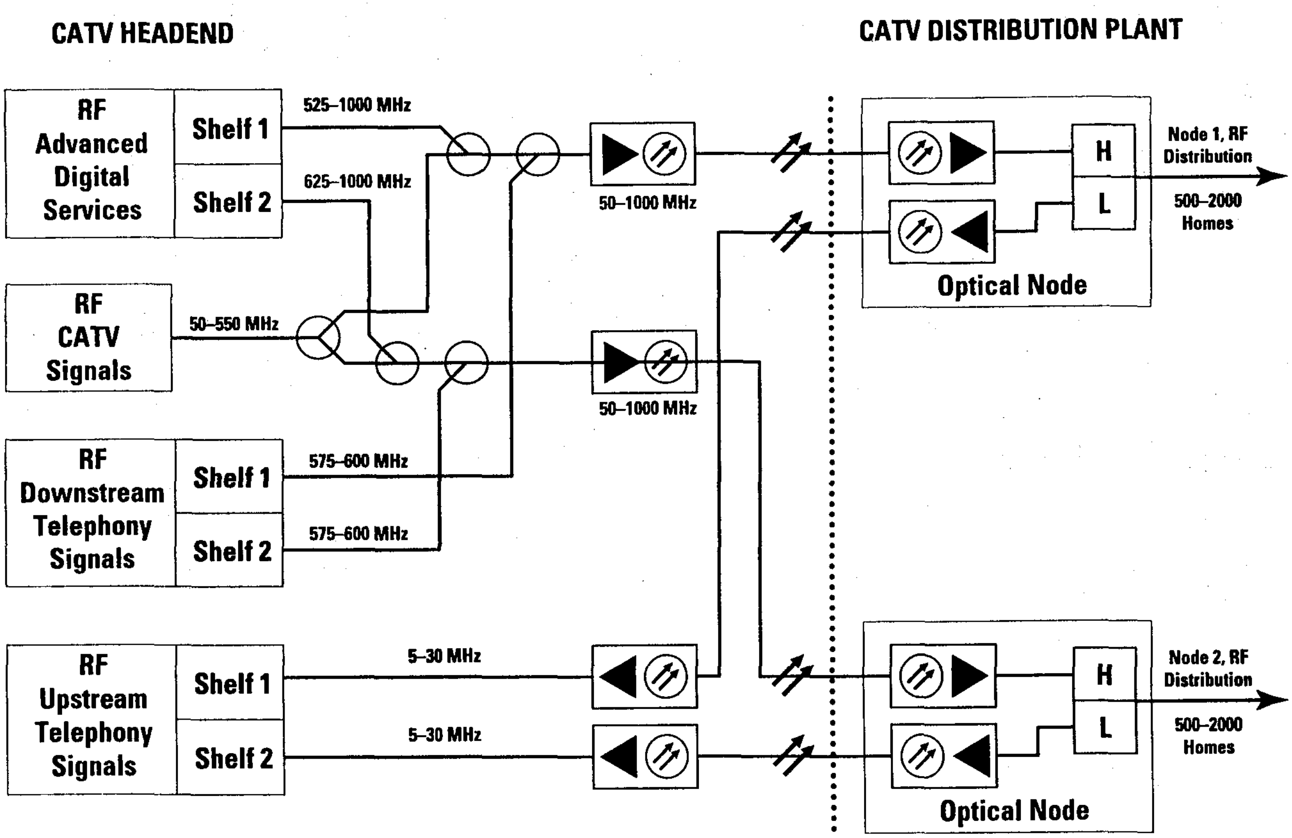


Figure 6. Advanced FSA Architecture with Digital and Telephony Services

Note that two downstream DFB lasers are now used, one for each node. Many envisioned advanced digital services (e.g., VOD) require unique data streams to be provided for each node, and this requires separate downstream lasers for each node (of course, these services could be combined with telephony on a separate laser per node as in Figure 5, providing this laser has adequate capacity).

Although DFB lasers are far more expensive than conventional Fabry-Perot lasers used for transmitting RF data carriers, using one DFB per node allows simplifying the remaining portion of the distribution architecture since these lasers can carry the telephony and advanced digital services as well, and all on one fiber. Of course, this means only one downstream receiver is required per node, as well. Comparing Figure 6 with Figure 5, this latter approach may make more economic sense from the outset (even if advanced digital services are not provided), certainly so if extra downstream fibers are not already present as required in figure 5.

Since the DFB lasers in Figure 6 are not being optically split as in previous examples, lower power lasers will suffice to cover the same distance to the node. Although the trend in the past has been toward higher power lasers to allow greater optical splitting, more systems will begin to require lower power lasers (but many more of them) as advanced architectures are deployed. Hopefully the increased volume and lower power requirements will drive down the cost of these relatively expensive DFB lasers. Long-term, the architecture in Figure 6 should be the more economical approach to delivering advanced services of all types in a hybrid fiber/coax system.

Network Reliability

Since the telephone serves as the fastest, best access a subscriber has to emergency services and information, any system delivering telephony must be highly reliable. In terms of the cable distribution system, many new demands will be placed upon the early detection and correction of potential faults (preferably before an actual failure occurs) and in locating failures and repairing them quickly when they do occur. This will require the deployment of advanced network monitoring and management systems.

However, appropriate monitoring and management systems are not enough to guarantee reliable service. For this reason, all powered devices in the network must have some form of emergency power for operation when utility power is not present. This power can be provided through battery back-up power supplies or gas-powered generators. In any case, switchover to backup power must occur quickly enough to guarantee no interruptions in service take place.

Finally, the distribution system itself must be made more reliable. In part, this can be accomplished through the use of better trade practices and higher reliability products in the construction of the system. Additional reliability can also be gained by providing other backup systems. In the case of fiber optic runs, duplicate fibers using route diversity should be used in case of an accidental fiber cut.

The active devices in the plant must also contain backup circuitry. All amplifiers and optical receivers must utilize circuitry which allows a failed component to be electronically bypassed or replaced by a functional equivalent. This will require all receiver and amplifier modules to be duplicated in each equipment housing, and an intelligent means of switching between these modules must be provided. Furthermore, this intelligent switching device must be tied into the overall network management and monitoring system so that any failure can be reported directly back to the office.

Fully redundant active devices must be deployed in a modular fashion. This allows failed modules to be replaced without bringing down the system, but it also allows optional initial installation of the active devices without the redundancy feature. This allows an operator to plan economically today for easy migration to full backup capability once telephony services are deployed. The backup modules should be operated as cold standbys (i.e., without power applied). This serves three purposes: power consumption is reduced, heat generation is decreased, and standby modules are less likely to be damaged by any surges which may be presented to the power supply. This increases module reliability while keeping operating costs lower.

SUBSCRIBER DROP ARCHITECTURE

Telephony Subscriber Terminal Inside Home

The subscriber drop architecture is critical for economical deployment of telephony services over a cable system. Costs added to maintenance, installation, or hardware at the drop are multiplied by the number of drops in the system. Several options exist for location, powering, and deployment of the Subscriber Terminal Unit (STU), and each of these options has unique advantages and disadvantages. However, regardless of whether the unit is inside the home or on the side of the home, the STU need only be installed when service is initiated. This ties capital expenditures to revenue generation and allows the operator to pay as he goes. A logical choice is to place the STU inside the subscriber's home (Figure 7).

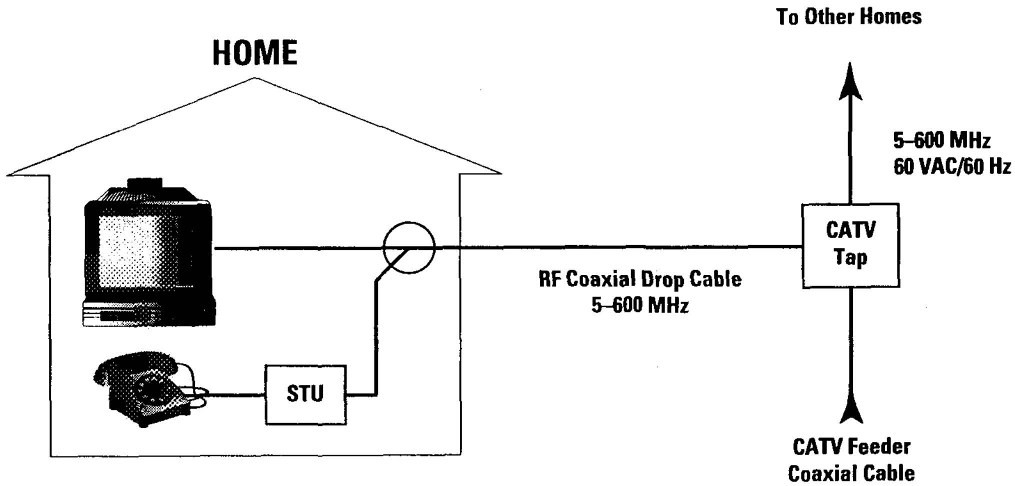


Figure 7. Telephony Subscriber Terminal Inside Home

The cable industry has a long history of placing cable television converters in the home, and many customers are used to their presence. Objections, when raised by customers, tend to focus not on the converter, but on the loss of functionality of consumer products attached to the converter. For an STU, no loss of functions are encountered since the subscriber's telephone equipment works as it always has. And since no interaction is required between the subscriber and the STU, the unit may be placed in a utility closet or any convenient location. Connection to the cable system may readily be accomplished with a directional coupler placed at any convenient point along the internal house wiring or even outside the home.

One strong advantage to placing the STU in the home is increased reliability and lower unit cost. The home provides a well-protected environment, free from temperature extremes and precipitation. An STU designed for in-home installation need not be temperature hardened or environmentally sealed.

Another logical choice in this situation is to power the STU from the home, and once again, the precedent exists for powering set-top converters from the home. The advantage here is that the subscriber absorbs the operational costs for powering the STU. Although an individual STU does not draw very much power by itself , the actual operator costs for powering the STU would not be trivial when one considers that thousands of the units will be operational.

Power failures also present a larger problem when delivering telephony service. When power fails in the home and the cable TV converter loses power, little is lost since the TV set has also lost power. But in the case of telephony, a power failure causing loss of telephone service could be catastrophic since access to emergency services is lost. Therefore, home-powered STU's have to provide some form of battery backup to maintain service during power outages. How long this battery needs to maintain service during a power failure is debatable, but most power failures tend to be relatively short in duration, typically well under four hours. Batteries also need periodic replacement, but modern sealed batteries have lifetimes of several years.

It is also possible to power an in-home unit from the network. However, network powering of the STU does not solve the battery problem, but simply moves the location of the batteries from the STU out into the network since the network itself now must provide battery backup for all the STU's. Fewer, but larger batteries will be required, but maintenance of the batteries should be easier since they will be concentrated at and co-located with the system's standby power supplies, and 24 hour access will be available. During extended power outages, gas-powered generators could be used to extend system operation beyond the batteries' capability.

Network powering, however, does create additional installation problems if power is carried on a separate cable from the drop and adds safety, regulatory, and reliability concerns if powered via the drop. If powered via the drop, power-passing passives must be used in the STU signal path, and care must be taken to block power down any other signal paths in the home. A step-down transformer would also likely be required to provide safer voltages in the home. Under these circumstances, any modifications the subscriber performs to his in-home wiring would likely generate a service call. For these reasons, powering an in-home STU from the network via the drop is not a viable option.

Telephony Subscriber Terminal on Side of Home

Another likely location to mount the STU is on the side of the home. This provides easy access should maintenance be required, but as indicated above, environmental hardening and sealing are now required. Standard twisted pair is used to feed the telephone service from the outside STU to the internal home telephony wiring. Locating the STU on the side of the home also makes it easy to clearly define the network termination point and where home wiring starts.

More powering options exist for a unit mounted on the side of the home. Powering is still possible from the home itself, but in this case a low voltage would be fed out to the STU from a plug-in wall transformer inside the home. This power could be routed to the STU either via a separate small-conductor power cable or by reverse feeding power to the STU up the coaxial cable providing RF to the TV. Of course, powering over any RF coaxial cable requires using power-passing and power-blocking passives where applicable. If powered from the home, separate cabling is the preferred method. But regardless of cabling, a backup battery must be provided with the STU. In this case the battery is always accessible to the operator should maintenance be necessary.

Powering may also be provided from the network via the coaxial drop (Figure 8), but in this case power may be blocked before it enters the subscriber's home, thus avoiding many of the safety and regulatory issues associated with providing 60 VAC down the drop. However, this introduces new problems, primarily at the tap.

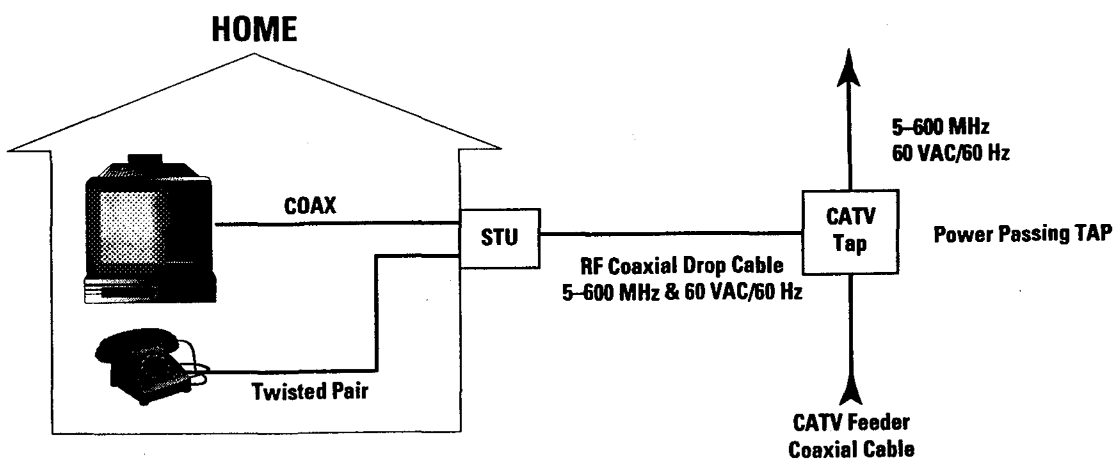


Figure 8. Telephony Subscriber Terminal Outside Home— System Powered on Drop

As new services are added to cable systems, more bandwidth will be required, and this requires higher performance components in the system. Taps are no exception to this, and although 1 GHz taps are now generally the standard when upgrades or rebuilds are performed, designing a tap to pass power down the drop without sacrificing RF performance (primarily in insertion loss and return loss) is next to impossible since power diplex filters must now be added to the main input/output ports as well as to each subscriber port. Decreased insertion loss performance can lead to frequency response problems or may require additional amplifiers to compensate for higher losses (or alternatively, fewer homes could be served per node). Poor return loss performance can potentially lead to intersymbol interference in digital signals.

Additional problems are likely to exist if powering is provided over the cable drop cable. First of all, the drop is required to be connected to the house electrical system ground near the point of entry into the home. The cable system itself is required to be bonded frequently to the power grid ground as well in the feeder system. Because of this, it is not uncommon to see sheath currents on the outer conductor of the coaxial drop cable. These currents can serve to add or subtract, depending on phase, from the power being provided over the drop to the STU. Additionally, several conductor interfaces exist in the drop, none of which have been optimized for power passing. First of all, the F-connector center conductor seizure mechanisms on most devices, while adequate for maintaining RF conductivity, do not provide much surface contact area or surface pressure to maintain good ohmic contact for power applications.

Next, several dissimilar metals are used in these interfaces which may cause corrosion under power-passing conditions. The outer braid and foil of the cable itself are aluminum, and as with aluminum house wiring, are subject to rapid oxidation and increased contact resistance. The F-connector itself is typically brass. The F-port on the tap is also typically brass, but is commonly nickel-plated. The center conductor of the drop cable is typically copper-clad steel, while the F-port seizure is typically tin-plated beryllium-copper.

Finally, moisture ingress at any of these connections will be even more critical to control in power-passing applications. Any decision to provide power via the drop cable will require careful analysis of all these factors if reliability is to be maintained.

Most of these problems can be avoided if network powering is delivered to the home via separate power conductors joined in a Siamese cable with the coaxial cable. This configuration is shown in Figure 9.

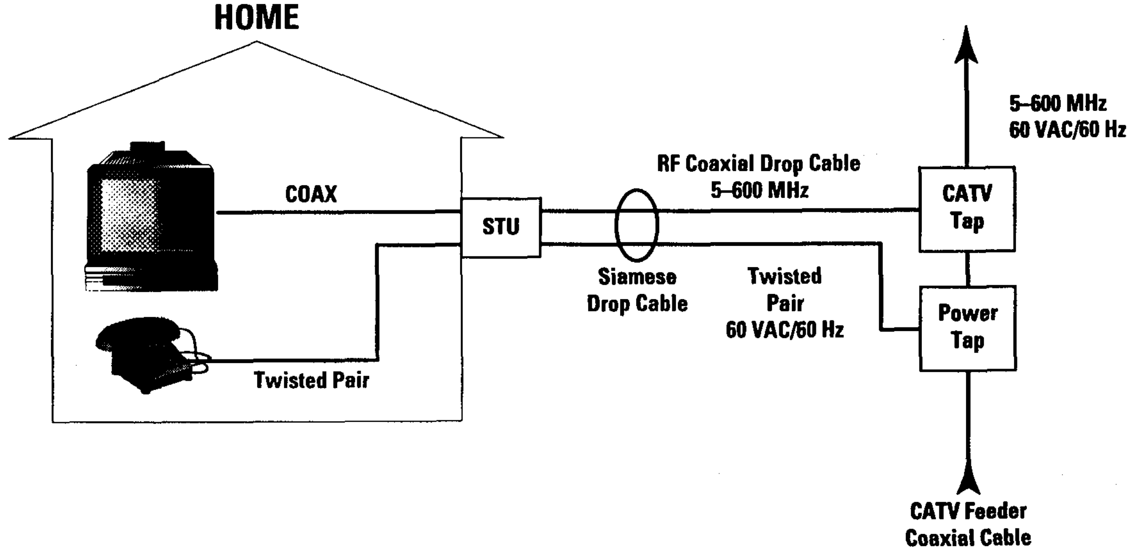


Figure 9. Telephony Subscriber Terminal Outside Home— System Powered on Siamese Cable

In this case power is extracted from the network via a power tap, which functions exactly as a power inserter does in today's systems to inject power into the cable system. This power tap need not be separate from the RF tap, and it is likely that taps will be available shortly which will include this function. Additionally, different regulations may apply regarding the maximum voltages which can be used if powering is not over the drop. If higher voltages can be used, power transmission efficiency can be improved through lower I^2R losses..

Telephony Subscriber Terminal at Curb Side

One other configuration exists for deploying the STU, but in this case the STU is located at the tap and supports multiple subscribers (typically 4, 8, or 16, though any reasonable number could be provided). Th is configuration is shown in Figure 10. In this case, telephony is served to the home over a Siamese coaxial/twisted pair cable.

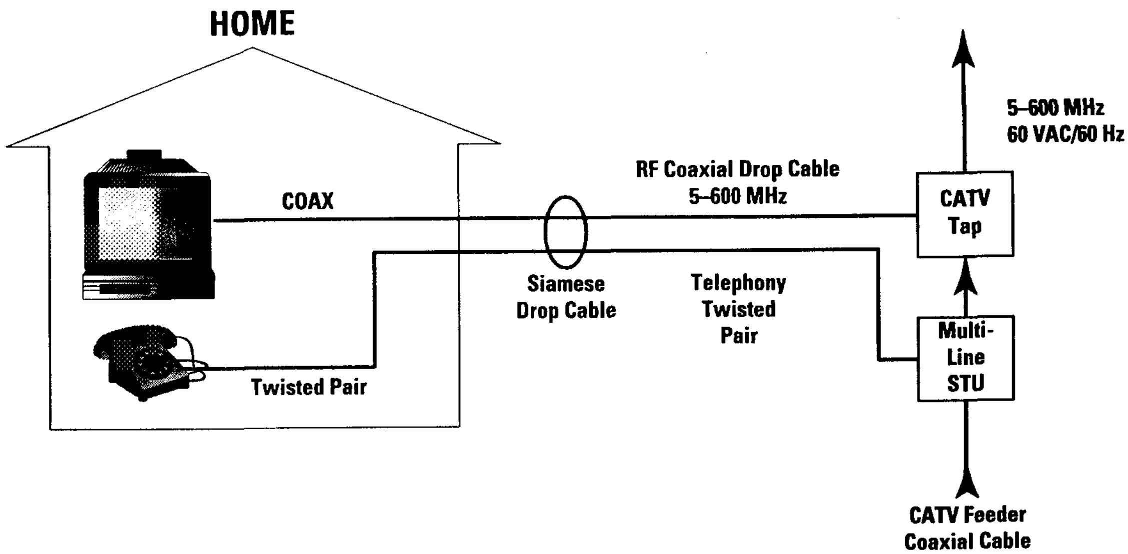


Figure 10. Telephony Subscriber Terminal at Curbside— Telephony on Siamese Cable

This approach has several advantages. First, reliable network powering is easily accomplished since this device is similar to a tap: it has an input and output directly connected to the feeder. Second, from a subscriber perspective this service appears exactly the same as his existing service since no unit need be installed on the subscriber's premises. Third, since multiple subscribers are served from one unit, significant hardware savings can be realized by eliminating the duplication of circuits required at each home when single subscriber units are used. For example, only one power supply is now required to serve several telephone subscribers whereas before one was required in each single-subscriber STU. Fourth, eliminating all these redundant circuits should lead to greatly reduced power consumption, thus lowering a system's powering costs. Last, those circuits which must still be unique to each telephone subscriber can be manufactured as plug-in modules and added to the STU only as new subscribers are signed up .

SUMMARY

Many advanced digital services will soon be deployed on hybrid fiber/coax cable systems. At the moment, this is the only system capable of delivering such services cost-effectively. Telephony will be one of the first services deployed. Many options and architectures exist for providing the signals needed to support telephony. Careful consideration must be made now to make sure both the distribution system and the subscriber drop are designed properly to be ready for this service and to be ready to evolve as more services come on-line.

Digital Backbone Network Applications for Inter-City and Intra-City Regional CATV Networks

Robert W. Harris
C-COR Electronics, Inc.
State College, PA

Abstract

This paper describes five working examples of how high speed digital fiber optic transmission systems are currently being applied within CATV networks to deliver video, audio and data services for distribution into the cable network. These digital networks provide high signal performance (RS250C medium haul and short haul) and can process and transport the entire range of signals present within the CATV environment including: baseband video and audio, BTSC subcarrier audio, baseband and IF scrambled video, 64 QAM and 16-level VSB signals and low-speed asynchronous data. Also presented here is a review of the digital building blocks that comprise the network and the justifications for selecting digital transmission in each network.

INTRODUCTION

Over the past four years, a notable number of MSO's have been installing and activating synchronous high speed digital fiber optic networks in CATV systems. These networks supply high quality video, audio and data signals to hub sites for distribution into their CATV networks.

Dedicated digital networks provide a low-cost, video-optimized, alternative to SONET standard systems which can cost as much as four times that of proprietary systems. In fact, 9 of the top 15 and 18 of the top 30 MSO's currently use dedicated PCM (pulse code modulation) digital transmission systems in their networks and gain many of the key advantages of a SONET based network at a fraction of the cost

These digital systems offer similar network functions as SONET networks by providing automatic self-healing redundancy, Drop/Add multiplexing techniques (Drop and Insert), Drop and Repeat functions, and transparent optical regeneration.

This paper will review the advantages of digital transmission and the building blocks required for developing networks with a variety of capabilities. This paper will then describe the flexibility and reliability of digital transmission through five examples of functioning digital networks.

DIGITAL BACKBONE NETWORKING ADVANTAGES

Digital backbone networks utilize synchronous time division multiplexing, uncompressed full linear pulse code modulation (PCM) and high data rate optical transmission for use in a variety of CATV, educational and broadcast applications. Some of the more well known advantages of digital backbone networking are:

- Signal performance unaffected by optical distance, optical splits or optical repeats
- Very high signal performance unaffected by system expansions or additions
- Robust transmission format
- Transparent drop and addition of video, audio and data channels
- Cost effective with AM Supertrunking [1,2]

Another advantage of synchronous time division multiplexed digital systems is its ability to accept and process a wide variety of signal formats found within the CATV environment.

These advantages work to enhance practical and realizable digital networks in CATV systems. Digital system functions such as switching, routing, multiplexing, drop and insert, drop and repeat and regeneration (repeating) allow flexible and expandable networks which transparently manipulate video, audio and data signals without degradation.

DIGITAL NETWORK BUILDING BLOCKS

Synchronous Time Division Multiplexing

Synchronous time division multiplexing (TDM) is used as a cost effective and practical way to achieve multi-channel operation and digital Drop/Add/Repeat capability. By optimizing the maximum data transfer efficiency in the multiplexer, a minimum amount of data "overhead" is needed to operate the multiplex structure. This allows the maximum amount of transmission channel capacity to be available for signal "payload". Maximizing channel capacity is significant since the main purpose of these networks is to transport multiple channels of high data rate video signals rather than lower rate voice and data channels.

Synchronous TDM also allows every channel to be fully independent of one another. Therefore, video, audio and data channels can be added or removed from the optical transmission channel without affecting any other signal or other part of the network.

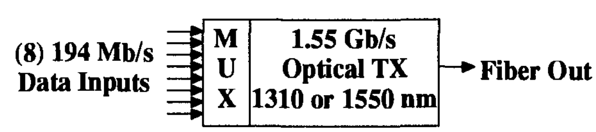
The synchronous TDM structure can also have a multi-level hierarchy (multiple TDM units operating within a system) each with a different but scalable rate. This is useful when the digital network is required to accept, process and transport a variety of signal formats.

Optical Transmission Terminals

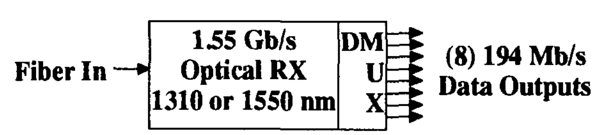
The optical transmission terminals are used to multiplex multiple data channels and launch the resulting high speed serial data stream onto the optical fiber (see figure 1a and 1b). As many as 16 video channels are transported on a single optical wavelength. Both 1310 and 1550 nm optical wavelengths are used and loss budgets are typically 30 dB. A typical range between the transmitter and receiver is <60 km, however, distances greater than 100 km can be achieved. Since both wavelengths are available, wavelength division multiplexing (WDM) can be used to operate optical carriers at both windows, thereby transporting as many as 32 uncompressed video channels per fiber.

Once the loss budget and/or the distance limitation is reached, either an optical receiver or optical regenerator may be used. Unlike a transmitter or receiver, an optical regenerator (repeater) does not perform any multiplexing or demultiplexing functions on the data stream and is used when it is not necessary to drop the signals at that location.

Multiplexer and Optical Transmitter
Figure 1a.



Optical Receiver and De-Multiplexer
Figure 1b.

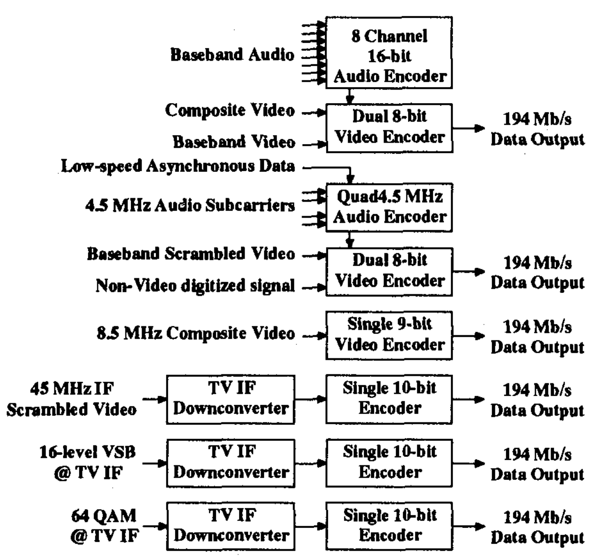


Signal Codecs

The signal codecs (acronym for COder/DECoder) provides the A to D (analog to digital) and D to A (digital to analog) conversion. These codecs have been designed to accept and process a variety of input signal formats found in the CATV environment. These include baseband video, composite video, baseband audio, 4.5 MHz audio subcarriers, baseband scrambled video, IF scrambled video and TV IF carriers. Because of the wide range of signal formats and performance requirements in the CATV network, the signal codecs require different input bandwidths, sampling frequencies, sampling accuracy, and coding formats.

Therefore, as part of a complete synchronous TDM system, the digital data ports are designed to accept any digital coding and/or framing pattern. This allows different types of signals to be multiplexed together and transported in the same optical channel. This could include video codecs with different sampling accuracy or coding formats (8-bit or 10-bit), digitized non-video signals and future signals such as digitized HDTV or digitally compressed signals.

Various Signal Codecs
Figure 2.

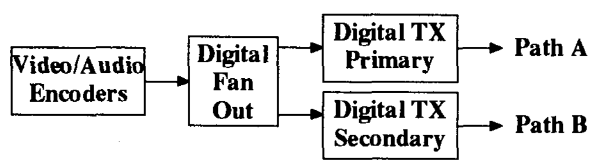


CONFIGURING DIGITAL BACKBONE NETWORKS

Redundant Transmitters

The configuration shown in figure 3 is for redundant transmitters. The output of the video encoders are routed to a "Digital Fan Out" device. The digital fan out simply duplicates the input data "n" times. In this example, the digital fan out is a 1x2 fan out and each output is routed to two separate transmitters. This configuration is usually done to provide redundancy on two paths (primary and secondary) to the receive site(s). Note also that both transmitters do not have to operate at the same optical wavelength.

Redundant Transmitters
Figure 3.

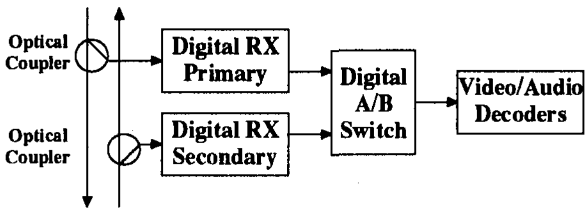


Redundant Optical Receivers

Figure 4 shows an application with redundant receivers. Each optical receiver accepts the optical signal from a separate path. The optical couplers shown are not necessary but do show how the signals can be dropped at the receive site and passively passed on the next receive site via the optical couplers.

The demultiplexed data from the outputs of each receiver (primary and secondary) are routed to a "Digital A/B Switch" device. This intelligent switch will pass the primary path signals to the decoders. However, in the event of a fiber cut, or a failure in the primary transmitter or receiver, the digital A/B switch will switch automatically and instantaneously to pass the secondary receiver output to the decoders.

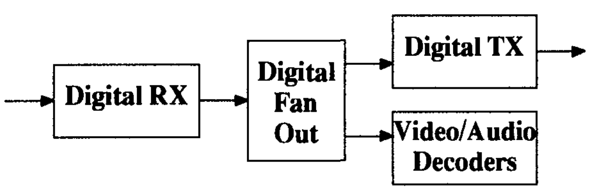
Redundant Optical Receivers
Figure 4.



Drop and Repeat

Figure 5 shows a network technique known as "Drop and Repeat". This allows for some or all of the transported channels to not only be dropped off at the receive site and decoded but also be repeated, *transparently*, to another receive site. The repeating process is transparent because the received data signals remain in a digital format as they are routed to the input of the repeating transmitter, hence, no signal performance degradation.

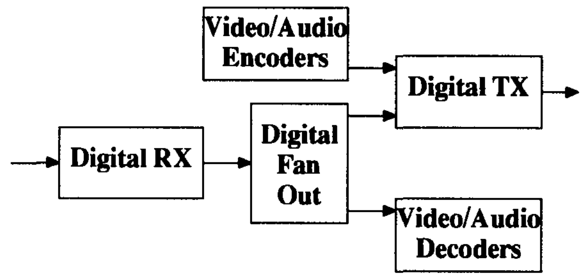
Drop and Repeat
Figure 5.



Drop/Add and Repeat

Figure 6 shows a network technique known as "Drop, Add and Repeat". Like drop and repeat, this too allows for some or all of the transported channels to be dropped off at the receive site, decoded and also be repeated, *transparently*, to another receive site. However, additional channels may also be added or inserted at this site for transmission on the network. Drop/Add multiplexing plays a significant role in digital systems where there are multiple signal origination points within the network.

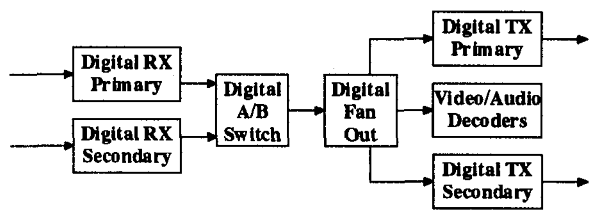
Drop/Add and Repeat
Figure 6.



Redundant Drop/Redundant Repeat

Figure 7 shows the maximum level of redundancy at a digital site using both redundant optical transmitters and receivers.

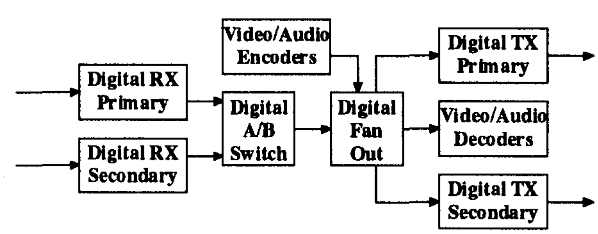
Redundant Drop/Redundant Repeat
Figure 7.



Redundant Drop and Add with Redundant Repeat

Figure 8 shows a site configuration similar to "Drop, Add and Repeat" except for the added redundancy in the optical terminals.

Redundant Drop, Add, Redundant Repeat
Figure 8.



DIGITAL FIBER OPTIC NETWORK EXAMPLES

The following five examples show the flexibility and reliability features of high speed synchronous digital fiber optic networks.

Network Example 1.

A fully redundant self-healing digital ring network transporting 64 channels to five hub sites around a metropolitan area and making extensive use of transparent Add/Drop Multiplexing and Drop/Repeat functions.

This network is shown in figure 9. The franchise area encompasses most of the suburbs around a medium-sized midwestern city. Originally, the sites were both standalone headends and AML microwave receive sites. The system engineers wanted a network with the following functions; a central headend for primary satellite signal reception, a single location for primary advertisement insertion and full, automatic redundancy via a backup headend which was already in place.

A digital backbone was chosen because of the relatively long distances between hub sites and for the ability to perform the "Drop/Add/Repeat" functions transparently.

The digital network uses a variety of configuration building blocks shown earlier. The network includes a primary and backup headend as well as two sites which serve as the primary insert for local off-air channels. The local off-air channels are inserted at these locations because the reception at these sites are much better than at the primary site.

The channels from the backup headend, which contain the entire channel line-up, are also carried on the digital network. Any channel that fails in the primary headend or fails anywhere on the primary path is automatically switched to the backed-up feed until the problem is repaired.

The majority of channels are processed as composite baseband video and 16 channels each are transported at 1.55 Gb/s per wavelength. A long distance path (>66 km) uses 1550 nm optical terminals. The rack space required for the digital equipment at each site is approximately one and half six foot racks. In two of the sites, all equipment (digital, modulators, processors, etc.) is housed in underground closed environmental vaults. The remaining sites use existing buildings.

Network Example 2.

A fully redundant self-healing Inter-City digital ring network transporting 80 composite IF channels to six hub sites at the 1550 nm optical wavelength within a large metropolitan area and providing direct RF outputs at the hub.

This CATV system serves over 4,000 miles of cable plant in one of the largest cities in the U.S. Originally, there were several standalone headends throughout the area but were eventually replaced with high power AML to seven hub sites.

As improvements in end-of-performance and network reliability became increasingly important, alternate methods of delivering high quality signals to these sites were analyzed. AM Supertrunking, based on lightly loaded AM transmission (LLAM), and digital transmission were both considered.

Digital transmission was chosen because of its consistent performance, its ability to transparently repeat the signals throughout the network and its ability to be configured in a fully redundant, automatically self-healing ring network. The network is shown in figure 10.

The network makes extensive use of redundant transmitters, redundant drop techniques and optical splitting. Should a failure occur at any point in the network (fiber break or optical terminal loss), the network automatically by-passes the fault so that each hub site remains "on-line".

All 80 channels are processed at TV IF. In other words, the IF output from all modulators and IF scramblers in the main headend are routed to the input of the digital equipment. Therefore, no scramblers are required at any of the receive hub sites and only IF to RF upconverters are required for on-channel frequencies. Planned future signal formats to be transported over this digital network will include digitally compressed video using either 64 QAM or 16-level VSB.

The entire network is operating at the 1550 nm wavelength at a data rate of 1.55 Gb/s. The network was designed as such to exploit the lower loss at 1550 nm which results in only requiring two signal regeneration points. If 1310 nm optical terminals were used, at least four regeneration points would have been required. Therefore, the use of 1550 nm terminals within the network reduced the overall network cost.

The required rack space is equivalent to about three full six foot racks. All receive site equipment is housed in existing environmentally controlled buildings.

Network Example 3.

A 60 channel point-to-multipoint network processing and transporting a variety of signal formats throughout an Intra-City Regional Network within a state to six separate hub sites. Employs Drop and Repeat & Add/Drop multiplexing techniques.

Figure 11 shows the layout of this network. The primary reason to install this network was to eliminate seven standalone headends. Each headend served a different CATV system in different cities, each with about 5,000 to 15,000 passings and all systems are owned and operated by the same MSO.

Each system in this region is to be upgraded from 300 MHz to 550 MHz and will, therefore, require the addition of about 40 channels per city. But the cost to add 40 channels at each standalone headend, including satellite receivers, satellite antennae, modulators, scramblers, etc., was difficult to justify.

A Regional Headend concept was determined as the most cost effective way to centralize the capital-intensive investments and spread these investments across a wide base. Digital transmission technology was chosen as the best method to deliver headend quality signals to these sites. Further, unlike an AM system, the digital network is "channelized" and therefore each digital receive site can provide its own channel line-up independent of the other hubs.

The digital network is a point-to-multipoint configuration operating at 1310 nm at a transmission rate of 1.55 Gb/s. Multiple signal formats are transported including; baseband video, composite video, IF scrambled video, 4.5 MHz audio subcarriers and RS232 data. Future signal formats to be transported over this digital network will include digitally compressed video using either 64 QAM or 16-level VSB.

The total cost per received channel for this network is about $2,000.

Network Example 4.

An Intra-City Regional Digital Network transporting 64 CATV channels from Toronto to Ottawa through a distance of 425 km using 1550 nm optical terminals and five optical regeneration points.

This network is shown in figure 12. A redundant ring network is in place around the city of Toronto. The ring network is extended to Ottawa via digital transmission and five regeneration points through a total fiber distance of 425 km. There is no signal performance degradation at the end of the link since the signal has remained in a digital format.

The regeneration points could be easily upgraded at a later date to a drop and repeat site if any of the sites were required to distribute the signals from that location.

All channels are processed at composite baseband video. Sixteen channels are TDM at 1.55 Gb/s and optically transported at the 1550 nm wavelength. Each digital regenerator is located within an environmentally controlled building.

Network Example 5.

A bi-directional, fully redundant, self-healing Inter-City digital ring network transporting 80 channels to nine hub sites within a large metropolitan area and providing direct RF outputs at each hub.

This CATV system serves over 4,500 miles of cable plant in one of the largest cities in the U.S. Originally, there was single headend and high power AML transmission to eight hub sites.

As improvements in end-of-performance and network reliability became increasingly important, alternate methods of delivering high quality signals to these sites were analyzed. AM Supertrunking and digital transmission were both considered.

Digital transmission was chosen because of its consistent performance, its ability to transparently repeat the signals throughout the network and its ability to be configured in a fully redundant, automatically self-healing ring network. The network is shown in figure 13.

The network makes extensive use of redundant transmitters, redundant drop techniques and optical splitting. Should a failure occur at any point in the network (fiber break or optical terminal loss), the network automatically by-passes the fault so that each hub site remains "on-line".

All channels are processed in the headend as baseband video (including baseband scrambled video) and separate 4.5 MHz audio subcarriers. Sixteen channels are transmitted per optical wavelength at a data rate of 1.55 Gb/s. The receive site video decoders use a patented technique for directly converting digital video to TV IF (45.75 MHz). The IF outputs are then routed to frequency agile IF to RF upconverters. All codecs, optical terminals, and IF/RF upconverters are driven by the same master clock frequency which yields highly accurate RF frequency outputs.

The cost for the entire digital network, including RF outputs, full optical redundancy and a single digital video return channel from each hub site is about $4,000 per channel.

CONCLUSIONS

Digital fiber optic transmission using synchronous TDM has been shown to be a cost effective and practical way to achieve multi-channel operation, Drop/Add capability and fully automatic self-healing redundant ring networks. Synchronous TDM not only facilitates multi-channel capability, but also allows every channel to be fully independent of one another. As a result, video channels can be added or removed from an optical transmission channel without affecting any other signal or other part of the system. The same advantages apply when TDM is applied to auxiliary services such as audio and data signals.

Another key characteristic of the digital systems described here is their ability to interface with a variety of input signal formats found in CATV systems. The synchronous TDM used allows any digital coding and/or framing patterns as long as its frequency synchronous with the system clock. This allows different types of signals to be multiplexed together and transported in the same optical transmission channel.

High speed synchronous TDM digital systems have been used to build video fiber optic networks from a set of basic building blocks consisting of optical terminals, signal codecs and simple processing elements such as digital switches and fan outs. These systems provide uniform signal performance independent of the number of channels transported as well as the types of video, audio and data services carried on the network. Additionally, the robust characteristics of digital transmission ensure system performance that is unaffected by link distances or repeats.

These characteristics have shown that flexible and reliable digital networks are realizable today and meet changing system needs such as system expansions, system additions or new signal formats.

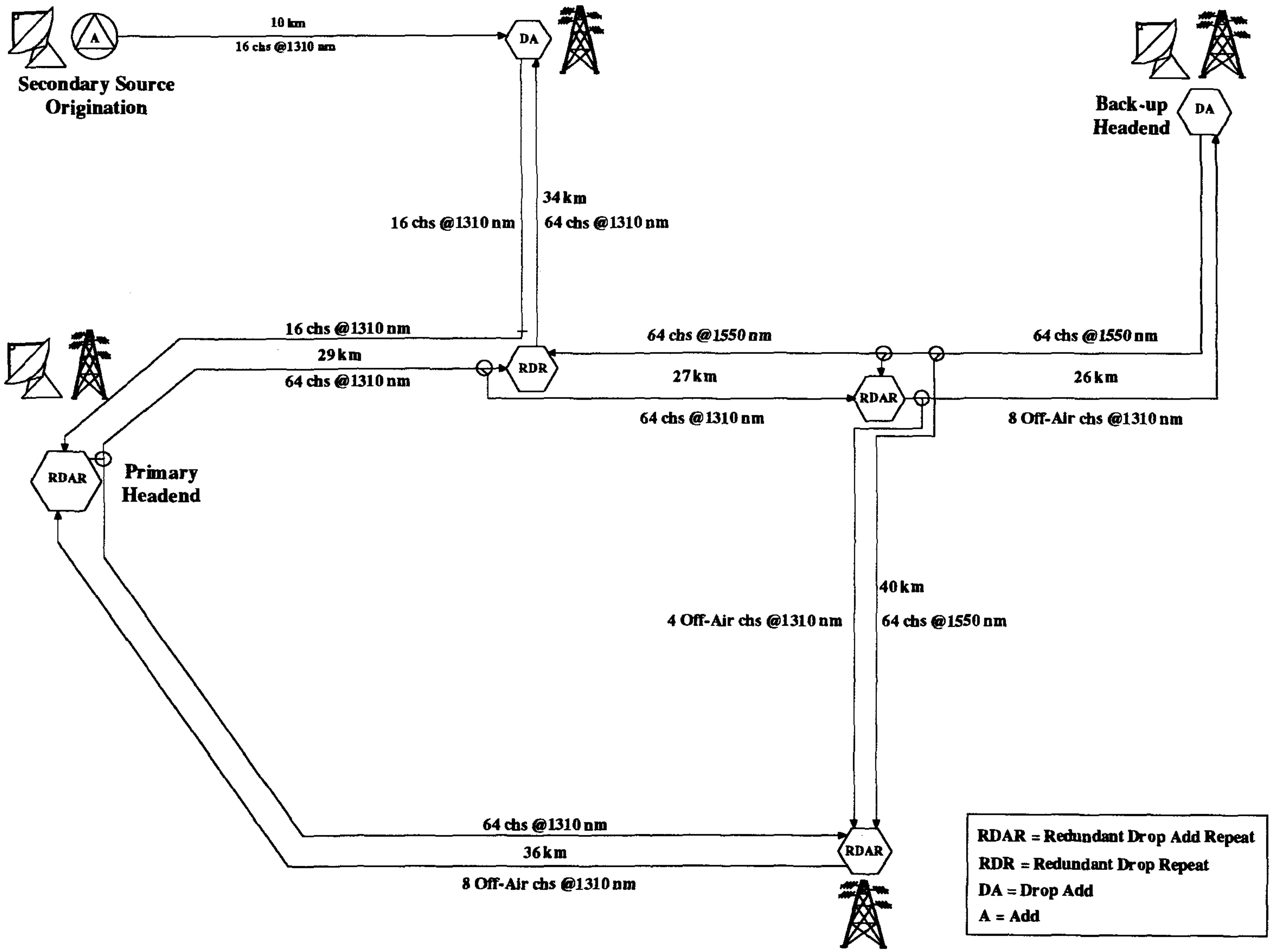
REFERENCES

[1] R.W. Harris, "Digital Video versus AM Supertrunking: A Cost and Performance Analysis", CED, July and Aug. 1993.

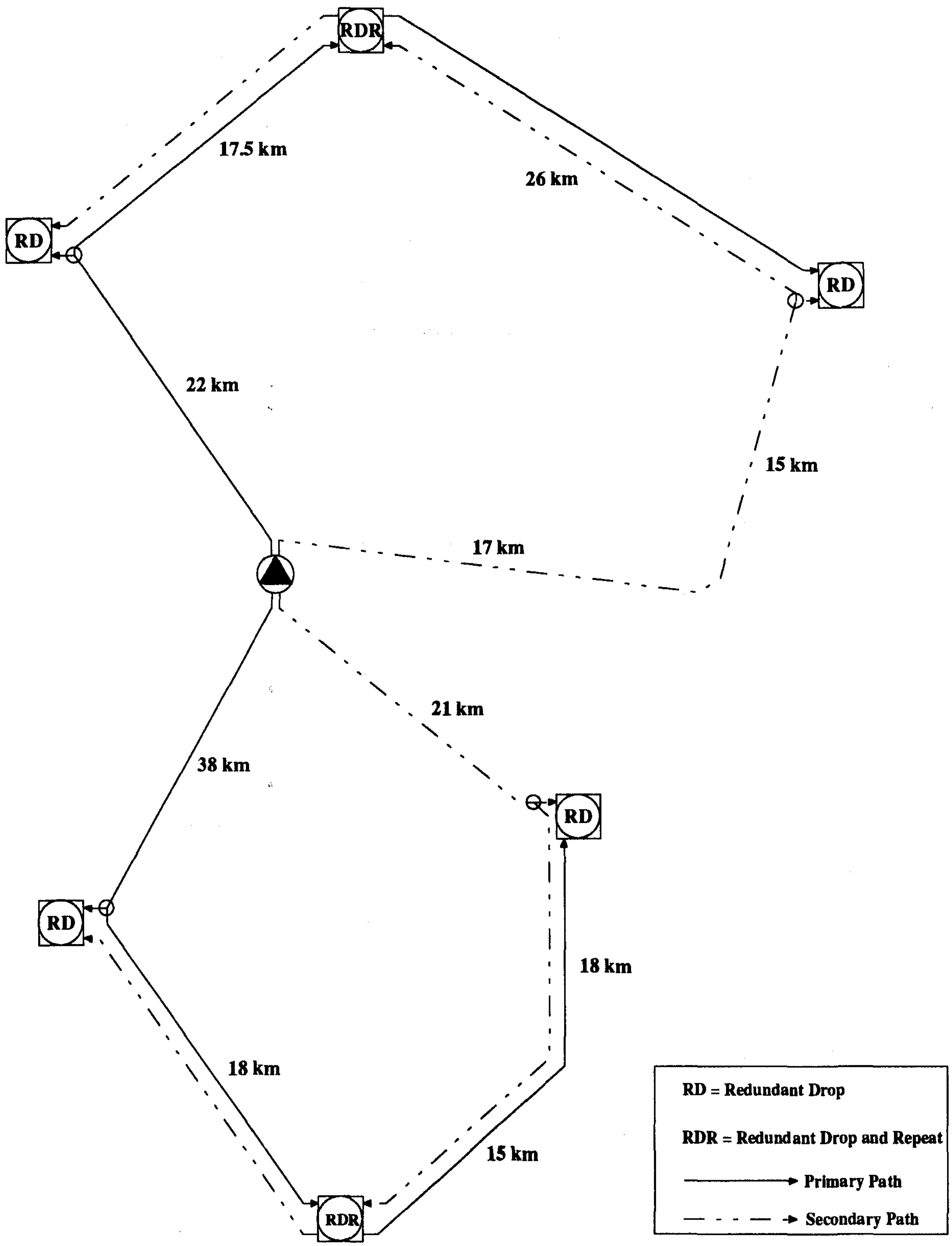
[2] R.W. Harris, "A Cost and Performance Comparison Between Uncompressed Digital Video and Lightly Loaded AM Supertrunking Methods within CATV Network Upgrades and Rebuilds", to be published in 1994 NCTA Proceedings, May 1994.

[3] S.D. Dukes, "Photonics for Cable Television System Design", Cable Television Laboratories, Inc.

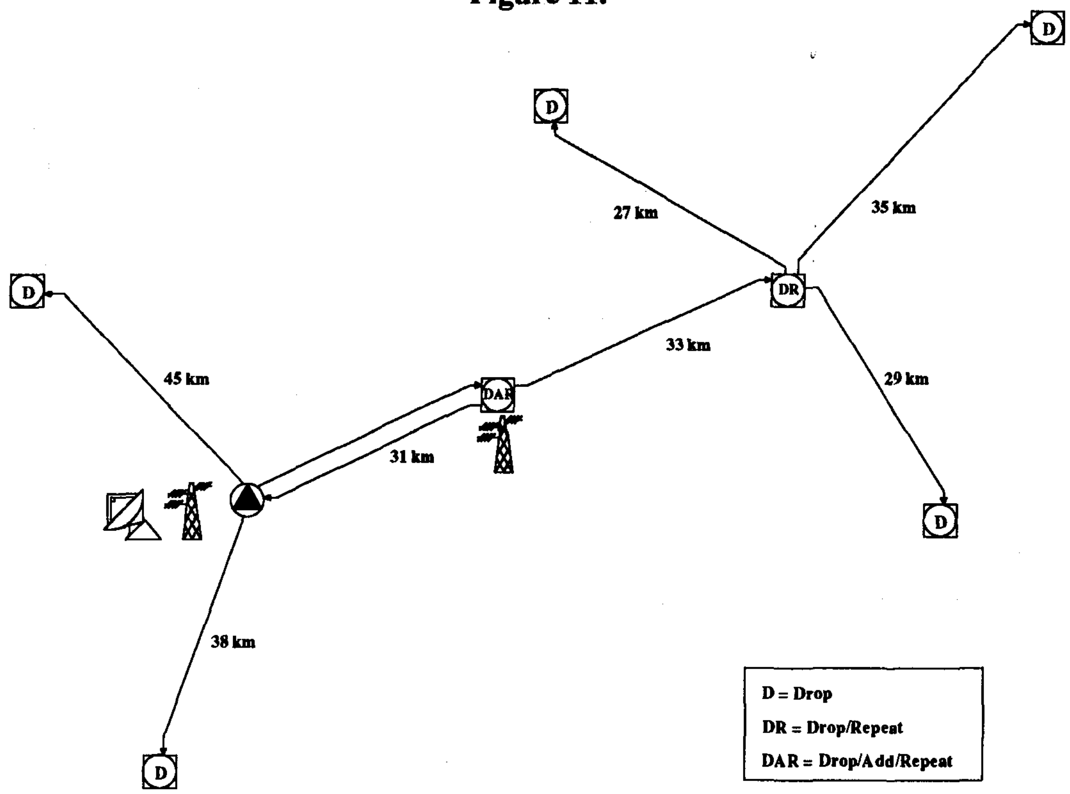
Inter-City Redundant Ring Network with Backup Headend, Drop/Add and Automatic-Self Healing Redundancy
Figure 9.



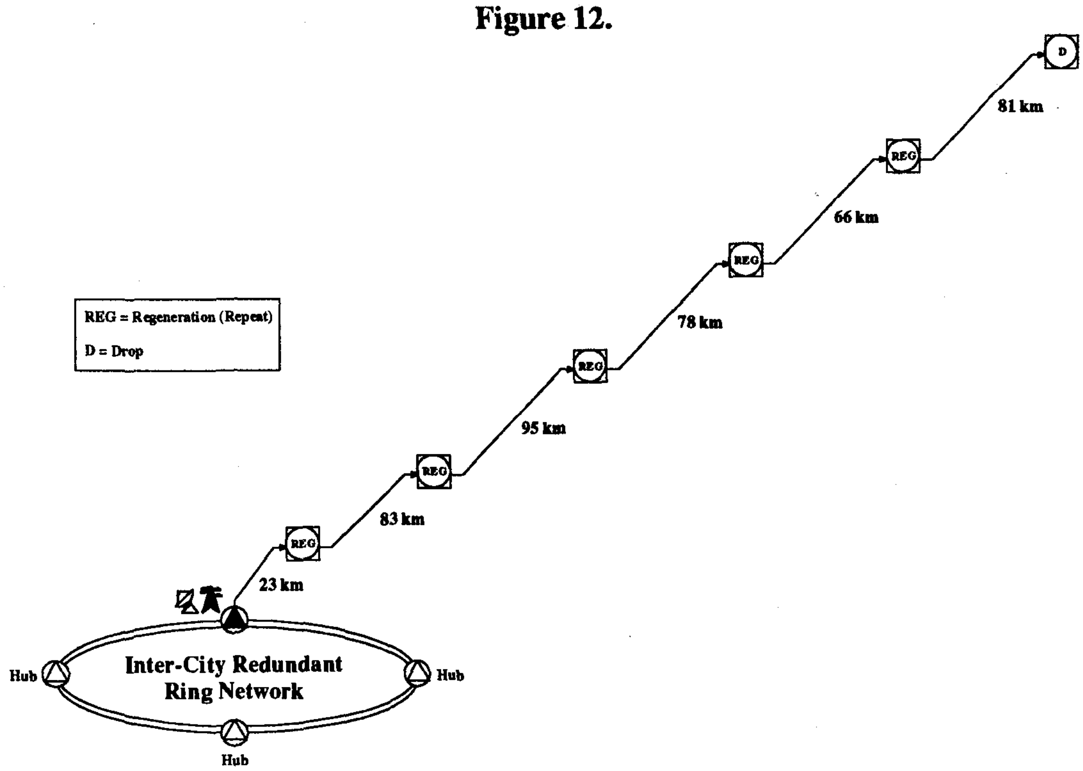
Inter-City Digital Network with Full Automatic Self-Healing Redundancy Transporting Composite IF Inputs and Providing Direct RF Outputs
Figure 10.



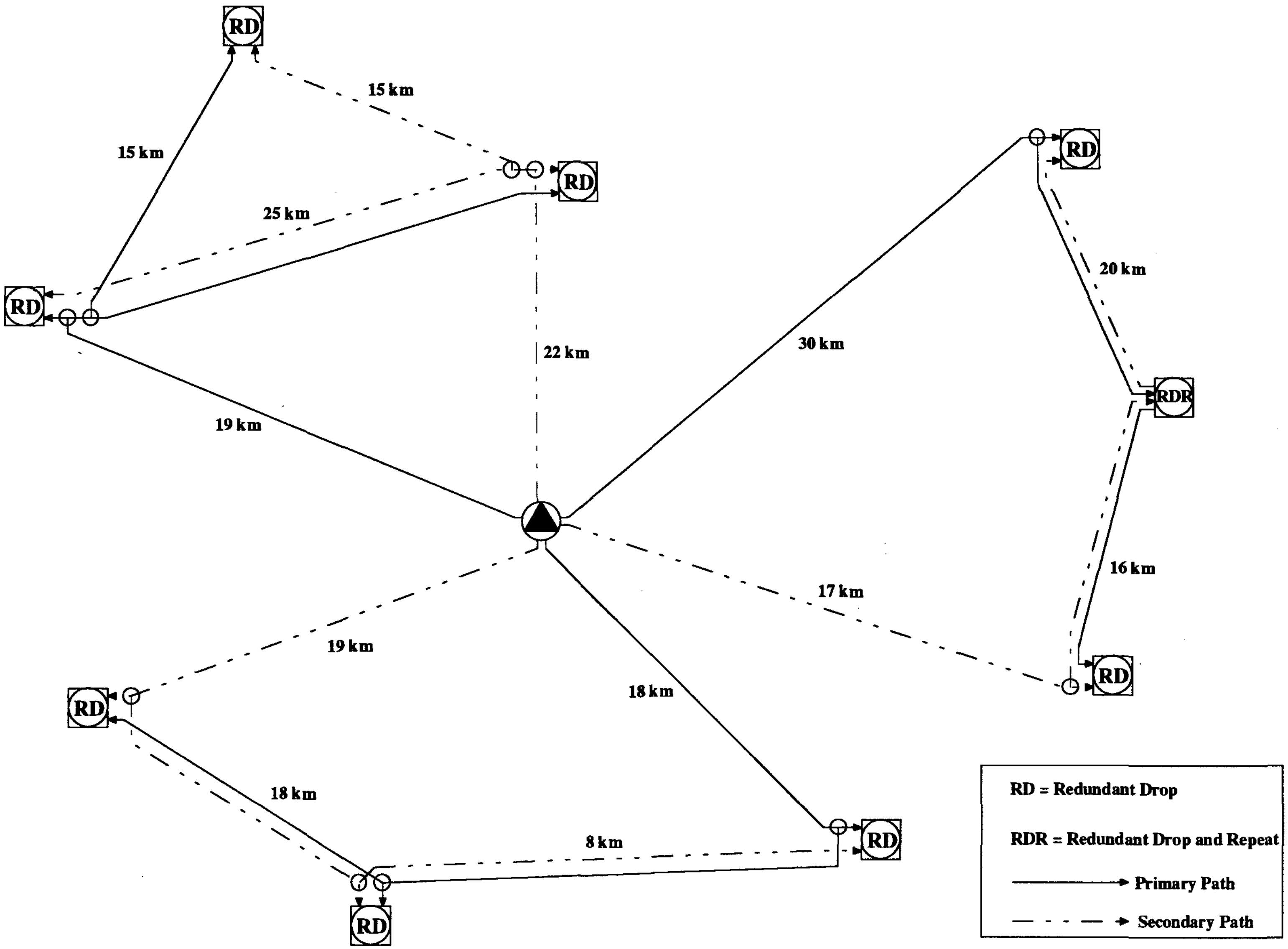
Intra-City Regional Network for Headend Replacement
Figure 11.



Intra-City Regional Network Through 425 km
Figure 12.



Inter-City Digital Backbone Network with Automatic Self-Healing Ring and Direct RF Outputs
Figure 13.



DIGITAL COMPRESSION—INTEROPERABILITY ISSUES FOR THE CABLE INDUSTRY

TV/COM International
16516 Via Esprillo
San Diego, CA 92127

Graham S. Stubbs
Consultant

Tony Wechselberger
Executive Vice President

Abstract

Standardization to achieve interoperability of digital video compression technology will require more than finalizing ISO MPEG-2 and resolving concerns about security. Interoperability goals encompass the seamless transmission of digital programming across equipment, market, and national boundaries, and include availability of equipment from multiple vendors to work on the same programs. The layered structure of digital video compression lends itself to consideration of standardization issues at the levels of compression encoding, multiplex assembly, conditional access and encryption, and processing for transmission.

This paper discusses the urgent need for the cable industry to take charge of the process of assuming that its long-term interoperability requirements will be met.

INTRODUCTION

The considerable progress in the development and adoption of ISO MPEG-2 working draft standards does not yet assure the cable industry that long term benefits of interoperability will be met. At a time when the industry is planning for, and in some cases implementing facilities to compress, multiplex, and distribute cable programming, these issues deserve urgent close attention. Interoperability and standardization are not just a matter of encryption standards—the issues are broader.

Compression Standards

Present plans for a 1994-5 roll-out call for a hybrid of MPEG and non-MPEG compression to be implemented for U.S. national distribution to cable subscribers. This approach, while appearing expedient, should receive close scrutiny regarding the long term implications for interoperability with other media, and with other consumer equipment. In addition, virtually the rest of the world is imposing MPEG-2 Main Profile Main Level edicts on systems and equipment to be used in similar applications. The alternative of a full MPEG-2 compliant approach is still achievable if the industry acts promptly.

Transport Stream Standardization

Although present U.S. cable plans do call for use of the MPEG-2 Transport Stream syntax, enough variations and designer-specific freedom is permitted within MPEG-2 that multi-vendor interoperability is by no means assured. Here the cable industry has an opportunity to narrow the options for variations within the MPEG-2 structure, such that an openly specified syntax can be used by all vendors with assurance of interoperability of equipment to which a clear (non-encrypted) MPEG-2 Transport Stream is delivered.

Encryption

To achieve broad agreement on an entire security methodology would be very difficult and possibly undesirable from the point of view of vulnerability to piracy. Nevertheless, it should be possible to develop an agreement on scrambling (i.e. the algorithm used to encrypt the transport stream packets), and yet leave the methods for key delivery (i.e. conditional access) open to continuing development, and allow employment of various proprietary methodologies, including experimentation with a variety of approaches for the use of smart cards.

In Europe, the Digital Video Broadcast Group (DVB) has already started to successfully tackle some of these issues with an objective of achieving broad interoperability. In Canada, also, there appears to be a determination to achieve a similar objective through the work of ABSOC (Advanced Broadcasting Systems of Canada).

The U.S. cable industry cannot afford to ignore the opportunity to determine its own digital future. This paper identifies the principal remaining interoperability issues in the digital arena, and suggests that industry take control of the process of assuring standardization to meet the industry's long-term needs.

INTEROPERABILITY GOALS

The context for discussing interoperability of digital compression systems cannot be as a single industry or a single country.

There will develop a *global* market for programming that is prepared and stored in digitally compressed form. Compressed programs distributed by satellite within one country will find markets in other countries within the satellite footprint. The technologies encompassed by digital video compression are global in nature, not the proprietary fiefdom of a single vendor.

Thus the principal objectives in interoperability relate to:

- Seamless exchange of programming services between media, whether Digital Storage Media (DSM) such as a file server, or transmission media such as satellite, cable, telephony, or others.
- Compatible operation of equipment that is provided by multiple vendors.

A standards-based "Open Architecture" is desired for North America. Open architectures ensure lowest cost seamless transitioning between media, and ultimately, consumer equipment. Cooperation between vendors will only occur through market forces that dictate interoperability for true multiple vendor savings and full realization of the benefits of an open architecture.

The majority of the world's consumer electronics companies have already committed to MPEG video and audio. North American cable operators, satellite program distributors, and program providers must also press for strict adherence to the standards developed under the

International Organization for Standardization (ISO) MPEG-2.

The cable industry must be allowed to realize the benefits of a truly seamless implementation of digital technology to ensure that digital video, audio, and data can be moved using the same protocols from program provider to satellite uplink, through cable systems to set-top boxes in the homes of cable subscribers. The industry cannot afford to have road blocks in the system such as proprietary, non-standard products mixed in with standard MPEG-2 elements. The industry can not at all really afford a later "upgrade to MPEG-2."

In some instances, conversion of protocol or signal format may be inevitable (for example, when converting between different modulation methods which are optimum for satellite and cable transmission). Standards choices should maintain protocol integrity, and should be consistent with minimizing the cost of such transformations.

Equipment provided by multiple vendors must be capable of operating with digitally compressed programming. **Multiple vendor cooperation should not, however, be equated with multiple licenses of a single vendor's proprietary system.** All vendors must agree to work within the MPEG-2 framework, and must further work to an open definition of all parameters necessary to achieve guaranteed interoperability. There is no place for proprietary compression /decompression technology or for compromise products which fail to deliver the full potential of MPEG-2. The proprietary approach ensures single vendor control over technological applications, stalls the development of a healthy, level, competitive playing field, and may bring only the limited resources of a single vendor to bear on problems as they develop, rather than the complementary resources of an entire industry.

LAYERS IN DIGITAL VIDEO COMPRESSION

In order to discuss interoperability and standardization issues in any detail, it is worthwhile to review the **layer concept** of the digital video compression system. Figure 1 illustrates the layer concept, starting with an inner core of program compression encoding and building through successive layers of transport transmission.

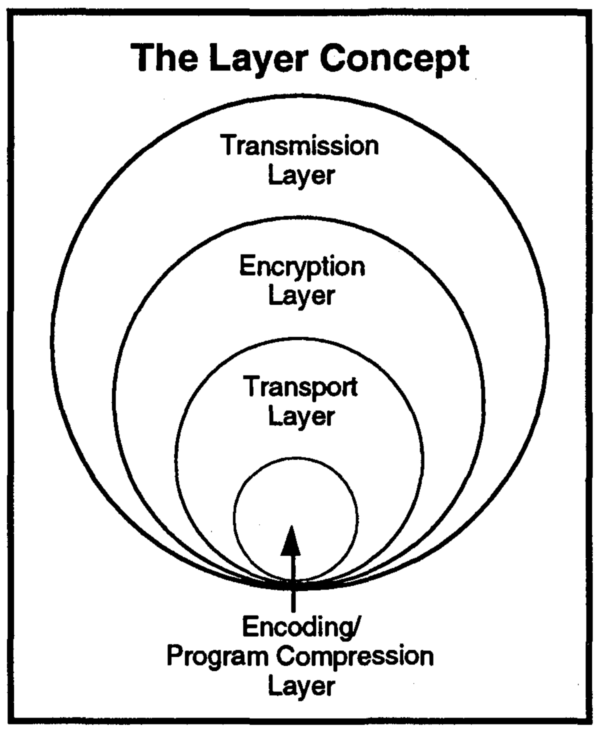


Figure 1

The **program compression encoding layer** comprises the following:

- Video compression coding, which may include intrapicture redundancy reduction and intraframe redundancy reduction/ prediction. Various coding rates achieve differing degrees of picture quality.
- Audio compression coding, based on psycho-acoustic masking. (Both Musicam and Dolby are of this type.) Various coding rates are used to achieve specific sound quality, and the number of audio channels may vary as a function of the number of languages desired and/or type(s) of audio such as mono, stereo, or surround.
- At the compression encoding layer, packetized elementary streams are formed corresponding to video, audio, and other data.
- Stream identities are created corresponding to the program source.
- Presentation time stamps are added to assure synchronized delivery of video and audio components of a given program.

In the **transport layer**, bitstreams are split into equal length packets for efficient sorting:

- Each packet carries header information to identify the content (payload) and purpose of the packet.
- Additional packets carry control and other data.
- Digitized compressed streams of transport packets can be built into a multiple program multiplex.

Security and conditional access functions are added in the **encryption and authorization layer**:

- The packetized, multiplexed transport bitstream is scrambled by performing a mathematical operation (encryption) involving another piece of digital information (an encryption key) which can be changed frequently. This operation can be applied differently to each successive transport packet.
- The keys necessary for descrambling are delivered only to authorized users and thus the processes of encryption/ decryption and program authorization and conditional access are intimately related.
- The process of key delivery/key management and general box entitlement is the "conditional access" function. Different conditional access systems can be used to deliver common descrambling keys.

The **transmission layer** prepares the compressed, packetized, multiplexed, encrypted bitstream for RF transmission:

- Since transmission channels introduce noise and distortions to digital signals, forward error correction (FEC) coding is applied to facilitate detection and correction of data errors.
- Efficient (and error-free) use of channel space is critically dependent upon the choice of RF modulation method for a given transmission medium.

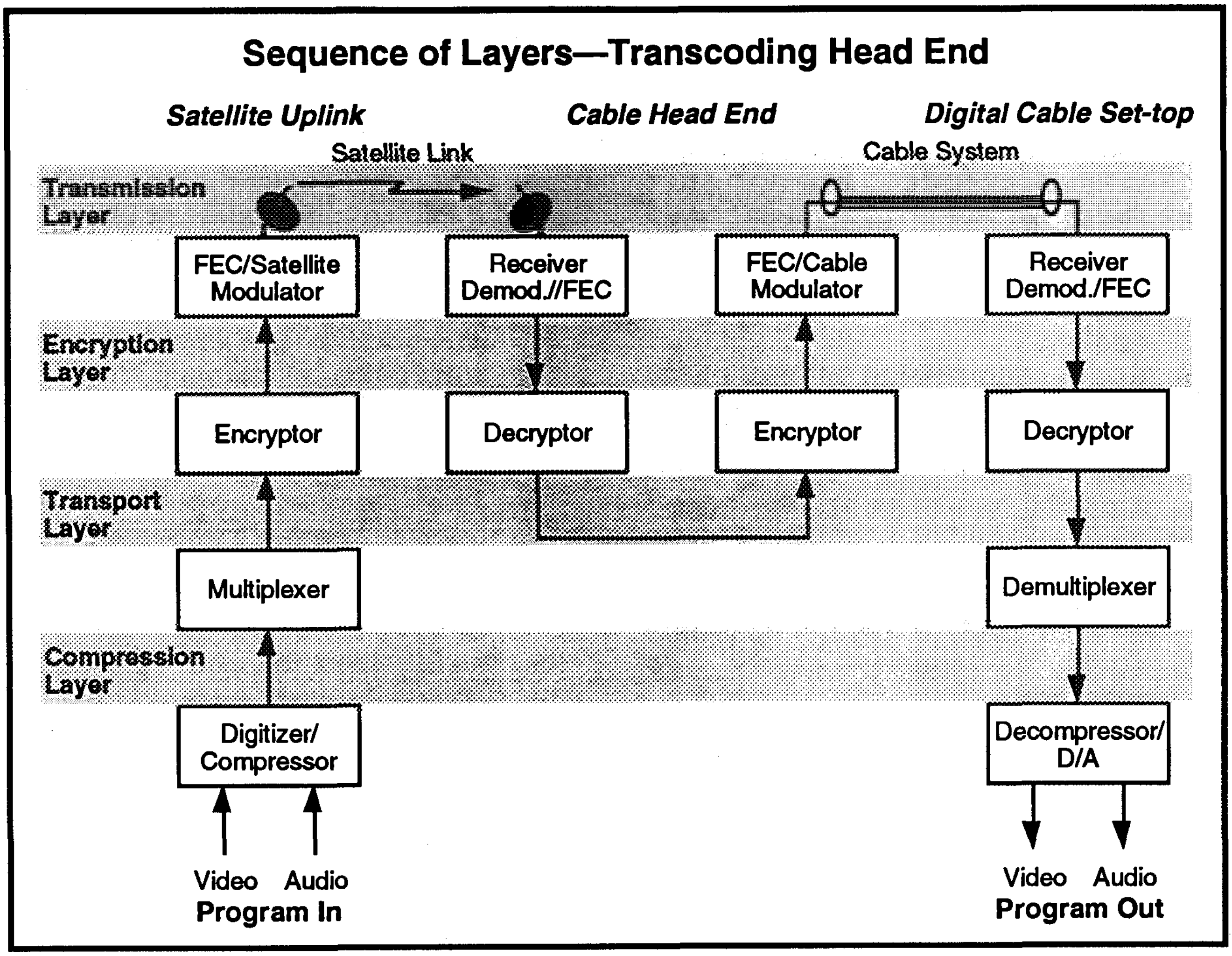


Figure 2

An illustration of the use of the layer concept in describing an end-to-end system is provided in Figure 2, which shows successive transmission through a satellite link and through a cable system. In this example, transcoding equipment at a cable head end converts from satellite modulation (e.g. QPSK) to cable modulation (QAM or VSB), and also provides for decryption and re-encryption. Note that the compression layer and transport layer as applied to the program material are unchanged.

Interoperability and standardization issues may be quite different as applied to the different layers and interfaces.

INTEROPERABILITY ISSUES

Compression Encoding

In order to assure interoperability, MPEG-2 Main Profile Main Level compression compliance is essential. The timing of MPEG-2 is no longer an issue. Cable and satellite delivery systems will, during 1994, have access to 100% MPEG-2 compliant systems. The ISO MPEG committees have now frozen all essential parameters of the world standard necessary for compression encoding, and chip vendors are nearing release of MPEG-2 compliant silicon components. The benefits of agreement on digital standards are now available and are being adopted throughout the world. Full compliance with MPEG-2 Main Level Main Profile requirements means that advantage can be taken of the maximum capacity and quality assured by MPEG-2 without the need for costly upgrades later. This is essential for cable's leading role in the national communications structure, and to compete as a global player. There is no longer any room or rationale for half measures or proprietary compression solutions.

Transport Multiplexing

Compliance with MPEG-2 is not sufficient on its own to assure seamless exchange of programs between media and interoperability of equipment supplied by different vendors. MPEG-2 can be thought of as a toolbox of rules from which selections can be made for specific purposes—particularly in the transport multiplex area. Vendor cooperation and standardization of certain of these parameters—for example, those specifying sources and destinations of programs, time synchronization of program contents, and the "hooks" for attaching encryption—are necessary to assure seamless operation.

An openly specified and complete syntax, within the MPEG-2 specification, is necessary for vendors to build equipment designed to reprocess and/or receive digitally compressed programs. The transport stream defined in the MPEG-2 standard provides coding syntax which is necessary and sufficient to synchronize the decoding and presentation of video information while ensuring that coded data buffers in the decoders do not overflow or underflow. The information is coded in the syntax using time stamps concerning the decoding and presentation of coded audio and visual data, and time stamps concerning the delivery of the data stream itself. However, the MPEG-2 rules allow the use of a number of user-selectable bits and fields. It is in this area of user-selectable data that vendor coordination is necessary—on an open, non-proprietary basis—in order to guarantee interoperability.

Encryption and Conditional Access

Encryption, security, and conditional access are not specified directly by MPEG-2, other than allowing data space with the transport for security-related and conditional access data. To achieve broad agreement on a single, overall security methodology would be both difficult, and potentially disastrous, in terms of vulnerability to piracy. Nevertheless, certain aspects of the security are candidates for standardization, including the method used to scramble the digital bitstream, and the "hooks" for retrieval of control data.

The methods used to secure satellite and cable transmissions can be decoupled (see Figure 2 for an example of decrypting the satellite signal and re-encrypting for cable distribution). The goal of a headend-in-the-sky (HITS) can still be achieved by distributing the cable addressable control data stream as a data service via satellite.

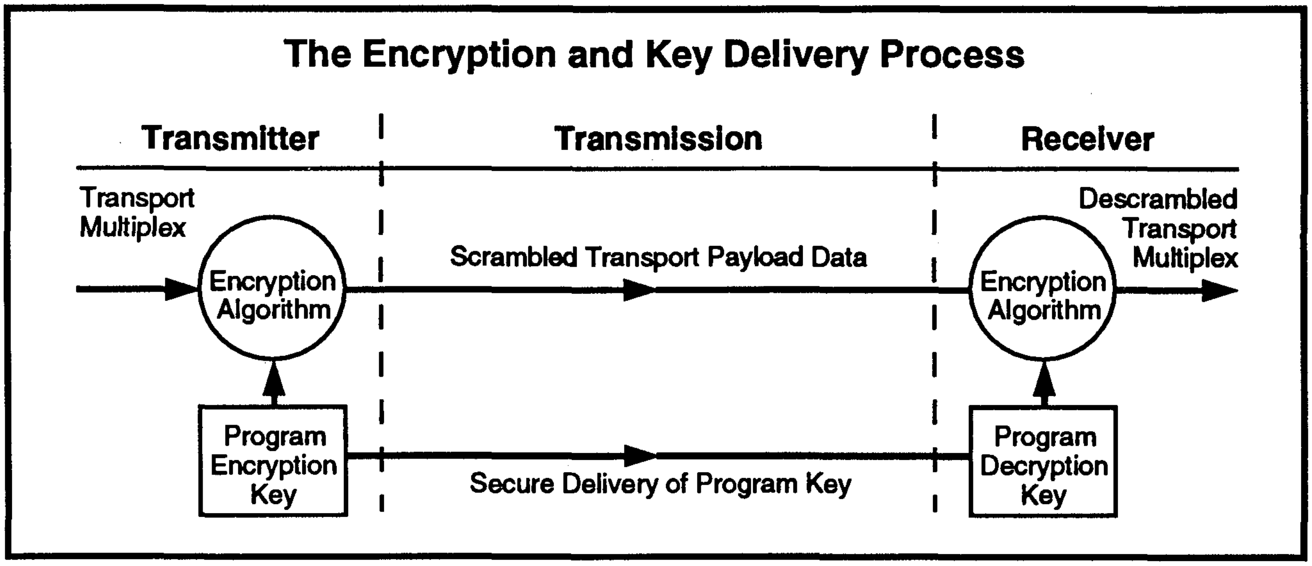


Figure 3

As can be seen from Figure 3, the process required to scramble the transport payload and the delivery of decryption keys can be considered separate. In fact, with a common scrambling method, more than one key delivery system may be employed (see Reference). Selective distribution of decryption keys is the preferred method of controlling access to programs in a digitally encrypted system. Thus key distribution can be thought of as an essential component of conditional access. Because of variances in market requirements and the potential benefit of added security, it may be desirable to allow multiple conditional access systems to coexist and operate in the same digital compression system.

The method of injecting and retrieving conditional access streams into and from the MPEG-2 transport is provided for, but not specified within MPEG-2. Two conditional access-related message types are anticipated:

- **Entitlement Management Message (EMM)**—which is subscriber specific (can also be global).
- **Entitlement Control Message (ECM)**—which is program specific.

The method for mapping EMM and ECM streams to a common program in the MPEG-2 tables must be defined, as well as agreement on the use of private data tables.

There are advantages to agreement on bitstream scrambling and on placement of EMM and ECMs. The European Digital Video Broadcast Group (DVB) has already made considerable progress in this area—its activity is a useful model for action here.

Transmission—Satellite and Cable Modulation

Transmission of high speed digital data via satellite or cable requires more than just the selection of an appropriate RF modulation method. Optimized systems apply advanced forward error correction (FEC) coding to transmissions in order that receivers can identify and correct errors caused by noise or distractions. Agreement of the transmission segment necessarily involves discussion of FEC as well.

The satellite case should appear relatively straightforward. There is broad agreement that QPSK is the satellite modulation method of choice. However, transponder bandwidths are not all the same and TDM data rates may well vary. Fortunately technology now exists to permit a receiver to work adaptively with a very wide range of received signal bandwidths and an equally wide range of data rates (e.g. 1 - 90 Megabits per second). Still, some agreement on boundaries would be helpful.

In the case of cable modulation, the case is not quite so clear. In making a selection of a modulation method for terrestrial broadcasting of Advanced Television (HDTV), the Grand Alliance and the FCC's Advisory Committee on Advanced Television Service (ACATS) selected 8 VSB based on testing of VSB and QAM techniques. These organizations also made a selection for higher data rate ATV service on cable of 16 VSB. The 16 VSB modulation is equally applicable to multi-channel digitally compressed programming. A 6 MHz bandwidth cable channel using 16 VSB modulation could be used to transmit up to 38.5 Mbps of data (useful payload).

There is also a push for the use of 64 QAM and higher order modulation such as 256 QAM for multi-channel cable transmission. A 6 MHz bandwidth cable channel using 64 QAM modulation could be used to transmit up to 28.8 Mbps of data (useful payload); approximately 35 Mbps for 256 QAM.

It may be thought that the selection of a modulation method can be dealt with on an individual cable system basis. But what happens to the ultimate possibility of putting the digital tuner and demodulator inside the subscriber's TV receiver? To achieve that objective, agreement on modulation (and associated FEC) is still necessary.

How Necessary Is Standardization?

In order to achieve:	*Compression Encoding*	*Transport Multiplex*	*Scrambling/ Scrambling Interface*	*FEC/ Modulation*
Pre-compressed delivery of programs	Essential			
Transmission between different media (cable/satellite, etc.)	Essential	Essential	Helps	
Multiple vendor sourcing of commercial satellite/cable head end equipment	Essential	Essential	Essential	Essential
Multiple vendor sourcing of consumer equipment	Essential	Essential	Essential	Essential

PRIORITIZING STANDARDIZATION

Can standards be prioritized? The need for standards is certainly most compelling at the compression encoding and transport multiplex levels. How essential is standardization to interoperability? The previous table indicates the relationship between standardization of each of the compression layers and the various needs for program exchange and the notion of multiple vendor sourcing of equipment.

CONCLUSION

The interoperability goals of seamless program delivery and multi-vendor equipment supply are achievable for the cable industry's next generation of digitally compressed video hardware. The world's leading electronics companies—including major U.S. cable equipment vendors—have made it possible for all the world's cable operators and satellite users to "drive on the same side of the road" on the information highway. Only 100% MPEG-2 compliance coupled with agreement on specific transport, transmission, and scrambling parameters will allow this to happen. Industry-wide pressure and cooperation is required to obtain the best possible technical and economic benefits in the adoption of digital video compression.

- Universal provision for MPEG-2 Main Profile Main Level.
- Openly defined and agreed upon Transport Parameters.
- Coordination of scrambling of the digital bitstream and the associated transport hooks.
- Agreement on cable modulation with the best multi-channel and HDTV transmission capacity.

REFERENCE

"Methodologies for Multiple Conditional Access Technologies in Digital delivery Systems." Tony Wechselberger and George Parkanyi. Paper presented at Canadian Cable Television Association Convention, May 1994.

About the Authors:

Mr. Stubbs is the president of Graham Stubbs Associates, a consulting firm specializing in cable television technology. He previously served in engineering and general management executive positions with Eidak Corporation, Oak Communications (now TV/COM International), and Jerrold, developing addressable technologies for the cable and satellite industries. He holds a B.Sc. in Physics and Electronics. During the past twenty years, he has authored numerous technical papers in the cable field, and holds several key patents in conditional access and signal security.

Mr. Wechselberger has held various positions with TV/COM International (formerly Oak Communications) since 1980. His staff supports all of TV/COM's cable, satellite, and compression engineering initiatives and he is responsible for a number of key patents. Before joining TV/COM, Mr. Wechselberger was with the General Dynamics Electronics Division. He holds B.S. and M.S. degrees in Electrical Engineering and has published and lectured extensively in the fields of TV signal security, and general conditional access requirements involving communications, and encryption.

Digital Video Servers: Storage Technology and Applications

Richard F. Annibaldi
Pioneer New Media Technologies, Inc.

Abstract

How can cable operators help prepare themselves for the introduction of compressed digital video into their cable systems? This paper describes the technologies available for one important area, mass storage used in digital video servers. With this information, operators can better evaluate their alternatives in relation to their planned applications.

INTRODUCTION

As cable system operators strive to implement the National Information Infrastructure in the coming years, they will face many new challenges. In particular, the use of digital compressed video requires that cable systems implement unfamiliar technologies. Although some compressed video sources will probably be provided by satellite, the opportunity exists to provide other sources either directly in the cable system, or within a multiple system interconnect. These locally provided sources of digital video will take the form of a digital video server.

The basic function of such a server is to provide appropriate streams of digital information to the subscriber of the cable system. Depending upon the specific services that the operator wishes to offer, each of these streams may be interactively controlled by a subscriber.

VIDEO SERVER ARCHITECTURE

Let's begin by examining a typical structure of a server. Although the exact form used may vary for each cable system, digital video servers will likely contain several common building blocks. The three most basic blocks are a central processing unit (CPU), an input/output (I/O) system, and digital storage.

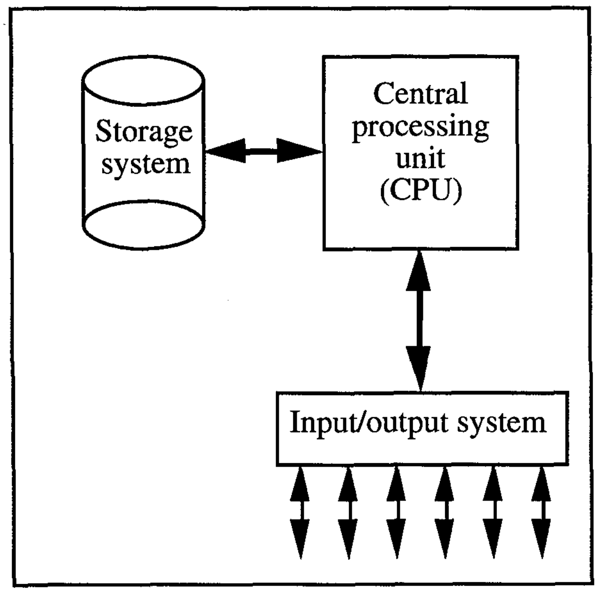


Figure 1: Basic Server Architecture

Our interest is in this third block used to hold large amounts of digital information. To allow for rapid movement of data into and out of the storage system, most devices use the Small Computer System Interface (SCSI). SCSI, pronounced "scuzzi", and the improved version, called SCSI-2, allow transfer of data at up to 20 mega-bytes per second (MBps), equivalent to 160 mega-bits per second (Mbps). Although this is substantially higher than the bit rates needed to support full motion video, remember that this interface must supply enough data to support multiple subscribers simultaneously.

DIGITAL VIDEO FORMATS

To evaluate storage systems, it is helpful to review the variety of digital video formats in use today. The storage system should be flexible enough to handle any combination of the video formats that will likely appear on the network.

Formats to be considered include:

MPEG1	Compressed full motion video at about 1.5 Mbps.
MPEG2	Compressed full motion video using several different profiles and levels at up to 60 Mbps - typically uses 3 to 4 Mbps for NTSC and 10 Mbps for HDTV
H.261	Video telephony and conferencing at bit rates up to 1.92 Mbps
JPEG	Compressed still pictures
Others	Proprietary standards including General Instruments' DigiCipher

Each of these formats may find a specific niche application on cable. For example, while movies would require an MPEG or equivalent format, interactive home shopping might use JPEG with freeze frames. Each format was developed based upon the desired compression algorithms, resolution and data rate. As a result, the actual data format stored on the media differs.

In fact, there is even more variety. With three profiles available at several levels each, MPEG-2 alone has many variations. The data may need to match a particular transmission format such as Asynchronous Transfer Mode (ATM). Although the video server processor can convert formats, storing the video in the final transmission format can increase throughput. The storage technology should be flexible enough to handle all these possibilities.

As a rule of thumb for later comparisons, consider the typical amount of data needed for a two-hour movie. With a 3 Mbps data rate, this movie requires 2.7 gigabytes (GB) of storage. Higher resolution, such as HDTV, would need much larger storage.

STORAGE TECHNOLOGIES

We can classify storage technologies into four basic groups: random access memory (RAM), magnetic tape, hard disk drives, and optical disc drives. Table 1 summarizes characteristics of each type described below. Because these types often complement each other, it is most likely that servers will use some combination of them.

Random Access Memory

The oldest, fastest and most flexible method of storage is Random Access Memory (RAM). Unfortunately, RAM is also the most expensive for large storage size. Nevertheless, its advantages make it desirable to use in conjunction with the other types of storage. After loading a RAM buffer from any other type of storage, the processor can quickly access bursts of data destined for multiple users.

Magnetic Tape

There are several digital video recording formats, such as D1 and D2, in use today. However, these machines record data in rigid formats and provide digital outputs primarily for duplication and editing. For more flexible data formats, magnetic tapes for computer applications are a better choice. Virtually all of these computer tape formats can store digital video. However, these tape formats share several disadvantages in video server applications. The two most important of these are long access time when positioning the tape to a random point, and the reliability effects of tape wear from repeated playback.

The availability of both hard disc and optical storage makes extensive use of magnetic tape in server applications less desirable. Nevertheless, magnetic tape can still be valuable for creating inexpensive archive copies of digital material that is no longer in regular use.

Table 1: Comparison of Storage Technologies

STORAGE TECHNOLOGY:	RAM	Magnetic Tape - Computer formats	Hard Disk	CD-ROM	DLD - Replicated	DLD - WORM
CHARACTERISTICS:						
Write capability	√	√	√			√
Erase capability	√	√	√			
Removable media for archive		√		√	√	√
Simultaneous outputs	1	1	1	1	4	4
Estimated total equipment cost per mega-byte per output (includes complete electronics, drive and all removable media, if any)	$25-$40	$20 on cartridge More for faster formats	$1-$2 on SLED $3-$16 on RAID	$50 for 1 to 6 disc changer Less for larger changers	< $1 on single disc drive <$0.10 on changer	< $1 on single disc drive
Average access time to any data location	< 1 µsec (processor access time)	50 secs for cartridge (tens of seconds for faster formats)	15.6 msec	< 500 ms within single disc	2 seconds within single disc	2 seconds within single disc
Sustained data rate for read operation in mega-bits per second (Mbps)	Limited only by processor speed	0.5 Mbps cartridge; up to 120 Mbps for others	16 to 40 Mbps; higher for RAID	1.2 Mbps	Up to 15 Mbps	Up to 15 Mbps
Typical storage increment in giga-bytes (GB)	Variable	0.13 GB per cartridge	Up to 1.6 GB per drive	0.54 GB per 4.75 inch disc	5.4 GB per 12 inch disc	5.4 GB per 12 inch disc
ADVANTAGES:	•Fastest	•Inexpensive media •Good for creating archive copies	•RAID improves access time and data rate	•Well suited for multiple copies	•Largest capacity •Well suited for multiple copies	•Largest capacity
DISADVANTAGES:	•Most expensive	•Very slow access time for random locations on tape	•Need RAID or backups for data protection	•Set-up cost for single disc copy	•Set-up cost for single disc copy	•Moderate cost for single disc copy

Hard Drives

This is now the dominant technology in the computer industry for large capacity storage, especially on personal computers. Hard drives combine the ability to read and write data at relatively high speed with non-volatility.

There are two general types of hard drive systems: the traditional Single Large Expensive Disk (SLED) and the newer Redundant Array of Inexpensive Disks (RAID). A SLED can now have capacities of 1.6 GB of data, with larger drives planned.

Groups of drives are combined with appropriate control software to form RAIDs. The different types of RAID configurations are classified as six different levels numbered 0 to 5. These "levels" do not indicate their relative merits; they simply identify different configurations that have different advantages and disadvantages. Table 2 summarizes the different levels. The characteristics of the various levels are improved speed and, more importantly, error detection and correction.

Table 2: Redundant Arrays of Inexpensive Disks (RAID)

RAID LEVEL:	RAID 0	RAID 1	RAID 2	RAID 3	RAID 4	RAID 5
CHARACTERISTICS:						
Data on original disk duplicated or mirrored on second disk		√				
Data striped across multiple disks using one byte per drive accessed	√			√		√
Data striped across multiple disks using full sectors on each drive accessed					√	
Error detection and correction codes stored on a separate check disk			√	√	√	
Parity interleaved with data and striped across several disks						√
ADVANTAGES:	•Increased speed	•Full redundancy of data	•Large data block efficiency	•Increased speed •Large data block efficiency	•Increased efficiency for small data blocks	•Allows multiple simultaneous writes
DISADVANTAGES:	•No error detection / correction	•Only 50% of disk capacity usable	•Unnecessarily redundant error detection / correction	•High overhead working with small amounts of data	•Slow writing of data due to shared check disk	•Most complex controller required

RAID redundancy for error detection and correction permits continued operation even after one drive fails. In addition, with many RAID controllers, an operator can actually replace a single failed drive without shutting down the system.

Besides the obvious approach of time sharing by multiple users, the operator also can configure some RAID levels to support simultaneous accesses to different drives. This would require careful handling of contention among users for the same individual drive. The specific advantages and disadvantages must be weighed against the desired applications to determine the best fit.

Optical Disc Drives

Compact Disc Read Only Memories (CD-ROM) and Digital Laser Discs (DLD) are the two primary types of optical disc technology. While CD-ROMs have been in use for many years, DLD is still under development. Details on DLD given in Table 1 are target specifications.

CD-ROMs: As the name states, CD-ROMs cannot be erased or rewritten. Like existing analog laser discs, CD-ROMs are replicated using a stamping process. Each 12 cm (approximately 4.75 inch) diameter CD-ROM can contain up to 540 mega-bytes of data. Drives are available for either a single CD-ROM or a magazine of six CD-ROMs. Changers for much larger quantities are also anticipated in the future.

The 540 mega-byte capacity of the CD-ROM is much less than the 2.7 GB required by our rule of thumb movie. Also, the CD-ROM data rate is only 1.2 Mbps. For these reasons, vendors are introducing several variations on CD-ROMs specifically targeted towards the video storage market. Some are spinning the disc faster to achieve two, three or four times the standard data rate. Others are developing products that differ more significantly.

As an example, Pioneer's α (alpha) Vision System uses high density recording and replicating techniques to place 2.12 GB on the same CD-ROM sized disc. The data transfer rate for α Vision is 4.7 Mbps, nearly four times the usual CD-ROM rate. The result is a disc which can store sixty minutes of video equivalent in quality to analog laser discs. In addition to the video, this disc stores two full stereo, or four monaural audio tracks. Other data can be included at a transfer rate of 130 kbps.

DLDs: A 30 cm (approximately 12 inch) diameter Digital Laser Disc can contain up to 5.4 GB of data, ten times that of a CD-ROM. DLD will support a variety of video formats including MPEG-2. A SCSI-2 interface carries the 15 Mbps data transfer rate.

A stamping process produces normal analog laserdiscs, called replicated discs. Currently planned playback drives for DLD will handle both replicated and Write Once Read Many (WORM) formats. WORM allows for digital storage of cable system specific content, while retaining all of the playback advantages of replicated laserdiscs. These advantages include non-contact reading to eliminate media wear, quick access to any location on the disc, and the ability to place seldom-used discs directly into archive storage.

Traditional analog laserdiscs contain video in either of two formats: constant linear velocity (CLV) and constant angular velocity (CAV). CLV allows for twice the storage capacity of CAV by varying the rotation speed depending upon read location.

By using the CAV format, however, DLD can use more than one pickup at a time. For example, locating a pickup at each of the four major compass points around the disc allows four simultaneous, yet independent, outputs. Figure 2 illustrates this four pickup drive, showing the four pick-ups or heads, each reading a different part of the data on

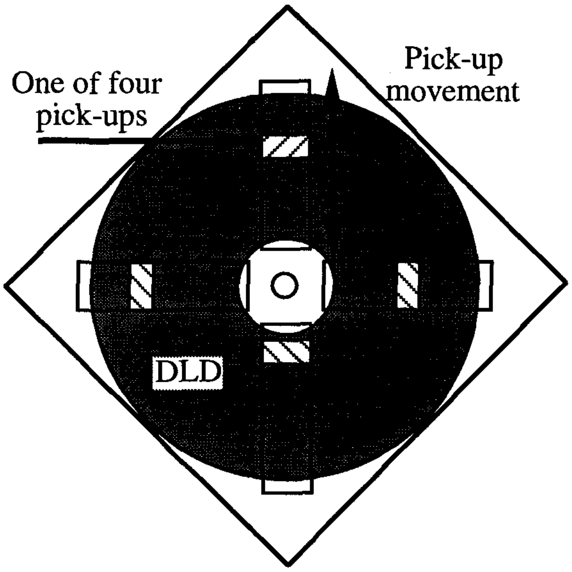


Figure 2: Four Pick-up DLD Player

the disc. Each pick-up moves across the spinning disc to the desired start position, from which it tracks the spirally recorded data.

In addition, there are also plans for automatic disc changers with two independent players and up to 252 DLDs. These changers allow easy access to huge libraries of information.

A variety of support equipment is also under development for DLD. First, a four channel MPEG2 video decoder allows use of the DLD drive as an analog video source. This makes for an easy introduction of the DLD into today's cable system. Once digital transmission begins, the operator simply removes this decoder from the headend.

Second, a four channel MPEG2 data synchronizer/multiplexer provides the steady data stream needed for reliable, flicker free video decompression at the set top decoder. Next, a combined encoding and authoring station handles preparation of both real time and non-real time data for recording. Finally, the WORM recording system writes the data onto the twelve inch disc.

CABLE SERVICES

Now that we understand the technologies available, we can apply this information to the cable system's video server. What services does the cable operator plan to offer? Let's look at each of these and see which technology fits best.

Near Video on Demand (NVOD) consists of providing the same movie on multiple channels with staggered start times. For example, one two-hour long movie could start at 8:00 PM on the first channel, 8:30 PM on the second, 9:00 PM on the third and 9:30 PM on the last. NVOD is ideal for first run movies because it can handle very large numbers of purchases using only a few channels. From our list of technologies, the four pickup DLD stands out as the best match for NVOD.

Video on Demand (VOD) provides individual access to a program. For example, each subscriber can purchase a particular movie, starting and pausing that movie whenever they wish. VOD requires rapid response to subscriber requests. Also, the number of programming choices to be offered to the subscriber has a large impact on the selection of technology.

Hard drives are a possible approach to VOD, although relatively expensive when offering many program choices. A DLD changer with as many as 252 discs can provide maximum choice, containing over 1300 giga-bytes of data. Because the DLD changer contains two players with four heads each, it can handle multiple accesses to this library. In addition, data retrieved from the DLD changer could be buffered in RAM while a different disc is loaded into one of the players for use by other subscribers.

High Definition Television (HDTV) can easily replace standard compressed video on the cable system. Both NVOD and VOD can work with these HDTV signals. The only requirements for the server are the higher bit rate and the specific digital format used for HDTV.

Interactivity can be very different from the full length feature films mentioned above. Films are stored sequentially in memory and require only occasional VCR-like controls to interrupt the normal flow. On the other hand, interactivity can involve random access to very short sequences, such as still frames of video or even short blocks of text. When interactivity requires many random accesses, hard drives are the logical choice. If video sequences are lengthy, however, DLD may have some advantages. In either case, the best approach is to store subscriber inputs in RAM to avoid disturbing the source data on hard disk or DLD.

EVALUATION STEPS

The following steps can help to determine the storage needs for a particular cable system. Simply answer each of these questions for your system:

1) Which services will you carry in the short term? What about the long term plans? Use these answers to match appropriate storage technology to each planned service. Most likely, you will want to phase in the new services gradually.

2) What is the expected popularity of each service? Use this answer to help identify the relative quantities of each storage type needed.

3) Which technologies can support the immediate applications without becoming obsolete when new services are later introduced? This will ensure that your system will grow and adapt to your needs.

Provide these answers to potential video server vendors so they can propose suitable solutions to your requirements. With your guidance and feedback, they can provide a much better system for you.

CONCLUSIONS

To be ready to face the challenges of the digital age, cable system engineers must gain an understanding of many applicable computer technologies. One such technology is the storage of large quantities of digital video and other information. The ideal storage technologies for any cable system will depend upon the services offered. Knowing your applications and the capabilities of various technologies should help in equipment selection.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of Koh Uchiyama of Pioneer Electronic Corporation in providing source materials for this paper. In addition, thanks to many individuals at Pioneer New Media Technologies for helping to review this work.

REFERENCES

(1) Alford, Roger C. "Disk arrays explained", *BYTE*, October, 1992, pp. 259-266.

(2) Casey, M. "Disk arrays: Hype vs. reality", *Computerworld*, August 19, 1991, pp. 54-55.

(3) Ranade, Sanjay. *Jukebox and Robotic Libraries for Computer Mass Storage.* Westport, CT: Meckler Corporation, 1992

(4) Stone, M. David. "Hefty Storage on a Budget", *PC Magazine*, October 12, 1993, pp. 247-264.

EDUCATION ON DEMAND

Jerry D. Neal
Pioneer New Media Technologies, Inc.

ABSTRACT

How can cable operators acquire higher visibility and first hand community involvement leading to a more positive image for a cable system? The answer is detailed in the following paper which describes the technology involved in providing an "Education On Demand" service to local school systems in a cable franchise area.

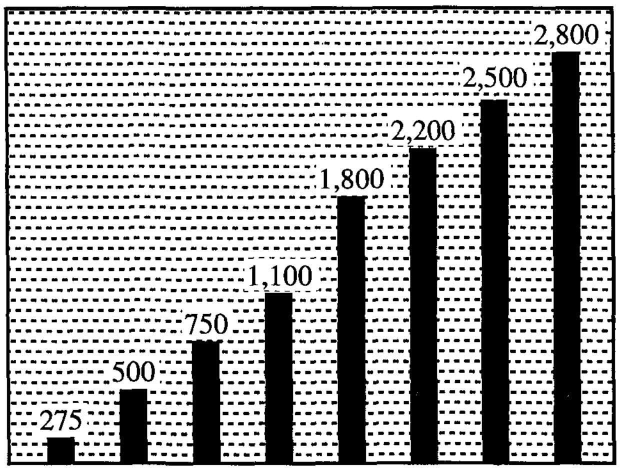
INTRODUCTION

Education in the classroom has changed dramatically over the years and continues to evolve today. Interactive learning systems, which have been available for more than ten years, are gaining wide-spread acceptance within the education community. Schools throughout the country that are using videodiscs for interactive educational training is the main reason this change has taken place. The advantages of videodiscs compared to traditional methods of teaching are overwhelming. The only disadvantage is the cost of the equipment per classroom.

The idea of Education On Demand overcomes this disadvantage by removing the individual equipment from each classroom and installing the equipment in a remote location at a cable system's headend or at a central location at the school. By installing the equipment remotely, the amount of equipment can be decreased and the equipment can be shared by every classroom. Specific teaching material and use of the equipment is then scheduled in advance. Not only can a single school have the use of the equipment in a local cable system, but an entire school system can schedule use of the interactive system.

EDUCATIONAL VIDEODISCS

The videodisc or laserdisc (the terminology is interchangeable) was first used in classrooms in the early 1980s. It has proven to be such a powerful method of teaching and learning that it has grown in popularity ever since. To support this popularity, production of videodiscs for educational use has increased from 275 different videodiscs in 1987 to approximately 2,800 titles produced by more than 275 different companies today. In a relatively short period of time, the number of available educational videodiscs has skyrocketed as depicted in Figure #1.



1987 1988 1989 1990 1991 1992 1993 1994

Figure #1 - Number of Educational Videodisc Titles

Educational videodiscs cover subjects including science and health (which are the categories that contain the most titles), geography, history, art, drivers training, mathematics, music, foreign languages, and even physical education. In addition, many videodiscs contain Spanish/English or French/English audio which is very useful in bilingual classes.

Videodiscs have many advantages over the traditional films or videotapes which are used in classrooms. For example, when a film or videotape is shown, the classroom is typically darkened and the instructor sits in the back. Using a videodisc player connected to a monitor or television, the room is not darkened and the instructor is usually in front of the classroom controlling the learning environment. Fast and random access to any specific section of a videodisc is available along with freeze frame, step motion, and slow motion; whereas, films and videotapes require time consuming rewinding and fast forwarding. Another advantage is that videodiscs are virtually indestructible and can be shown indefinitely. How many teachers have been frustrated by broken filmstrips while showing a film or videotape?

INTERACTIVE SYSTEMS

A typical interactive instructional system includes a videodisc player, a monitor or television, and either a remote control device, a barcode reader, or a computer to control the player.

The handheld remote control is probably the most cumbersome method of player control, because the instructor must enter the frame number of an image or clip to be shown.

Using a computer to control a videodisc player is presently being used in some classrooms, which is the most versatile, but costly.

Thus, the most popular interactive instructional system in use today controls the player using a barcode reader. A barcode reader simplifies the access method by reading barcode labels of the section to be played and sends the command to the player automatically. Also the instructor can produce their own barcode labels, which allows greater flexibility for personally designed lesson plans. An example of this type of interactive system is shown in Figure #2.

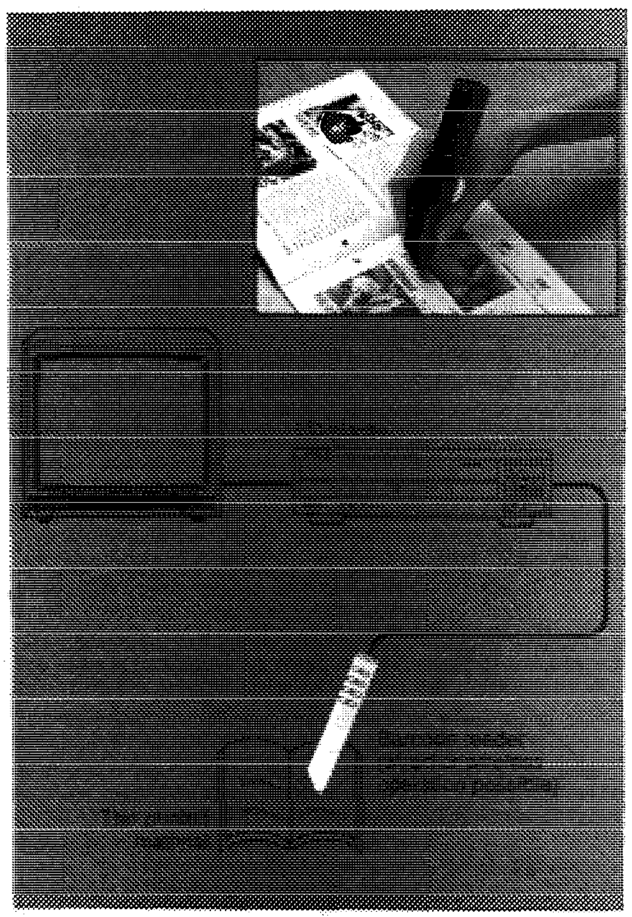


Figure #2 - Typical Educational Interactive System

Because videodisc players are so easy to use, the student instead of the instructor can become the operator. It is a proven fact that most people are visual learners. They grasp concepts and comprehend new ideas more readily if they are presented with something they can see and interact with. To support this claim, some textbook publishers are beginning to make use of the videodisc technology. Many publishers are inserting barcodes into their textbooks, and even including the videodiscs as part of the printed material.

If interactive education is the best learning tool available, why isn't it used in every classroom in every school? The answer to this question is cost. Federal and state funding, private corporate and individual donors along with the ever growing school district bond issues, cannot provide enough money to accomplish this goal.

EDUCATION ON DEMAND THE SOLUTION TO THE COST

To remove the cost of installing equipment in every classroom, the idea of Education On Demand (EOD) was born. EOD installs equipment for an entire school, or even a complete school system, in a single location at the school or at a cable system's headend. How will such a system work? The answer begins with cable. As of today, over

60,000 schools (61% of the 99,432 total schools in the U.S.) are provided with free cable services by their local cable companies. Four years ago a program called "Cable in the Classroom" was started and as Figure #3 shows, the number of schools connected to cable has increased an incredible ten times.

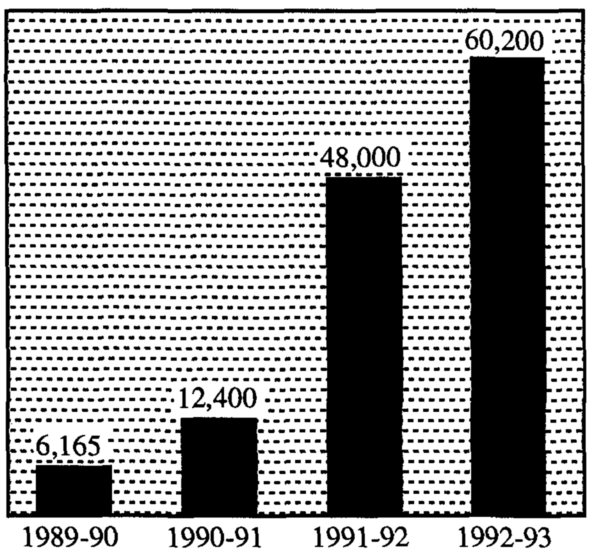


Figure #3 - Cable in the Classroom

With cable already in the classroom, the major obstacle of EOD is solved. The interactive classroom in the EOD system contains only a barcode reader and an IR transceiver. The barcode reader can be directly connected to the IR transceiver or the instructor can point the reader at the transceiver and send the barcode command. The IR transceiver then translates the barcode command and transmits the command over the cable system to the videodisc players in the cable system headend. The videodisc player responds to the command received and the classroom watches the result on the television.

Figure #4 depicts a very basic EOD system configuration. The 5 to 30 MHz path is the path on which the IR transceiver transmits the barcode commands. The type of path that is shown would be feasible only in a two - way cable system or if the equipment is located at the school.

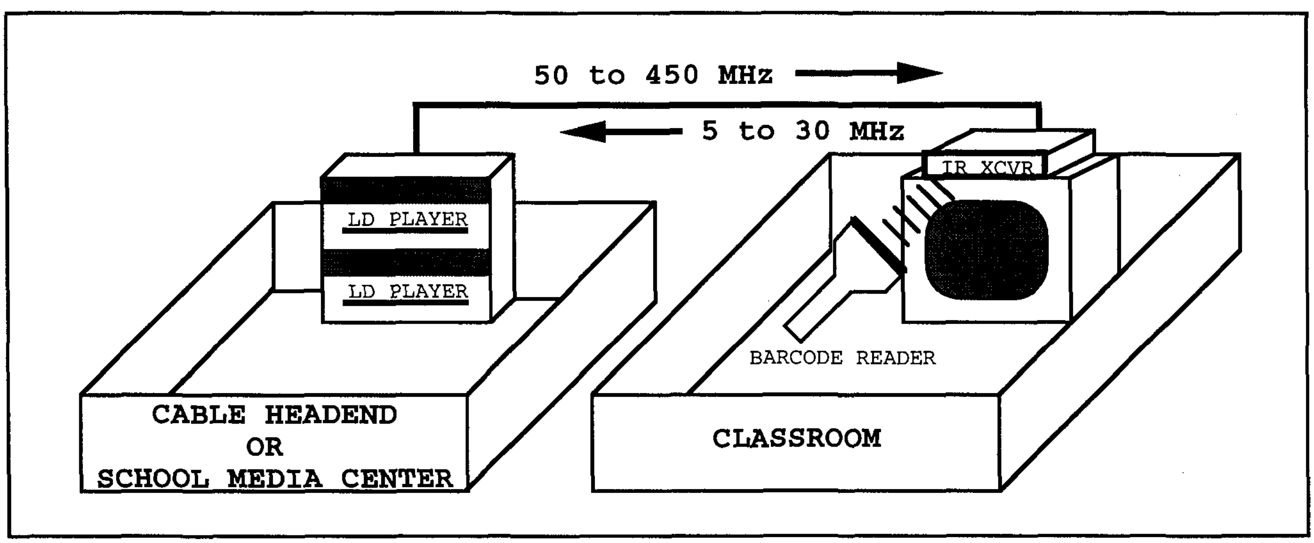


Figure #4 - Basic Education On Demand System

This return path would be accomplished using RF modems. For cable systems that are one-way, the return path can be via telephone modems. For the entire school only one modem or return path would be necessary, although multiple modems can be added for larger EOD systems.

The IR transceiver interface is a standard RS-232C type of interface; thus, common communication equipment such as A/B switches, modems multiplexors, and line sharing devices can be used to construct the return path

Figure #5 is an EOD system diagram that shows multiple classrooms connected to multiple players. To accomplish this an A/B switch is added to each classroom and then connected to a multiplexor (MUX). The A/B switch selects which LD player the barcode commands will control. The multiplexor is used so that multiple return paths are not needed.

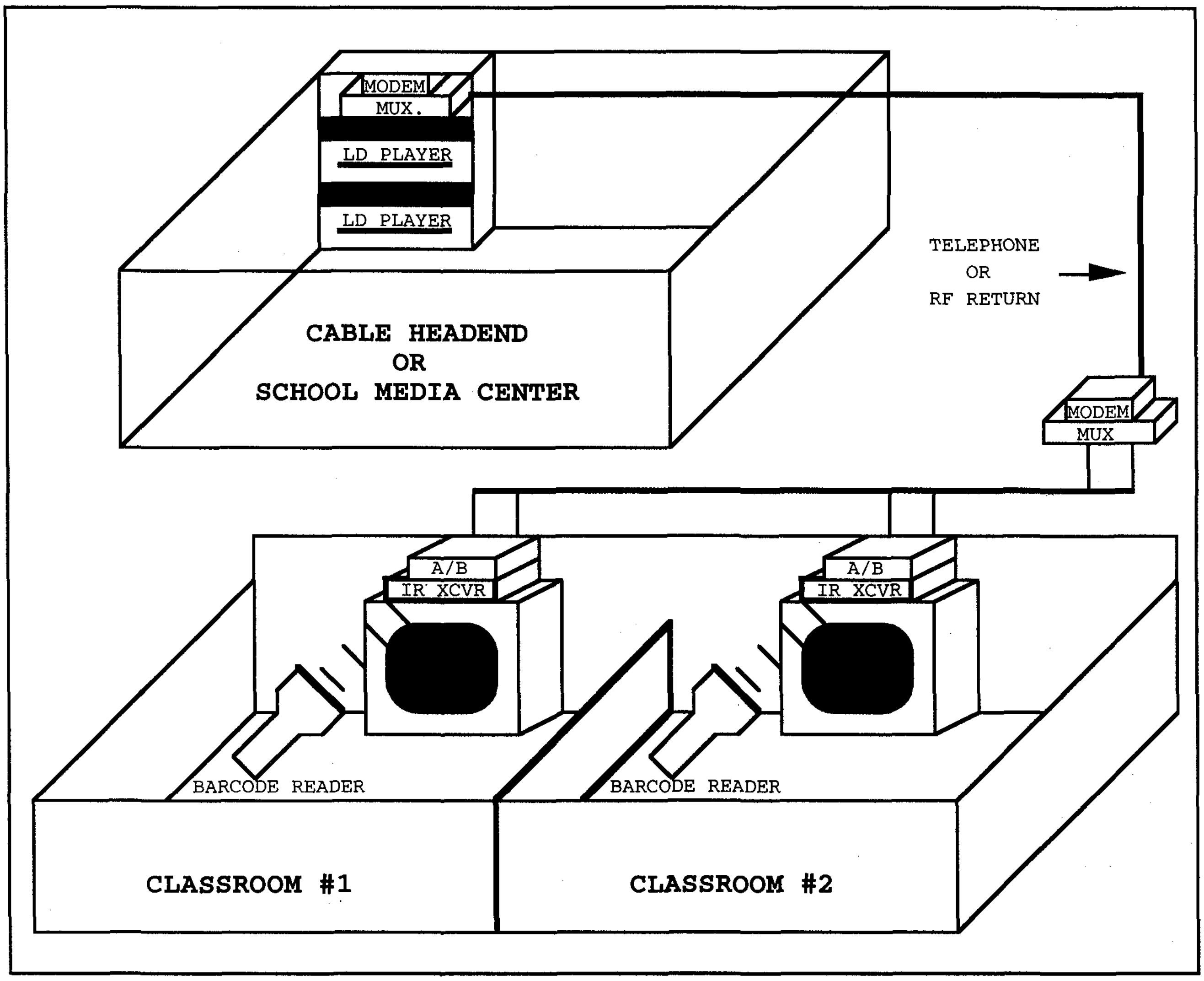


Figure #5 - Multiple Classroom EOD System

Figure #6 is an EOD system diagram that shows multiple classrooms connected to multiple players from multiple schools.

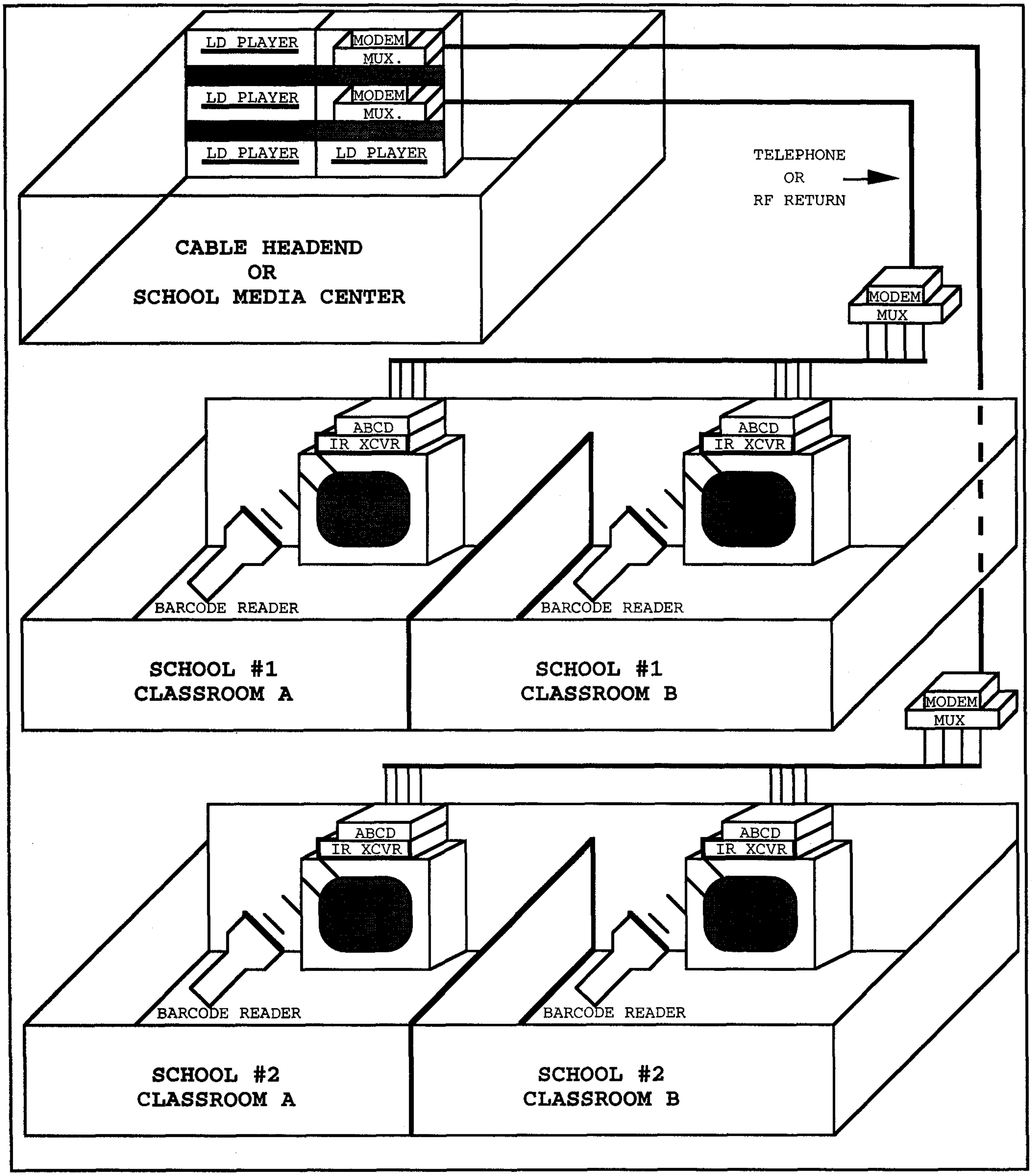


Figure #6 - Multiple School EOD System

With multiple schools the return path for each school must be separate. For a two-way cable system, this is done by offsetting each schools return path RF frequencies. For a one-way system, multiple phone lines must be used. Again, multiplexors and ABCD switches are necessary to allow access to any of the LD players from any of the classrooms. ABCD switches are used because four LD players can be selected for control.

Configurations for EOD are unlimited along with the number of videodisc players that can be controlled. Using multiplexors, ABCD switches, or line sharing devices allows multiple videodisc players from each classroom or from each school system to be controlled.

The only problem encountered with numerous players or multiple school access, is the scheduling of player usage. This problem is easily solved with the addition of a scheduling computer that automatically handles player usage. This scheduling computer can also contain a database listing any available educational videodisc in the system.

Not only can single videodisc players be connected to the cable network, but autochangers that can contain up to 72 different videodiscs can be connected. This piece of equipment truly makes EOD very versatile when configuring a system.

SUMMARY

Interactive learning systems are gaining wide-spread acceptance within the classroom. The advantages of reduced learning time, increased retention, flexibility in teaching style, ease of use, and probably the most important, the enjoyment of interactive learning, have led the way to further advancements in this type of training.

Education On Demand is an example of this type of advancement. Cable operators that install EOD systems in schools in their franchise areas, will benefit from higher visibility and an overall greater community awareness. Instructors will have a wide variety of current user definable information available that will allow flexibility in their teaching style. And most important of all, the student will gain the most by decreasing learning time and promoting greater information retention while having fun.

REFERENCES

Curriculum Product News, Vol. 27, No. 9, April, 1993, pp. 4 - 6.

Multimedia and Videodisc Compendium, Emerging Technology Consultants, Inc., 1994 edition.

Cable in the Classroom Members / Quality Education Data, Nielsen Research Data, March 1993.

Electronic Program Guide Applications
The Basics of System Design

William L. Thomas
TV Guide On Screen

Abstract

Some of the earliest features promised as part of the information highway will be embodied in Electronic Program Guide applications. These applications will provide an integral element of the navigation systems being developed to support new services which will be offered to consumers. In this paper, the system design requirements for a generic program guide application will be described.

EPG SYSTEM OVERVIEW

Electronic Program Guide (EPG) systems are currently under development for first deployment in cable, satellite and consumer electronics products.

On the surface, these applications seem simple in comparison to much of the sophisticated computer software now available to consumers through personal computers. However, due to the cost constraints inherent in mass produced consumer products, such as the first advanced analog and digital compression based cable converters, there are challenges in designing, implementing and supporting these new EPG products.

Program guides will be the first widespread interactive software applications to be seen directly on the TV screen along with television programming. To perform the necessary functions, the cable based systems must be comprised of at least the following elements:

- Method for gathering TV listings
- Extracting system specific information
- Sending the data to the local cable system
- Coordinating with the system controller
- Transmitting the data to the converter
- The guide application itself

In addition to having the basic elements in place, system technical issues relating to downstream transmission capacity limits must be taken into account. These factors include updating the program listing database and application software downloading

The level of user functionality must be carefully considered, and will be implemented with different characteristics for advanced analog, first generation digital and future advanced digital platforms.

Naturally, the user interface is receiving much attention as the first systems are being developed. To consider what is possible for upcoming EPGs, general user functions such as displays associated with channel changes, and listings by time, channel or category will be described. In addition, favorite channel, program restrictions and messaging features will be introduced. A critical part of the interface is presenting a variety of premium services such as subscription channels, pay per view, or Near Video On Demand (NVOD) information in a compelling fashion. These and other "look and feel" issues will be described and illustrated with example screens from a prototype application.

ELEMENTS OF EPG SYSTEMS

As described previously, there are a number of basic elements which must be present to have a complete EPG system. Each of these must be designed both to be efficient in their own right, and to work well with the other parts of the system. Let's consider each in turn...

Gathering TV Listings

There are only a few organizations that currently gather nationwide TV listings information. They include TV Guide, TV Data, and Tribune. Currently this data is primarily used to support printed guides, daily newspaper listings, and scrolling electronic channel guides used in many cable systems. In today's broadcast and cable environment, collecting listings information requires significant manpower.

Listing information includes broadcast time and channel, and program name. Various lengths of program name must be composed, so that they can be used in different printed or video display situations. For each program there may also be additional information including descriptive copy, list of actors, MPAA ratings, year of production, star rating, language, category (sports, movie, children), closed captioning and stereo broadcast mode.

On a nationwide basis, listing information must be collected from over 1100 VHF and UHF high power broadcast stations. In addition, low power stations must also be collected for cable specific guides. All cable networks, including national PPV must be collected. In the future, national NVOD schedules must also be captured.

System Specific Information

Given a complete set of listings information, including local broadcast, cable networks and PPV a system specific set of information must be made. This may be done before the data is transmitted to a cable system headend, or may be done locally at the cable system from a nationwide data feed including all listings information. In either case, an accurate channel lineup is required as the basis of the selection process.

Local listings information, such as the local community access channel programming should be included. If offered, local PPV schedules including pricing information must also be collected in the database.

Sending Data to Local Cable Company

As mentioned above, the listings information must be sent to the local cable system where the guide application is offered. Typically this will be via a satellite link, although in the early stages of deployment, there may be telephone based methods of data delivery.

In an EPG application program, it is important that listings information be accurate and up to date, as seen by the viewer. Clearly, this is one of the major benefits of an electronic delivery system. Given this requirement, the frequency of update is a critical design factor, as programming changes occur on a frequent basis. Printed publications are typically closed once a week. A goal for electronic publishing, such as EPGs, is for a daily close and transmission to the cable system. In the future, almost real time updating of the data used in the EPG is a desired target.

Interface with System Controller

The program schedule data used in a local cable company must be synchronized with the cable converter system controller for proper operation of the application. This is required for several reasons.

First, any pay services which are offered must be reflected in the user interface of the EPG. If a customer is authorized to receive a subscription service, the guide must list the service and allow tuning. If not, a marketing opportunity presents itself for the operator and the guide application is an excellent vehicle for selling the service.

Second, any PPV listings must tie back to the proper internal security tags used for order management. Pricing information is an obvious data element which must be coordinated between the EPG, system controller and billing system used by the cable operator.

Third, if shared channels are present, this must also be handled in an appropriate fashion by the guide.

Transmitting Data to Converter

The actual transmission of the listings data to an individual converter will be handled differently by specific hardware vendors. Both "in channel" (VBI) and "out of channel" (RF Carrier) approaches will be used. In most cases the converter vendor provides the headend equipment to transmit the listings data in a secure fashion. In other cases, there will be more of an open architecture approach.

For the majority of advanced cable converters, either analog or digital, it is also possible to download the EPG application program itself. As such, a separate function is required to manage the proper version of software which needs to be running in each model of the cable converter.

The Guide Application

There are many features which can be offered in an Electronic Program Guide. A prototype system has been developed to illustrate a full featured application under development by TV Guide On Screen. Several features from this system will be discussed to illustrate key aspects of the user interface.

The first mode of operation is a simple, yet powerful extension to the current channel up/down function provided on cable converters. This is the "Flip™" mode, which is illustrated in Figure 1. In Flip™ mode, a legend is added to the bottom of the screen showing channel number, call letters, current program name and running time. This allows the consumer to change channels and immediately determine what program is running, even during a commercial break. After a few seconds the display will go away.

A more advanced mode of operation is "Browse™", as illustrated in Figure 2. In this mode, it is possible to change the legend without tuning to a new channel. Both the channel listing and time can be changed. In effect this mode allows the consumer to scan programming that is being offered at the current time on other channels, or look into the future, without disrupting the current program which is being viewed. To navigate this feature, the user must learn to use left and right arrow keys, as well as the up and down arrow keys that they are already familiar with. Also illustrated on this screen is a mail waiting icon which is used by a system messaging function.

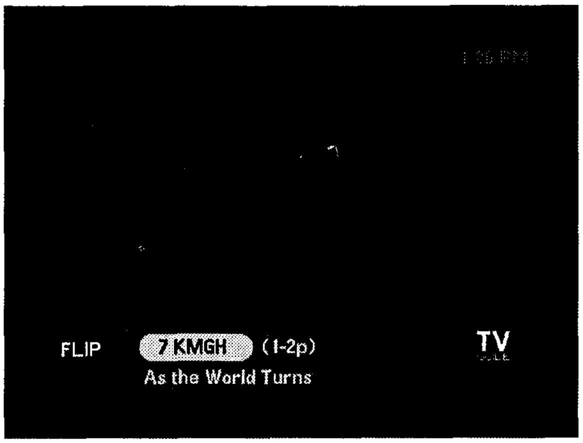


Figure 1. Flip™ Mode

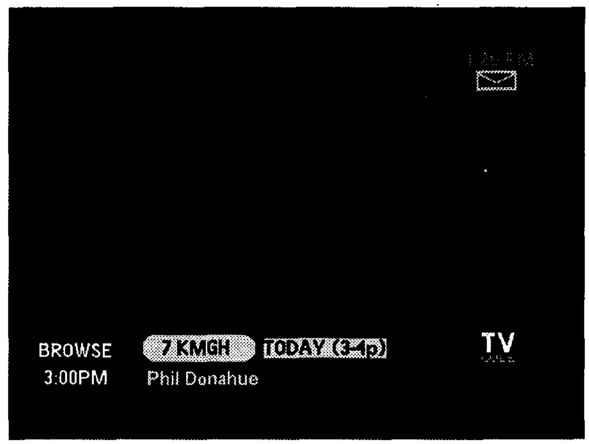


Figure 2. Browse™ Mode

To support the ever growing number of channels that are being offered on rebuilt cable systems it is desirable to provide a "Locator™" function, as illustrated in Figure 3. In the prototype this function is reached by tuning to Channel 1, or by the main menu as described below. This feature lets the user find any channel by broadcast, cable, premium or PPV category. By the user navigating to the desired channel, and then choosing the entry, direct tuning occurs.

In addition, the Locator™ screen can be used to program a favorite channel list, which can be invoked in any mode of operation of the EPG. Favorite channel will be of growing importance to the consumer as the number of channels grows.

To access other EPG features the user chooses the "Main Menu" which provides three categories of features. First, TV listings by Time, Channel or Category under the TV Guide area. Second, PPV and Premium Services under the home theater. Third, customer service features such as messages, Lockout™ and help would be provided.

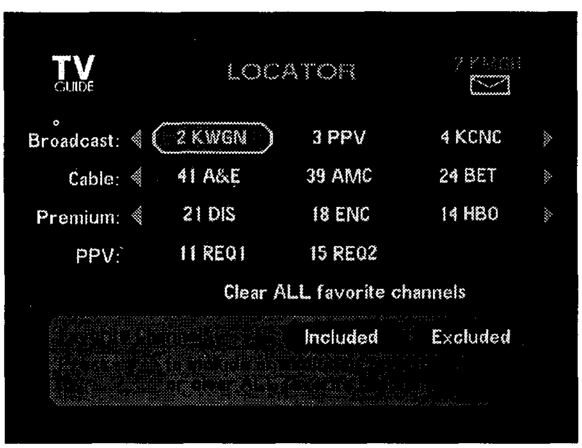


Figure 3. Locator™ Screen

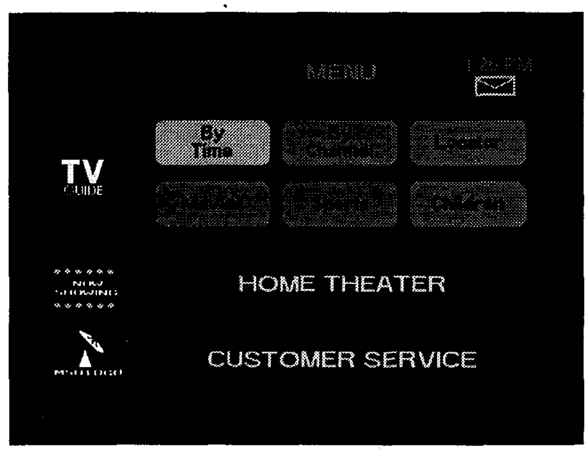


Figure 4. Main Menu

Looking at the Listings by Time feature, as shown in Figures 5 & 6, it can be seen that channel number, call letters, program name and running time can be provided. The "i" icon allows the user to see more information on the program, such as movie descriptions, or sports event details. If the program is currently available it can be tuned immediately by navigating to the desired program by using the up and down arrow keys and then choosing the program.

Future times can be examined by using the right arrow key to shift to the next half hour time interval. As shown in Figure 6, a "Reminder™" can be set by the customer if a future program is desired for viewing. If set, a Reminder™ legend will appear at the bottom of the screen shortly before the program is scheduled to be broadcast.

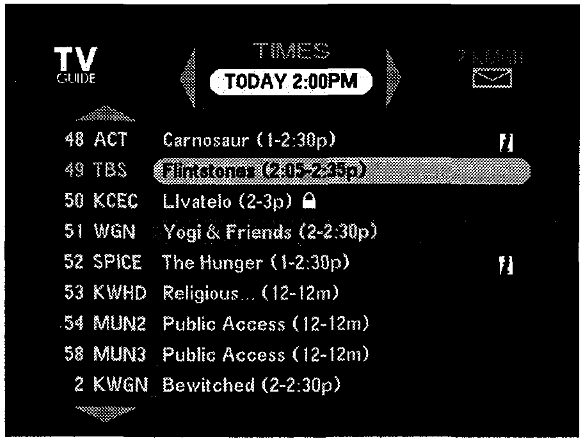


Figure 5. Listings by Time

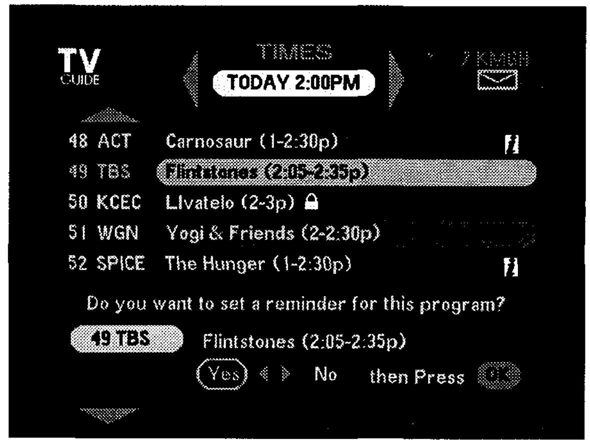


Figure 6. Setting a Reminder™

Other important features of the EPG are to provide support for various premium services offered by the cable operator. A screen which would allow premium services to be ordered from the EPG is illustrated in Figure 7. Impulse subscription to new services is a benefit which will certainly be explored in more detail as these applications are rolled out. Note that in this screen, HBO has already been ordered, as the check mark indicates that the customer is already a subscriber.

The last feature to be described in this paper is the "Lockout™" capability of the guide. Lockout™ features can be very powerful, as they can include limiting access by rating, theme, channel, or a specific program name. In all cases, a password or PIN code would be required to set Lockout™, and to access programs which have been restricted. Industry consensus on ratings codes would be beneficial to the success of such a feature.

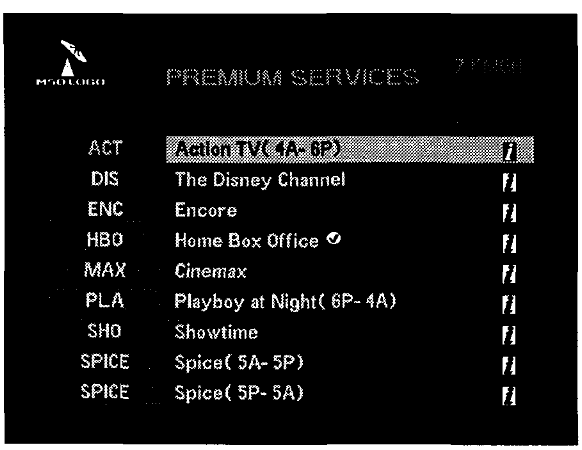


Figure 7. Pay Services

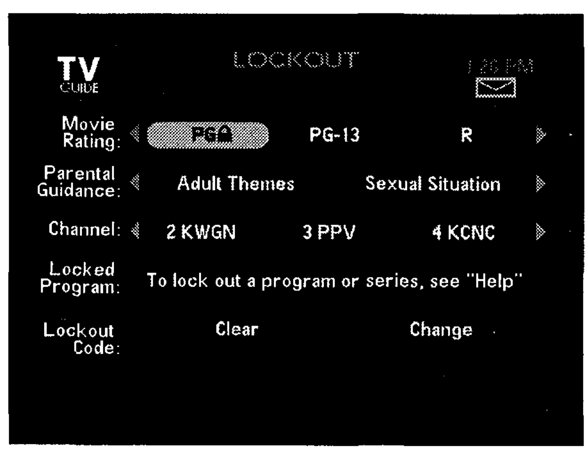


Figure 8. Lockout™ Screen

It is expected that many more features will be provided in a guide application as these products mature. However, it is important to keep the initial products easy to use, and yet offer powerful features that appeal to the consumer. This is the challenge of designing the user interface, as well as other aspects of the system.

The Future

This paper has explored the basics of an Electronic Program Guide. As with most products, there is much more that goes on behind the scenes to complete the implementation of the system.

Such topics as how to design the system to withstand power failures, perform in field installation and training of the customers will need to be determined as these systems near deployment to the field. The frequency of listing data updates, application program download and how these relate to the data transmission channels which are available also need to be resolved for initial rollout of these systems.

As the first of a new breed of interactive products, EPG applications will have cable customers moving into new experiences in using their television sets. Based on initial research, market tests and first hand experience, there will be no turning back as this class of product will be an important new service for the industry.

Engineering Requirements for Hybrid Fiber - Coax Telephony Systems

Charles Eldering†
Marc Kauffman†
David Grubb†
Steve Warwick‡
George Hawley‡

†General Instrument Corporation
GI Communications Division
2200 Byberry Rd.
Hatboro, PA 19040

‡DSC Communications
1310-C Redwood Way
P.O. Box 750699
Petaluma, CA 94975-0699

Abstract

Deployment of telephony services over the hybrid fiber-coax plant requires an understanding of the key system design parameters including bit error ratio, round trip delay, available bandwidth, system capacity, and system reliability. This article examines these parameters and discusses ways in which the cable operator will be able to provide telephony and advanced services over the hybrid fiber-coax plant.

Introduction

The rapid convergence of the cable and telephony industries poses the challenge to develop systems that can support traditional voice and data services, traditional broadcast services, and provide a migration path for advanced two-way video and multimedia services. The motivation for designing an integrated system over hybrid fiber-coax is to produce a single system which provides all of these services at a lower cost than separate systems that provide the same services.

Development of equipment to provide these services requires hybrid fiber-coax plant of suitable quality with node sizes of 500 - 2000 homes, as well as recognition of the key system design parameters such as bit error ratio and round trip delay which will need to be considered when deploying systems that support telephony services over the fiber-coax plant.

In this article we discuss a number of engineering requirements for hybrid fiber-coax systems, including the applicability of Bellcore Fiber In The Loop (FITL) requirements, requirements for the distribution system, channel capacity and bandwidth allocation, encryption requirements, round trip delay requirements, and system reliability. Finally we show how telephony services will eventually be integrated with advanced video and data services on the fiber-coax network.

Figure 1 illustrates an architecture for initial telephony services deployment, in which Coaxial Network Units (CNUs) are served off of the feeder part of the network by power passing taps and drop cables, and which provide telephony services to customers via twisted pair drops. The CNUs support 18-64 customers, and are similar in functionality to the Optical Network Units (ONUs) used in FITL applications. If the CNUs are supported by a Host Digital Terminal (HDT) as shown in Figure 1, fibers can be run to high density areas or businesses which have a service demand which will support a dedicated fiber.

Figure 2 illustrates a more advanced architecture in which Coaxial Termination Units (CTUs) are deployed at the side of the subscriber residence and provide basic telephony and advanced telecommunications (N-ISDN, videotelephony, and data) services. In this architecture, the Integrated Services Host Digital Terminal (ISHDT) has added functionality which provides media access control and data packet routing as well as accessing the public switched telephony network. The ISHDT is obtained by augmenting present HDT equipment with additional equipment to provide this functionality: replacement of the HDT is not necessary.

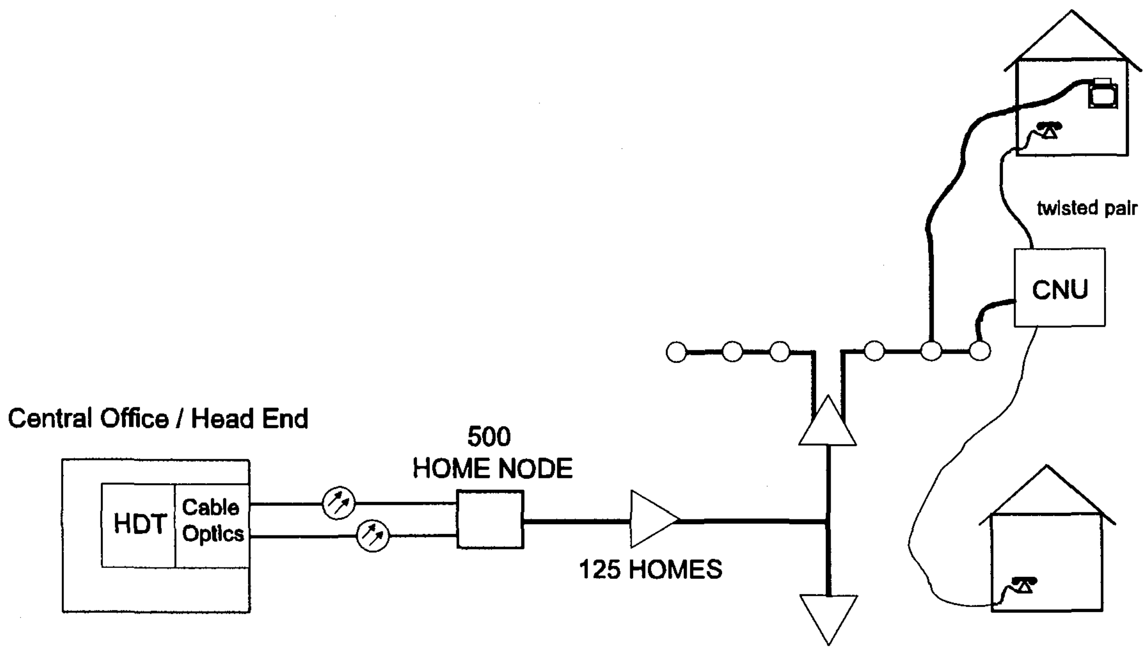


Figure 1. Initial telephony deployment, in which Cable Network Units (CNUs) are supported by the distribution network and the Host Digital Terminal (HDT) which is deployed at the head-end or central office. The CNUs provide service to 18-64 customers via twisted pair drops.

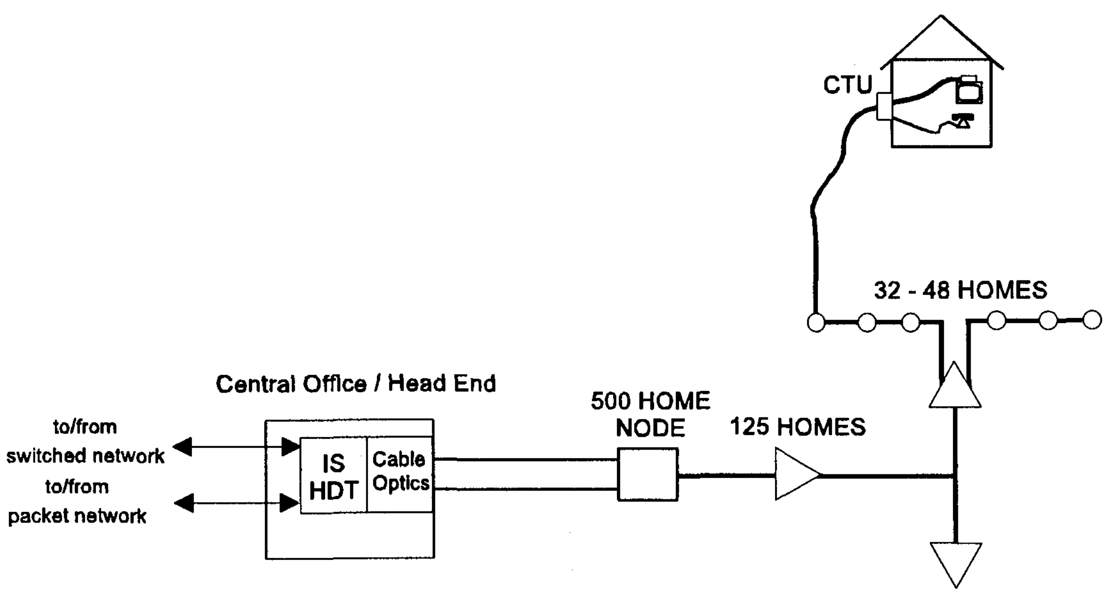


Figure 2. Advanced telephony deployment, in which Coaxial Termination Units (CTUs) are deployed at the residence, and are supported by the Integrated Services Host Digital Terminal (ISHDT) deployed at the head end or central office.

Applicability of Bellcore Fiber In The Loop requirements to hybrid fiber-coax systems

Telecommunications services provided over the cable network must ultimately be competitive with FITL services on a cost and performance basis, and must be compatible with FITL platforms. Although there may be exceptions for certain business scenarios, the long term requirement will be to develop cable telephony systems which meet the service objectives of Bellcore[1] TR-NWT-000909:

Service availability of 99.99%, corresponding to a downtime of 53 min/yr/line, excluding the local switch and the customer premises equipment,

Round trip delay not exceeding 2.1 ms, (for universal FITL systems)

BER performance for a DS1 signal transported through a cable telecommunications system of less than $1x10^{-9}$, excluding all Burst Errored Seconds in the measurement period,

Errored Seconds[2] not exceeding 0.14% for a DS1 signal transported through the cable telecommunications system. This is equivalent to no more than 10 Errored Seconds during a two-hour, one-way (loopback) test,

An average frequency of Burst Errored Seconds[3] for a DS1 signal transported through a cable telecommunications system of no more than 4 per day.

Although there are a number of other performance requirements related to transmission, as well as the switch and subscriber interfaces, the parameters listed above will be the most challenging for cable operators to meet. The ability to meet these requirements in the long term will assure that cable telecommunication services are performance competitive with FITL services.

The Bellcore switch interface requirements will be directly applicable to fiber-coax systems, as it will be necessary to support both older TR-008 and TR-057 interfaces, as well as the TR-303 interface, which will become the standard interface in the US. When fully supported by switch manufacturers, the TR-303 switch interface will allow a number of Operations,. Administration and Maintenance (OA&M) functions to be performed by the switch, as well as significantly lowering the costs of providing N-ISDN services.

Distribution plant requirements

For broadcast video services, the distribution plant has shown a continual evolution from a large number of subscribers (20,000 and greater) per head end transmission point over a purely coaxial system, to fiber-coax systems with four 500 - 2000 home nodes served from a single optical transmitter, followed by a coaxial transmission system in which the number of actives is highly reduced or in some cases even eliminated. This trend will facilitate the deployment of 2-way telecommunications services including data, but additional effort will be required to make the distribution plant reliable and secure in order to meet the reliability and service objectives presently met by FITL systems.

Figure 3 shows a Broadband Telecommunications Architecture (BTA) which uses a single optical transmitter for four 500 home nodes, and a dedicated optical transmitter for each return path. The fiber to coaxial transition is at the mini-bridger, which connects to additional mini-bridgers and serves subscribers via feeders and taps to drop cables.

Because the fiber-coax network is a point-to-multipoint system, the use of the return path for telecommunications services will require additional network security to prevent intentional or non intentional interference from rendering the network inoperative. A network security device, as shown in Figure 3, can perform this function by removing one or more subscribers from the network once it is determined that the source of interference originates from that part of the network. One approach to network security will be to develop the ability disconnect a small number (e.g. 4-12) subscribers from the network if the source of interference is located in any part of the distribution network. If necessary, it will also be possible to use addressable taps to individually disconnect a subscriber at the tap. It should be noted that the CTU will also contain a network security device for removing the subscriber from service if that subscriber is identified as an interferer, thus the distribution network security device is additional protection which allows disconnection of that part of the network from the CTU on back in which the interference originates. Techniques for detection of interferers and methods of locating them are presently being developed.

Bandwidth on the return path, and more importantly, usable bandwidth on the return path, is an important issue for deployment of telephony services on the fiber-coax system. Present systems typically have the 5 -30 MHz region available for return path transmission, and systems are now being upgraded for a 5-40 MHz return path. Settops presently operating in the 8-12 MHz region may require additional bandwidth for multimedia services and extension of the settop return path frequencies to the 8-15 MHz region is foreseen. The functions of status monitoring systems operating in the 5-8 MHz range are likely to increase, and this much bandwidth may be necessary for these systems in the future. As will be discussed, the frequency allocation on the return path will be important, as will reduction of the amount of ingress in the system. The achievable signal to noise ratio on the return path is adequate for modulation formats of 64QAM, but interference is likely to limit the modulation format to Quadrature Phase Shift Keying (QPSK) or 16 state Quadrature Amplitude Modulation (16QAM). However, it appears possible to make a portion of the return path spectrum reliable for telecommunications services. If we consider allocation of 25 MHz in the 15-40 MHz range to telecommunications services, there will be sufficient bandwidth for telephony services, but addition of advanced services such as videotelephony, data, and wireless services will undoubtedly require increased return path bandwidth. This can be accomplished by the use of frequency upconversion in the branches of the return path, as shown in Figure 4. In this scenario, all but one of the return path branches are upconverted, and the return path laser is modulated above 40 MHz. Since there is a potentially large bandwidth available from the return path transmitter at the fiber to the headend, (e.g. 750 MHz if a high bandwidth laser is used on the return link) upconversion provides the possibility of greatly increasing the return path bandwidth from the subscriber by translating the 5-40 MHz return spectrum to a region above 40 MHz. A possible frequency allocation utilizing upconversion will be discussed in a later section. The use of a high-split return (900 MHz - 1 GHz) is also possible, but the large transition region required for low cost diplex filters and high gain return path amplifiers make it a generally less desirable approach. However, if ingress and interference problems in the 5-40 MHz region cannot be overcome, the high-split return will become a suitable alternative to upconversion for increased bandwidth.

Powering is a critical issue for both FITL and fiber-coax systems. With the deployment of digital loop carriers, telcos were able to save money by running a fiber to a serving area and deploying the subscriber line card remotely from the central office, instead of running a large bundle of twisted pairs from subscriber line cards in the central office to the subscriber. However, this has complicated the powering scenario in the sense that the remote deployment of line cards will require back up power (e.g. batteries and connection to a remote powering source for generator backup) which was previously located at the central office. The deployment of smaller remote terminals, whether in FITL or fiber-coax scenarios, requires re-engineering of the power distribution network. A detailed discussion of all of the powering alternatives is beyond the scope of this paper, but it is clear that careful analysis of the cost and reliability tradeoffs will be necessary to determine the optimum location of battery backup and remote powering access for generator backup in the fiber-coax network. It is clear however, that in the case of CTU deployment (Figure 2) it will be necessary to power the CTU from the network to insure service during power outage. While power can be distributed from the node through the feeder network, it is not clear that supplying power to the CTU directly via the coaxial drop is optimum. An alternative to powering through the coaxial drop is the use of twisted pair from the tap for powering only. A comparison of the advantages of each of these powering schemes is given in Table I. From this table it can be seen that the advantage of coaxial drop powering is that the simple coaxial connection is compatible with present systems. The principal disadvantages are that the diplexing functions which enable power to be combined and later separated from the RF signal add RF signal loss and require additional diplexing filters, and the center conductor of the coaxial drop cable may exhibit higher I^2R losses than a twisted pair drop. For the twisted pair drop from the tap, the diplexing problem is avoided, but there is additional cost and complexity in terms of connectors and additional cable. Both solutions are technically feasible.

In summary, powering will be a major system consideration for fiber-coax telephony, and while it is likely that an AC, current limited scheme can be used to power the CTU, the tradeoffs between coaxial powering or a twisted pair power drop from the tap need to be carefully examined. For the CNU, coaxial powering will be possible, with minimal impact on the presently used powering schemes. For CTU powering, the presently used distribution network can be modified to support powering, but optimization of the powering scheme will be necessary to reduce cost and maximize reliability.

Table I
Comparison of advantages and disadvantages for coaxial vs. twisted pair drop powering

Coaxial drop powering	Twisted pair drop powering
Advantages:	Advantages:
Installer friendly: does not require separate connection at tap or at subscriber unit	Eliminates power/RF diplexing points at tap and at subscriber unit
Backward compatible with existing drop cables, if cables are in good condition	Low power loss
Disadvantages:	Disadvantages:
Added RF signal loss at tap and CTU	Additional cost in cable and connectors
Potentially high I^2R losses for power distribution	Not backwards compatible with existing cable plants

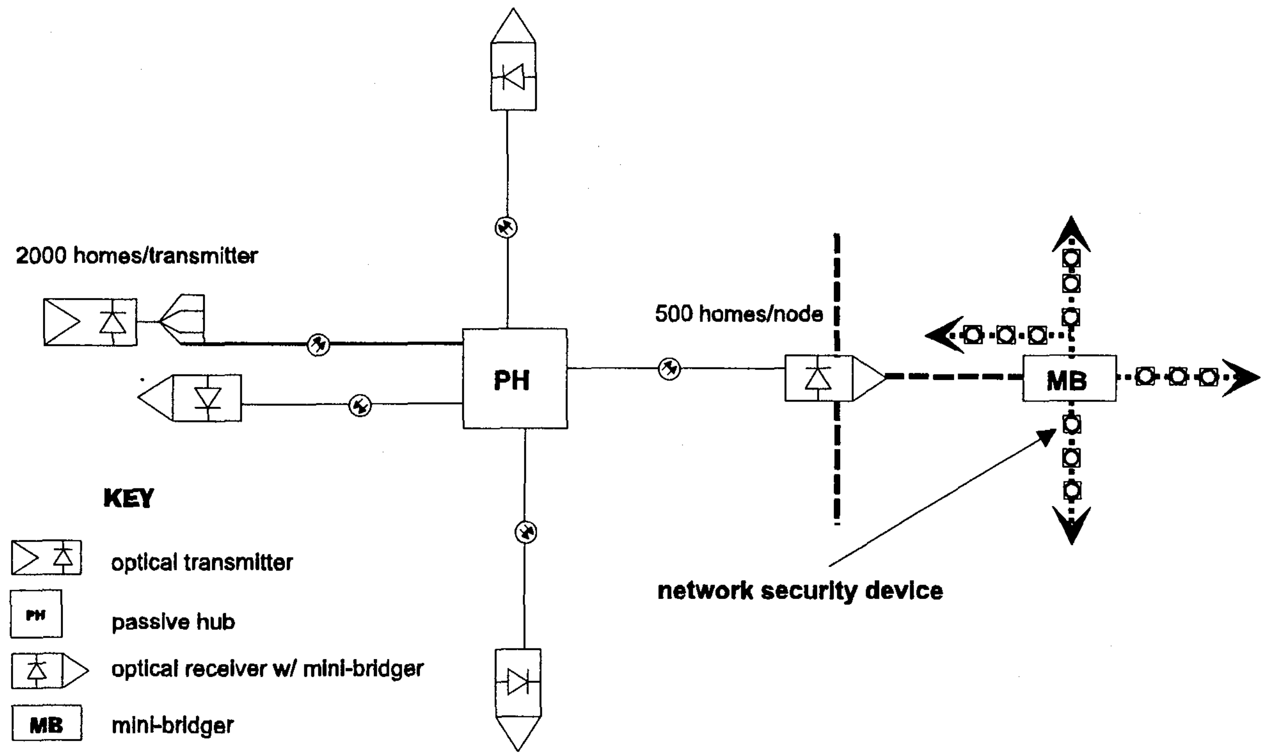


Figure 3. Broadband Telecommunications Architecture (BTA) using 500 home nodes and 2000 homes per optical transmitter.

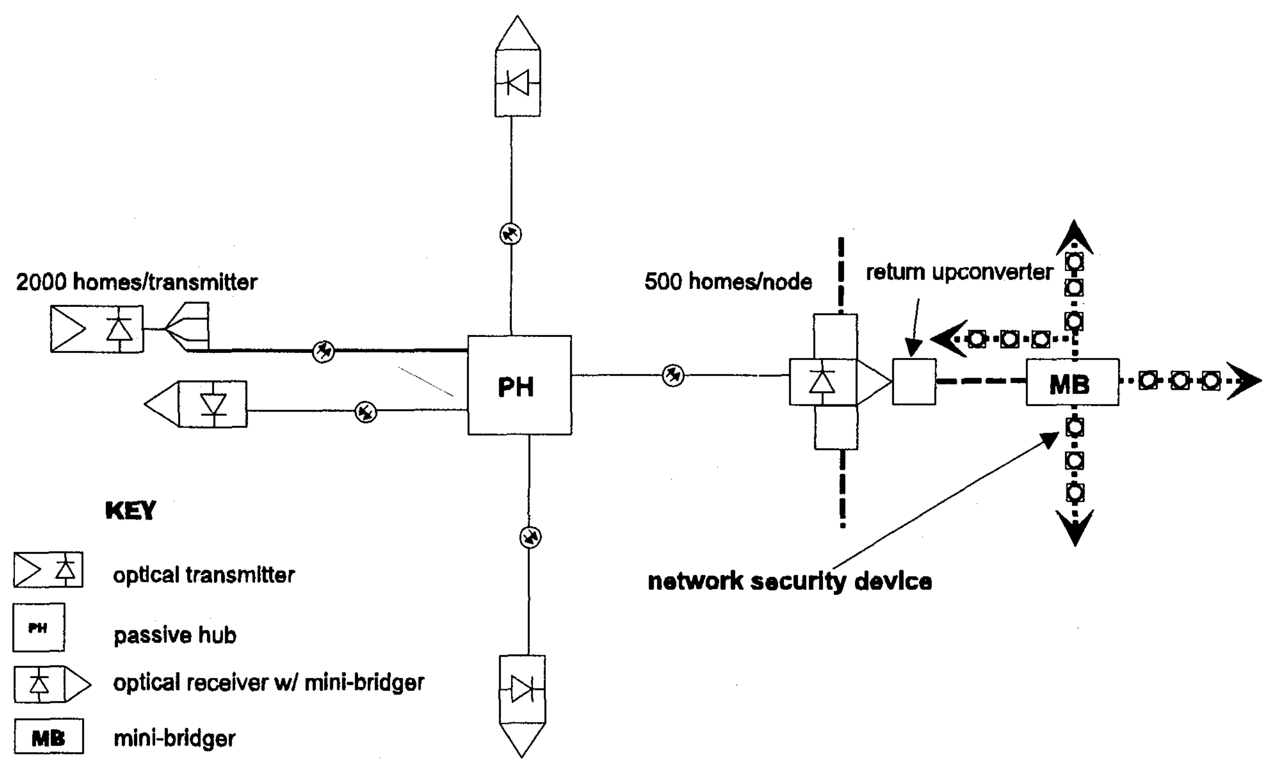


Figure 4. Broadband Telecommunications Architecture (BTA) using 500 home nodes and 2000 homes per optical transmitter, with frequency upconversion at the node for increased return path bandwidth.

Capacity and bandwidth allocation

While it is envisioned that modulation formats such as QPSK will initially be used on the return path, the ability to control ingress will eventually allow the use of a higher spectral density format such as 16QAM. In addition, the use of upconversion will allow tremendous expansion of the return path bandwidth and will provide sufficient bandwidth for a number of services in addition to telephony. For CNU service, QPSK modulation using frequency division multiplexing will be used to establish logical point-to-point connections between the HDT and CNUs in both the downstream and upstream. For the deployment of CTUs, downstream channels are expected to be transported in 27 Mb/s channels which fit in a 6 MHz bandwidth and which are transmitted using a 64QAM format. Table II shows both the upstream and downstream channel requirements and approximate system capacity for both QPSK and 16QAM modulation formats, assuming the use of 64QAM on the downstream. The channel capacity, while expressed in 64 kb/s (DS0) circuit switched connections, does not imply that the entire capacity of the system is dedicated to the switched telephony network. It is likely that in the future a substantial portion of the system capacity would be dedicated to packet based services which multiplex bandwidth on a statistical basis, and do not have circuit based DS0 or n x DS0 connections.

A proposed allocation of frequencies on 5-30 or 5-40 MHz return paths is shown in Figure 5. This allocation would allow status monitoring and settop services to coexist with telecommunications services. The downstream and upstream frequency allocations for this scenario are shown in Figure 6. Figure 7 illustrates a possible frequency allocation which would be used if upconversion at the node is used to expand the return path bandwidth. Figure 8 shows the downstream and upstream frequency allocations for this scenario.

Table II
Upstream and downstream channel requirements and capacity

upstream modulation technique	return path bandwidth	maximum number of 1.2 MHz upstream channels	number of 64 QAM (6 MHz) downstream channels required to support upstream requirements*	approximate number of DS0s supported per 500 home noae
QPSK	25 MHz (15-40 MHz)	20	8 (2)	480
QPSK	100 MHz (15-40 MHz with upconversion)	80	28 (8)	2106
16 QAM	25 MHz (15-40 MHz)	20	12 (3)	864
16 QAM	100 MHz (15-40 MHz with upconversion)	80	48 (12)	3744

* numbers shown here assume 2000 homes per optical transmitter, and 500 homes per transmitter (in parentheses)

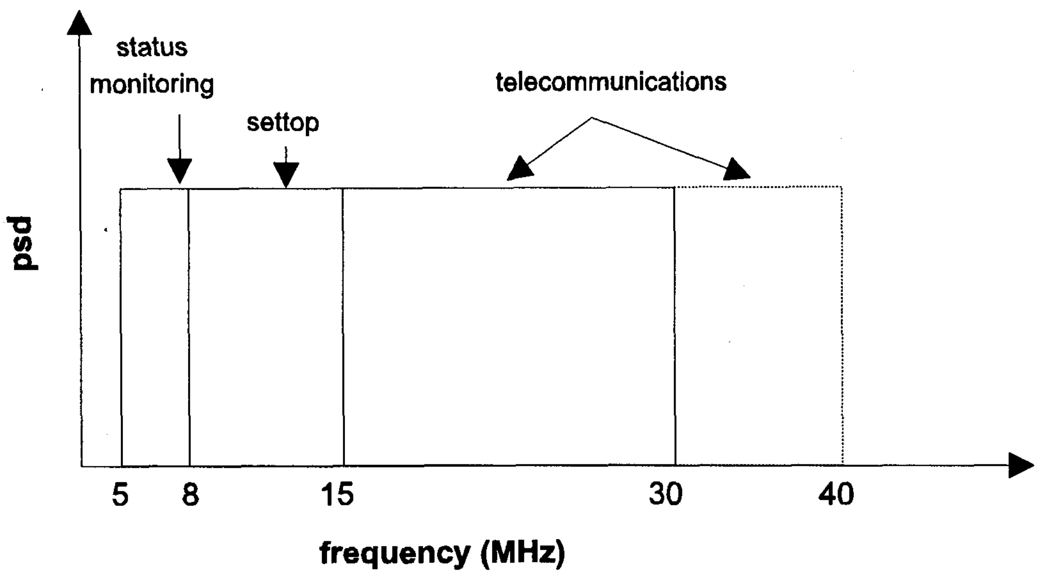


Figure 5. Proposed initial frequency allocations for upstream services on 5-30 and 5-40 MHz return paths.

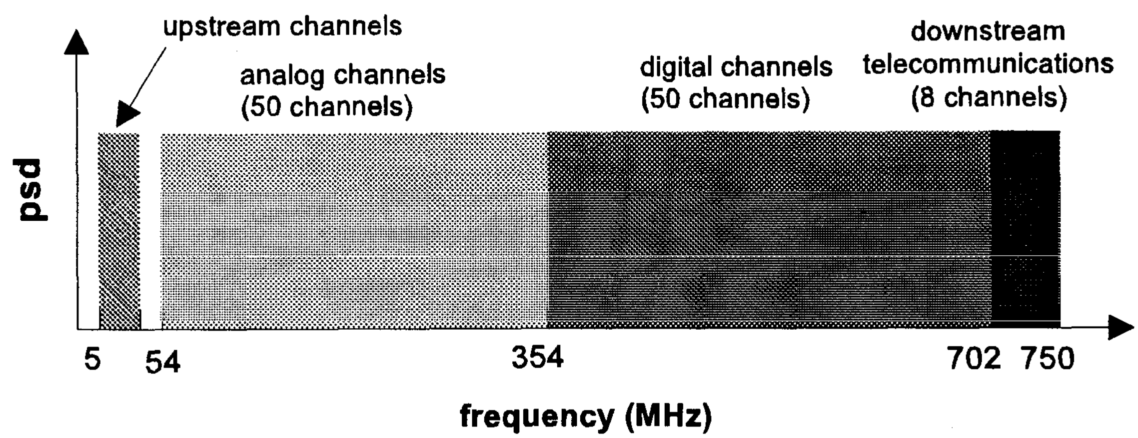


Figure 6. Downstream and upstream frequency allocations based on the use of a Broadband Telecommunications Architecture (BTA) using 500 home nodes, 2000 homes per optical transmitter, and QPSK transmission on the return path.

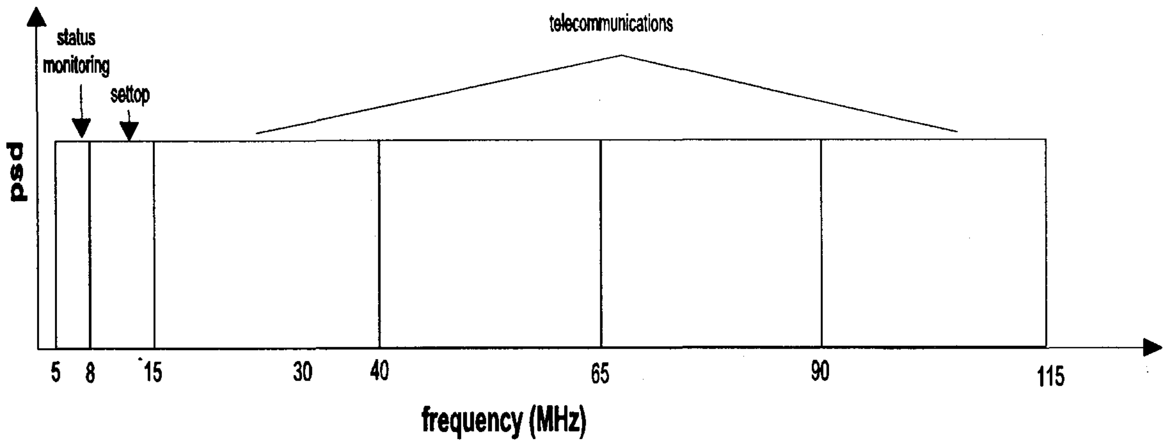


Figure 7. Proposed initial frequency allocations for upstream services using frequency upconversion at the node.

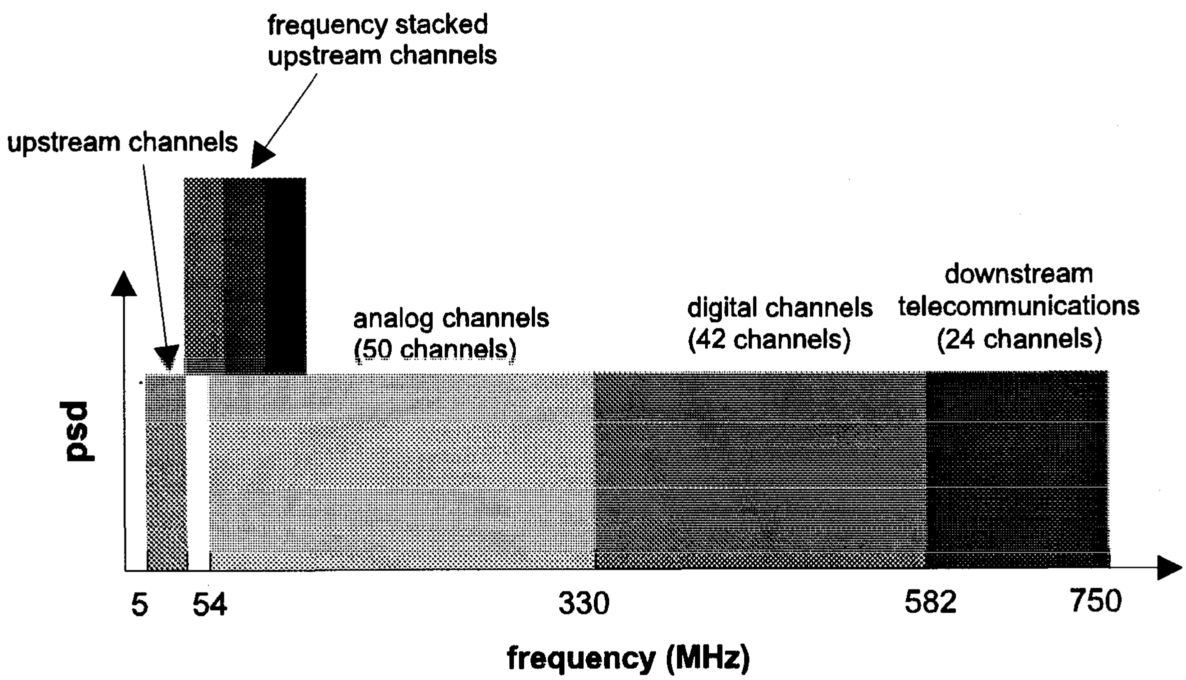


Figure 8. Downstream and upstream frequency allocations based on the use of a Broadband Telecommunications Architecture (BTA) using 500 home nodes, 2000 homes per optical transmitter, and QPSK transmission on the return path.

Encryption requirements

Unlike the current point-to-point topology of the residential telephony plant, the fiber-coax topology is point-to-multipoint. This will result in a potentially significant cost savings, but will require additional security measures to ensure privacy and resistance to unauthorized access. Encryption is currently used in video distribution products to prevent unauthorized reception of programs, but even if the code is broken the service to other customers is not affected since unauthorized reception does not affect the value of the service to the other customers. For telecommunications services this is not the case since even the perception that calls made over the facility are not secure devalues the service. In addition, system security is much more important for telecommunications services since users are billed on a direct usage basis. These reasons make the use of encryption for telecommunications services a must.

Encryption can be best supported on the network through the use of a key hierarchy, which includes both physical and algorithmic security. Such systems use multiple levels of key protected by distinct mechanisms and varied periodically to provide an extremely high level of security. Presently used broadcast video encryption methods can be modified to make encryption of telecommunications services possible.

The use of data encryption for telecommunications services has become a political and regulatory controversy due to the fact that government agencies may require access for authorized tapping. However, services such as call forwarding will make accessing the subscriber after the switch and signal transfer point practically useless, and it would appear that the only reasonable point for tapping will be at the switch. This would allow the use of high security encryption techniques on the access part of the network (either fiber or fiber-coax).

Round trip delay requirements

One of the basic performance parameters of voice communications systems is the round trip delay, due to the fact that excessively long delays can result in a degradation of service due to perceivable echo. In general, one-way delays of greater than 25 ms will cause any existing echoes of the speakers own voice (speaker echo path) to be annoyingly perceived by the speaker. Excessively long delays (e.g. 500 ms) can result in difficulties due to pure speech delay, even when echo cancellors are used to remove talker echo.[4] Because of these reasons, it is necessary to limit the total transmission delay, including the contribution of the delay in the access part of the network (switch to subscriber). CCITT recommendations[5,6] only provide recommendations for the one-way delay between speaker and listener, but the delay contributed by the fiber-coax part of the network must be included in this calculation. Bellcore recommendations[1] include a limit on the round trip delay from subscriber to subscriber which has a limit of 2.1 ms for a universal FITL system, and 1.5 ms for and integrated FITL system. In systems which do not meet this specification, active echo cancellation must be incorporated.

The 2.1/1.5 ms specification includes both the physical delay of the facility and the processing delay for the digitally encoded signals. In general, the physical delay is a lesser contributor in the loop plant. The primary contributors to processing delay include:

frame buffering of digital streams

upstream TDMA frame formation

symbol-to-binary conversion (modem processing delay)

error detection and correction

As is to be expected, manufacturers will need to develop means to perform these functions in a manner which meet the delay specifications, or incorporate echo cancellation.

System reliability

The deployment of fiber optic components in the local loop has raised the concern that system reliability will decrease with respect to the mature but bandwidth limited twisted pair solution. Because of this, Bellcore has paid particular attention to local loop reliability for FITL and have established downtime objectives for FITL systems.[1,7] Similarly, hybrid fiber-coax systems which support telephony will need to support downtime objectives, although it is probable that there will be different initial and long term objectives to allow for upgrading of the distribution plant and the development of methods to reduce ingress which may initially limit the availability of the return path. Table III shows the initial and long term downtime objectives for hybrid fiber-coax systems. These calculations include a MTTR of 2 hours for the central office/headend equipment, 4 hours for the distribution equipment and 6 hours for the CNU/CTU.

These calculations illustrate that the active distribution elements and ingress in the return path may initially make the unavailability of hybrid fiber-coax systems greater than the 53 min/yr FITL objective. However, the observed reliability of some distribution elements is much higher than the predicted reliability, and it is likely that the impact of actives in the distribution plant on overall system reliability will be minimal, due to the fact that the active elements contain mature and highly reliable technology. The forward and return path lasers are the exception to this, and laser reliability will be an issue which can be addressed by redundancy as well as increased component reliability. Methods of reducing ingress and the resulting interference in the return path are actively being studied to maximize the availability of the return path channel. It should also be noted that the long term goal for downtime of the CNU/CTU is lower than that for ONUs (26 min/yr) which should be achievable since there are no optical components in the CNU/CTU.[7]

Table III
Initial and long-term downtime objectives for hybrid fiber-coax telephony systems

Network Segment/Element	initial downtime (min/yr)	long term downtime (min/yr)
Digital Cross-Connect & Transmission Facility		
- hardware	2	2
- media	6	6
HDT	10	10
CableOptics	10	2
Distribution Components		
-receiver and mini-bridger	8	5
- mini-bridger	6	1
- line extenders (3)	18	2
- return path optics	12	5
CNU/CTU	26	10
Return path ingress	60	1
Unassigned	9	9
Total	167	53

Integration of telephony and advanced video/data/wireless services

While telephony coupled with analog and broadcast video services are expected to provide services in a cost effective manner, advanced services such as video on demand, data and wireless services are expected to generate additional revenue, if they can be deployed on the network at a reasonable incremental cost. Figure 9 illustrates the integration of these services, and illustrate the use of a network switch for routing of packet based video and data services, and Figure 10 shows a possible frequency allocation. The bandwidth allocated to each of the services will be dynamically allocated, to ensure best use of the entire spectrum. Integration will provide a cost effective means of not only providing, but provisioning, and administrating these services.[8]

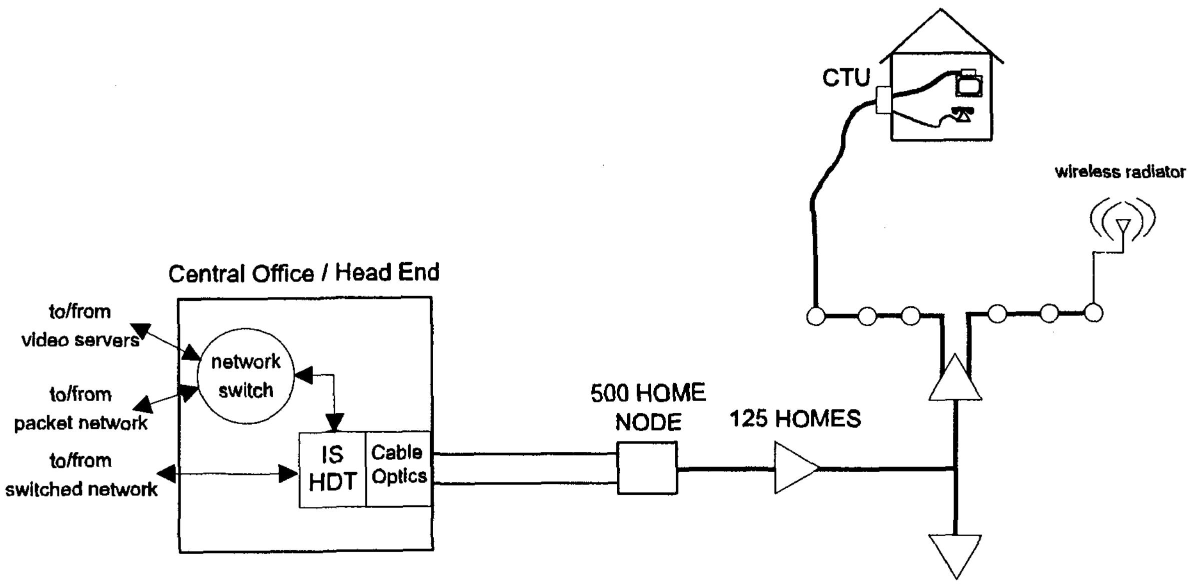


Figure 9. Integration of telephony, advanced video, wireless, and data services.

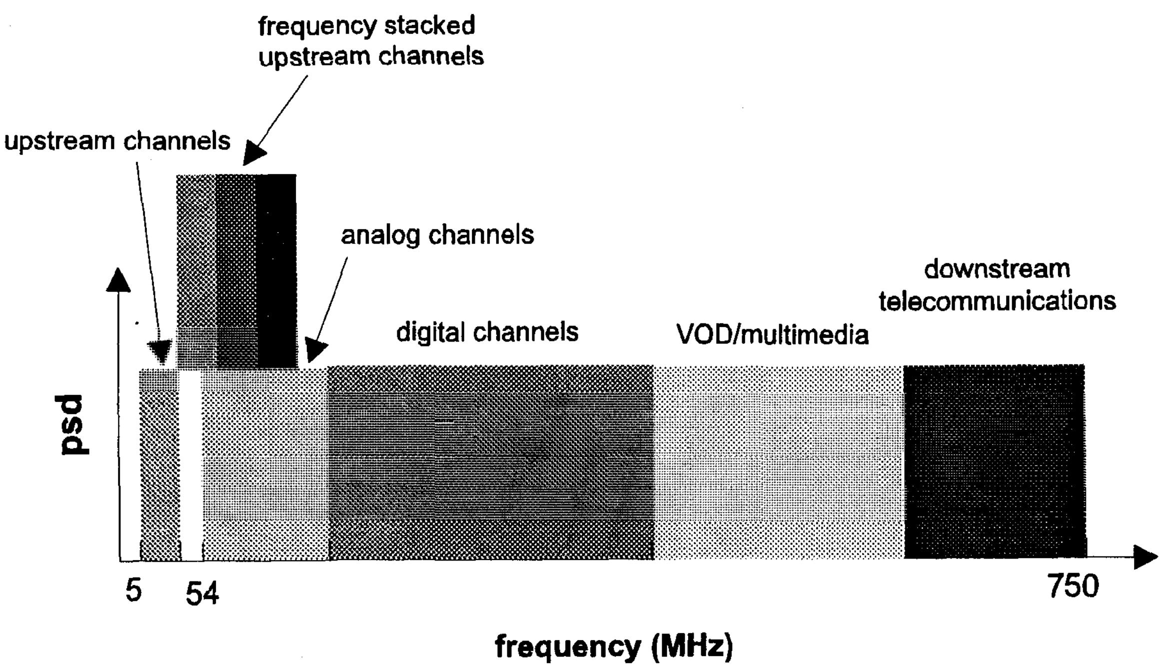


Figure 10. Frequency allocations for a fully integrated video and telecommunications system.

Conclusions

In this article we have discussed the primary engineering requirements for deployment of telephony over hybrid fiber-coax systems. Meeting these requirements and making the hybrid fiber-coax environment compatible and performance competitive with Fiber In The Loop will insure successful deployment of telephony services over hybrid fiber-coax. Particular concern will have to be paid to bit error ratio, round trip delay, and system reliability in order to match FITL service objectives. The ability to meet these objectives, coupled with the advanced video, data and wireless services which can be deployed over the broadband plant will insure the success of hybrid fiber-coax as the local loop transport means for all communications services.

Acknowledgments

The authors would like to acknowledge H. Lustig, P. Keeler, M. Capuano, R. Safadi, V. Ragavan, G. Roman, R. Young and J. Glaab for their technical and marketing assistance in the area of telephony over cable.

References

[1]TR-NWT-000909, *Generic Requirements and Objectives for Fiber In The Loop Systems*, Issue 1 (Bellcore, December 1991).

[2] An Errored Second is defined by Bellcore as a second in which at least one bit error was received, stated as a percentage over some time interval.

[3] A Burst Errored Second is defined by Bellcore as any Errored Second containing at least 100 errors.

[4] N. Kitawaki, K. Itoh, "Pure delay effects on speech quality in telecommunications," *IEEE J. on Selected Areas in Communications*, vol. 9, no. 4, pp. 586-593, May 1991.

[5] CCITT Recommendation G.131, "Stability and Echo," Fascicle III.1, pp. 143-155 (1988).

[6] CCITT Recommendation G.114, "Mean One-Way Propagation Time," Fascicle III.1, pp. 84-94 (1988).

[7] B. Unger, J. Spencer, "Fiber in the loop reliability: the impact for fiber optic components," *Optical Engineering*, vol. 30, no. 6, pp. 815-820 (June 1991).

[8] R. Safadi, "An optimal approach to a full service broadband communications network," *Proceedings of the 1994 NCTA Convention & Exposition*, (New Orleans, May 1-5, 1994).

FCC Technical Standards and Baseband S/N

J. Samborski and T. M. Straus
Hughes Aircraft of Canada Limited

Abstract

It is well known that S/N depends directly on C/N, but its dependence on other parametërs affecting picture quality such as C/CTB has not previously been fully explored. Since a measurement of C/N is in any case also required, it is suggested that a simultaneous measurement of S/N will yield a worst case number for C/CTB. This technique would not only satisfy the letter of the FCC requirement, but also, since S/N is an excellent indicator of overall picture quality, assure the cable operator that he is supplying a quality product to his customers without the onerous need of service interruption.

BACKGROUND

Proof of performance tests to demonstrate compliance with the recently established minimum FCC technical standards would normally require a measurement of C/CTB. This test is to be performed twice a year at a minimum of 6 test locations on at least 4 separate channels. Typically this could add up to over 100 channel service interruptions per year if the test is performed in the traditional manner. With the increased emphasis on quality of service, the cable industry is seeking means to minimize service interruptions while still assuring that the technical standards are maintained. A number of test methods are under consideration. One possible method, herein proposed, is the measurement of baseband S/N and the correlation between this parameter and C/CTB.

The NCTA Recommended Practices describes three methods of measuring C/CTB. The first method is the one that has been most commonly used but requires that the carrier be turned off during the second part of the measurement. A spectrum analyzer is used, and the value recorded is read directly from the face of the analyzer unless the thermal noise floor is close to the CTB level, in which case a correction is made for the proximity of this noise. Note, however, that the definition of C/CTB does not include the correction of the noise-like CTB distortion for the error in spectrum analyzer reading of absolute noise level[1].

Both the second and third NCTA methods of C/CTB measurement avoid the necessity of turning off the video signal. One drawback of the second method is that it cannot account for C/CSO. A drawback of the third method is that it must depend on the assumption that the variation of C/CTB with channel frequency remains constant so as to permit the calculation of C/CTB for various channels from a single out of band measurement.

Yet another method of measurement has been suggested[2]. However, the measurement of S/N is based on test equipment which is readily available. One caution should be noted before applying the results reported in this paper since the characteristics of the test equipment employed could differ in such a way as to affect the relationships suggested herein. In general, however, it is proper to observe that both CSO and CTB can be expected to influence the baseband S/N.

FACTORS INFLUENCING S/N

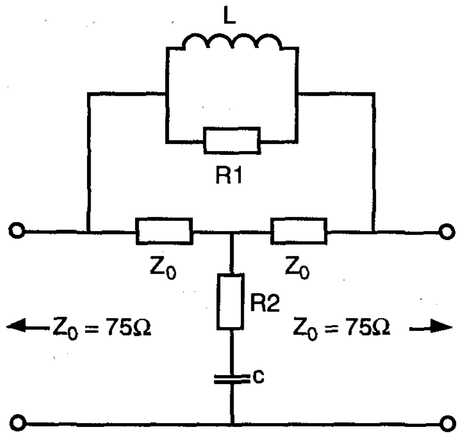
The relationship between baseband S/N and NCTA C/N has previously been analyzed[3,4]. The CCIR adopted a unified noise weighting network[5], shown in Figure 1, which supersedes the descriptions given by Figures 1 through 4 of reference (3). The resulting relationship, now also valid for the EIA[6], is

$$S/N = C/N + 0.6\ \text{dB} \qquad (1)$$

The influence of phase noise on S/N has also undergone analysis[7]. It is shown that

$$S/N = C/N_p \qquad (2)$$

where N_p is measured in a 1 KHz resolution bandwidth at a 20 KHz frequency offset from the carrier and it is assumed that the phase noise falls off at a rate of 6 dB per octave. On a spectrum analyzer, the distinguishing feature of phase noise is that it varies with frequency whereas thermal noise is essentially flat over a 6 MHz channel width. When measuring phase noise in this manner, it is important to keep in mind that the spectrum analyzer contribution is not necessarily insignificant. The analyzer contribution can be measured by utilizing a known very low phase noise signal such as from a crystal oscillator. In any case, the shape of the phase noise is in general more complex than the assumed 6 dB/octave roll off and therefore its contribution to S/N is best established by a single channel measurement (no CTB or CSO) with very high C/N.



$$A = 10 \log \frac{1 + \left[\left(1 + \frac{1}{a}\right)\omega\tau\right]^2}{1 + \left[\frac{1}{a}\omega\tau\right]^2}$$

$L = Z_0 \cdot \tau$ $\qquad R2 = \frac{Z_0}{a}$

$C = \frac{\tau}{Z_0}$ $\qquad a = 4.5$

$R1 = a \cdot Z_0$ $\qquad \tau = 245$ ns

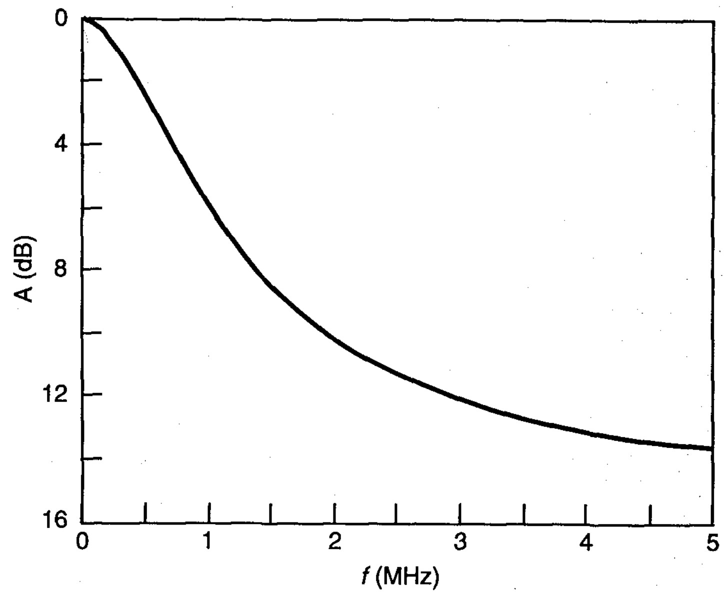


Figure 1. Unified weighting network for random noise.

To assess the influence of CTB on S/N, consider the impact of any narrow-band noise-like signal centered on the VSBAM carrier frequency. In particular, assume a 30 KHz rf bandwidth and a total lack of noise at any other frequency. Following the notation of reference (4), the integration to determine the video noise power, N_v, is greatly simplified since over the 30 KHz interval the response is essentially constant and the video noise is just $N_v = (2GV_n)^2/R \times 0.03/4$. Substituting this result into the ratio of rf and baseband S/N, one obtains C/N_{rf} = S/N + 1.1 dB. However, keeping in mind the correction of reference (1) and the definition of CTB, C/N_{rf} = C/CTB - 2 dB, so that

$$S/N = C/CTB - 3.1 \text{ dB} \qquad (3)$$

The noise weighting filter does not play a role since the frequency is so close to the carrier. However, for CSO, 7.4 dB weighting corresponding to 1.25 MHz offset must be taken into account. On the other hand the VSBAM receiver response is now at its maximum, adding 6 dB to the previous result so that a net 1.4 dB change is applied; i.e.

$$S/N = C/CSO - 1.7 \text{ dB} \qquad (4)$$

EXPERIMENTAL RESULTS

The measurements of S/N were made using a Tektronix 1450-1 demodulator and a Rohde and Schwartz UPS-F2 S/N meter. The standard method of S/N measurement not only involves the noise weighting network, but also a 10 KHz high pass filter intended to exclude low frequency noise contributions. Since CTB noise falls primarily in this low frequency regime, the most interesting result is obtained with the high pass filter disabled. Thus the data presented here is with this filter disabled, although it is worth mentioning that when only broadband thermal noise was present, the S/N with the high pass filter on measured only 0.2 dB greater than with the filter off.

Figure 2 shows the measured dependence of S/N on C/N. At very high levels of C/N, other noise terms, such as internal phase noise in the measurement system, predominate. Since the S/N deviates by 1 dB from a linear dependence at 64 dB, an absolute, test equipment back to back limit of 70 dB is implied. On the other hand, the region of linear dependence shows

$$S/N = C/N + 1.9 \text{ dB} \qquad (5)$$

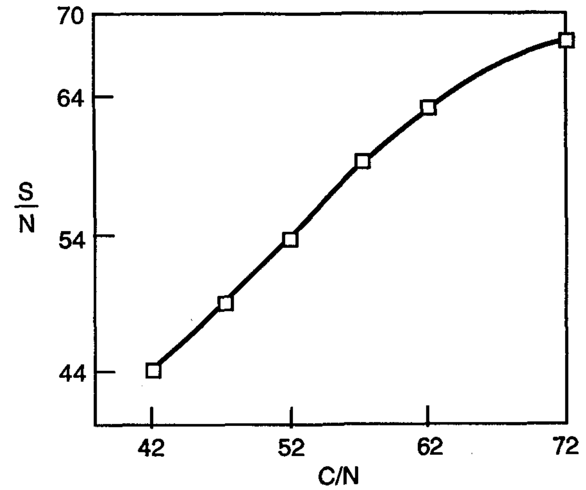


Figure 2. S/N vs C/N.

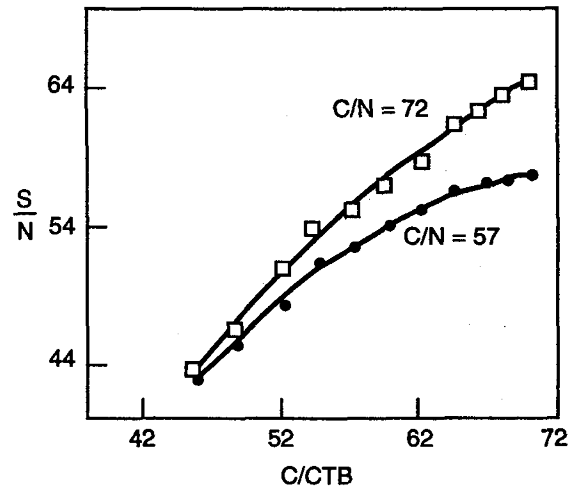


Figure 3. S/N vs C/CTB.

There is a clear 1.3 dB discrepancy compared to the theoretical result stated in equation (1). Whether this discrepancy is due to test equipment calibration error or divergence from ideal VSBAM receiver response is unknown. For this reason, readers are cautioned to "calibrate" their test setup under controlled conditions before applying these results to field measurements.

No attempt was made to vary the system phase noise in a controlled fashion. Nevertheless, this measurement illustrates how the phase noise contribution to S/N, for instance of a typical converter to be supplied to a customer, could be obtained; i.e. the test is performed with only 1 channel on to avoid intermodulation products, and the level is increased to obtain data at very high C/N when the thermal noise contribution is relatively unimportant.

In the next series of tests, summarized by Figure 3, C/CTB was varied in a controlled fashion. A cw Matrix generator provided the multiple frequency tones which resulted in the generation of CTB in a VHF amplifier. When CTB dominates compared to thermal noise, the result is

$$S/N = C/CTB - 2.6 \text{ dB} \qquad (6)$$

This is in fairly good agreement with equation (3) considering the uncertainty in the spectrum analyzer correction factor as well as random measurement errors. The 10 KHz filter of course had a major impact on the measured S/N, but as previously indicated, the results presented are with the filter off. It should also be noted that the CSO was negligible compared to the CTB.

The next series of tests centered on the output of two microwave systems, one utilizing a broadband high power SIBT transmitter[8], and the other a new low cost short range 18 GHz system dubbed the Streetcrosser. When fully loaded with rated channel capacity C/N was the dominant noise contributor but other terms including phase noise entered into limiting the S/N. For the SIBT based system, a measured C/N of 56.4 dB was associated with an S/N of 55.4 dB. The Streetcrosser, depending upon net link loss, showed C/N ranging from 55.9 down to 51.1 dB with corresponding S/N of 54 to 51.8 dB. These C/N and S/N relationships may be contrasted with a recently measured cable system[9] for which a C/N of 45.6 dB was coupled with an S/N reading of 47.3 dB. Another cable system test resulted in C/N of 46.1 and 43.5 dB on two widely separated channels. The corresponding S/N were 47.4 and 44.9 dB respectively.

ALLOWABLE S/N

The minimum C/N required under the FCC regulations is presently 40 dB but will rise to 43 dB by June 1995. The worst C/CTB or C/CSO permitted is 51 dB. By measuring both C/N and S/N, and assuming the relationship given by equation (5), a calculation of the combined S/N contribution of all other sources can be made. For instance, consider the above measured cable system C/N and S/N of 45.6 and 47.3 dB respectively. Application of equation (5) gives S/N due to thermal noise as 47.5 dB. This is only 0.2 higher than the measured total S/N. This implies a S/N contribution of 60.5 dB from whatever is the contributing noise source. If it were C/CSO, equation (4) would predict a value of 62.2 dB. Similarly, from equation (6), C/CTB can be no worse than 63.1 dB. Thus the FCC requirements are easily met. Indeed, the distortion on this cable system was immeasurably small.

Application of equations (5) and (4), together with the 51 dB limitation, permit one to calculate a minimum acceptable S/N corresponding to a measured C/N. The result is given in Table 1. It is of course possible that the FCC requirements are met even though the S/N is slightly below the minimum indicated. For instance converter phase noise could contribute enough to S/N to tip the scale. If that should be the case, a separate measurement of CSO and CTB must be made to assure compliance.

TABLE 1
MINIMUM ACCEPTABLE S/N

C/N	S/N
43	43.6
44	44.3
45	44.9
46	45.5
47	46.1
48	46.6
49	47.0
50	47.4

SUMMARY

It has been shown that many factors influence the reading of S/N, particularly if the 10 KHz high pass filter is turned off during the measurement. In particular, both CTB and CSO will have an effect and therefore a measurement of both C/N and S/N will provide a worst case limit for these distortions without necessitating turning off the channel. It is therefore suggested that the measurement of S/N provides the cable operator with another alternative to assure that the FCC requirements are met and that a high quality signal is being provided to the customers.

REFERENCES

1) Hewlett Packard Spectrum Analysis Application Note 150, pp 31-33, November 1, 1989.
2) R.K. Ward, private communication.
3) T.M. Straus, "Defining S/N", CED, February 1986.
4) NCTA Recommended Practices, 2nd Ed., Appendix C., 1993
5) CCIR Recommendation 567-1, Vol. XII, p 23, Geneva 1982
6) EIA-250-C, Television Transmission Systems Standard.
7) Rezin Pidgeon and Dan Pike, "Oscillator Phase Noise and its Effects in a CATV System", 1988 NCTA Technical Papers.
8) John Hood and Tom Straus, "AML Technology Review", CCTA Technical Papers, 1992.
9) Ted Hartson, private communication.

FM Induced Noise in Analog Fiber Optic CATV Links

Yaron Simler, Moshe Nazarathy and Yishai Kagan
Harmonic *Lightwaves*, Inc.
3005 Bunker Hill Lane
Santa Clara, CA 95054

ABSTRACT

This paper derives a model for the CNR in the presence of un-modulated FM and TV carriers on an analog fiber optic CATV link. Empirically, the inter-modulation distortion generated by the FM carriers mixing with themselves and with the TV carriers cannot be distinguished from noise and therefore in the presence of FM carriers, a new noise contribution should be added to the CNR formula denominator, namely distortion induced FM noise. It is the specific objective of this paper to derive the a modified CNR equation taking into account the FM effects.

DISTORTION INDUCED FM NOISE

The conventional carrier to noise ratio (CNR) formula for an analog lightwave link is given in Eq. 1:

$$\mathrm{CNR} = \frac{\mathrm{Carrier}}{\mathrm{Noise}} = \frac{\mathrm{Carrier}}{\mathrm{Shot} + \mathrm{Thermal} + \mathrm{RIN}} \quad (1)$$

In the presence of FM carriers, a new noise contribution should be added to the CNR formula denominator, namely the distortion induced FM noise:

$$\mathrm{CNR} = \frac{\mathrm{Carrier}}{\mathrm{Shot} + \mathrm{Thermal} + \mathrm{RIN} + \mathrm{FM}} \quad (2)$$

The normalized photocurrent in the receiver (including up to third order distortion terms and normalizing by the DC current generated by the CW light) is given by:

$$i = \phi + \alpha_2\phi^2 + \alpha_3\phi^3 \quad (3)$$

where:

$$\phi = m_{am}\sum_i \cos(\omega_i t + \psi_i) + m_{fm}\sum_j \cos(\omega_j t + \psi_j) \quad (4)$$

is the RF input (AM+FM carriers). Typically, the per carrier AM modulation index (m_{am}) is 4 dB higher than its FM counterpart (m_{fm}). For clarity we will refer to the modulation index simply as *m*, the nature of which (m_{am} or m_{fm}) should be clear from the context.

The powers of the various distortion components (proportional to the squares of the currents) are given by expressions proportional to α_2 and m^4 for the CSO and α_3 and m^6 for the CTB terms, where α_2 and α_3 are the second and third order Taylor coefficient of the memoriless nonlinearity modelling the link, respectively. While it would be possible to evaluate α_2 and α_3 directly from CSO and CTB measurements at a given modulation index for a known TV frequency plan, this is really not necessary if the composite distortions due to the AM and FM+AM frequency plans are known.

The general functional dependence of the new FM "noise" term is given by:

$$\mathrm{FM} = F_2 m^4 + F_3 m^6 \quad (5)$$

where F_2, F_3 are constants to be determined below and the two terms correspond to the CSO and CTB involving the FM+TV channels. Once the general functional dependence of the various terms on *m* is understood one can write a general expression for the total CNR (renaming F2 and F3 - CNR_{cso} and CNR_{ctb}, respectively):

$$\mathrm{CNR}_{\mathrm{tot}} = 10Log_{10} \frac{m^2}{m_o^2\left(\mathrm{cnr}_o^{-1} + \mathrm{cnr}_{\mathrm{CSO}}^{-1}\left(\frac{m}{m_o}\right)^4 + \mathrm{cnr}_{\mathrm{CTB}}^{-1}\left(\frac{m}{m_o}\right)^6\right)} \quad (6)$$

In this expression m_o is a nominal modulation index at which the CNR in the absence of FM effects is known to be equal to the nominal CNR_o at the modulation index m_o as obtained using the regular CNR formula (Eq. 1).
The next two terms correspond to the distortion power contributions represented as equivalent noise, i.e. $cnr_{\mathrm{CSO}} = 10^{\mathrm{CNR}_{\mathrm{CSO}}/10}$ and $cnr_{\mathrm{CTB}} = 10^{\mathrm{CNR}_{\mathrm{CTB}}/10}$ are the FM-CSO and FM-CTB induced CNR contributions (in linear scale), respectively. Since m_o^2 is the (normalized) carrier power, then these expressions represent the (normalized) noise power contributions due to FM-CSO and FM-CTB respectively in the video channel bandwidth.

To determine these terms theoretically, one makes use of a beat counting or distortion modelling program and runs it at an arbitrary modulation index for both TV channels alone and FM+TV channels. The two distortions distributions are subtracted from each other to find the contribution of the FM terms alone (this is done for CSO and CTB separately).

Let $\mathrm{ctb}_{\mathrm{TV}}(\omega_i)$, $\mathrm{ctb}_{\mathrm{FM+TV}}(\omega_i)$ be the two (linear scale) CTB power profiles for frequency plans involving TV channels and FM+TV channels, respectively. For a given channel frequency ω_i, Eq. 7 represents the "FM alone" CTB contributions, actually comprising all beats which involve at least one FM channel, e.g. in the CTB case, triple beat products involving one or two or three FM channels and two or one or no TV channels respectively.

$$\mathrm{ctb}_{\mathrm{FM}}(\omega_i) = \mathrm{ctb}_{\mathrm{FM+TV}}(\omega_i) - \mathrm{ctb}_{\mathrm{TV}}(\omega_i) \quad (7)$$

The contribution of all triple products involving TV channels alone is deleted by the subtraction. A similar procedure is performed to find the "FM alone" CSO contributions. The net "FM alone" composite beat profiles are now integrated over the bandwidth of the observation channel, using a noise equivalent bandwidth B_{TV} of 5 MHz for PAL or 4 MHz for NTSC.

These procedures yield the total distortion power in the channel band. In order to calibrate the scale, the beat profile program distortion results for the TV channels alone are used in the following way (for definiteness the CTB case is discussed, a similar reasoning applies to the CSO as well):
Let $\mathrm{ctb}_{\mathrm{TV}}(\omega_o)$ be the (linear scale) CTB power as generated by the beats counting program at the reference channel frequency ω_o where the CTB performance is specified to be equal to $\mathrm{CTB}_{\mathrm{TV}}(m_o)$ [dB] at the nominal modulation index m_o. Then

$10\log_{10}\left(\frac{\mathrm{ctb}_{\mathrm{FM}}^{(\mathrm{ch})}}{\mathrm{ctb}_{\mathrm{TV}}(\omega_o)}\right)$ is a figure describing the level of the integrated FM-CTB relative to the TV-CTB. The integrated FM-CTB is therefore (in dBc):

$$-\mathrm{CNR}_{\mathrm{CTB}} = \mathrm{CTB}_{\mathrm{TV}}(m_o) + 10\log_{10}\left(\frac{\mathrm{ctb}_{\mathrm{FM}}^{(\mathrm{ch})}}{\mathrm{ctb}_{\mathrm{TV}}(\omega_o)}\right) \quad (8)$$

A similar expression applies to the CNR contribution due to CSO.

In the next section, we will use an example to demonstrate how to apply the analysis to a specific AM+FM channel plan.

EXAMPLE

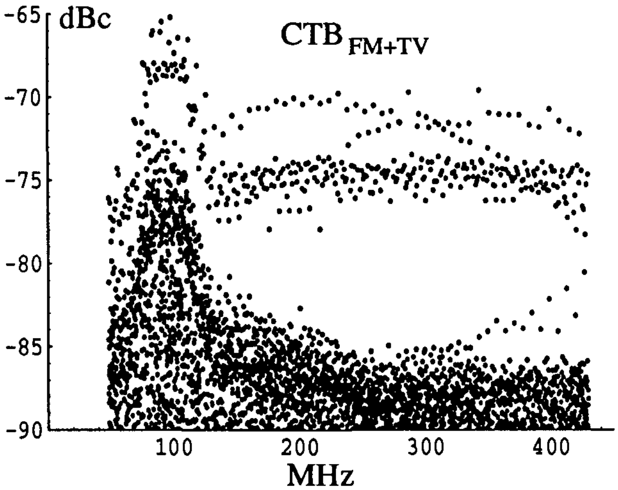


Fig. 1: Total induced CTB vs. frequency (35 PAL channels and 30 FM channels)

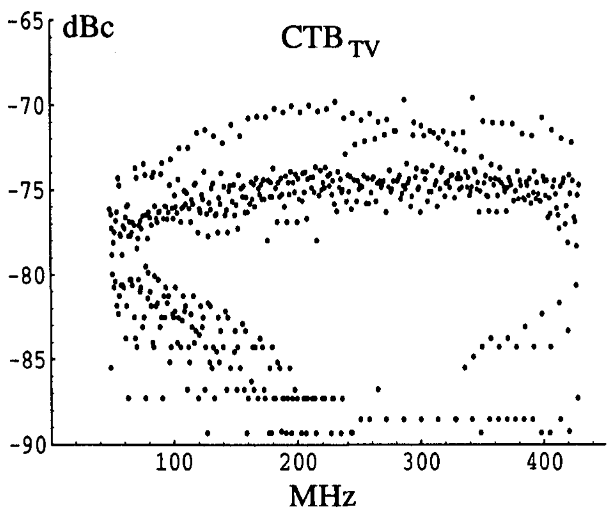


Fig. 2: CTB vs. frequency (35 PAL channels)

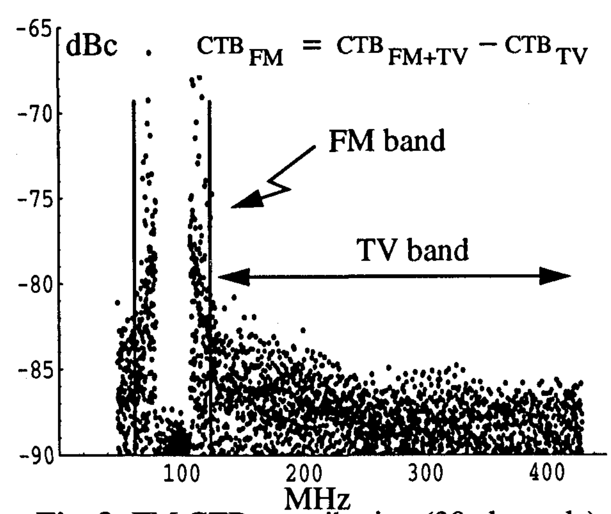


Fig. 3: FM CTB contribution (30 channels)

We illustrate the foregoing analysis by using an example. The TV plan chosen includes 35 PAL channels at 3.5% index modulation (m_{am}) each and 30 FM carriers (inserted between 80 and 108 MHz) with modulation index (m_{fm}) of 2.2% (4 dB below the AM). A beat counting program was used to generate the CSO and CTB profiles of the combined TV+FM channels and of the TV channels only. The FM channels contribution to either case (CSO and CTB) was obtained by calculating the difference between the two outcomes.

Fig. 1 through Fig. 3 depict the calculation results for the CTB case. First, the total induced CTB profile of both the FM and TV channels is calculated (Fig. 1), Fig. 2 depicts the CTB profile of the TV channels only and the difference between the two which yields the CTB contribution of the FM channels is shown in Fig. 3. The contribution of the FM channels to the CTB profile at the TV channels frequency band is clearly shown in the Figure.

The noise-like nature of the FM contribution can be seen in a close-up view of Fig. 3. To achieve that, an observation channel is arbitrarily chosen at mid-band at a frequency of 294.25 MHz. The CTB contribution of the FM channels over the 5 MHz filter bandwidth is shown in Fig. 4. Notice that the CTB contribution is scattered

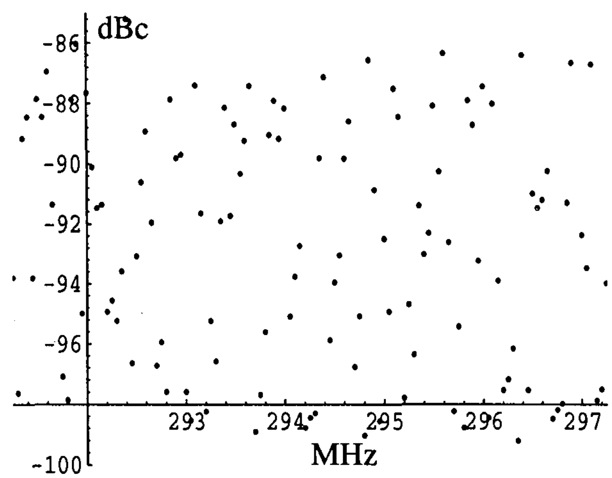


Fig. 4: CTB due to FM over the PAL observation channel (carrier at 294.25 MHz).

in random manner over the entire channel band at a level which is about -85 dBc. A similar noise-like behavior is observed for the CSO contribution of the FM channels.

The CNR equivalent degradation of the FM channels due to their CTB (CNR_{ctb}) is obtained by using Eq. 8. Integrating the CTB contribution over the channel filter (containing in this case 100 discrete points) and using a CTB_{TV}(294.25 MHz)=-71 dBc (see Fig. 2) we obtain (following similar procedure for CSO):

$$\begin{aligned} CNR_{ctb} &= 62.1\,\mathrm{dB} \\ CNR_{cso} &= 70.4\,\mathrm{dB} \end{aligned} \tag{9}$$

At the distortion levels shown above (CTB=-71, CSO=-70), the CNR_0 at PAL filter bandwidth of 5 MHz is expected to be 51.4 dBc.

To understand the significance of the above numbers let us plot the CNR expression of Eq. 6 using the numbers obtained in Eq. 9.

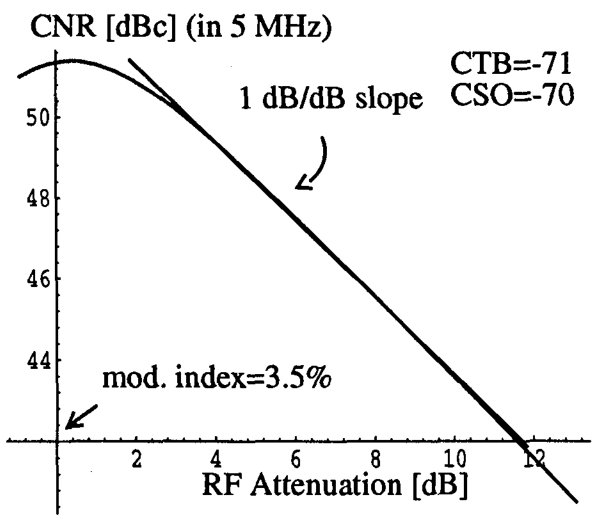


Fig. 5: CNR vs. RF attenuation including FM induced noise. Notice the deviation from straight line as modulation index increases.

As is shown in the Figure, the predicted degradation of the FM induced noise will cause a compression in the CNR performance with increase in modulation index. In contrast, the theoretical expression for CNR in Eq. 1 predicts that the CNR will follow the RF attenuation on a dB/dB basis. From Fig. 5 for example, the FM induced noise is expected to degrade the theoretical CNR performance ("FM less") by as much as 0.5 dB at a modulation index of 3.5%

COMPARISON WITH EXPERIMENT

To confirm the theoretical prediction of the FM noise contribution, an externally modulated YAG transmitter (Harmonic Lightwaves model HLT 6720) was loaded with the same 35 PAL channels and its CNR was measured as a function of the RF input pad with and without FM channels loading. The results and a comparison with the theoretical prediction shown in Fig. 5 are summarized in Fig. 6.

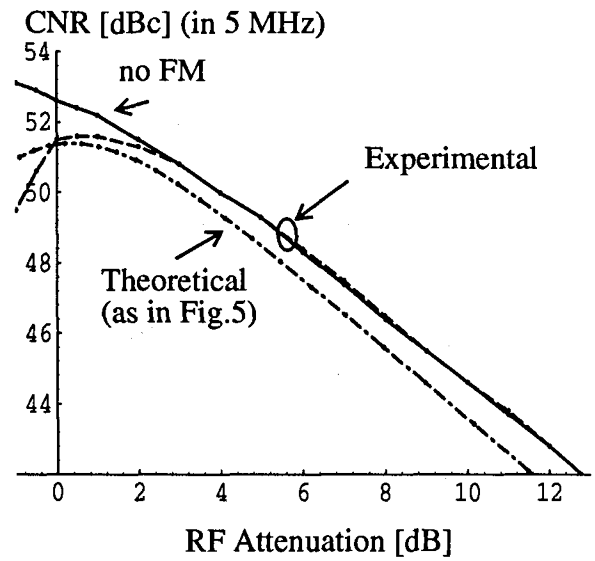


Fig. 6: A comparison between experimental and theoretical CNR vs. RF attenuation. Two experimental curves are shown, one including FM channels and the other "FM less".

The experimental results for the "FM less" case follow a straight line but the CNR curve starts bending in the region corresponding to modulation index of 4.5%. This seems to indicate clipping induced noise, a mechanism not taken into account in the present analysis. The predicted effect of FM loading on the CNR performance is clearly seen in the Figure by the deviation of the FM loading experimental curve from its "FM less" counterpart. There is an excellent agreement between the theoretical and experimental results over most of the RF range. The excessive compression of the FM loading experimental results evident at high modulation index (at attenuation levels less than 0 dB), can be attributed to higher then third order effects. Recall that the assumptions of the model presented here include only the effects of second and third order distortion, no higher orders or clipping distortion are taken into account.

CONCLUSION

Using non-linear (second and third order) analysis we have developed a model describing the effects of FM channel loading on a fiber-optic analog link. The FM channels loading is shown to degrade the CNR performance due to the non-linear mixing of the FM channels with themselves and with the AM channels resulting in an increase in the noise floor of the optical link. A corrected CNR equation taking into account the distortion induced FM noise is proposed.

To confirm the theoretical prediction a test which was comprised of 35 PAL channels and 30 FM channels was carried out and produced excellent agreement. As predicted, the CNR deviated from its 1 dB/dB slope due to FM channels loading as the optical modulation index increased. At very high modulation indexes the CNR was compressed even further than theoretically predicted indicating higher order non-linear contributions, contributions not included in the presented model.

HIGH DATA RATE VSB MODEM FOR CABLE APPLICATIONS INCLUDING HDTV: DESCRIPTION AND PERFORMANCE

Ronald B. Lee
Zenith Electronics Corporation
Glenview, IL. 60025

Abstract

The Digital HDTV Grand Alliance selected Vestigial Sideband Modulation over Quadrature Amplitude Modulation for their broadcast and cable modulation subsystem. This paper describes a 16-VSB modem which transmits 43 Mb/s over a 6-MHz channel. The VSB modem also has various modes of operation, trading robustness for data rate. The symbol rate, data rate, data frame, forward error correction (FEC), and synchronization pulses are all the same as those proposed and tested by the Grand-Alliance. Negligible additional hardware is required in future HDTV modems to receive and decode data transmitted by this modem.

INTRODUCTION

Based on comparative testing of digital modems for both broadcast and cable, the Digital HDTV Grand Alliance selected Vestigial Sideband Modulation (VSB) over Quadrature Amplitude Modulation (QAM) for their broadcast and cable modulation subsystem. Comparisons were made of an 8-VSB trellis coded modem to a 32-QAM trellis coded modem for broadcast applications and of a 16-VSB modem to a 256-QAM modem for cable applications. In those tests, the VSB modems out performed the QAM modems in both the broadcast mode and the cable mode. The VSB system was also subsequently accepted by the FCC Advisory Committee on Advanced Television Systems (ACATS).

This paper describes a 16-VSB cable modem, with various modes of operation, that is compatible with both the broadcast and cable modems proposed by the Grand-Alliance. Compatibility is increasing important especially in future applications where a common infrastructure is likely to exist for all digital services into the home. Digital televisions, computers, telephones, and other consumer equipment will accept data from this common infrastructure in order to support a variety of interoperable services into every home. Among these services are video-on-demand, video and audio telephony, home shopping and banking, computing, working at home, and information access.

There are two keys elements to the success of a common infrastructure for digital delivery into the home. The digital link into every home must be standardized, and the link must support a variety of flexible and reallocable services. The decision to use VSB for HDTV provides both of these essential elements. Not only will VSB be a prolific standard, existing in every high definition TV and able to receive digital signals from broadcast and cable, but it is also designed to accommodate flexible and reallocable services which will be necessary to share the data link between multiple applications. For maximum interoperability, it makes sense that all digital data delivery into the home also use VSB modems and thereby maintain compatibility with the HDTV's of the future. The high definition TV, with its display and VSB modem, will be the center for all kinds of home services and applications.

The 16-VSB modem transmits 43 Mb/s over a 6-MHz channel. The symbol rate, data rate, data frame, forward error correction (FEC), and synchronization pulses are all the same as those proposed and tested by the Grand-Alliance. Negligible additional hardware is required in future HDTV modems to receive and decode data transmitted by this modem. The data rate of the 16-VSB modem matches that required to transmit two HDTV programs over one 6 MHz cable channel.

HDTV BACKGROUND

In order to understand how this modem is compatible with the HDTV modem, a block diagram of the 8-VSB broadcast HDTV receiver tested by the Grand-Alliance is shown in Figure 1. For comparison purposes, a block diagram of the 16-VSB receiver is shown in Figure 2. By comparing the two diagrams, it is easy to see the similarities between the two; except for the Trellis decoder, the blocks are the same. Table I shows a detailed comparison between the functional differences of each block. In every case, the broadcast receiver is either identical, can be adapted with little or no change, or is a superset of the cable receiver.

16-VSB MODULATOR

A conceptual block diagram of the 16-VSB modulator is shown in Figure 3. A 16-level signal, representing 4-bits of information, is generated by applying the 4-bit input to a D/A converter. A new input appears at a 10.7 MHz rate and thus results in a transmission rate of:

10.7 MHz x 4 bits = 43 Mb/sec.

A small amount of DC is added to the output of the D/A to create a small amplitude pilot which is transmitted to aid the receiver's

BLOCK	**8-VSB Broadcast Receiver**	**16-VSB Cable Receiver (with 2, 4, & 8-VSB modes)**	**Comments**
Tuner	Tunes Broadcast and Cable Bands	Tunes Cable Bands	Cable ready broadcast tuner tunes to superset of channels
SAW Filter	5.38 MHz BW	5.38 MHz BW	Identical
IF Demodulator	Locks onto Pilot	Locks onto Pilot	Identical
A/D converter	8-Bits @ 10.7 MHz	8-Bits @ 10.7 MHz	Identical
Sync & Clock Recovery	Locks to Sync Signals	Locks to Sync Signals	Identical
Adaptive Equalizer	Feedforward + Feedback	Feedforward	Broadcast equalizer has more capability
Phase Tracker	Tracks 2 & 8 level signals	Tracks 2,4,8,&16 level signals	Broadcast receiver needs additional look-up table to receive cable modes
Trellis Decoder	Yes	None	Broadcast receiver can bypass to receive cable modes
De-interleaver	Yes	Yes	Broadcast receiver has more memory
R-S Decoder	(208,188) T=10	(208, 188) T=10	Identical

COMPARISON OF BROADCAST AND CABLE MODEM

TABLE I

carrier recovery. This signal is modulated to an IF band centered at 46.7 MHz and filtered to remove the upper sideband. The filter is a 11.5% roll-off root-Nyquist filter which leaves a small vestige of the upper sideband, hence the name Vestigial Sideband Modulation (VSB). The IF signal is subsequently modulated to a 6-MHz RF channel for transmission.

Because the transmitted data is random, the spectrum of the transmitted signal is flat throughout the band except at the band edge where the root-Nyquist filter rolls it off. A small pilot 310 kHz away from the lower band edge is also present. The pilot only contributes 0.3 dB to the signal's total power. A spectrum plot of the transmitted signal is shown in Figure 4. In this figure, the pilot is visible because the resolution bandwidth of the plot is only 100 kHz.

All data in the VSB system are sent in 188 byte blocks with 20 additional bytes of Reed-Solomon (T=10) parity for a total 208 byte block. Before transmission, the data blocks are placed into a fixed frame structure which aids in receiver synchronization. The fixed data frame is shown in Figure 5. A data frame consists of two data fields of 313 data segments each. Each data segment is made up of 836 symbols. Every data field begins with a one segment "Field Sync" signal which include several pseudo random sequences used by the receiver as a training reference for the adaptive equalizer. A 4 symbol segment sync is also transmitted at the beginning of every segment as part of the data frame.

The top of Figure 5 illustrates how blocks are packed into segments for the various VSB modes. For the 16-VSB mode, two blocks are sent per segment. In 8-VSB mode, 1½ blocks are sent per segment, in 4-VSB mode, one block is sent per segment and in 2-VSB mode, ½ block is sent per segment. Different VSB modes are simply transmitted by applying a different number of bits to the input of the D/A converter each symbol time. One bit for 2-VSB, two bits for 4-VSB, three bits for 8-VSB, and four bits for 16-VSB.

In order to protect against burst noise, data within the data frame are interleaved before transmission. The frame syncs and segment syncs are unaffected by interleaving.

VSB MODES

The different VSB modes supported by the modem enables the flexibility to trade-off data rate for robustness and may be exploited in various ways. The more robust modes may be used for poorer quality channels as might be found at the upper roll-off region of a cable system. Different VSB modes may also be used where low power transmission is desired. During equipment failures which cause a loss of carrier-to-noise ratio, the robust modes can be switched in to continue a level of service until the fault has been corrected. The VSB receiver automatically adjusts to the transmitted VSB mode.

Table 2 shows the trade-off for the various VSB modes between data rate and white noise performance.

VSB Mode	Xmitted Data (Mb/s)	Usable Data (Mb/s)	C/N Threshold
2-VSB	10.8	9.7	10 dB
4-VSB	21.5	19.3	16 dB
8-VSB	32.3	28.9	22 dB
16-VSB	43	38.6	28 dB

VSB Modes

TABLE 2

16-VSB RECEIVER

The following is a description of the blocks of the 16-VSB receiver shown in Figure 2 .

Tuner

The tuner for a cable modem will require better phase noise performance than standard NTSC tuners. The phase noise threshold of the 16-VSB modem has been measured by CableLabs at -82 dBc @20 kHz. For best performance, the tuner phase noise should be approximately 10 dB less than this.

SAW Filter

The IF filter is a SAW filter made of lithium niobate. Its response is a 11.5% roll-off root-Nyquist filter. In applications where the 16-VSB signal is run lower in power compared to adjacent NTSC signals, it is desirable to have at least 40-50 dB attenuation at the adjacent sound and picture carrier frequencies.

IF Demodulator

The IF demodulator recovers the carrier using the transmitted pilot. It consists of a frequency phase-locked loop (FPLL) which can pull-in at least 100 kHz. The PLL bandwidth is 2 kHz and can track out phase noise within this bandwidth.

The IF demodulator is one IC which includes IF and tuner AGC.

A/D Converter

The 16-VSB system uses an 8-bit A/D converter sampling at the symbol rate of 10.7 MHz.

Synchronization and Clock Recovery

Synchronization and clock recovery of the VSB system are very robust because they take advantage of the repeated segment sync signals. Synchronization and clock recovery are possible down to a 0 dB carrier-to-noise ratio, which is 10 dB beyond the 2-VSB threshold.

Adaptive Equalizer

In digital modems, the adaptive equalizer is one of the largest pieces of hardware in the receiver. The VSB adaptive equalizer is designed as a 63 tap feedforward equalizer using only real taps. In contrast to QAM systems which use two samples per symbol and require complex filters, this translates into a four to one reduction in complexity. In addition, the VSB equalizer adapts on a training sequence repeating at 24 Hz; this further simplifies the hardware complexity by only requiring relatively slow tap calculations. The VSB equalizer only uses one low speed multiplier for all the tap calculations.

Like the sync and clock recovery, the equalizer will operate down to a carrier-to-noise ratio of 0 dB.

Phase Tracker

The phase tracker is a first order decision feedback loop used to track out phase noise in the signal. The bandwidth of the phase tracker is approximately 60 kHz and tracks residual phase noise left by the IF PLL. The concatenated combination of IF PLL and phase tracker gives the VSB system excellent immunity to phase noise.

A fast gain and offset loop is also incorporated into the phase tracker to remove gain modulation and DC errors which might arise from various sources.

Deinterleaver

The VSB system uses interleaving at the transmitter and de-interleaving at the receiver to combat impulse noise. In order to minimize memory, a convolution structure is used. The 16-VSB system was measured at CableLabs to withstand 47 uS noise bursts repeating at a 10 Hz rate.

Reed-Solomon Decoder

The FEC used in the VSB system is a Reed-Solomon (R-S) T=10 code with a block size of (208,188). In comparison to trellis decoders whose complexity grow almost exponentially with the code length, the R-S decoder is very hardware efficient. The complexity of R-S decoders grow less than linearly with increasing T. With the R-S code, the 16-VSB modem achieves a white noise threshold of 28 dB.

Diagnostic Capability

A valuable feature of the VSB system is its ability to acquire even under extremely adverse conditions. The fact that the receiver can acquire and equalize the signal even under signal conditions where data is totally lost allows the receiver to supply diagnostic information to the cable operator or customer even after the system has failed. Information available from the VSB receiver include the received carrier-to-noise ratio and the condition of the equalizer taps.

This diagnostic information has proven useful in all the field tests of 16-VSB. All the proposed QAM cable modems cannot supply any information after data is lost.

PERFORMANCE

The performance of the 16-VSB system has been measured and publicly demonstrated on various occasions. Hardware measurements have shown that the 16-VSB system is able to achieve virtually theoretical performance. The use of a pilot and synchronization signals effectively eliminate implementation loss. Table 3 shows a performance summary of a 16-VSB system measured at CableLabs.

Besides laboratory measurements, the 16-VSB system has been field tested extensively. It has been tested on 3 cable systems including links with AML's, fiber, and up to 35 amplifiers. In all cases the 16-VSB system operated with margin and no plant modifications were necessary. The 16-VSB system has also been tested on hybrid fiber-coax links.

Parameter	Performance
Carrier-to-Noise	28 dB
CTB (CW)	43 dB
CSO (CW)	35 dB
Phase Noise Threshold	-82 dBc
Residual FM	4.9 kHz
Pull-in Range	+90 to -130 kHz
Burst Error @ 10 Hz Rep.	47 uS

16-VSB Performance Summary

TABLE 3

CONCLUSION

A 16-VSB modem which transmits 43 Mb/s has been demonstrated to operate over real cable systems. Laboratory measurements have shown that the modem achieves theoretical performance. The VSB modems have functional and complexity advantages over QAM systems and are compatible with the proposed Grand-Alliance VSB system. Future high definition TV's can demodulate the 16-VSB signal with minimal additional complexity.

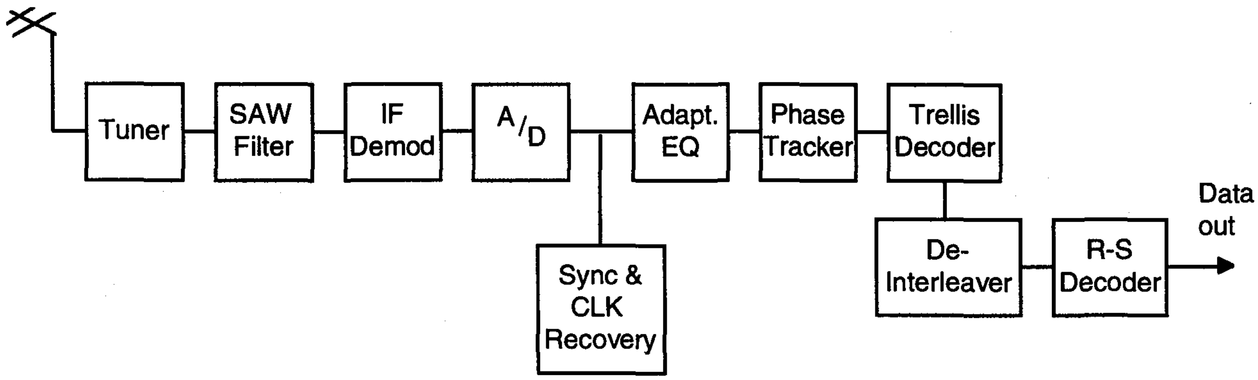
The complexity of the receiver is low enough that the FPLL is being integrated into one analog IC while the sync and clock recovery, adaptive filter, phase tracker, de-interleaver including memory, and R-S decoder are being integrated into one digital IC. VSB technology offers a high performance, low cost solution which will be compatible with HDTV.

REFERENCES

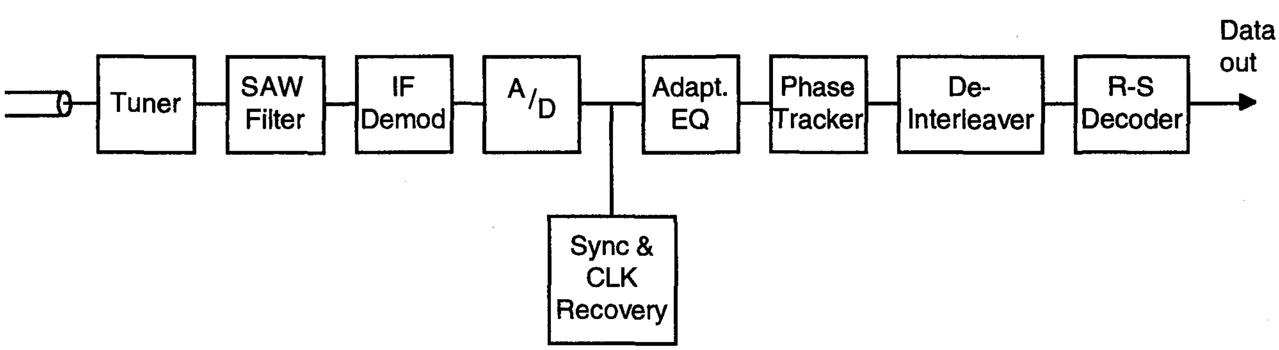
[1] "VSB TRANSMISSION SYSTEM, GRAND ALLIANCE, Feb. 18, 1994 TECHNICAL DETAILS" - Monograph submitted to the FCC Advisory Committee on Advanced Television Systems (ACATS).

[2] "PRACTICAL IMPLEMENTATION OF A 43 MBIT/SEC (8 BIT/HZ) DIGITAL MODEM FOR CABLE TELEVISION" by R. Citta and R. Lee, presented at NCTA '93 in San Francisco.

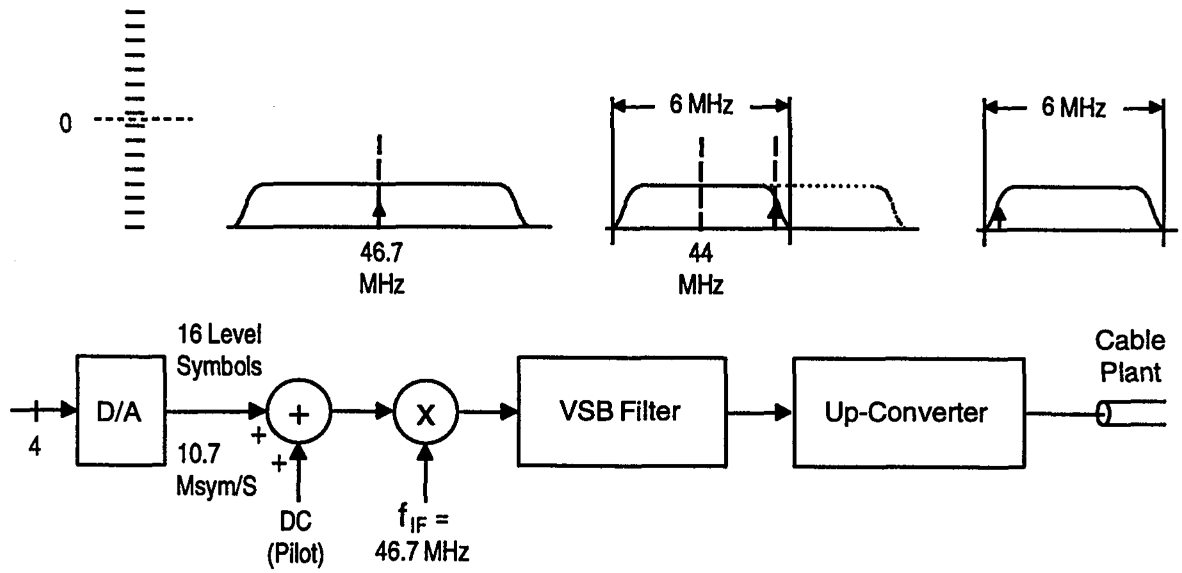
[3] "SUMMARY OF THE ZENITH 16-VSB MODEM FIELD TEST ON VIDEOTRON'S CABLE SYSTEM IN MONTREAL"; Technical Report by G. Sgrignoli, 6/21/93. Available from Zenith Electronics Corporation, Glenview, IL 60025.



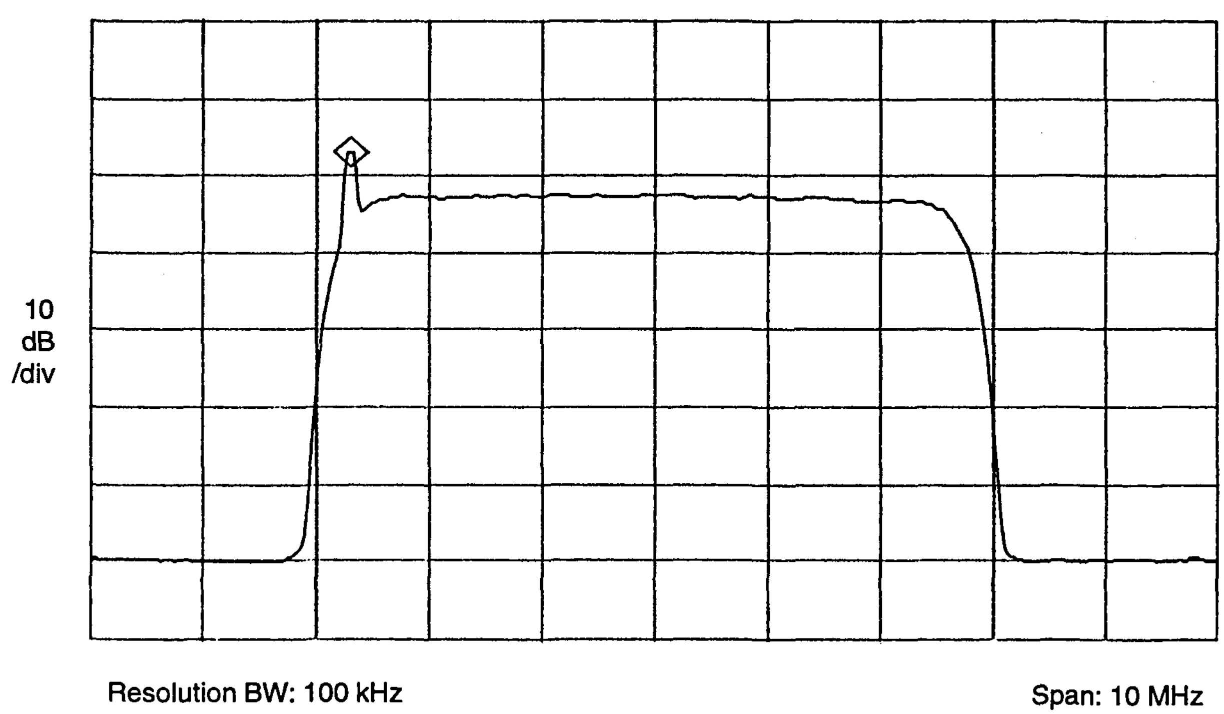
Grand Alliance 8-VSB Receiver
Figure 1



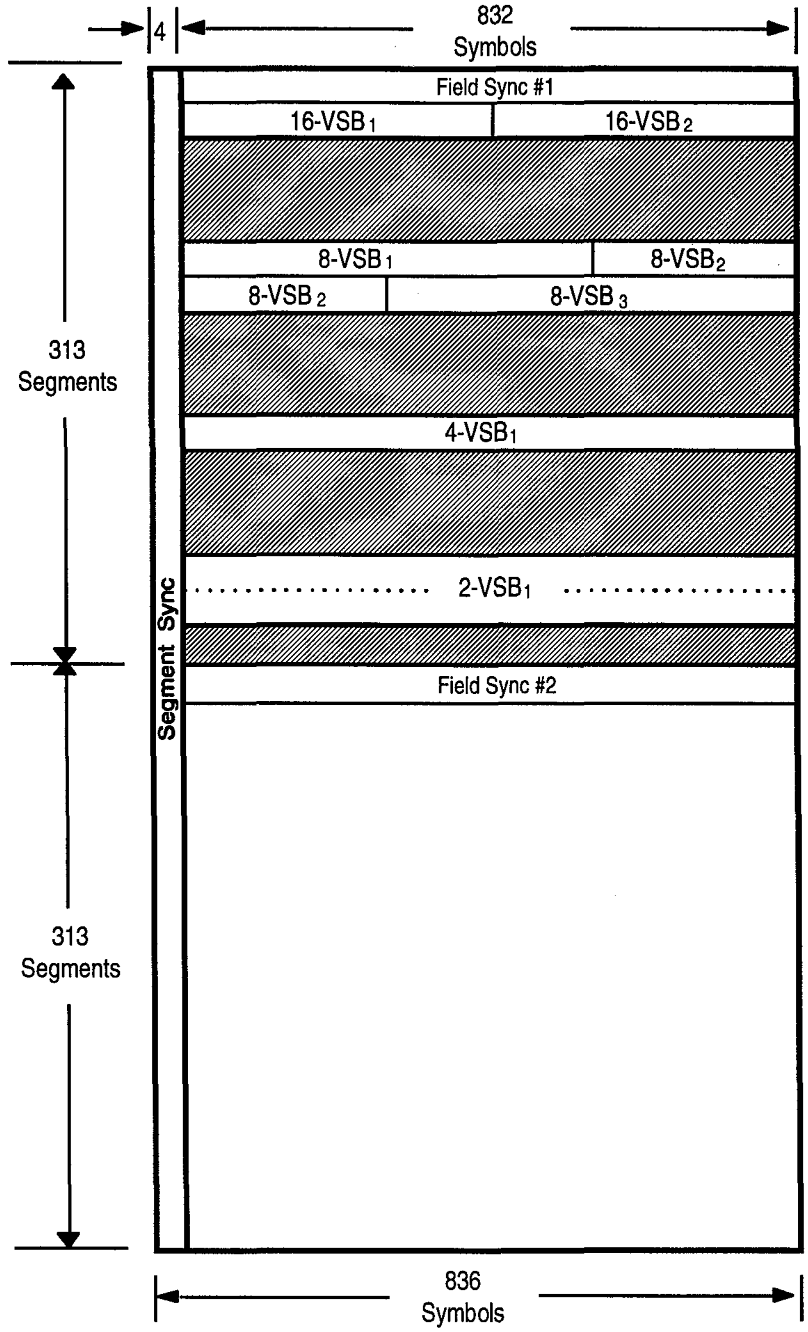
16-VSB Receiver
Figure 2



16-VSB Modulator
Figure 3



16-VSB Spectrum
Figure 4



1 Packet = 188 bytes + 20 bytes FEC

Variable Rate Data Frame
Figure 5

HIGH DEFINITION TELEVISION - DEFINING THE STANDARD

Brian James
Director, Advanced Television Testing
Cable Television Laboratories, Inc.
Alexandria, VA, USA

Abstract

In February, 1993 the Federal Communications Commission's Advisory Committee on Advanced Television Service recommended that the four digital ATV transmission systems have proposed improvements made and be re-tested to determine the best system for North America. The proponents were also encouraged to form an alliance to combine the best parts of the individual systems into one superior system. The four proponents subsequently formed the "Grand Alliance" to develop a single advanced television system that would incorporate the best parts of the various proposed systems. By late fall of 1993 all aspects of the system had been determined except which transmission system to use. The Alliance proposed that the Advisory Committee participate in the testing of the proposed transmission systems to minimize the time required to review the results and agree on the recommended standard.

The tests started in early January, 1994 and were completed in early February. The results of the tests showed the vestigial sideband implementation proposed by Zenith to be better, in most respects, than the QAM implementation proposed by General Instrument. The VSB system was recommended as the ATV standard with the QAM system retained as the backup. This paper reviews the results of the cable portion of the tests.

BACKGROUND

The Advisory Committee on Advanced Television Service was appointed by the Federal Communications Commission in 1987 oversee the development and testing of potential advanced television systems. Of the initial twenty plus proposals only four digital transmission system proposals made it through the first round of tests and were recommended for a second round of tests. The proponents were given time to incorporate proposed improvements with final tests originally scheduled to begin in the spring of 1993. At the same time the proponents were encouraged to try to come together with a single proposed system which would incorporate the best portions of the various systems and result in the best possible system.

The proponents announced the formation of the "Grand Alliance" on May 24, 1993, after the first system had arrived at the Advanced Television Test Center for re-testing but prior to the beginning of the tests. The Grand Alliance, consisting of: AT&T, the David Sarnoff Research Center, General Instrument, Massachusetts Institute of Technology, North American Philips, Thomson Consumer Electronics and Zenith Electronics, indicated that the system to be proposed would consist of the best parts of all the individual systems, selected to provide the best possible system.

The Advisory Committee accepted the proposal from the Alliance and gave it the time necessary to develop the final system. The Technical Subgroup of the Special Panel of the Advisory Committee was charged with overseeing the development of the final system. By the end of November, 1993 all but the transmission system had been specified. The Alliance had narrowed its proposals down to two modulation systems, the QAM system, originally proposed by GI/MIT, and the VSB system proposed by Zenith. Each of these systems had a basic, single ATV channel data rate and a second, high data rate option for cable which would carry data for two advanced television channels within one 6 MHz channel.

The Alliance initially planned to perform system selection tests privately and to have the results reviewed by the Advisory Committee. Later, the Alliance decided that it would speed the process to have the Advisory Committee participate in the testing. The tests were scheduled to begin at the Advanced Television Test Center in Alexandria, VA in January 1994. The equipment arrived in early January with a couple of changes to the approved transmission systems. Zenith withdrew their 4 and 6-VSB systems and replaced them with a single 8-VSB system which had a superior trellis coding implementation. The data capacity of the GI 256-QAM system was increased from 32.9 Mbps to 38.2 Mbps.

The tests began in early January, 1994 and were completed in just over a month. The cable portion of the tests were performed by CableLabs during the week of January 24th. The systems tested were the 32-QAM standard data rate and 256-QAM high data rate systems proposed by GI and the 8-VSB standard data rate and 16-VSB high data rate systems proposed by Zenith. Only the cable portion of the tests were performed on the high data rate systems since they are not intended for over-the-air use.

TEST PROCEDURE

Only transmission modems were to be tested, not complete HDTV systems, which made it impossible to observe interference levels by looking at the picture. This problem was overcome by estimating the bit error ratio (BER) which would result in errors first becoming visible in the picture. That error ratio, 3 x 10-6, was then defined as the threshold of visibility (TOV). The impairment level for TOV was the highest level of impairment which would result in a BER equal to or better than 3×10^{-6}. This was determined by increasing the level of the impairment until the BER was worse than 3×10^{-6} then reducing the impairment in small increments until the BER was at or just better than 3 x 10-6. The BER was obtained by observing errors for 20 seconds and calculating the BER. The BER had to remain below the threshold value for three consecutive 20 second periods to be classified as the threshold.

Four of the tests were repeated as multiple impairment tests in which the main impairment, CTB, CSO, phase noise or residual FM, was reduced in level below threshold and Gaussian noise was added to the signal until a new threshold was found. This test showed the ability of the system to operate in the presence of multiple impairments and gave an indication of the trade-off that took place between Gaussian noise and the main impairment under test.

TEST RESULTS

Signal-to-Noise (S/N)

Signal-to-Noise is the basic test to determine the threshold level of digital systems operating in the presence of thermal (Gaussian) noise. It is a repeatable test and it is one of the main constraints in designing cable television systems. In the digital ATV systems the signal power is specified as the average power of the signal within the 6 MHz band and the Gaussian noise is measured in the same 6 MHz band. This is slightly different from the NTSC measurement practice where the noise bandwidth is only 4 MHz.

The low data rate systems (32-QAM and 8-VSB) produced equal threshold S/N values of 14.8 dB. The high data rate systems showed some differences with the 256-QAM having a threshold value of 29.3 dB while the 16-VSB had a somewhat better threshold value of 27.6 dB.

Composite Triple Beat (CTB)

A second major area of interest for cable operators is composite triple beat, a distortion created each time the cable signals are amplified. The CTB product generated by the NTSC visual carriers on a cable system falls 1.25 MHz above the lower band edge, i.e. at what would normally be the location of the NTSC visual carrier. Each of the digital modulation systems reacted differently to CTB products falling at this location.

The 8-VSB system, with a threshold of 12.6 dB, performed much better than the 32-QAM system, which had a threshold of 32.0 dB. The superior performance of the VSB system was due, in part, to the presence of a comb filter which is turned on when the receiver detects co-channel interference or CTB products. CTB products fall at the same frequency as a co-channel visual carrier resulting in the receiver interpreting the CTB interference as a co-channel interference and turning on the comb filter. There is a noise tolerance penalty when the comb filter is turned on, therefore, the filter is used only when co-channel is present at a sufficiently high level.

The 16-VSB modem, with a C/I threshold of 44.0 dB, performed better than the 256-QAM system which had a C/I of 46.5 dB. The comb filter is not present in the 16-VSB modem as it is intended for cable system use only and should not need co-channel protection.

A multiple-impairment test was performed to determine the trade-off between CTB and Gaussian noise. It was expected that as the amount of CTB was reduced the receiver tolerance to

Gaussian noise would increase. The CTB interference level was reduced by 1, 2, 3, 6, 9, and then 12 dB below the threshold level and Gaussian noise was adjusted until the receiver threshold was again reached and the S/N ratio at this threshold was measured. See Table 1.

The results for the 32-QAM, 256-QAM and 16-VSB modems were as expected with the noise tolerance increasing as the CTB interference was reduced. The 8-VSB system acted opposite to the expected manner and became less tolerant of noise as the CTB level was reduced. This was attributed to the comb filter having trouble deciding whether to switch in or out due to the large changes in the level of the CTB interference as contributing carriers came in and out of phase. In production receivers there may be a switch to turn off the comb filter and the selection algorithm may be better optimized.

Table 1
CTB vs. Gaussian Noise

	32 QAM	8 VSB	256 QAM	16 VSB
CTB TOV C/I	32.0 dB	12.6 dB	46.5 dB	44.0 dB
S/N @ CTB TOV-1 dB	—	—	32.9 dB	28.0 dB
S/N @ CTB TOV-2 dB	—	—	31.7 dB	27.9 dB
S/N @ CTB TOV-3 dB	16.8 dB	19.3 dB	30.7 dB	27.8 dB
S/N @ CTB TOV-6 dB	15.5 dB	19.0 dB	29.6 dB	27.7 dB
S/N @ CTB TOV-9 dB	15.2 dB	22.8 dB	29.4 dB	27.7 dB
S/N @ CTB TOV-12 dB	15.2 dB	25.9 dB	29.4 dB	—
S/N with no CTB	14.8 dB	14.8 dB	29.3 dB	27.6 dB

Composite Second Order (CSO)

Push-pull amplifiers were introduced to reduce the level of second order beats when cable systems expanded beyond 12 channels but the introduction of AM fibre links has returned second order distortion to an interference which must be considered in system design. The second order distortion products fall 1.25 MHz above the NTSC visual carrier or 2.5 MHz above the lower band edge.

The 32-QAM modem performed very well in the presence of CSO interference with a C/I of 10.6 dB while the 8-VSB reached threshold at 28.5 dB. Among the high data rate systems, the 16-VSB modem had the better performance with a threshold C/I of 33.4 dB while the 256-QAM high data rate modem threshold was reached at 37.0 dB

The multiple impairment test was conducted by reducing the CSO level and increasing noise until a new threshold was found with both noise and CSO present. All modems behaved as one would expect with noise tolerance increasing as CSO was reduced. The full results are shown in Table 2.

Table 2
CSO vs. Gaussian Noise

	32 QAM	8 VSB	256 QAM	16 VSB
CSO TOV C/I	10.6 dB	28.5 dB	37.0 dB	33.4 dB
S/N @ CSO TOV-1 dB	22.0 dB	15.5 dB	35.8 dB	33.7 dB
S/N @ CSO TOV-2 dB	19.6 dB	15.2 dB	32.8 dB	31.7 dB
S/N @ CSO TOV-3 dB	18.4 dB	15.2 dB	31.8 dB	30.3 dB
S/N @ CSO TOV-6 dB	18.1 dB	14.8 dB	30.4 dB	28.6 dB
S/N @ CSO TOV-9 dB	18.0 dB	14.7 dB	29.7 dB	28.2 dB
S/N @ CSO TOV-12 dB	17.2 dB	14.7 dB	29.6 dB	27.8 dB
S/N with no CSO	14.8 dB	14.8 dB	29.3 dB	27.6 dB

Phase Noise

The presence of phase instability on the local oscillators used in channel conversion will introduce phase noise into the signal. Synthesizers are used in many modulators and heterodyne processors to create the local oscillator signals. NTSC signals are very tolerant of phase noise and can tolerate the use of local oscillators with relaxed specifications for the phase noise. Unfortunately, digital modulation systems are less tolerant of phase noise and it will be a constraint in equipment designs. Phase noise level is measured 20 kHz from the carrier in a 1 Hz bandwidth and referenced to the level of the carrier on which it is measured.

The 8-VSB modem was more tolerant of phase noise than the 32-QAM modem with a C/I of 77.1 dB compared to an 81.3 dB C/I. The 16-VSB modem performed slightly better than the 256-QAM system with a C/I of 83.0 dB compared to an 84.2 dB C/I.

A multiple impairment, phase noise vs. Gaussian noise trade-off test was performed. All systems displayed a normal reaction by increasing their tolerance for Gaussian noise as the phase noise level was reduced. Full results are given in Table 3. The 32-QAM, 8-VSB and 256-QAM systems were within one dB of the random noise threshold when the phase noise was reduced 6 dB below threshold level while the 16-VSB modem was within one dB with only a 1 dB reduction in phase noise.

Table 3
Phase noise vs. Gaussian Noise

	32 QAM	8 VSB	256 QAM	16 VSB
Phase Noise TOV C/I	81.3 dB	77.1 dB	84.2 dB	83.0 dB
S/N @ Ø Noise TOV-1 dB	17.4 dB	22.9 dB	34.5 dB	28.4 dB
S/N @ Ø Noise TOV-2 dB	16.3 dB	21.1 dB	31.8 dB	28.4 dB
S/N @ Ø Noise TOV-3 dB	15.9 dB	19.0 dB	30.6 dB	28.3 dB
S/N @ Ø Noise TOV-6 dB	14.8 dB	15.8 dB	29.7 dB	28.2 dB
S/N @ Ø Noise TOV-9 dB	—	15.1 dB	29.6 dB	28.1 dB
S/N @ Ø Noise TOV-12 dB	—	14.9 dB	29.4 dB	28.1 dB
S/N with no Ø Noise	14.8 dB	14.8 dB	29.3 dB	27.6 dB

Residual FM

A very small amount of ripple on the DC power supply feeding the oscillator can introduce residual FM into local oscillators. While the NTSC signal is very tolerant of residual FM, digital signals are much less tolerant. Tight design specifications for residual FM may be necessary on oscillators designed for use with digital signals.

The threshold levels of residual FM were similar for the two low data rate modems. The 32-QAM system tolerated an 8.4 kHz peak FM signal while the 8-VSB modem performed slightly better with a threshold of 8.8 kHz. The 256-QAM modem was capable of handling an impressive 70 kHz peak residual FM while the 16-VSB modem could only tolerate a 4.7 kHz signal.

The modems were tested to determine the trade-off between random noise and residual FM. The 32-QAM modem tolerance for noise increased as the residual FM was decreased but the 8-VSB tolerance remained constant as the residual FM was decreased. Both of the high data rate modems were within 1 dB of the noise threshold when the residual FM was reduced to 50% of the threshold value.

Table 4
Residual FM vs. Gaussian Noise

	32 QAM	8 VSB	256 QAM	16 VSB
Res. FM TOV	8.4 kHz	8.8 kHz	70 kHz	4.7 kHz
S/N @ 0.75 Res. FM TOV	23.0 dB	21.1 dB	34.2 dB	34.5 dB
S/N @ 0.5 Res. FM TOV	19.1 dB	21.0 dB	30.9 dB	28.5 dB
S/N @ 0.25 Res. FM TOV	16.1 dB	21.0 dB	30.0 dB	28.2 dB
S/N with no Res. FM	14.8 dB	14.8 dB	29.3 dB	27.6 dB

Pull-In Range

During the life of a receiver there can be some drift in the local oscillator frequency which the receiver must be able to overcome to tune a desired channel. In addition, some channels are offset in frequency to improve interference performance or to meet federal regulations. The modems were tested to determine their ability to tune a signal that was offset from nominal frequency allocation. A value of 100 kHz was determined to be a reasonable offset which the modems should be capable of tuning and was selected as the maximum pull-in to be reported.

The 8-VSB, 256-QAM and 16-VSB modems were capable of tuning at least a ±100 kHz offset. The 32-QAM modem could only tune a +74 kHz and -80 kHz offset.

Hum Modulation

As signals move down the cable system they are amplified a number of times. If the DC power supplies are not properly regulated some power line frequency amplitude modulation of the signal can occur. With NTSC signals, this interference becomes visible as bars moving up the picture when the modulation reaches about 3%. In the presence of high hum modulation the digital pictures may exhibit block errors or the picture may freeze.

All of the modems were capable of operating with hum modulation greater than 3%. The 32-QAM continued to operate at 15.2 % modulation, the highest amount available on the test bed. The 8-VSB and 16-VSB were equivalent at 7.7% and 7.6% respectively. The 256-QAM modem reached threshold with the modulation at 5.7%.

Summation Sweep

Cable operators commonly use some form of summation sweep to determine and adjust the frequency response of the systems. High level summation sweep systems were expected to cause problems with the digital signal as data is lost when the sweep signal passes through the channel. The test was performed using a Wavetek sweep set 10 dB above the digital signal and programmed to spend about 0.2 msec in each 6 MHz channel. The amount of interleaving used in the encoder helps determine if the amount of data lost exceeds the threshold value when the sweep signal passes through the channel.

The two QAM systems both had error ratios in excess of the threshold levels which would result in errors being visible in the picture. The error ratios of the two VSB systems were below the threshold level and no errors would be visible in the picture. All of the modems showed zero errors with the Calan type low level sweep system.

Channel Change Time

The common usage of remote control devices for channel selection and the large number of TV channels available on cable has resulted in the sport of channel surfing where viewers rapidly tune through channels on the system to find one(s) they desire to watch. The picture must be displayed very shortly after channel selection or the subscriber is likely to become frustrated. The digital receivers are more elaborate than analogue receivers and, it is expected, will take longer to lock up and begin to deliver data. In addition, the MPEG decoder is expected to require about half a second to produce a picture after it begins receiving data, therefore, the time required for the receiver to lock up and begin delivering data must be kept to a minimum.

The 32-QAM receiver required 1.1 seconds to begin delivering data while the other three receivers required just over half a second.

Peak-to-Average Power Ratio

Signal power of a digital signal is normally specified as the average power measured within the 6 MHz channel while NTSC power is determined by measuring the power during the synchronization pulse. The digital signal, depending on the input data, will at times exceed the average power by some amount. The systems were measured to determine maximum peak-to-average ratio that could be expected 99.9% of the time. The peak signal levels determine the worst case distortion created by the digital signals and must be considered when determining operating levels and designing cable systems.

The 32-QAM signal peak-to-average ratio was 5.8 dB, the 8-VSB and 256-QAM peak-to-average ratio was 6.4 dB and the 16-VSB was 6.5 dB.

Burst Error

Digital systems use various techniques to minimize the impact of loss of data caused by ignition noise, intermittent power line noise, loose connections, etc. The implementation used in a system determines whether a receiver will tolerate a very brief interruption or one with a much longer period. The systems were tested by introducing high level noise for increasing periods of time at a 10 Hz rate until the threshold error ratio was obtained. A second test was then performed with the duration of the noise burst held at 20 µsec while the rate was increased until threshold was reached.

The 8-VSB system performed better than the 32-QAM system and the 16-VSB system worked better than the 256-QAM system. See Table 5 for complete results.

Table 5
Burst Noise Performance

	32 QAM	8 VSB	256 QAM	16 VSB
Burst length (µsec) @ rep rate (Hz)	60 @ 10	190 @ 10	27 @ 10	150 @ 10
Burst length (µsec) @ rep rate (kHz)	20 @ 1.5	20 @ 1.6	20 @ 0.03	20 @ 2.4

Data Rate

Both the QAM and the VSB systems offered a high data rate cable option. This option is intended to allow two ATV signals to be carried within one 6 MHz channel.

The data rate of the 32-QAM modem was 19.2 Mbps while the high data rate 256-QAM modem data rate was 38.2 Mbps. The data rate of the 8-VSB modem was 18.8 Mbps while the 16-VSB modem rate was 37.5 Mbps. Both high data rate modems accepted two standard data rate inputs and provided two standard data rate outputs.

CONCLUSIONS

The VSB systems performed better than the QAM signals in the majority of the tests. The theoretical performance of the two types of modulation formats is very close and details of each system's implementation determined which system operated better. The VSB implementation was better than the QAM implementation in the majority of the tests.

The FCC Advisory Committee's Technical Subgroup accepted the findings in the cable and broadcast lab tests and the Grand Alliance's own internal recommendation, which favored VSB modulation, and selected the VSB system as the transmission system for Advanced TV. The 8-VSB system will be used for terrestrial broadcast and cable distribution while the 16-VSB system will be used for cable transmission where the double data rate is deemed desirable or necessary.

The next steps in the selection of the advanced ATV system are to field test the VSB modems, then to build a complete encoder, modulator, receiver, and decoder and perform the appropriate lab and field tests on the complete system. As of the time of this writing, the modem field tests were scheduled to begin in April 1994 and the complete system lab tests are expected to take place in the fall of 1994 . Complete system field tests are to take place in the spring of 1995 with the final recommendation for the system occurring shortly after the field tests.

CableLabs will conduct the cable portion of each of the next three phases of ATV system testing, and will participate in the subsequent standard documentation process.

HOW TO TEST COMPRESSED DIGITAL TELEVISION

Guy W. Beakley, Ph.D.
StellaCom, Inc.
Arlington, VA 22209

Abstract

Conventional analog video test measurements are not adequate for compressed video signals. This is because digital video distortion and artifacts often are nonlinear, discontinuous and depend on picture content. Most analog measurements assume that errors are of a continuous, linear nature. The only alternative to date has been subjective testing. Formal subjective tests (e.g., CCIR 500) can provide reliable, relative measures of video quality. However, such testing is time-consuming and expensive. Objective testing methods are needed to provide efficient, repeatable measures of video quality. Presently, no standardized objective measures exist for digital video. We have implemented measures from a number of sources and others of our own design on a low-cost workstation. These measures utilize complex digital image processing techniques to analyze differences between source and processed video sequences. This paper presents some of these measurements and describes our implementation of an automated system to capture and test digital video quality.

I. Why Compress Television?

Television is rendered in digital form for the inherent quality and lack of degradation in replication and transmission. It is compressed so that it can easily be transmitted and stored. For example, the Armed Forces Radio and Television Service (AFRTS) intends to procure and operate a compression system for the delivery of six video channels with associated stereo audio, three to six additional stereo audio channels, 1.5 Mbps T-1 data and two FM talk radio channels via 36 Mhz satellite transponders to AFRTS downlink locations worldwide. Before compression they transmitted only one video and four audio channels in a transponder.

II. Why Test Video Quality of Digital Television Systems?

The limitations of standard video testing methods for the evaluation of codecs have been demonstrated so frequently that an alternative is often employed. The codec evaluator simply looks at the video using a favorite test tape and makes a judgment based on experience and intuition. For this to be effective the test tape has to be extensive and contain a number of scenes that may be difficult for the codec to process. In addition, the codec evaluator must have the correct viewing conditions. The only alternative has been to assemble a panel of viewers and run subjective tests under controlled conditions.

Objective measures, which are repeatable and do not depend on viewing conditions or the mood of the viewer, need to be used in addition to subjective testing to compare compression systems. Codecs not only remove redundant information (lossless compression), but also modify the picture (lossy compression). Unfortunately, the picture distortion may result in perceptible impairments. Objective measures aid the codec evaluator in identifying and understanding codec artifacts. The complex processing of today's codecs requires complex testing that can be done only with a computer[1]. Also, the computer allows flexibility for objective test evolution as standards evolve for testing codecs.

Computer tests[1] can be done for spatial and temporal distortion, horizontal, vertical, and diagonal resolution, image build-up, bit-error-rate and various detail and motion impairments. These tests are based upon work done by NASA, NTIA, ACATS, CRC and others to quantify the quality of digital video codecs. The computer is used to predict how the average viewer rates a sequence on the CCIR-500 Impairment Quality Rating scale. The measures have been correlated with subjective testing by the NTIA. In addition, an absolute error signal is calculated, so that the computer user can see what has actually been lost or gained by codec processing. The computer allows the user to view the original video, codec processed video and error signal side-by-side with the error measures. These measures are also useful for analyzing bit-error and concatenation effects. In fact, the impairment quality rating measure has been correlated with subjective tests that include impairments due to bit errors and concatenation.

While conventional analog video measurements are needed, they prove little about a codec's performance with complicated motion. Dynamic tests are essential because most high-performance codecs use motion compensation as part of the compression algorithm. The need to measure resolution in several dimensions is new to the testing of digital codecs. The standards for such tests are not yet established. The use of moving zone plates to measure resolution is successful in revealing information not obtained from any other test. Other measures not widely known, such as spatial distortion and temporal distortion, are also necessary for describing a codec's performance.

The only way to completely test a codec is to play a wide range of real video materials through it. Test materials must be selected to represent the full range of video content for which a particular codec (or class of codecs) is intended. Test signals and related measurements can still serve a purpose, but their utility is more limited than with analog systems for the reasons outlined above. The tester needs to compile an extensive test tape consisting of many short video sequences, play the video through the compression system, and record the video output. Then, the output and input video can be compared, both visually and by computer analysis.

A combination of subjective and objective test methods is the most effective way to test codecs. Subjective methods are most useful for rapid, qualitative assessment of video quality. Formal subjective testing[2] is useful for quantitative quality assessment but is often impractical due to manpower, cost and time considerations. Objective testing with computer-based measurements can provide repeatable, quantitative results for comparisons of video quality. A suite of comparative distortion measurements is needed to fully characterize video quality. Statistics can be compiled automatically for large amounts of video and the computer can be programmed to identify poor-quality sequences for further analysis (subjective and objective) and for documentation. Moreover, the objective methods described below provide additional insights into codec operation and impairments that cannot be easily discerned using subjective methods.

An important benefit of testing with real video is that common source materials can be used for both subjective and objective testing. Many codec developers and end users already have their own *killer* test tapes for codec evaluation. The Ad-hoc Group on MPEG-2 Verification Testing[3] has assembled a set of six test sequences of 525-line/60-fields/sec video plus a few 24-frames/sec film-based sequences for subjective quality testing. The National Telecommunications and Information Administration, Institute for Telecommunication Sciences (NTIA/ITS) uses a set of 36 sequences[4] for their work with the Interexchange Carriers Committee on video quality (T1A1.5). Standard libraries exist at the CCIR as well. All such video materials can be used directly for objective quality measurements as described below.

There are many requirements for testing the quality of compressed digital video. The codec manufacturer's objective is to produce the best picture quality at a minimum encoded data rate. End users need to compare the video quality of competing codecs for purchase decisions. In addition there are a number of reasons for continual testing. One reason is identification and documentation of picture artifacts. Video quality varies with scene content. Users will want to document troublesome scenes and feed these back to the codec manufacturer. Video quality also varies with the use of a statistical multiplexer. A low-priority channel may not have the bit-rate it needs for good picture quality if the high-priority channels are carrying sports programs. What happens if the network degrades because of attenuation of carrier power? What happens when the packets are grossly out of order? Also, compression can be used at the signal origination or at the destination in addition to the communications or storage system. Then concatenation effects become important and should be tested. During the early stages of compression the vendors will make encoder improvements and there will constantly be new codec vendors with "better" products. These claims will need to be verified.

It is commonly believed that MPEG-2[5] will solve most of the users' video quality issues. However, MPEG-2 is a flexible syntax or tool kit allowing over thirty variables that affect picture quality. The use of MPEG does not guarantee picture quality. Examples of MPEG variables affecting picture quality include data-rate, use of I-, P- and B-frames, allocation of bits to I-, B- and P-frames, field/frame adaptive prediction, method and range of motion estimation/compensation, slice size, type and scale of quantizer, downloadable quantizer matrices and buffer size. Also, many different error detection, correction and concealment strategies can be employed in an MPEG decoder. Furthermore, filtering and resampling can occur before and after the encoding/decoding process. Variability also exists in the quality of the NTSC encoding/decoding process. MPEG-2 provides a flexible framework for allowing varying picture quality depending on communications and storage requirements and the intended application.

2.2. Why analog tests are not sufficient

Codecs employ complex algorithms that incorporate discrete spatial and temporal processing. They exhibit non-linear responses to changing scene content and operating conditions. Typically, codecs introduce distortion into the picture. Also, artifacts can be introduced in any of the encoding, transmission and decoding stages of a digital system. Analog measurements are not intended to characterize these types of systems and phenomena. Most analog tests employ predefined test signals and are intended to measure continuous, linear response characteristics of analog devices. For digital systems, the response to a particular test signal only describes the performance of the system for the test signal and often has little or no relation to how the system will perform on real video. For example, the commonly used method for determining signal-to-noise-ratio (SNR) employs a static, flat gray input signal. Most DCT codecs can produce very high SNR of 60 dB or higher on static scenes -- but SNR on real scenes depends on internal quantization levels for the DCT coefficients as well as motion prediction and other coding techniques. Many codecs tested at StellaCom have exhibited significant variation in SNR depending on the pedestal level used to make the SNR measurement. A measurement that varies so much with gray level is not useful for measuring the quality of codecs. In addition, typical noise and distortion for real scenes are higher than

SNR measurements imply and can vary widely by scene and by codec. Few conventional analog video measurements are useful for testing codecs because they prove little about a codec's performance with complicated motion. Dynamic tests are essential because most high-performance codecs use motion compensation as part of the compression algorithm.

Our testing system incorporates a suite of distortion measures and viewing tools on a graphics workstation. The Impairment Quality Rating (*IQR*) measurement developed by Wolf and others at NTIA/ITS[6,7,8] is used to predict how an average viewer rates a sequence on the CCIR-500 impairment quality scale. The *IQR* measurement has been correlated with subjective testing using statistical methods. In addition, a digital error signal is computed. It can be viewed as video and used for statistical measurements such as RMS Error and Signal-to-Error Ratio (*SER*). The testing system allows the user to view video (source, codec-processed video and error signal) side-by-side with plots of measurement results on the computer display. These and other measures are described below. An important consideration is that the measures be as general as possible. The approach is to treat the encoder, transmission system, storage system and decoder as a single black box system that can introduce both analog and digital distortions. Analog distortions are typically introduced in the processes of analog-to-digital and digital-to-analog conversion, prefiltering, and NTSC encoding and decoding. The encoder can introduce analog and digital distortions as well as encoding artifacts. Transmission errors or *bit errors* for digital systems can also cause artifacts. Error correction and concealment processing in the decoder will ultimately determine the output video quality in the presence of such bit errors. Concatenation effects can be tested by analyzing the recorded output of multiple passes through the system.

III. <u>Computer Measurements</u>

A. Video Content Measures

Video content measures form a foundation for testing with real video. Their primary function is for classification of source video materials. Then, the overall performance of a compression system can be related to different classifications of source materials. This is important because codecs' video quality can vary dramatically depending on spatial detail, motion, colorimetry, and other content factors

Many measures can be useful for characterizing scenes. We use measures of spatial and temporal information content for luminance and chrominance. Measures designed to detect the presence of specific features may also be useful for testing purposes. Features such as scene cuts, frame repeats, 3/2 pulldown, and others can be used to organize video materials and testing results. Scene cuts are particularly important because codecs with interframe processing and motion prediction often exhibit increased distortion after a cut. Glenn[9] shows that the presence of scene cuts can mask visual sensitivity to spatial detail and motion immediately before and after the cut. Ideally, any distortion measure designed to predict perceptible image quality would need to test for the presence of scene cuts in the source video. Wolf et.al.[6] have recently proposed a distortion measurement that incorporates scene cut masking effects using a threshold on the frame-to-frame increase of a temporal content measure to indicate the presence of a cut. Other researchers have developed scene cut detection schemes based on colorimetry statistics[10,11].

B. Test-Signal Measurements

The main purpose for test signals in this testing methodology is to determine the magnitude of those effects that are invariant to scene content and operating conditions. Invariant effects such as gain and level distortion, spatial filtering, spatial misregistration, and time delay may be introduced by either analog or digital processes. Gain and level distortions can be made in the processes of analog-to-digital and digital-to-analog conversions. Frequency response attenuation (i.e., horizontal resolution) can be affected by pre- and post-spatial filtering (analog or digital). Spatial cropping and misregistration are common in multimedia and teleconferencing codecs.

Test signals can also serve a more traditional role to determine the upper and lower limits of performance. Examples of these kinds of tests include resolution measurements[12], zone-plate loading[13], noise loading[14] and SNR measurements. We have found these types of measures to be useful for analysis of codec performance. But, in general, they do not provide quantitative predictions of video quality for real video.

1. Colorimetry

Gain and level distortions are defined as linear scaling and translation of color values in the YUV color space. These distortions are generally introduced by analog processes and so they are linear time-invariant effects. Gain and level are measured using a statistical comparison of source and degraded video test sequences in a manner similar to Wolf[6]. Gain and level are computed independently on the luminance and chrominance channels. Luminance gain and bias are analogous to contrast and brightness. Chrominance gain is analogous to color intensity. Chrominance level is usually fixed, but could potentially be modified. Gain and level are reported as separate distortion measures. The degraded video is preprocessed to compensate for the measured gain and level shifts before other distortion measures are computed. Thus, the other distortion and quality metrics do not penalize a codec for linear impairments that can be corrected by conventional analog adjustments (e.g., brightness, contrast, color, etc.).

2. Spatial Cropping And Registration

Spatial cropping and spatial misregistration are common artifacts of digital video compression systems. DS-3 codecs (45 Mbps) typically encode and pass the entire active picture. However, MPEG-2 defines a 704x480-pixel cropped subregion of the full 720x483 CCIR-601 active picture for encoding. Other low-data-rate encoders crop to a greater extent. Many codecs introduce horizontal and vertical misregistration in the picture. This problem has been seen even on 45 Mbps codecs. Measures of cropping and spatial registration are useful characterizations in their own right, and also they are necessary inputs for preprocessing prior to error signal computation.

ITS has recommended the practice of cropping to an assumed *viewable* region of 672x448 pixels to approximate typical consumer horizontal and vertical overscan. Any cropping outside the viewable region is ignored, while cropping within the region is treated as distortion. Manual inspection of the output signal (either with a digital oscilloscope or image viewing tools on a computer) is used to determine the actual active picture.

C. Comparative distortion measures

Comparative distortion measures involve direct comparisons between source video and degraded output video imagery. These measures include the discrete error signal, RMS error and signal-to-error ratio, and a CCIR-500 impairment quality rating (*IQR*) with three associated distortion measures.

Distortion measures can be characterized as either local or cumulative. A local measure is a discrete spatial-temporal sample of video distortion. The most fundamental local measure is the error signal, defined as the difference of the source and degraded digital imagery, and sampled at discrete pixel locations. A cumulative measure is defined over some spatial-temporal block of video. Common examples of blocks include an 8x8 coding block, multiple blocks, field, frame or short video sequence. Root-mean-squared error (RMS error) is an example of a cumulative distortion measure that can be computed over any arbitrary block of video. Typically, local measures are used to form the kernels of cumulative measures. Combinations of digital filtering, weighting, integration, and statistical techniques are used to define cumulative measures.

1. Error Signal and *SER*

The error signal is a basic diagnostic tool for testing digital video systems. The error signal is computed as a pixel-by-pixel difference of the digitized CCIR-601 source and degraded video. The error is computed separately on each of the three YUV channels, where Y is luminance, U is the blue color difference (also called C_b), and V is the red color difference (also called C_r). Figures 1 and 2 show a degraded frame and the absolute value of its luminance-error signal (enhanced for printing). Examination of the error signal is often useful for detailed analysis of compression artifacts.

The simple definition of the error signal is complicated by two real-world factors. First, the degraded imagery must be accurately corrected for temporal alignment, spatial cropping and registration before the differences can be taken. Spatial cropping and registration are performed as described above. For video systems that have a fixed delay, temporal alignment is performed by visual inspection with compensation for the delay in the video capture system. However, some systems can introduce variable frame dropping/repeating (especially low-data-rate codecs). For these, a technique developed by ITS is used to determine a fixed delay to best align each pair of source and degraded sequences[15]. Second, the pictures are aligned spatially, and subpixel registration is performed. Third, the pictures are cropped.

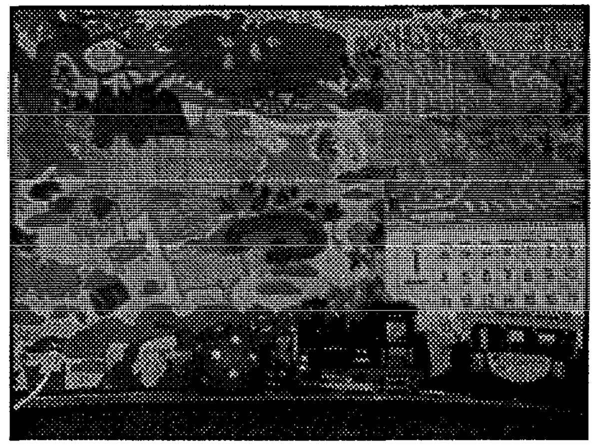


Fig. 1. Degraded version of *Mobile & Calendar* sequence,frame4.

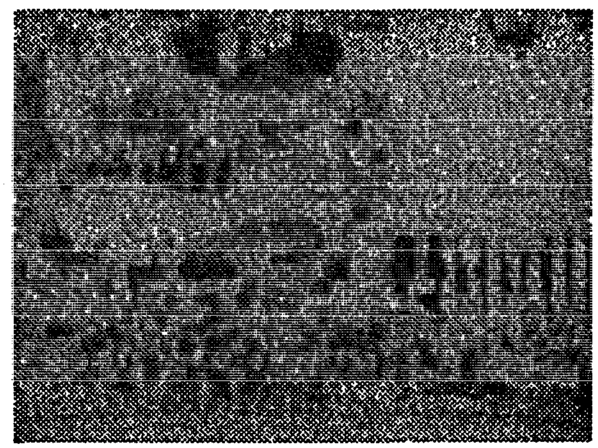


Fig. 2. Absolute Y-error (values x25) of degraded *Mobile & Calendar* sequence,frame4.

Fourth, the degraded video is corrected for any invariant gain and level distortions introduced by the video compression system. The degraded degraded imagery is corrected according to the method previously described, and the gain and level are reported separately. Finally, the error signal reflects only those distortions that are not correctable by constant analog adjustments to the signal.

Signal-to-error ratio (*SER*) is similar to the traditional signal-to-noise ratio (*SNR*) but generalized to apply to real video. It is defined using the log of the ratio of peak-to-peak signal versus the RMS error and is expressed in dB.

The SER is computed by field and overall for each video sequence as shown below.

$$SER_{field}(t) = 20\log_{10}\left\{\frac{V_{pp}}{rms_{space}\left(S(x,y,t)-D(x,y,t)\right)}\right\} \tag{1}$$

$$SER_{sequence} = 20\log_{10}\left\{\frac{V_{pp}}{rms_{time}\left(rms_{space}\left(S(x,y,t)-D(x,y,t)\right)\right)}\right\} \tag{2}$$

Signal-to-error ratio is defined in a similar way for each of the luminance and color-difference chrominance channels in the YUV color-space. Note that SER can be measured for any video signal and for any type of impairment -- for a gray input it approximates the traditional unweighted SNR measurement.

SER is an absolute measure of image reproduction accuracy. It is most significant for assessment of contribution-quality compression where it is desirable to minimize distortion of any kind. Also, SER is very useful for identifying occurrences of discrete impairments and artifacts within a codec's output sequence. Major spatial and temporal impairments cause recognizable signatures in field-by-field SER plots that are unique to each codec and test sequence combination. SER is less useful as an overall measure of perceptible impairment quality because different compression-induced distortions are perceived differently by the human visual system.

1. Impairment Quality Rating

The following is an approach developed by researchers at NTIA/ITS[6,7,8] to predict subjective impairment quality ratings (IQR) for pairs of degraded versus source real-video sequences. The method produces a rating on the CCIR-500 five-grade impairment scale for double-stimulus tests[2] (5 = *Imperceptible*, 4 = *Perceptible But Not Annoying*, 3 = *Slightly Annoying*, 2 = *Annoying*, 1 = *Very Annoying*). The IQR measurement formulation is based on a functional combination of distortion measures, optimized for correlation with available subjective test results. ITS uses a linear combination of distortion measures of the form:

$$IQR \approx c_0 - c_1 m_1 - c_2 m_2 - c_3 m_3 \tag{3}$$

The c_i are weighting coefficients determined to give the best fit to subjective test results using a least-squares-error criterion. The three distortion measures were selected from an exhaustive search of combinations of over 100 different distortion measures. These three measures, with coefficients of $c_0 = 4.7485$, $c_1 = 0.9553$, $c_2 = 0.3331$, and $c_3 = 0.3341$ produced the best correlation with the average ratings of a large subjective test involving 48 viewers, 36 nine-second test sequences, and 27 different types of analog and digital impairments. Analog impairments included NTSC encode/decode, VHS record/play, and a noisy RF channel. Digital impairments included video codecs operating in the range of 56 Kbps to 45 Mbps with simulated digital networks and controlled error rates. ITS showed correlation with CCIR-500 subjective test results with an expected accuracy for individual ratings of about 0.5 rating points.

Conceptually, m_1 is a measure of spatial distortion, m_2 is a measure of reduced motion, and m_3 is a measure of added motion. An important aspect of the NTIA/ITS approach is that these three measures do not involve direct comparison of the source and degraded video imagery. Rather, the source and degraded sequences are each characterized by small feature sets from which the measures are computed. In the case of m_1 , the feature set is the set of spatial variance values of the Sobel-filtered images of each frame in the video sequence -- essentially a measure of frame-by-frame edge content.

The method can be used to predict subjective ratings for arbitrary codecs, scenes and impairments. Predicted ratings have roughly a two-thirds probability of being accurate within 0.5 CCIR-500 rating points based on ITS's statistical analysis. *IQR* ratings are interpreted in the context of the range of impairments and viewing conditions present in the correlated subjective test. In addition, the magnitudes of the c_1m_1, c_2m_2, and c_3m_3 distortions (we denote the c_im_i by M_i) can provide insights into the nature of impairments that are generally not available from subjective tests. The M_i are indicators of the relative contributions of spatial and temporal distortions to the overall *IQR*. The overall *IQR* results and M_i distortions are plotted by sequence for comparisons between two or more codecs. Fig. 3 shows *IQR* ratings for the same codecs and sequences shown earlier.

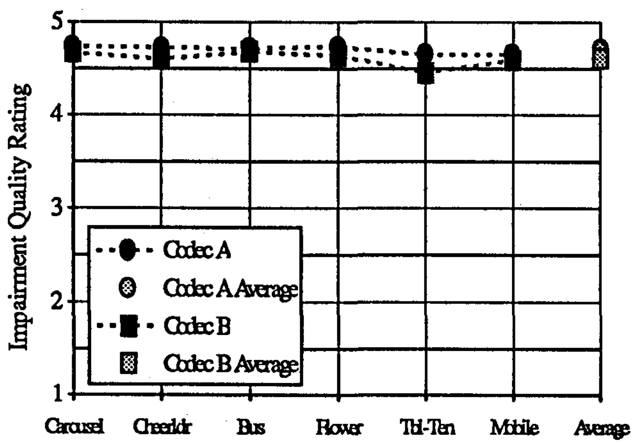


Fig. 3. Overall *IQR* for two codecs and six sequences.

The approach is well suited to overall assessment of one codec versus another on a common set of test sequences. Voran[16] has shown that average ratings, taken over several sequences give significantly better correlation with subjective tests than do individual ratings. Also, the approach is amenable to remote, real-time computation since only a minimal data set must be transmitted from the source site to the receiver site for comparison and computation of the underlying measures.

a. SER Results

SER tests have been made on a number of commercial codecs. Luminance and chrominance SER measures are calculated using the six video sequences described previously. Overall Y, U, and V SER are plotted in Figures 4, 5, and 6 for two 45-Mbps codecs.

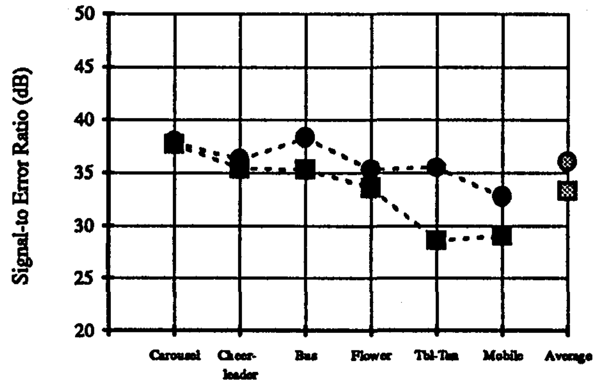


Figure 4. Luminance (Y) Signal-to-Error Ratio, Overall by Sequence.

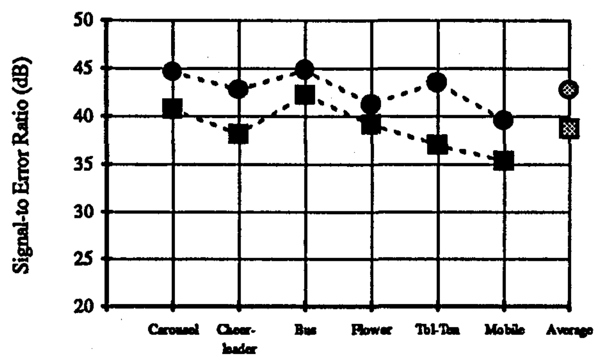


Figure 5. Blue Color Difference (U) SER, Overall by Sequence.

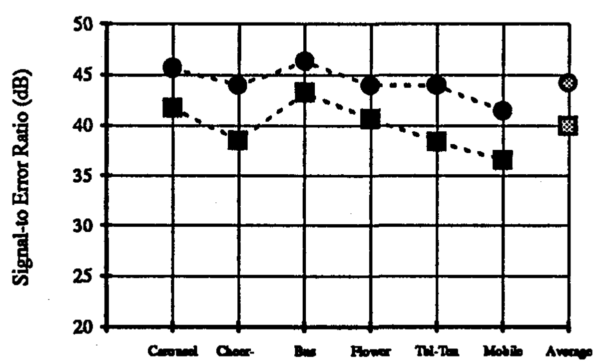


Figure 6. Red Color Difference (V) SER, Overall by Sequence.

Codec A has a distinct, consistent advantage in SER for all channels. Averaged over all sequences, codec A showed advantages of 2.7 dB, 4.0 dB, and 4.3 dB across the Y, U, and V channels, respectively. Codec A produced a better SER than codec B on all sequences. Codec A had the greatest advantage on the scenes with high spatial-temporal content, especially *Table Tennis* where it had a 7 dB advantage in luminance SER. Both codecs produced higher signal-to-error in the chrominance channels than in luminance.

SER values in the range of 30 dB to 40 dB are typical of good DS-3 codecs on these sequences. Distortions characterized by 30-40 dB overall SER are often, but not always, imperceptible. Distortions characterized by less than 30 dB are more likely to be perceptible. Distortions in the low-20's dB are likely to be perceptible and at least somewhat annoying.

The testing system also produces frame-by-frame images of codec-induced distortions that can be viewed interactively on the computer. These "error sequences" in conjunction with the field-by-field SER plots provide a valuable tool for identifying and analyzing codec artifacts. In the present case, they did indeed help to locate a particular type of artifact induced by these codecs.

IV. Instrument Measures

A. Measurement of Multi-Dimensional Codec Resolution

In conventional NTSC tests, horizontal resolution is measured as frequency roll-off using the multiburst or continuously swept test waveforms. However, vertical resolution and diagonal resolution are seldom measured. Vertical resolution in NTSC is set by the number of visible scan lines in the system, and by the Kell factor which depends upon the camera and the monitor. Transmission and processing seldom change the vertical resolution in conventional TV. However, the introduction of compression in video signal transmission changes this situation. Vertical resolution can be seriously affected by the compression and decompression processing, as can horizontal and diagonal resolution. Therefore, all three resolutions need to be measured. Systems that use predictive coding are even more complicated to test, as the various resolutions can depend upon the speed at which the scene is changing. A method of measuring these three resolutions using moving zone plates[17] can be used.

1. Signal Source and Test Equipment

The signal source is a Tektronix TSG-1001 multi-format zone-plate generator. For testing resolution, one-dimensional moving zone plates are used, with a specific set-up for each of the three directions. A single zone-plate parameter is used for measuring resolution in the horizontal or vertical direction, and a pair of parameters is used for the measurement in the diagonal direction. The equipment is configured as in Figure 7. Modulation-transfer-function data is captured by a Tektronix 2232 digital storage oscilloscope with samples externally clocked by a pulses, one per frame, from the zone-plate generator. The data is then plotted on a laser printer. Although picture monitors are present, they are not involved in the measurement, making this an accurate test of the codec resolutions.

2. Results Obtainable From Moving-zone-plate Tests

Plots taken with the setup in Figure 7 show the modulation-transfer-function (MTF) as a function of spatial frequency in each of the three principal directions. Resolution (cycles per picture height) in each direction is defined as the spatial frequency at the 3-dB point on each MTF plot. Resolution in TV-lines is double the resolution number in cycles/picture-height. The results from making these plots with many codecs are as follows: (1) Horizontal MTF for luminance and chrominance is identical to that obtained with the stationary swept sine wave test signal or the multiburst test signal applied separately to the luminance and chrominance channels. (2) Vertical MTF is generally flat to the Nyquist frequency and beyond for both luminance and chrominance, but as there is generally no vertical prefiltering provided prior to the 2:1 decimation of the chrominance channels, chrominance signals are not protected against vertical aliasing with most codecs. (3) Diagonal MTF is likely to have ripple, caused by the quantizing matrix in DCT-based codecs.

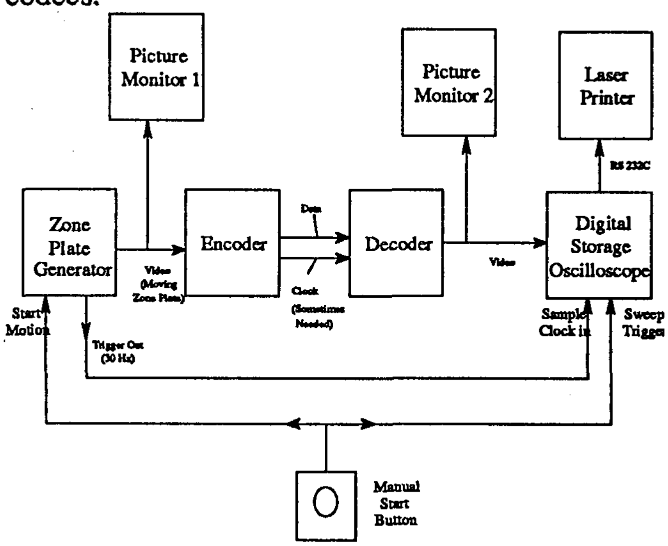


Figure 7. Equipment Configuration for Resolution Testing.

1. EIA-250-C Testing

Since no standard exists for measuring the picture quality of compressed digital video signals, the question is frequently asked, "Does the codec meet the EIA-250-C short, medium or long-haul specification?," even though it is known that the EIA measurement set does not adequately characterize a complex digital video compression codec. The Tektronix TSG-1001 multi-format video test signal generator and the VM-700A video measurement can be used to determine if there are any problems in the NTSC coding,, but will not find problems in the compression coding.

B. Subjective Testing with Zone-Plate Loading

Circular zone plates with spatial frequencies up to full Nyquist detail at the top and bottom of the picture and with various rates of motion can be applied to these codecs. The outputs are observed subjectively on a monitor. The artifacts of NTSC are always very prominent in this subjective test, making it difficult to recognize small artifacts of codec processing for zone plates with high-density features. However, some codecs have artifacts that are easily found by this subjective test, making it a worthwhile part of the testing plan.

V. Implementation of a Digital Video tester

We have developed a system for testing digital video compression systems based on a Silicon Graphics workstation. The system computes the error signal, *SER*, *IQR* and associated distortion metrics by field, by frame and by sequence. In addition, it computes temporal alignment, spatial registration, and gain and level. The tester incorporates viewing tools that allow interactive viewing of source, degraded, and error video imagery on the computer monitor. Graphs of *SER*, *IQR* and M_1, M_2, and M_3 can be displayed along side the video imagery. We use D1, D2 or Betacam tape to record source video and play it through a codec. Currently, source and degraded video are captured on an Accom WSD disk recorder. We are in the process of fully integrating video capture into our testing system using the standard video input capabilities of an SGI Indy workstation. A diagram of this system is shown in Fig. 8. The Indy system is a low cost approach and does not have many of the Accom features including capture of 30 seconds or more of 720 x 486 video.

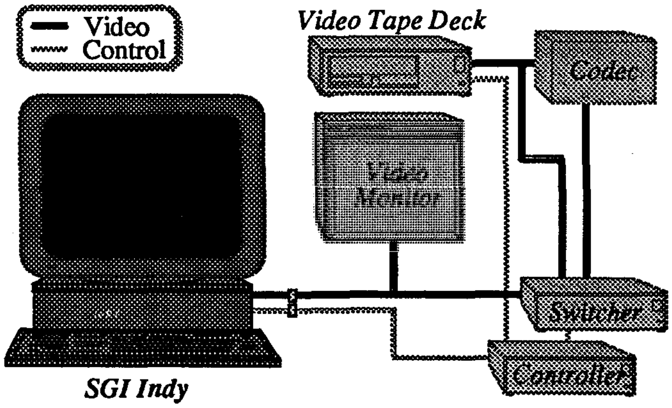


Fig. 8. Diagram of prototype testing system.

The Indy incorporates a number of video technologies that enable this system. The Indy has a standard video input subsystem (VINO) that allows burst capture at 30 frames-per-second video directly to system RAM. Video is captured at 640x480 pixel resolution in YUV 4:2:2 format. This format is similar to CCIR-601 but uses square pixels and therefore a slightly lower sampling rate. Video input is via the standard NTSC or S-video input ports. The input video is digitized one field at a time in the VINO subsystem, then transferred real-time across the Indy's internal GIO-64 bus (33 Mhz, 64-bit) to a RAM buffer. The burst capture is limited by the size of the RAM buffer -- the Indy's maximum RAM configuration of 256 MB allows capture of six-second sequences. The source and degraded imagery are downloaded to files on an internal SCSI-2 disk for later access during analysis and viewing.

Our system performs image processing computations for the various measurements completely in software. The Indy incorporates a MIPS R4000 100 Mhz RISC CPU that is capable of image processing at a usable level of performance for our applications. We use SGI's object-oriented ImageVision library for the development of our image processing software. Computation of all measures for a five-second sequence currently takes about 40 minutes. With automation, an overnight test can compute a full set of measures for about 20 five-second test sequences. We believe that with careful selection, 20 sequences of varying content is sufficient to provide adequate characterization of a codec's performance.

We expect to improve the performance in a number of ways: 1) Tuning our software should give about a 2x increase in performance, 2) We can test using one-second sequences. This allows a greater variety of test material (i.e., 100 sequences versus 20) because computation time is proportional to video length independent of the number of sequences. The main reason to test with five-second sequences is to allow subjective assessment. However, one-second sequences can be selected for objective testing from a much larger test tape used for subjective assessment. Alternatively, short objective test sequences can be played in a continuous loop or swing mode for subjective assessment. 3) Computer price-performance is advancing rapidly. Our ImageVision software will take advantage of faster CPUs and parallel processing when they become available on low-cost workstations. 4) Dedicated image processing hardware (e.g., DSP boards) can accelerate performance by 100x or more. DSP boards are available on other platforms, though not yet on the Indy. CPU performance is the main bottleneck in the existing systems performance. The GIO-64 bus provides extremely fast transfers of image data between RAM and CPU. SCSI transfers of image data from disk to RAM can limit image processing and video image display performance in some circumstances. However, the ImageVision library incorporates a innovative look-ahead, multi-tasking caching scheme that eliminates most of these bottlenecks.

Figures 10 and 11 show typical displays from StellaCom's tester. Fig. 10 shows the source video, the error signal, and the degraded video imagery being displayed simultaneously in separate synchronized windows. The three pictures are presented either still or in motion at up to 20 frames per second. A graphical user interface control panel provides VCR-like controls for playback with random access. Graphs of *SER* and *IQR* and associated spatial and temporal distortion measures can be displayed in separate windows. Fig. 11 shows the degraded picture along with plots of the *IQR* and M_1, M_2, and M_3 frame-by-frame kernels. The overall *IQR* for the sequence appears as a dashed horizontal line. A cursor tracks the current frame on each plot during playback. The range of frames for playback and graphing may be reset to any subset of the full test sequence. Also note that standard desktop tools on the SGI allow for interactive magnification and enhancement of the video imagery for detailed assessment of distortion and artifacts.

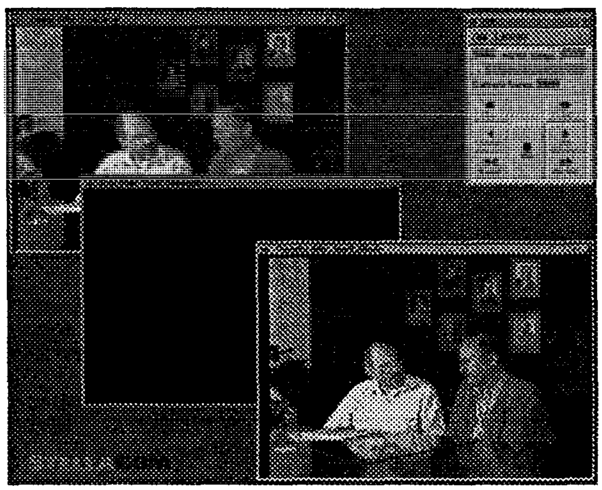


Fig. 10. Display of source video, error signal, and output video on StellaCom's tester.

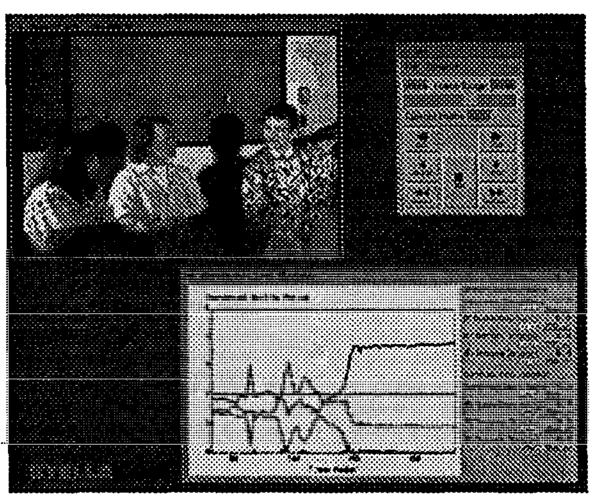


Figure 11. Display of degraded video and *IQR* measurements

VI. Summary

Compressed digital video systems require new testing methods. We have demonstrated that a graphics workstation which has been programmed with appropriate measures can suitably test such systems. We have developed a testing system that incorporates a suite of distortion measures and viewing tools for the workstation. The Impairment Quality Rating (*IQR*) measurement developed by Wolf and others at NTIA/ITS can be used to predict how an average viewer rates a video sequence on the CCIR-500 impairment quality scale. The *IQR* measurement has been correlated with subjective testing using statistical methods. Associated with the *IQR* are a spatial distortion measure and two temporal distortion measures. In addition, we compute an error signal that can be viewed as video and used for statistical measurements such as RMS Error and Signal-to-error ratio (*SER*). Our system allows the user to view video (source, codec-processed video and error signal) side-by-side with plots of measurement results on the computer display. It permits in-depth analysis of the quality of digital video compression, transmission and storage systems.

The quality of compressed video depends on a number of factors including data rate. Codecs employ complex algorithms that incorporate discrete spatial and temporal processing. They exhibit non-linear responses to changing scene content and operating conditions. Codecs working at low data rates and processing video with significant detail and motion can introduce distortion into the picture. Consider the impairment quality rating (*IQR*) for degraded versus source video sequences. Analog VHS material normally yields a rating between 3.0 and 4.5 (between slightly annoying and barely perceptible) depending on whether the source material readily shows the imperfections of VHS tape. The average rating for VHS is usually between 3.5 and 4.0. The rating for a 6 Mbps digital codec is normally above 4 depending on the source material. Hence, a good 6 Mbps digital codec can be superior to an analog VHS tape by this measure. Another measure of picture quality is resolution. VHS video has a horizontal resolution of approximately 220 TV lines per picture height (TVL/PH). Static horizontal resolution on most 6 Mbps codecs exceeds 450 TVL/PH. The average resolution of a 6 Mbps digital codec on moving objects is less than the static resolution, but can exceed that of an analog VHS tape.

Compression of video signals is cost-effective now and will be pervasive in the future. Engineers and engineering standards bodies had better learn how to measure the quality of compressed video systems. Otherwise, we will not know the ultimate quality of the material we are producing, transmitting and displaying.

Acknowledgments

The author would like to thank Mr. Ted Dienert, president of StellaCom, for his support of this project. Appreciation is expressed to Mr. Steve Wolf of NTIA/ITS for providing video test materials, subjective and objective test results, and numerous discussions. Thanks are expressed to Chris Cressy, Dr. Elliott Kohn, Jeff Van Pelt and Javier Vega of StellaCom for significant contributions to this effort.

References

1. C. Cressy and G. Beakley," Computer-based testing of digital video quality," IS&T/SPIE Symposium, San Jose, CA, February 6-10, 1994.
2. CCIR Recommendation 500," Method for the Subjective Assessment of the Quality of Television Pictures", 1990.
3. T. Hidaka, "MPEG-2 Verification Test," ISO/IEC JTC1/SC29/WG11/MPEG93/732, August, 1993.
4. S. D. Voran and S. Wolf, "The Development and Evaluation of An Objective Video Quality Assessment System That Emulates Human Viewing Panels," *International Broadcasting Convention*, Amsterdam, July, 1992.
5. ISO/IEC CD 13818, "Coding of Moving Pictures and Associated Audio," November, 1993.
6. S. Wolf, et al, "A Summary of Methods of Measurement for Objective Video Quality Parameters Based on the Sobel Filtered Image and the Motion Difference Image, T1A1.5/93-152, November, 1993.
7. A. W. Webster et al, "An Objective Quality Assessment System Based on Human Perception," *SPIE 1993 International Symposium on Electronic Imaging: Science and Technology*, June, 1993.
8. S. Voran and S. Wolf, "An Objective Technique for Assessing Video Impairments," *Proceedings of the IEEE Pacific Rim Conference*, 1993.
9. W. E. Glenn, K. G. Glenn, and C. J. Bastian, "Imaging System Design Based on Psychophysical Data," Proceedings of *the SID*, Vol. 26/1:71-78, 1985.
10. K. Otsuji and Y. Tonomura, "Projection Detecting Filter for Video Cut Detection," *Proceedings of ACM Multimedia 93*, 1993, pp. 251-257.
11. H. Ueda, T. Miyatake and S. Yoshizawa, "IMPACT: An interactive natural-motion picture dedicated multimedia authoring system," *CHI'91 Conference Proceedings*, 1991, pp. 343-350.
12. E. S. Kohn and G. W. Beakley, "Resolution Testing of An HDTV Codec with Moving Zone Plates," 134th SMPTE Technical Conference, Toronto, November 1992.
13. G. W. Beakley, C. R. Caillouet, E. S. Kohn and J. L Van Pelt, "Testing of a Digital HDTV Codec Through the NASA Communications Network," *1992 HDTV World Conference Proceedings*, National Association of Broadcasters, April 1992.
14. G. W. Beakley and E. S. Kohn, "Transmission of Compressed Digital High-Definition Video Through the NASA Communications Network," *Proceedings of SPIE Conference*, San Diego, July 1992.
15. S. Wolf and A. Webster, "Objective Performance Parameters for NTSC Video at the DS3 Rate," T1A1.5/93-60, April 28, 1993.
16. S. Voran, "The Effect of Multiple Scenes on Objective Video Quality Assessment," T1A1.5/92-136, July, 1992.
17 E.S. Kohn and G.W. Beakley, "Resolution Testing of An HDTV Codec with Moving Zone Plates," 134th SMPTE Technical Conference, Toronto, November 1992.

Hybrid Multichannel Analog/Digital CATV Transmission via Fiber Optic Link: Performance Limits and Trade-Offs

Qun Shi and Robert S. Burroughs
Panasonic Technologies, Inc.

***Abstract* - It is highly likely that a hybrid analog/digital system will be the format for future Cable transmission systems. The motivation, for using hybrid systems, is to maximize the tuning compatibility with existing consumer electronic devices, and such a mixed system may provide excellent digital video pictures, while still maintaining the quality of conventional analog (AM) signals. However, our research results have shown that in such mixed systems the digital signals may suffer a significant performance degradation due to the nonlinear, especially the clipping, operation of the laser transmitter when mixed signals are transmitted by a fiber optic link. Careful considerations should be given to available trade-offs in such hybrid system designs and realizations.**

1. INTRODUCTION

Research Target

The intent of this research was to develop cost-effective hybrid, analog/digital, transmission systems for new Cable, fiber-to-the-node or feeder architectures, (as shown in Fig. 1).

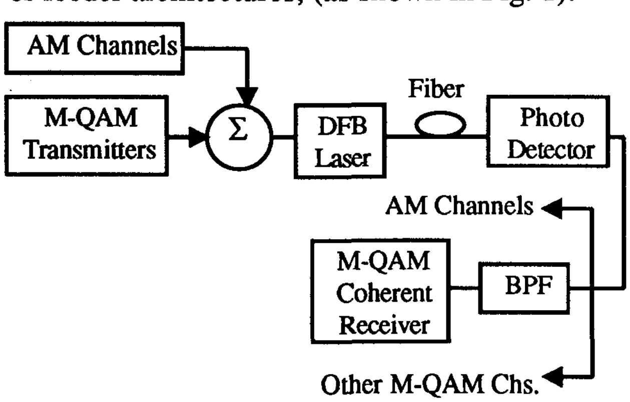


Fig. 1 Basic Configuration of the Hybrid SCM AM/QAM Fiber Optic System

General design targets are:

- 1.0 GHz., hybrid fiber optic/coax. transport to each Service Area, (~ 150 channels, 70 analog and 80 digital).
- One laser /Service Area, used to deliver all analog and digital downstream channels.
- Small Service Areas, (~ 500 homes ea.).
- Independent transmission of digital channels to each Service Area.

These targets would be required to implement cost-effective systems. All targets appear realizable, considering the recent developments in Cable technology:

- availability of high performance lasers,
- deployment of 1.0 GHz. coaxial transport,
- demonstration of compression of several television programs within 6 MHz. of bandwidth,
- development of digital video servers, and the
- possibility for telephony and interactive services by major Cable operators.

Problems for the Research.

Although, laser diodes have been shown to be able to transmit 80 analog channels with satisfactory Cable performance specifications, adding 80, or so, digital channels to the AM channels, using a single laser diode and fiber, would require further investigation. Theoretically the digital signals could be carried at power levels significantly lower than the analog AM signals.

Two initial problems were:

- A better model was needed for laser diodes used for analog multi-carrier Cable transport.
- A clear understanding of how laser diode induced multi-carrier nonlinear impairments would affect AM/digital transmission was needed.

Few experiments have been done and not many articles have been published for multicarrier transmission of standard 6 MHz. AM Cable channels by lasers and fiber optic cables. Several models for the transmission of several analog channels over a single laser diode appeared [1]-[3]. These early models made several simplifying assumptions and the resulting models did not appear to agree closely with experimental results [6], [7]. This variation in experimental results and the theoretical models suggested that a more representative model should be developed.

Distortion - What is the Problem ?

The initial problem inhibiting the transmission of Cable quality signals by laser diodes was associated with the linearity of the laser diode. The major problems seemed to be low power and a lack of linearity in the operating region of the diode.

The classic distortion products produced by amplifier nonlinearities in multi-carrier analog Cable systems are composite second order distortion, (CSO), and third order distortion or composite triple beat, (CTB). These types of distortion are also produced by laser diodes. But, there were two new aspects to optical transmission by laser diodes, which were not prevalent in existing coaxial systems with standard trunk amplifier cascades:

- laser diodes are single ended and clip multi-carrier signals,
- the nature of these clipped signals is impulsive in nature and its effect on hybrid AM and digital transmission has not been throroughly investigated.

Limiting Factors

It has been shown that for laser diodes the limiting factor for multi-carrier transmission of VSB-AM signals is no longer the nonlinear operating region of the diode; it is the clipping effect [1]-[5]. This effect is caused by the clipping of the combined signals at the laser diode threshold when the modulation index is increased.

Further it has been shown that laser diode clipping creates a probabilistic impulse noise whose average value is small but, whose peak value is quite large [8]-[12]. These large peak values of impulse noise do not significantly affect AM transmission because of their low probability of occurrence, making their average value small. These large peaks, although they occur infrequently, do greatly affect the digital transmission portion of the hybrid system.

What Has Been Done

Several trade-offs are required to be able to do cost effective hybrid system designs that use the full capabilities of modern Cable components and architectures. In this paper we develop the basic analytic tools for analyzing hybrid systems and then show an iterative approach that can be used to design cost effective hybrid transmission systems.

In Section 2 we show a laser diode model that includes the contributions of clipping distortion, which are modeled as a Poisson impulse train. This model has been shown to agree closely with experiments done at other laboratories [6][7].

In Section 3 the effect of this clipping noise on the BER performance of M-QAM is modeled. Clipping-induced distortion is shown to be the most significant noise affecting BER, under specific conditions, and a noise floor is shown which limits the performance of digital transmission in hybrid fiber optic systems. Three cases are presented which show the BER performance of 16-QAM and 64-QAM in relationship to the impulsive noise generated by laser diode clipping.

In Section 4 we show several performance limits and trade-offs using 64-QAM and 256-QAM as examples. A design example is given to show how trade-offs can be made using typical components and system parameters.

In Section 5 techniques are discussed for reducing the effects of clipping distortions, with channel coding. Examples of common, Reed-Solomon codes are selected for one and two symbol correction. The effects of these codes on clipping distortion noise is then shown, respectively, for 64-QAM and 256-QAM signals combined with AM signals in hybrid systems.

2. MODEL AND CHARACTERISTICS OF CLIPPING DISTORTION AND NOISE

Distortion (Intermodulation) Products

It is known that the clipping-induced distortion in a multi-carrier fiber system generates many higher order in-band intermodulation products in addition to the classical CSO and CTB products [4][5]. Several models have been developed to calculate the in-band distortion power. Earlier models [1]-[3] have, however, made simplifying assumptions and appeared not to agree with experimental results [6][7].

It is this discrepancy between the earlier models and experiments that have led our effort to develop a model more accurately accounting for the clipping distortion. The model, as described below, can predict the clipping distortion power at any channel for all orders of distortion, including the CSO and CTB [4].

Briefly stated, for an N-channel signal input to an ideal laser, the model predicts that the total carrier to distortion power ratio (C/NLD) for a channel is given by [4]:

$$(C/NLD)^{-1} = \frac{1}{2\pi h_1^2} \sum_{n=2}^{\infty} \frac{K_n}{N^{n-1} 2^n} \cdot e^{-2/m^2 N} G_{n-2}^2(m\sqrt{N}) \qquad (1)$$

where K_n is the product count of the n-th NLD at that channel, m is optical modulation index (OMI) per channel, $h_1 = [1 + \mathrm{erf}(1/m\sqrt{N})]/2$, erf(·) being the error function, and

$$G_{n-2}(m\sqrt{N}) = \sum_{k=0}^{[(n-2)/2]} (-1)^{n-k} \frac{(n-2)!}{(2k)!(n-2-2k)!} \cdot \left(\frac{\sqrt{2}}{m\sqrt{N}}\right)^{n-2-2k} C_k, \quad n \geq 2 \qquad (2)$$

with $C_k = 1$ if $k = 0$ and $C_k = 1 \cdot 3 \cdot 5 \cdots (2k-1)$ if $k \geq 1$. The CSO and CTB are then obtained from (1) by letting n = 2 and n = 3, respectively, i.e.,

$$CSO = \frac{K_2}{8\pi N h_1^2} e^{-2/m^2 N} \qquad (3a)$$

$$CTB = \frac{K_3}{8\pi N^2 h_1^2} e^{-2/m^2 N} \qquad (3b)$$

To give an example, in Fig. 2 we show the CSO obtained from (3a) for a 42-channel system, and compared with experimental results [6][7]. The figure also includes the simulation results based on the earlier models [1][2].

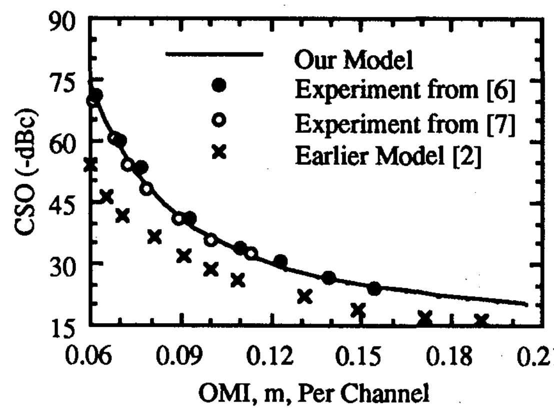


Fig. 2 CSO vs. OMI for 42-Channel Case

The figure shows that the CSO given by our model closely agrees with the experimental data, and is 5-15 dB less than that estimated by the earlier models.

Clipping Distortion as "White" Noise

From the frequency viewpoint, the clipping-induced distortion is wideband, and may be considered as "white" noise to the system. This can be verified by the power spectral density (PSD) of the clipping noise which is given by [4][8]:

$$P_w(f) = \sum_{k=2}^{\infty} \frac{h_k^2}{k!} \int_{-\infty}^{\infty} \left[\frac{R_{in}(\tau)}{R_{in}(0)}\right]^k e^{-j2\pi f\tau} d\tau \qquad (4)$$

where $R_{in}(\tau)$ is the autocorrelation function of the laser input signal and h_k is given by [4]

$$h_k^2 = \frac{R_{in}(0)}{2\pi} e^{-2/m^2 N} G_{n-2}^2(m\sqrt{N}), \; k \geq 2 \qquad (5)$$

Fig. 3 shows a plot of $P_w(f)$ when the input signal consists of 70 AM and 80 digital (64-QAM) signals. Fig. 3 shows that the spectrum of the clipping noise is roughly flat below 1 GHz - the band of interest for CATV. The in-band noise power at a channel is then given by $NLD_{in} = P_w(f_c) \cdot B$, f_c being the carrier frequency of that channel and B is the bandwidth. Note that this NLD should be the same as that given by (1); that is, that we have two alternative ways to calculate the NLD induced by clipping.

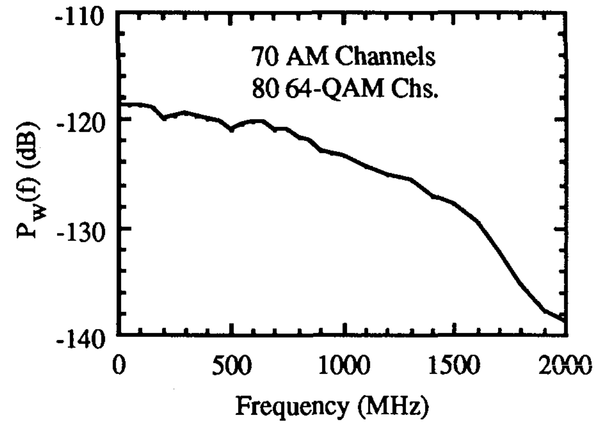


Fig. 3 PSD of the Clipping Noise

<u>Clipping Distortion as "Impulse" Noise</u>

While the clipping noise exhibits wideband spectrum, it really behaves like an impulsive noise infrequently occurring over random time intervals as experimentally demonstrated in [9] [10] and theoretically proved in [11][12]. In particular, for small probability of clipping, the clipping noise is shown to possess the following statistical properties [11][12]:

1. It is modeled as a Poisson impulse train.
2. Its pulse duration follows a Rayleigh distribution.
3. Its pulse shape is approximated by a parabolic arc.
4. Its mean pulse width is less than a nanosecond.
5. It is non-Gaussian distributed; its probability density function (PDF) is of the type [8][11]:

$$f(n) = (1 - \gamma)f_g(n) + \gamma f_w(n) \tag{6}$$

where $f_g(n)$ is the PDF of the Gaussian component, $f_w(n)$ is the PDF of the non-Gaussian component, and γ is known as the clipping index (or impulsive index) which is equal to the product of the average number of impulses in a second and the mean impulse duration.
6. Its variance can be obtained from (1) or (4) or from the simplified equation [8][11]:

$$\sigma_w^2 \doteq \frac{8\bar{\tau}}{3} (\pi\rho)^{-3/2} e^{-\rho} R_{in}(0) \tag{7}$$

where $\rho = 1/m^2N$ and $\bar{\tau}$ is the mean impulse duration. The in-band noise power is then $NLD_{in} = \sigma_w^2 \cdot B \approx P_w(f_c) \cdot B$.

To illustrate the impulse behavior of the clipping noise, in Fig. 4 we conceptually show how the impulse noise train is generated by the clipping phenomenon.

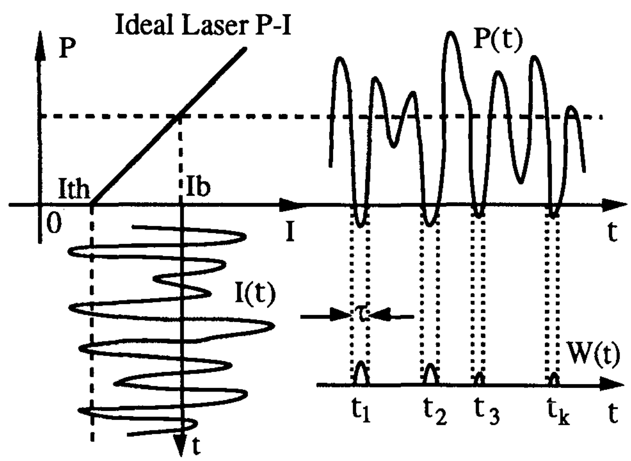


Fig. 4 Clipping-Induced Impulse Noise Train

The figure illustrates that the laser output signal which has been clipped may be decomposed into a sum of the attenuated input signal and a noisy impulse train w(t), viz.:

$$P(t) = KI(t) + w(t) \tag{8}$$

where attenuation coefficient K can be obtained from the laser P-I curve and is equal to $K = h_1$. We note that this impulse-like time domain waveform responsive to clipping has also been observed by recent experiments employing the hybrid SCM signals [9][10].

3. BER PERFORMANCE OF M-QAM SUBJECT TO CLIPPING NOISE

BER Model

Having described the impulsive behavior of the clipping noise, we now study the performance of a digital M-QAM system impaired by the clipping noise in conjunction with a Gaussian background noise. First, we devise a BER model of M-QAM that can be used in such events. To this end, we use an analytic model of Fig. 1 which, as shown in Fig. 5, specifically indicates an M-QAM system embedded in both the clipping and Gaussian noises.

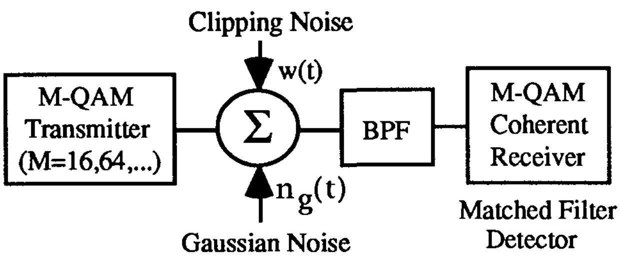


Fig. 5 Analytic Model for M-QAM Signal in the Presence of Clipping and Gaussian Noise Sources

In Fig. 5, the Gaussian noise refers to the additive background noise sources existant in the optical link including laser intensity noise RIN, shot noise, and thermal noise i_n; its variance is given by [7]-[11]:

$$\sigma_g^2 = i_n^2 + 2q(rP_0) + RIN(rP_0)^2 \qquad (9)$$

where q is the electron charge, r is the photodetector responsivity, and P_0 is the received optical power. The clipping noise has a variance given by (7). For the hybrid system, the input correlation function $R_{in}(\tau)$ consists of both the AM and M-QAM signals. Thus (7) can be rewritten as [8][12]:

$$\sigma_w^2 = \frac{4\bar{\tau}}{3}\pi^{-3/2}\mu^5 e^{-1/\mu^2}(rP_0)^2 \qquad (10)$$

where $\mu^2 = N_a m_a^2 + N_q m_q^2 = \mu_a^2 + \mu_q^2$, N_a and N_q being the number of AM and QAM channels, and m_a and m_q being the OMI's of AM and QAM, respectively. From (10), we see that the clipping noise has contributions from both the AM and QAM clipping distortions. If μ_q is small compared to μ_a, which usually is the case, the AM distortion dominates the clipping noise.

For M-QAM corrupted by an additive noise, the BER is generally expressed by [13]:

$$BER = \frac{1}{\log_2\sqrt{M}}\left(1 - \frac{1}{\sqrt{M}}\right)\int_{d_m/2}^{\infty} f(n)dn \qquad (11)$$

where d_m is the minimum distance between adjacent states in the M-QAM constellation and is given by $d_{min}/2 = \sqrt{[3T_sP_{av}/(M-1)]}$, $T_s = 1/B$ is the symbol time and $P_{av} = m_q^2(rP_0)^2/2$ is the average signal power of M-QAM, and f(n) is the PDF of the noise. Hence, for the hybrid system, using the joint PDF of the clipping and Gaussian noises as represented by (6), the BER of M-QAM may be devised as follows [8]:

$$BER = \frac{1 - 1/\sqrt{M}}{\log_2\sqrt{M}}\left\{(1-\gamma_T)\mathrm{erfc}\left(\sqrt{\frac{1.5\Gamma_g}{M-1}}\right) + \gamma_T\left[\mathrm{erfc}\left(\frac{\Delta_1}{\sqrt{2}}\right) + \frac{9\gamma_T}{4}\cdot\frac{\Phi_3(\Delta_2)}{(2+\gamma_T G)^2} + \frac{21\gamma_T}{4}\cdot\frac{\Phi_5(\Delta_3)}{(3+\gamma_T G)^3}\right]\right\} \qquad (12)$$

where $\gamma_T = T_s\lambda = \gamma T_s/\bar{\tau}$ is the clipping index per symbol interval, $G = \sigma_g^2/\sigma_w^2$ is the variance ratio of the Gaussian to clipping noise component, $\Gamma_g = P_{av}T_s/\sigma_g^2$ is signal to Gaussian noise ratio, $\Delta_i = \sqrt{[3\Gamma_g/(M-1)(1+ i/\gamma_T G)]}$ (i=1,2,3), and $\Phi_k(z) = H_k(z)\exp(-z^2)/\sqrt{2\pi}$ (k = 1,2,3), $H_k(z)$ being the Hermit polynomials.

The first term of (12) indicates the effect of the Gaussian noise only, while the second term indicates the effect of both the clipping and Gaussian noises. Further, Eq. (12) shows that the BER of M-QAM in this case is specified by three parameters: Γ_g, γ_T, and G. From (9) and (10), we note that these three parameters all depend on system parameters such as the number of channels and OMI's of both the AM and M-QAM signals. Hence, the BER model

(12) can be utilized to evaluate the digital system performance under various conditions.

BER Performance

We now present several example cases, and also compare our results with the reported experimental results [9][10].

In general, when many M-QAM signals are added with AM signals, especially with higher-level modulations such as 64- or 256-QAM, the effect of the clipping of the QAM signals can not be ignored; the effect will not only affect the QAM signals themselves, but also the AM signals. This multi-effect will be addressed in the next section. In the following examples, we consider the AM to QAM interference only. That is, we consider the scenario in which a few M-QAM channels are multiplexed with the AM signals. For this case, the effect of clipping is mainly due to the clipping of the AM signals, since QAM requires much lower power, (e.g., 10~20 dB below), than AM for reliable transmission. This AM dominance on the clipping noise components has also been verified by one of the recent experiments [10].

In case 1, we use 42 AM carriers (55.25-337.25 MHz) and a 16-QAM located above the AM carriers with 6 MHz bandwidth. The BER results are shown in Fig. 6 versus the OMI of 16-QAM for three different AM OMI's: $m_a = 0$ (no AM carriers), $m_a = 5\%$ (C/NLD = 69 dB) and $m_a = 6\%$ (C/NLD = 53 dB). The solid BER curves are the analytic results obtained by (12), and the black dots are the experimental data.

Fig. 6 shows that for 42 AM carriers, even with a typical 5% AM OMI, the effect of clipping on BER is significant; the BER increases greatly as compared to the no AM case. For a higher 6% AM OMI, the BER begins to show a floor. Increasing the OMI of 16-QAM, i.e., increasing the signal power will no longer reduce the BER. For this case, other means may have to be applied, such as coding techniques.

Fig. 6 also shows a strong correlation between the model and the experiments; both show a similar BER behavior and the two appear to agree for small OMI's of 16-QAM and for the high clipping case. The difference between the two results may possibly lie in the hardware limitations and other effects that are not dealt with by the model.

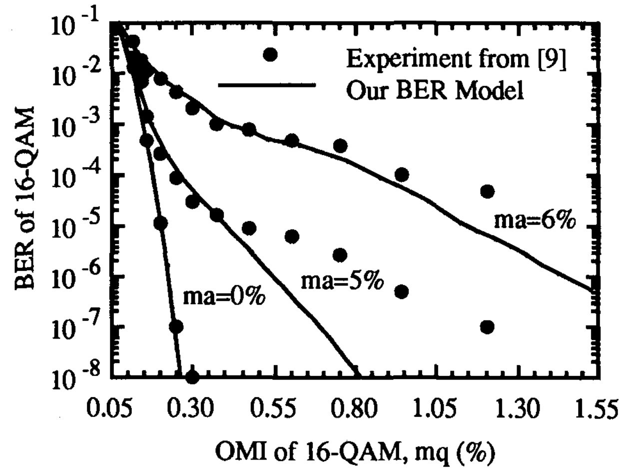


Fig. 6 BER Performance of a 42-AM/16-QAM System. RIN=-150 dB/Hz. i_n=35 pA/$\sqrt{\text{Hz}}$. r = 0.87 A/W. P_0=-1 dBm.

In case 2, we use 60 AM carriers (55.25-439.25 MHz) and a 64-QAM with 15 MHz bandwidth. Again, three different values of AM OMI's are used: $m_a = 0$, m_a=4%, and m_a=4.6%. The BER's obained from (12) along with the experimental results extracted from [10] are plotted in Fig. 7 versus the OMI of 64-QAM.

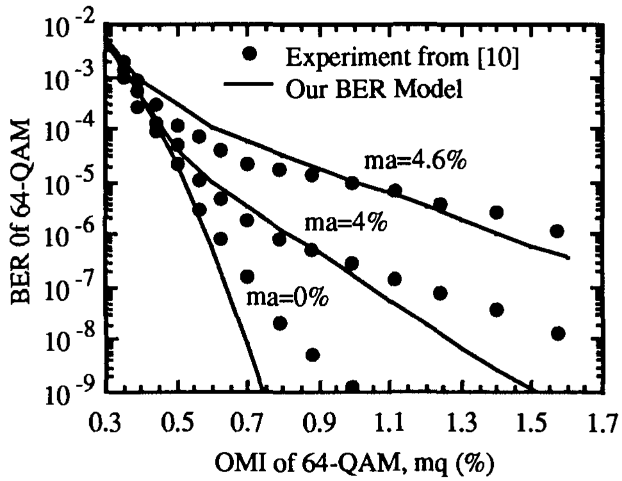


Fig. 7 BER Performance of a 60-AM/64-QAM System. RIN=-155 dB/Hz. i_n = 30 pA/$\sqrt{\text{Hz}}$. r = 0.9 A/W. P_0 =-1 dBm.

Fig. 7 shows the similar BER behavior to Fig. 6; i.e., the BER degrades significantly with the increase of the AM OMI. Again, the analytic and experimental results exhibit the similar BER decreasing trend, and agree reasonably well for small OMI's of 64-QAM.

Notice that from Figs. 6-7, one can also find the (average) signal level of the 16/64-QAM relative to AM, at a certain BER. As an example, in Fig. 7 consider BER=10^{-8} and m_a = 4%. Then, from Fig. 7 the model gives m_q = 1.25%, about 10 dB ($20 \cdot \log_{10} m_q/m_a$) below the AM carriers, whereas the experiment [10] gives m_q = 1.53%, about 8 dB below the AM carriers.

In case 3, we show how the BER of M-QAM is further influenced by the OMI of AM. The case is shown in Fig. 8 using 64-QAM, with m_q = 0.75% which gives a BER of 10^{-9} with no AM carriers and is 13 dB below the AM carriers (m_a = 3%). For comparison purpose, Fig. 8 also includes the results obtained by assuming the clipping noise as Gaussian noise; the results are indicated by dashed lines.

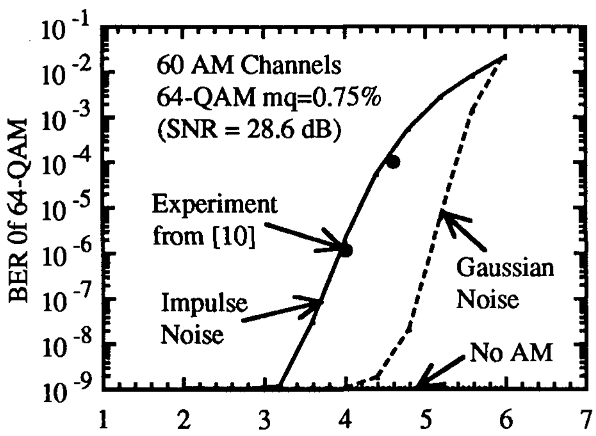


Fig. 8 BER of 64-QAM vs. the OMI of AM .

Fig. 8 shows that to have a required BER, the OMI of AM needs to be operated at a lower level than it would otherwise; e.g., the OMI of AM reduces from 4% to 3% for a 10^{-9} BER. One may also keep the AM level but increase the QAM level to obtain a desired BER. The trade-offs are that the former will reduce the CNR of the AM signals, while the latter will limit the number of QAM channels that can be applied. These trade-offs will be studied in the next section. Finally, we see that the BER obtained based on the impulse noise model agrees more closely to the experiments, and is much higher in value than predicted by the Gaussian noise model.

4. PERFORMANCE LIMITS AND DESIGN TRADE-OFFS

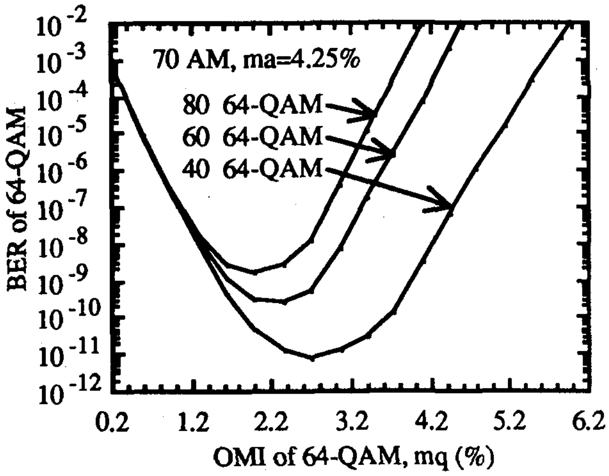
We have shown that our BER model is capable of predicting the BER behavior of M-QAM under various clipping conditions, and agrees reasonably well with the experiments. Here, we show that the model can be used to determine the transmission limits of M-QAM and perform trade-off studies.

Transmission Limits of M-QAM

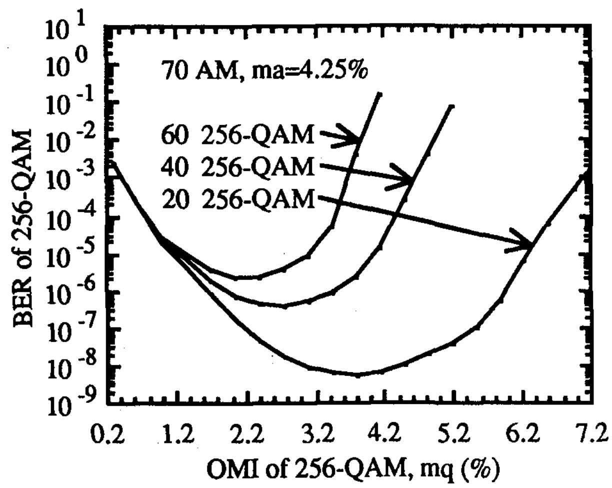
For Cable engineers, the limiting conditions of system parameters, such as the number of channels, OMI, and required optical power, are essential to system designs. In analog systems, CNR and C/NLD (or CSO/CTB) are used to obtain these transmission limits [1]-[4]. In digital systems, the BER is also required, in addition to CNR and C/NLD [8]. Analytically, the transmission limits may be obtained by minimizing the BER model of (13) with respect to, say, the OMI of M-QAM for a specified CNR and C/NLD. However, due to the complexity of the equation, there may be no closed-form solution. A simple means would be a graphic solution, as illustrated below.

In Figs. 9(a)-(b), we show the BER's of 64-QAM and 256-QAM, respectively, for different numbers of channels, multiplexed with 70 AM carriers. In each case, the AM OMI m_a = 4.25% which yields a CNR of 52.5 dB and a C/NLD of 60 dB for the analog channels with digital signals absent [4]. Both cases show that a minimum BER exists for each channel number. The reason is due to the fact that for small digital OMI's, the clipping effect from AM signals dominates the BER, and for large digital OMI's, then the clipping effect from QAM signals dominates the BER, as indicated by the figures. Hence, a minimal BER exists when the two effects are equal.

Thus, one way to determine the channel capacity is to first specify a minimum BER, and then to search by means of the BER model for an optimal OMI of QAM and a maximal channel number that meets the specified BER. For example, given a minimum BER of 10^{-9}, Fig. 9(a) shows a limit of eighty 64-QAM channels operating at 2% OMI, and Fig. 9(b) shows that less than twenty 256-QAM channels can be accommodated.



(a) 64-QAM



(b) 256-QAM

Fig. 9 BER of M-QAM as a Function of the Number of M-QAM Channels for (a) M = 64 and (b) M = 256. RIN = -155 dB/Hz. i_n = 12 pA/$\sqrt{\text{Hz}}$. r = 0.9 A/W. P_0 = 0 dBm. B=6 MHz.

Trade-Offs Among System Parameters

The trade-off study is another crucial step in system design. In a hybrid analog/digital SCM system, the trade-offs exist when we consider the interference of analog to digital signals or digital to analog signals.

For the effect of AM on QAM, as we have shown, the clipping of the AM signals yields significant BER degradations. To meet a specified BER, the OMI of M-QAM must be operated at a higher value than the no clipping case. This sets a limit for the number of digital channels to be used. For example, Fig. 10 shows that for an OMI of 256-QAM as small as 0.55%, there is virtually no BER difference between the one and one hundred channel cases. But, if a 2.5% OMI is used, the BER difference is visible even when 20 channels of 256-QAM are applied.

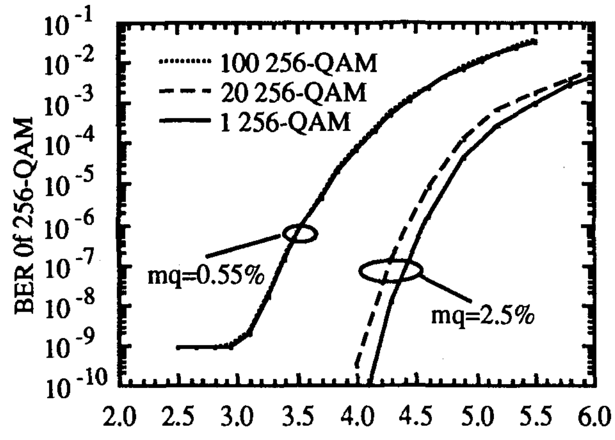


Fig. 10 BER of 256-QAM vs. OMI of AM (70 AM)

For the effect of QAM on AM, operating the M-QAM power (or the OMI) at large values in order to meet the required BER, will inevitably reduce the specified C/NLD (or CSO/CTB) of the AM signals. On the other hand, if the OMI of QAM is kept small, the signal level of AM (i.e., the OMI of AM) must be reduced to obtain the desired BER, as shown in Fig. 10. This will affect the CNR of the AM signals. Considering, again, Fig. 10, for BER = 10^{-9} a 20-channel 256-QAM operating at 2.5% OMI, which is only 4.5 dB below AM, would give a 5 dB C/NLD (or similarly CSO/CTB) reduction in AM channels. If, however, a 0.55% OMI is used, Fig. 10 shows that the OMI of AM must be reduced to 3% from the previous 4.25% to have the same BER of 10^{-9}. This implies a reduction of 3 dB in CNR.

In short, the above mentioned trade-offs for both analog and digital channels must be taken into account in a hybrid system design. One approach would be to first specify the CNR and C/NLD (or CSO/CTB) requirement for AM with a margin of a few dB's and the BER requirement for M-QAM, and then determine the other system parameters by means of the

BER model. We will illustrate this idea via the following design cases.

Design Study Cases

The architecture we consider is the popular fiber-to-the-feeder (FTF) architecture, with one fiber per node and one laser per fiber.

The objective is to provide for system designers and/or Cable engineers the reference values of system parameters involving both the analog and digital channels and the trade-offs among various parameters, given certain requirement specifications.

Specifically, we here consider the following scenario: given a link budget and the number of channels for both AM and M-QAM systems, how would a designer choose a laser diode and what sacrifice he has to make in order to meet the system specifications?

First, we look at what is given (the numbers are for illustration purpose):

1. **Link budget**; assume 8 dB link budget which would give a 20 km transmission with 0.4 dB/km link loss including fiber and splice loses.
2. **Channel capacity**; assume 70 AM channels and 80 M-QAM channels, making a total of 150 channels.
3. **Optical detector**; assume responsitivity r = 0.9 A/W and thermal current density $i_n = 12$ pA/√(Hz).

Next, we set up requirement specifications:

1. **Analog**: CSO/CTB ≤ -70 dBc (C/NLD ≥ 55 dBc).
2. **Digital**: (Uncorrected) BER ~ (on the order of) 10^{-5}.

Note that the first requirement is set based on the clipping distortion only; it is intended to leave some room for normal nonlinearity of a laser diode which when considered would bring down the CSO/CTB to -65 dBc or -60 dBc. The second requirement is set without any error correction such as coding applied.

Finally, based on the given conditions and the requirement specifications, we can then determine the desired analog and digital system parameters by means of the developed clipping distortion and BER models. The parameters may include the CNR and OMI of AM and the SNR and OMI of M-QAM. Note that iteration steps may have to be used to obtain the desired parameters that satisfy all requirements simultaneously. Table 1 lists a set of these parameter values obtained for two DFB lasers with different output powers and RIN's, respectively, using 64-QAM as an example.

8 dB Link Budget. 150 AM/64-QAM Channels

Lasers Opt. Received Power	4 mw -2 dBm		8 mw 1 dBm	
RIN (dB/Hz)	-150	-155	-150	-155
CNR of AM (dB) OMI of AM (%)	50.7 4.45	52.5 4.45	52.3 4.45	55.2 4.45
SNR of QAM (dB) OMI of QAM (%) Level Below AM (dB)	33.6 0.84 14.5	34.92 0.84 14.5	34.77 0.84 14.5	36.3 0.84 14.5

Table 1. System Parameters for Two DFB Lasers

From Table 1, we observe that (a) the OMI's of AM and QAM are unchanged in all cases, (b) for same RIN, the higher the laser power the higher the CNR (or SNR), and (c) the 4 mw laser with -155 dB/Hz RIN gives a slightly higher CNR (SNR) than does the 8 mw laser with -150 dB/Hz RIN. This implies that a low power laser with smaller RIN may be traded off for a better performance with a high power laser with a larger RIN.

Other trade-offs can also be made, such as reducing the analog or digital requirements. Table 2 gives two other cases in which the CSO/CTB ≤ -65 dBc or BER ~ 10^{-3} using the 8 mw laser with a RIN of -155 dB/Hz. As expected, we see that in both cases, the CNR and OMI of AM do get improved. In the former case, the SNR and OMI of 64-QAM also improve.

8 dB Link Budget. 150 AM/64-QAM Channels

COS/CTB BER	≤ -70 dBc ~ 10^{-3}	≤ -65 dBc ~ 10^{-5}
CNR of AM (dB) OMI of AM (%)	55.36 4.53	55.5 4.6
SNR of QAM (dB) OMI of QAM (%) Level Below AM (dB)	27.5 0.3 23.6	37.0 1.1 12.4

Table 2. Different Trade-off Conditions

As a final remark, we point out that the above cases use one set of parameters and conditions to illustrate the idea. Practically, one may need to use different system parameters and requirements to satisfy the designer's needs.

5. CODING TO REDUCE THE EFFECT OF CLIPPING DISTORTION/NOISE

All the BER's of M-QAM presented above, as an effect of clipping noise, were assumed uncorrected. We now show how coding would improve the performance.

Channel Coding - Error Correcting Codes

Channel coding employing error correcting codes are widely used in digital communication systems to reduce the effect of noise. Traditionally, coding has been used in an additive Gaussian noise channel. Here, we demonstrate that coding is also effective for reducing the impact of the clipping noise.

Specifically, we consider the use of a Reed-Solomon (RS) code [13], which is developed for multi-level (nonbinary) digital systems and is capable of correcting both the random and burst types of symbol errors. Consider a RS code with N total symbols and K information symbols. Then, the code is guaranteed to correct up to t = [(N-K)/2] symbol errors [13]. We denote this code as RS(N,K,t) code. The code rate of a RS(N,K,t) code is defined as $R_c = K/N$. Note that in a coded digital communica-tion system, the information data rate or the bandwidth of the system is increased by an amount of $1/R_c$ since channel coding introduces redundant symbols in the data. This poses another trade-off factor for the hybrid system design, as demonstrated below.

Decoded BER and Coding Gain

For RS codes, if hard decision decoding is used, the decoder output BER can be expressed by [13]:

$$BER_c = \frac{N+1}{2N^2} \sum_{i=t+1}^{N} \binom{N-1}{i-1} SER^i (1 - SER)^i \quad (13)$$

where SER is the symbol error probability at the decoder input. For M-QAM, the SER and BER (uncoded) are related by [13]

$$SER = \log_2 M \cdot BER \quad (14)$$

One important parameter to measure the benefit of coding is the coding gain, defined as the difference of the signal to noise ratios before and after coding, for the same specified BER. In a system design, this quantity can be utilized to determine a proper error correction code or codes to compensate certain BER loss.

Example One - BER Improvement

In this case, we consider a hybrid system of 70 AM carriers and one 64-QAM channel. We use a RS code with N = 63, t = 1 (single-error correction) and t = 2 (double-error correction), respectively. Both uncoded and coded BER's are shown in Fig. 11. It shows that even a one-error correcting RS code can reduce the BER significantly. For instance, at $m_q = 0.8\%$, the BER is reduced from 10^{-6} to 10^{-9}. The higher the error correction capability, the lower the decoded BER. However, the BER improvement will be traded off with an expansion of bandwidth as mentioned earlier.

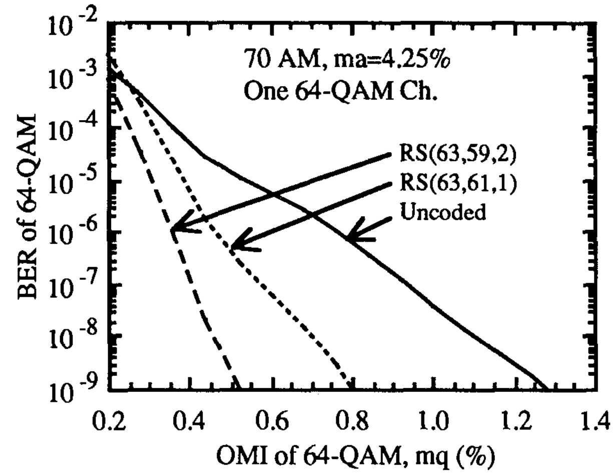


Fig. 11 Uncoded/Coded BER's of 64-QAM with 70 AM. The optical parameters are the same as those in Fig. 9.

From Fig. 11, we can obtain the coding gain at a certain BER. For example, at BER = 10^{-9} the RS(63,61,1) and RS(63,59,2) codes give 4 dB and 7 dB coding gains, respectively.

Example Two--Coding to Increase Capacity

In this case, we show that coding also increases the (digital) channel capacity. We consider a hybrid 70-AM and sixty 256-QAM system. For this case, we use a RS code with N = 255. Fig. 12 shows the results using double-error correction (t = 2). Also shown is the case of 75 channels of 256-QAM with the same 2-error correcting RS code.

We observe that the coding reduces the minimum BER, enlarges the workable BER range, and increases the number of channels. For example, the figure shows that the minimum BER is reduced from larger than 10^{-6} (uncoded) to less than 10^{-9} (coded). If a coded BER of 10^{-9} is required, seventy-five (75) 256-QAM channels can be used, implying a 15 channel increase compared to the uncoded case. Increasing the t value will increase the channel capacity. For example, with t = 3, i.e., a triple-error correcting RS code, then 125 channels of 256-QAM can be accommodated given the same BER of 10^{-9}; this gives more than twice capacity increase over the uncoded case. Again, the benefit has to be traded-off with the bandwidth expansion or an increase in the information data rate.

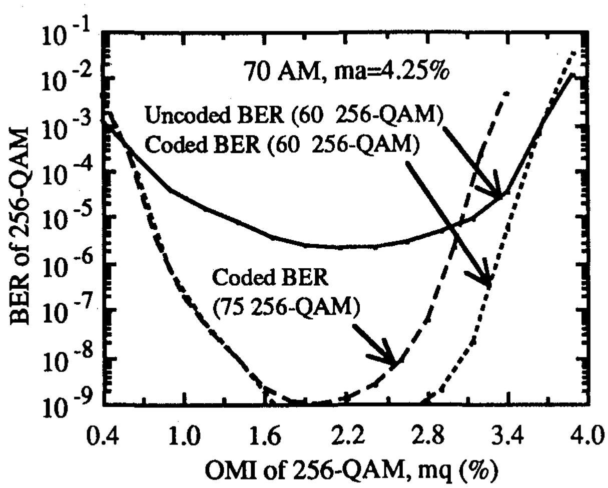


Fig. 12 Uncoded/Coded BER's of 256-QAM in hybrid AM/QAM. The parameters are the same as Fig. 9.

Finally, we note that concatenated codes [13], comprised of inner and outer codes, may also be employed in such environments. Using a concatenated code, particularly with a trellis code as inner code and a RS code as outer code [13], gives an advantage of no further bandwidth expansion in the inner code part and yet produces a lower BER, a higher coding gain, or a higher channel capacity than that using the RS code alone.

6. CONCLUDING REMARKS

We have developed a model that can be utilized to characterize the clipping-induced distortion impairments encountered in a hybrid SCM analog/digital fiber optic system. Specifically, the model is developed for Cable engineers to estimate both the analog and digital performances when implementing a hybrid fiber optic system. In terms of analog (AM) signals, the model estimates the C/NLD and CSO/CTB induced by the clipping distortion. In terms of digital signals, the model is then used to predict the BER of a digital channel, subject to the clipping-induced noise.

The results herein presented have shown that the model agrees reasonably well with the experimental results involving both the analog and digital signals. The results have also shown that in a hybrid AM/M-QAM system, the M-QAM signals may suffer a significant (BER) performance degradation due to the presence of the impulsive clipping noise. In certain cases, BER floors have been obtained. These effects have also been demonstrated in recent experiments involving hybrid optical transmission [9][10].

A major benefit of an analytic model is that one can use the model to evaluate system performance under various conditions, without conducting extensive experiments and computer simulations. To illustrate this benefit, the model, particularly, the BER model herein developed, has thus been employed to determine the transmission limits, such as channel capacity and OMI's, and to perform trade-off analysis for both the AM and M-QAM signals. A design example has also been presented using practical system parameters and requirements. The results presented indicate that as a result of the impulsive non-Gaussian clipping noise, only a limited number of M-QAM channels can be accommodated if high-level M-QAM signals, such as 64-QAM or 256-QAM, are used. Further, the transmitted power of M-QAM must also be higher than in the Gaussian noise case. The trade-off analysis and the design example show that both the analog and digital requirements (e.g., CNR, C/NLD or CSO/CTB, and BER) have to be carefully balanced for doing a hybrid optical transmission system design.

Means have also been sought to reduce the effect of the clipping distortion/noise. In particular, we have shown that applying coding techniques, i.e., error-correction techniques, can effectively reduce impact of the clipping noise on M-QAM signals and therefore improve the BER performance significantly. The examples presented using the RS codes show that the BER of M-QAM (e.g., 64-QAM) can be reduced significantly even with a single-error correcting RS code. The results further show that coding also increases channel capacity and enlarges the workable range of BER (see the 256-QAM example). The benefit of coding must, however, be sacrificed with an expansion of bandwidth or information rate increase. The trade-off should also be considered in system design stages.

Finally, we note that another digital modulation scheme known as multi-level vestigial sideband modulation (M-VSB), has been developed [14] and is becoming popular in Cable community, due to some advantages over the M-QAM system. The M-VSB system still employs amplitude modulation and exhibits similar BER behavior to M-QAM, though a slight BER improvement is observed [14]. Hence, M-VSB will still be subject to the impairments of the clipping distortion/noise when it is used in a hybrid AM/M-VSB fiber optic environment. Hence, our model herein developed for M-QAM may also be modified without great difficulty to deal with the M-VSB signals. This work is currently under way in our laboratories.

REFERENCES

[1] A.A.M. Saleh, "Fundamental limit on the number of channels in subcarrier-multiplexed lightwave CATV system," Electron. Lett., 1989.

[2] T.E. Darcie, "Subcarrier multiplexing for lightwave networks and video distribution system," IEEE J. Selec. Areas. Commun., 1990.

[3] K. Alameh and R.A. Minasian, "Ultimate limits of subcarrier multiplexed lightwave transmission," Lectron. Lett. 1991.

[4] Q. Shi, *et al.*, "An alternative model for laser clipping-induced nonlinear distortion for analog lightwave CATV systems," IEEE Photonics. Technol. Lett. 1992.

[5] N.J. Frigo and G.E. Bodeep, "Clipping distortion in AM-VSB CATV subcarrier multiplexed lightwave systems," IEEE Photon. Technology Lett. 1992.

[6] M. Tanabe, *et al.*, "80 channel AM-FDM TV signal optical transmission system," Nat. Tech. Rep., Japan, 1990.

[7] C.J. Chung and I. Jacobs, "TV channel capacity of lightwave multichannel AM SCM systems as limited by laser threshold nonlinearity," OFC'92, Tech. Digest.

[8] Q. Shi, "Performance limits on M-QAM transmission in hybrid multichannel AM/QAM fiber optic systems," IEEE Photon. Technol. Lett. 1993.

[9] K. Maeda, *et al.*, "BER characteristics of 16-QAM in an AM/16QAM hybrid optical transmission system," OFC'93 Tech. Digest, also in Electron. Lett. 1993.

[10] X. Lu, *et al.*, "Impulse-induced bit-error impairment in AM-VSB/64-QAM hybrid optical transmission," OFC'94 Tech. Digest.

[11] Q. Shi, "Asymptotic clipping noise distribution and its impact on M-ary QAM transmission over optical fiber," to be published in IEEE Trans. Commun. 1994.

[12] J.E. Mazo, "Asymptotic distortion spectrum of clipped, dc-biased Gaussian noise, " IEEE Trans. Commun. 1992.

[13] J.G. Proakis, *Digital communications*, 2nd Ed. McGraw-Hill, 1983.

[14] R. Citta and R. Lee, "Practical implementation of a 43 mbit/sec (8 bit/Hz) digital modem for cable television," NCTA Paper, 1993.

IMPLICATIONS OF TELEPHONY SERVICE ON BROADBAND RF CABLE PLANT DESIGN

Marty Cowen
Scientific-Atlanta

Abstract

Recent events in the telephone and cable industry have included plans for the distribution of telephone services on broadband RF cable plants. This paper will discuss the effects that this service will have on new and existing plant designs. Implications on such things as system architecture, fiber node sizes, system powering, return bandwidth, drop integrity, and plant distortion requirements will be reviewed.

TELEPHONY REQUIREMENTS

System Components

A cable telephony system consists of headend interface units (HIU) and customer interface units (CIU). (See Figure 1) The HIU must interface with a telephony switch, transmit signals in the downstream path, and receive signals in the upstream path. In addition, this is where the control system, used to provision and monitor the system, would reside. The CIU must receive and process the downstream signals, transmit in the upstream path, provide a standard 2 wire telephone interface to the customers in house wiring, and provide ringing voltage.

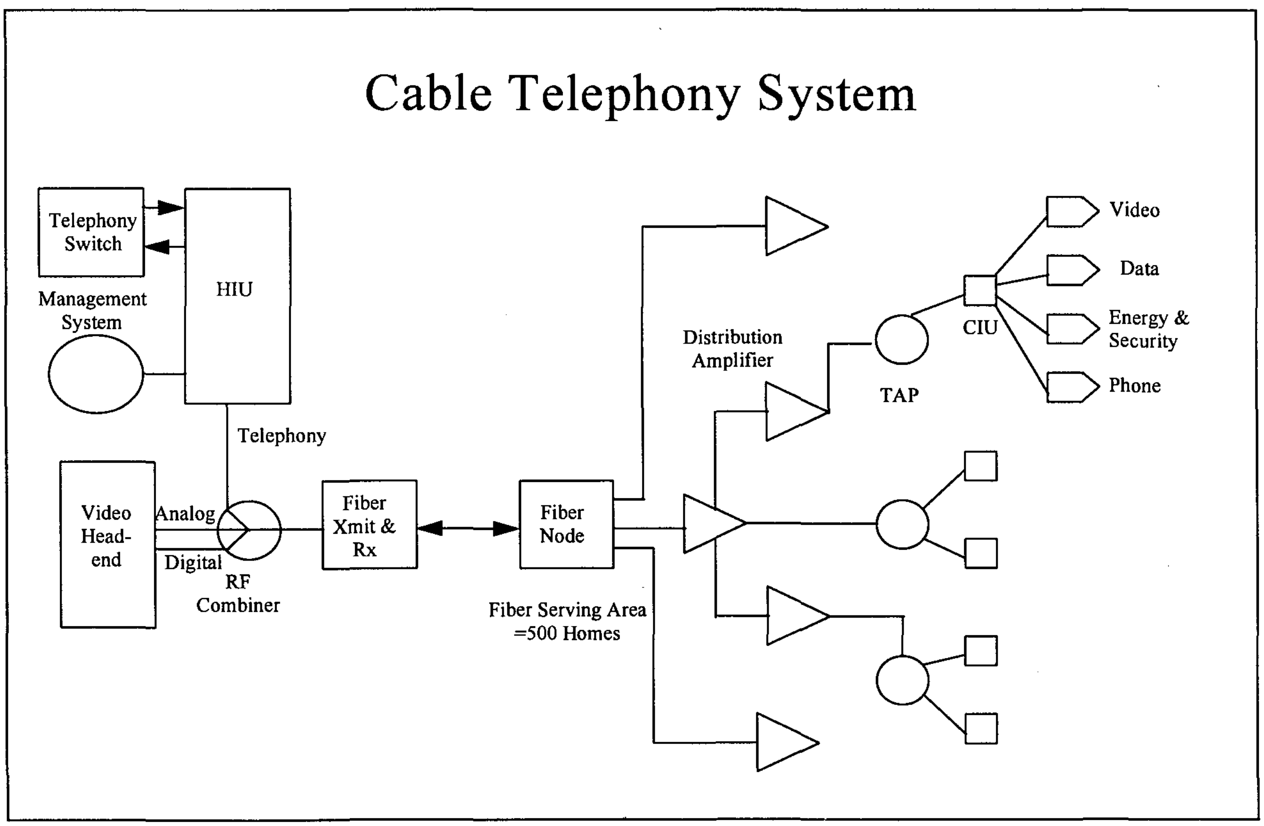


Figure 1

In between these two components is a Fiber to the Serving Area (FSA) distribution system. Since the reverse band of each fiber node can be processed separately at the headend, this architecture allows for the required amount of reverse bandwidth per subscriber.

Type of Service

Before deciding on cable plant requirements, the first question that must be answered is what level of telephony service is required. The Regional Bell Operating Companies have strict requirements in the areas of service downtime and line concentration, among others. A cable operator who desire's to offer the primary phone service may or may not have to meet these requirements. Teen lines, Fax lines, and modem lines could be offered, which might not fall under the same restrictions as the primary phone service due to the implications of emergency services. These requirements will also vary from country to country. Another possibility is T1 service for business applications. The integration of a cable telephone system with utility management functions would also be a natural fit. This paper will assume that the operator desires to offer a level of service equal to the current phone service in the United States.

Some of the main requirements placed on the carriage of Plain Old Telephone service (POTS) are in the area of downtime. Today's twisted pair local loops have extensive backup networks and typically maintain phone service for 4 hours or more after the power is first interrupted. This would require an extensive investment in backup power supplies to achieve this in a broadband plant.

SYSTEM REQUIREMENTS

Architecture

Of the recent technological advances in broadband RF plant designs, the switch from tree and branch style architecture to FSA is the one most responsible for making two way communications possible. This is because it brings with it the ability to separate the reverse paths by fiber nodes and reduce the amount of noise funneling that occurs in trunked systems. Another advantage of this type of system is the increased reliability brought about by decreased cascade lengths. Recent studies have shown that FSA architectures have achieved downtimes of less than 25 minutes per year per subscriber.[1]

Downstream Path

The additional requirements on the downstream path are minimal compared with the upstream path. Cable systems have been transmitting signals in this direction for many years and advances in equipment bandwidth and video compression are allowing more and more data to be carried in this direction. In a cable telephony system, each fiber node will carry only the calls for that node. Since the downstream path does not have inherent problems with ingress and interference, a higher order modulation type can be used in order to get a higher bits-per-Hertz ratio. An approach using QPR modulation could fit 480 DS0's in 18 MHz. (A DS0 is a standard single call, 64 kB/sec, signal format.) This 18 MHz would not have to be contiguous but could be split into 3 MHz bands so that two could fit in any available video channel. The effect on the distortions of the downstream path is minimal since these signals can be carried 7 dB below the level of video carriers.

Upstream Path

The reverse path frequency bandwidth will be the limiting factor in determining the fiber serving area node size. Due to the harsh realities of the return band, it is desirable and necessary to use a robust modulation scheme such as QPSK. As you can see from Table 1, this type of modulation does not offer the bits-per-Hertz

Modulation Type	Required E_b/N_0*	Bits per Hertz
QPSK	15.7 dB	2
16 QAM	22.6 dB	4
64 QAM	28.8 dB	6

* E_b/N_0 is defined as bit energy to noise power spectral density and is proportional to C/N based on the noise BW and bit rate of a given system. [2]

Table 1

performance of some higher order modulation schemes, however it does offer excellent performance with minimal carrier to noise (C/N). Also, since certain portions of the reverse band may not be useable at all times, it is desirable to use many relatively small bandwidth carriers. This will minimize the impact of any interference on the total number of carriers available for use at any one time. These carriers should also be agile in nature in order to deal with the variable nature of the reverse band ingress. Assuming all these things to be true, the reverse path then consists of agile QPSK modulated carriers, each dedicated to one customer call in progress, that are spread across the available reverse bandwidth. The amount of available reverse bandwidth is then directly proportional to the number of calls on the system at one time.

Recently cable distribution equipment manufacturers have announced plans to offer amplifiers with an increased amount of reverse bandwidth. The standard subsplit offered 25 MHz of bandwidth from 5-30 MHz. The expanded subsplit will offer 32 MHz of bandwidth from 5-42 MHz, a 48% increase. This does not come without tradeoffs. Decreasing the filter crossover region will require higher order filters which will cause increased channel 2 rolloff and increased group delay. But, since the FSA designs have short cascades of amplifiers this should be a manageable problem. Any other options that require reverse bandwidth also need to be considered. Services such as Status Monitoring, Impulse-Pay-Per-View, and any of the expanding number of

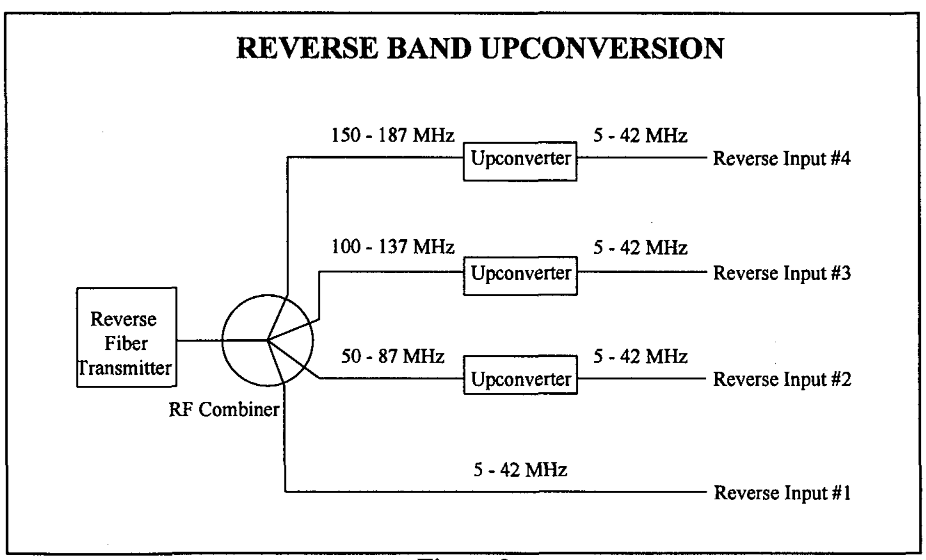


Figure 2

interactive services will affect the available bandwidth, and, therefore, the available number of calls.

Reverse Band Upconversion

One possible way of increasing the available return bandwidth per fiber node, is to separate the return bands that are being fed into the return laser and to block convert them up in frequency (see figure 2). This would allow a 2000 home node to be able to utilize its return band in 500 home pockets. The return bands at the headend would then have to be downconverted before being processed. This would require extra hardware at the fiber node and the headend, but this equipment could be less costly than a fiber link. The required downstream bandwidth in this scenario would increase four times, from 18 MHz, to 72 MHz per node.

Noise Funneling

As was stated earlier, the reverse path is a harsh environment. This has as much to do with interfering carriers as it has to do with noise funneling. The noise funneling that occurs is governed by the following equation:

$$\text{Noise} = 10\log N + C/N$$

where N = number of amplifiers per node and C/N = the carrier to noise of each amplifier (assuming they are all the same type of amplifier set up in the same configuration). For a typical 500 home node, the number of amplifiers = 4.5. This yields 6.5 dB of additional noise due to just noise funneling of the upstream coaxial plant.

Return Lasers

Due to the high cost of DFB lasers, Fabry-Perot lasers are currently being used in the return path. This type of laser has shortcomings in the area of noise and distortions. The C/N of a typical upstream system is set by the Fabry Perot laser. The C/N could be improved with the use of a DFB laser but this would cost approximately twice as much. However, if reverse band upconversion, as described earlier, is used, this may be required due to the increased number of carriers transmitted. In the future an intermediate grade of laser may be required with performance and cost between that of a DFB and Fabry-Perot laser.

Drop Integrity

Another source of noise in the return path is the customer premise drop. To minimize ingress in the reverse band, special care will need to be taken to ensure that the drops are of good integrity. This means the use of adequately shielded cable with tightly installed connectors. Recent studies have shown that by placing reverse band stop filters on the areas of the customer premise drop that do not require reverse transmission capability, the average noise returned to the headend was decreased by greater than 7 dB.[3] While this may not be feasible due to the increasing use of the reverse band in the home, it does show that if care is taken to limit undesirable ingress from the home, significant gains could be made.

POWERING

System powering will be a very important issue to any cable telephony system. The system could be customer powered or network powered, with or without backup capability. A strong case can be made for a battery backed up network powered solution for any system that is truly going to offer the type of service expected by consumers in the United States. Home powering will be resisted by consumers because it means a box in their home and because they do not need it with the present telephone service. Home powering is unacceptable without battery backup because it fails to offer lifeline service. The first time someone is unable to call 911 when the power is out will be the beginning of the end of that type of system. Home powering with battery backup would provide an unimaginable nightmare in replacement and disposal of rechargeable batteries. A network powered solution would require the same type of backup systems that are currently offered for cable systems today. Network power also would require the device to reside outside the home in order to meet the National Electric Code.

To get the power to the box located on the side of the home a twisted pair/coaxial cable combination could be used. (See figure 3) This would eliminate the contact corrosion that can occur when power is carried on the coax drop cable. While this would require the replacement of the drop cables, or the addition of a twisted pair cable in current systems, it would allow for the operators to ensure the integrity of these drops. This would also require the addition of a power converting device at the tap locations or a replacement of the RF only taps with taps that pass RF and Power.

To estimate the increased power demands for a cable telephony system, we will assume the following power requirements for the Customer Premise Equipment.

Ringing	7 Watts
Talking	5 Watts
Sleep Mode	2 Watts

A system with 500 home nodes, 100% penetration, and all subscribers off hook at

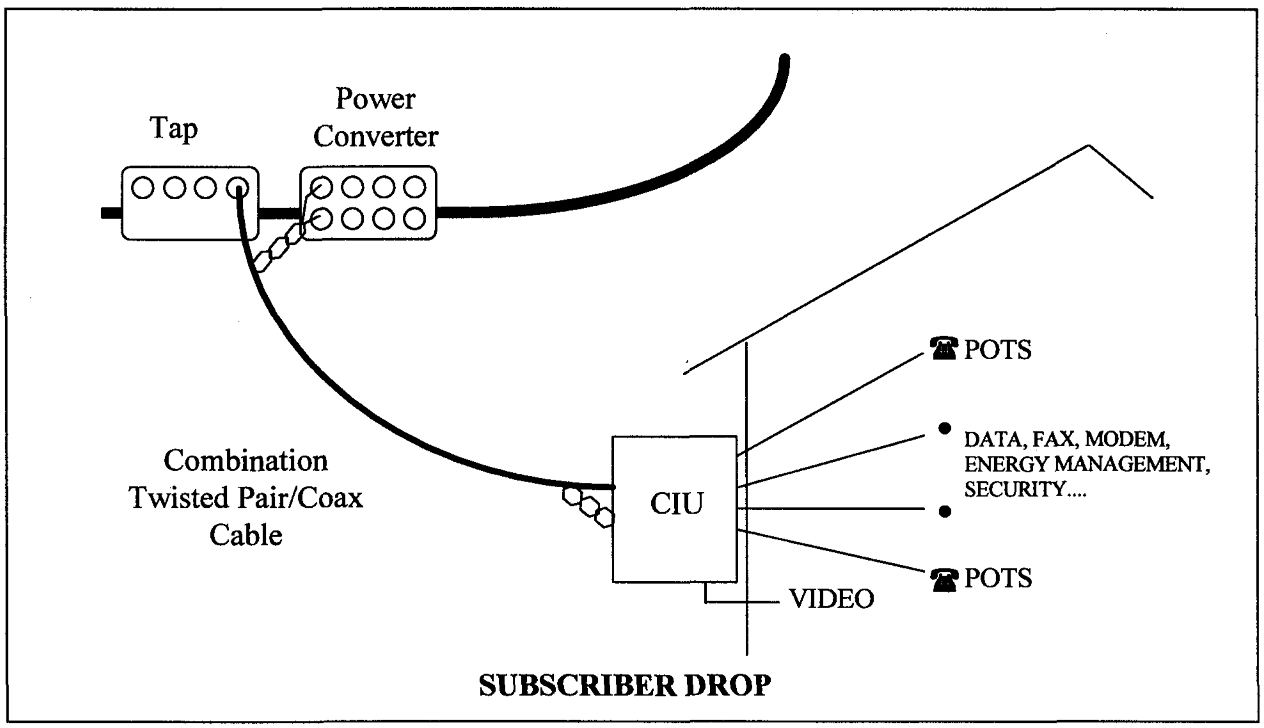


Figure 3

once, would then require 2500 more watts per node. This would require three additional 15 Amp backup power supplies.

CONCLUSION

There are several requirements placed on the design of a broadband RF cable system in order for it to be able to carry telephone signals. The primary requirement is the use of a Fiber to the Serving Area distribution system. Other requirements include:

- Available bandwidth in the forward and reverse paths
- An upstream band with sufficient integrity
- A power system capable of powering the customers premise equipment with battery backup

All of these requirements have feasible solutions and their implementation will result in a quality service that will be accepted by the consumer.

REFERENCES

[1] R. Pinkham, "Combining apples and oranges: The modern fiber/coax network", Telephony, pp. 28-32, February 7, 1994.

[2] B. Sklar, "Digital Communications Fundamentals and Applications", 1988, Prentice Hall, Englewood Cliffs, New Jersey 07632

[3] M. Dzuban, AT&T Return Path Study, CableLabs Return Path Conference, February 15, 1994.

IN-HOME INTELLIGENCE FOR INTERACTIVITY

DANIEL M. MOLONEY
VICE PRESIDENT, ADDRESSABLE SYSTEMS

ABSTRACT

Set-tops have played a critical role in providing entertainment to the home in the cable industry. As the advent of advanced technology has created an opportunity to provide additional services to consumers, the need for a fully integrated two-way network to support these services has arisen. The issue of intelligence in the system is a key element of interactivity, and one which has been at the center of a great deal of debate.

This paper will look at the various components of the two-way network and will focus on the advantages of placing intelligence in the home to provide interactivity. The data communication network will be presented, along with the technology required to enable all levels of interactivity in the home. The potential interactive services will also be covered, with focus on introducing interactivity in the analog environment and migrating to the digital environment.

INTRODUCTION

The race to evolve the information highway from a one-way broadcast oriented network to a two-way interactive infobahn has begun. Fueled by the ongoing improvements in fiber optic technology, the advent of cost effective, high quality digital compression technology, and the reduction in the costs of computing power and graphics, the movement to the information superhighway has gained significant momentum. As a result, the broadband "cable" network has moved to center stage and decisions made today about what technology to deploy will have significant ramifications for years to come. While few people believe this evolution will occur overnight, the failure to recognize and respond to this paradigm change now will be tantamount to standing at the dock after the boat has left.

Because there are many divergent forces pulling the interactivity movement in different directions, the evolution path is not crystal clear. Tests are being conducted in various markets throughout the world in an attempt to find the "holy grail", the ideal compilation of on-line, off-line, and home control applications. Many of the tests are approaching the situation from the marketing standpoint of searching for the "killer application", but no one has yet to find it (if one exists at all).

This paper will also focus on issues surrounding the evolution to interactivity and the requirements of the network and in-home technology. A brief summary of the network topology will be given, followed by a discussion of digital compression technology. The focus will then turn to the use of computing power in a distributed intelligence architecture as a means to provide the level of support for interactivity with the TV. While the essence of the paper will be on interactivity with the TV, many of the concepts also apply in multimedia to the PC.

NETWORK ARCHITECTURE

Traditional networks constructed over the past several years have been designed and built to serve current levels of programming, that is a one-way broadcast delivery of entertainment services to the home. This typically entailed the use of some fiber optic technology in a backbone configuration coupled with a standard coaxial based distribution network. While the economic benefits of this approach were quickly evident to network operators, long-term architectural growth capability was not provided. Expanding beyond this level of fiber optic utilization typically required a major rebuild due to the need to splice fiber and relocate distribution amplifiers.

As operators began to look at the potential for adding new revenue generating services such as two-way impulse pay per view, a new architecture took hold. This type of distribution network , commonly known as Fiber -to-the-Feeder (FTF), provided the reduced node sizes necessary to effectively carry data upstream in a sub-split system. While the cost of such an approach could exceed the traditional network costs by 10 - 20 %, the long-term benefits in flexibility justified the additional investment. It also required the network could be rebuilt only one time, with an easy migration path to a more advanced network.

A number of FTF architectures have been proposed, each with its trade-offs between cost and flexibility. One such approach, known as the Broadband Telecommunications Architecture (BTA), enables the system to be designed for further segmentation by strategically placing the node locations initially and laying enough dark fibers to handle the potential traffic in the future (see figures 1 & 2). As the traffic volume from the two-way interactive services grows, the BTA system can be segmented further by reducing the node size both downstream and upstream. Node sizes down to 500 homes downstream and 125 homes upstream can easily be facilitated to ensure no contention on the network, regardless of the type or level of services available.

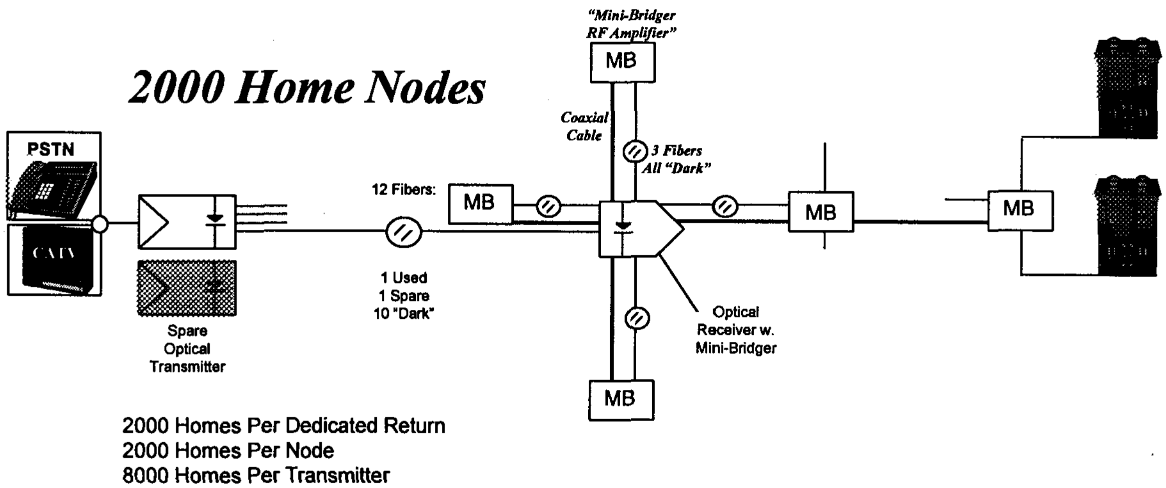


Figure 1 - BTA Stage I

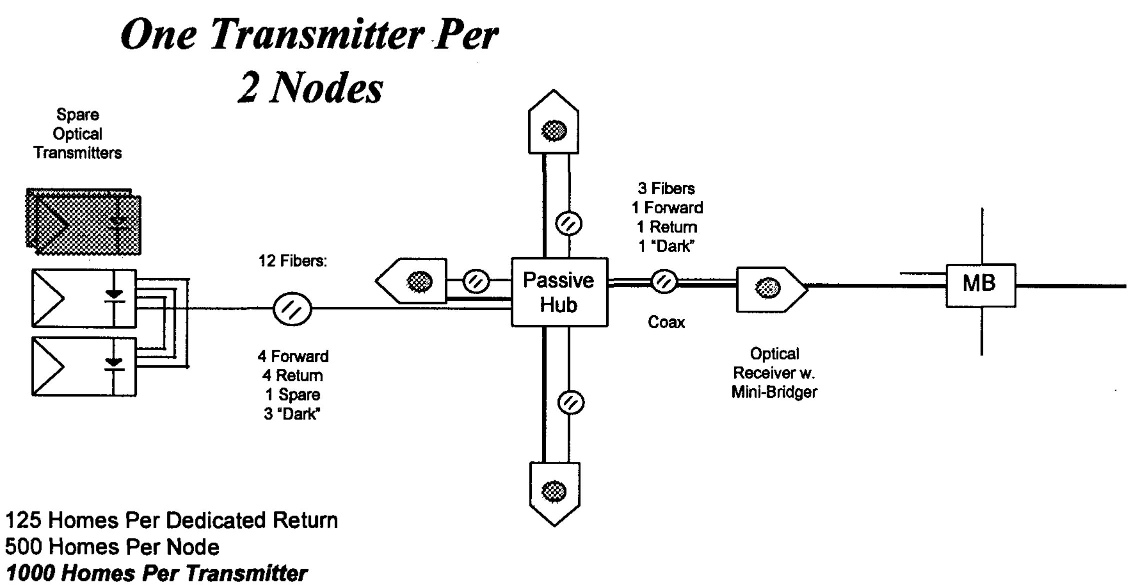


Figure 2 - BTA Stage V

By reconfiguring the network, the evolution to a full range of interactive services can be facilitated. When analyzing the requirements of interactivity, it is useful to develop a system for categorizing them. For simplicity, four types of interactivity have been defined. Type one interactivity is broadcast in nature and requires no upstream signaling. Basic applications such as electronic program guides fall into this category. Type two interactivity adds a non real-time upstream path (14/64 Kbps) for the collection of purchase data on impulse pay per view and impulse home shopping.

Type three interactivity introduces the concept of a client / server network, but still in an asymmetric environment. This requires real-time signaling downstream (1.5 Mbps) and upstream (64 Kbps) and intelligence at both ends of the network. Movies and information on-demand are examples of this level. The final level of interactivity adds the dimension of symmetry to the network as is found in telephony and video telephony. The complications of supporting this level of service will not be fully addressed in this paper in order to maintain the focus on more near-term applications, but in general they represent a more complex overlay of level three interactivity.

DIGITAL COMPRESSION

While the development of the network architecture will facilitate growth in interactive services, the advent of digital compression provides the high quality, bandwidth efficient transmission method necessary to carry all the video, audio, and data information to the consumer. Digital compression is not a necessary element for basic interactivity, but it provides a robust structure to facilitate the vast array of advanced level three services such as information on demand and downloaded video games.

Digital compression offers the benefits of bandwidth efficiency (2 - 10 programs per channel), along with CD quality digital audio and several high speed data pipes per channel. Real-time on-demand interactivity requires the ability to retrieve and playback information as required from a central server. With the use of digital compression technology for storage of the entertainment and information programming, true video on demand becomes feasible. These massive storage banks of information servers will enable the consumer to gain control of a virtual channel and simulate the point to point communication found on telephone networks, but at significantly higher speeds.

COMPUTING TECHNOLOGY

The third element in building the interactivity infrastructure, computing power and advanced graphics, has benefited from the rapid acceleration in the PC industry cost curve. Questions have been raised, however, regarding the need for advanced in-home intelligence. Several proponents have proposed a centralized intelligence network only, with only "dumb" data modems in the home for communication purposes. While this model appears more cost effective at first glance, a balanced approach between in-home and central intelligence offers the best long-term solution.

By examining the usage patterns across the network, the proper balance becomes evident. The first application evident to the user is the navigator, or user interface to the system. Because this service is likely to be accessed by many people simultaneously, at least a portion of the functionality must be maintained within the set-top to minimize the impact of the network. The need to wait for the set-top to "boot up" this application will not be tolerated by the consumer.

As the level of functionality expands beyond basic interactivity, this level of client/server balance continues to be critical. For example, it is not economical for a consumer to store a complete movie within their home (>1 GIGA byte). Therefore, in this application, only the navigator for the VOD service would be stored locally, while the movie itself is playing real-time on the network.

Games represent another area where a balance in intelligence between the client and server is critical Video graphics and processing technology is becoming more economical, but with the advent of 32 bit games, too much memory will be required in the home to store an entire game. Therefore, it is likely that a portion of the game will be downloaded to the set-top. As the user progresses through the various levels of the game, different portions will be downloaded transparent to the user, thereby eliminating any of the latency concerns on the network.

SET-TOP EVOLUTION

Traditionally, the set-top served as a device to enable access to a variety of broadcast services and limited two-way services such as PPV. High speed computing power and advanced graphics were elements of the PC environment and not the TV entertainment environment because the costs were too significant to justify the investment by the consumer and the typical applications did not require such capabilities. By coupling this technology with video from a digital compression network, however, consumers can now have access to a vast array of applications, from level one to level four.

Because virtually none of the services envisioned on this information superhighway have a proven track record, the capital required at both ends of the client / server network for mass deployment cannot be justified today. Therefore, the need for a proper architecture in the set-top from initial deployment is mandatory. To handle this requirement of multiple featured set-top products in the same system, a modular approach to the design of the set-top is mandatory. By developing a basic platform, both for the analog product as well as the digital product, and then enabling a PC technology upgrade module, the set-top can continue to grow and cost effectively change as various applications justify the incremental capital costs.

COMPUVERTER APPROACH

The CompuVerter combines the low cost analog and digital set-top platforms with a family of computer technology upgrade modules (see figure 3). The basic platform contains expanded bandwidth tuning (1 GHz), bit mapped on-screen display graphics, expanded two-way communication and navigation systems, network security, and digital compression The modular interactivity upgrade then adds the high capability processor, an optimized graphics subsystem for video games, synthesized speech, downloadable operating system(s) and applications, and an interface to in-home peripherals, both wired and wireless.

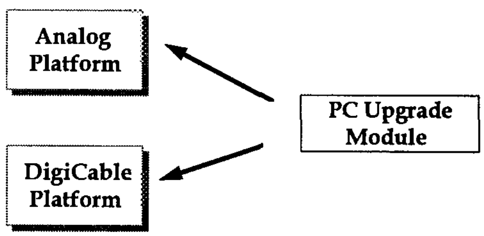


Figure 3 - CompuVerter Platforms

The CompuVerter enables a vast array of interactive applications to be provided to the consumer, from the basic one-way broadcast type through the full client / server interaction type. Services likely to be provided on the network initially include a navigational gateway combining live video and sophisticated graphics, augmented video programming such as play-along games, enhanced home shopping (virtual shopping malls), information and video on demand, and downloaded video games.

The rollout of the CompuVerter in the network will clearly be driven by the need for advanced capability. Most people agree the initial application will be a navigator, probably starting in the form of an electronic program guide. With the proper planning, however, the intelligence in the home for the guide can also serve as the basis for other interactivity By coupling this with the basic type 3 applications such as VOD, an economic justification for in-home intelligence can be made (see figure 4)..

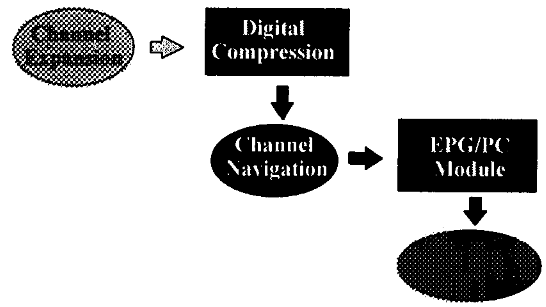


Figure 4 - Technology Evolution

CLIENT / SERVER NETWORK

The CompuVerter will serve as the foundation of the client / server network for interactivity. Several other elements of the distributed intelligence architecture are the servers, switching devices, network controllers, and downstream data generators. The following diagram outlines the logical connection of these various devices in the network.

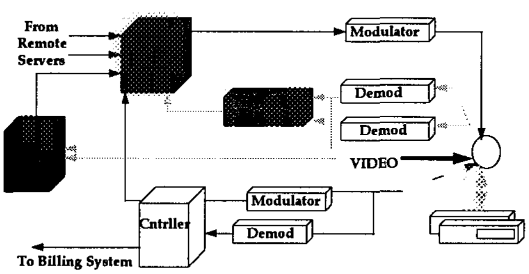


Figure 5 - Interactive Network

While this paper does not cover the functionality of each device in detail, it is important to realize the complexity of the headend interfaces in the client/ server environment. In order to facilitate seamless interactivity to the consumer, network control management is critical. More intelligence in the home will alleviate some of these issues, but the timing considerations between the set-top and headend will demand a fully integrated network.

CONCLUSIONS

The race toward the interactive world has begun. While some people have stated this will be a revolution, the technology required has already begun to be deployed in a more cost-effective evolutionary path. First, fiber rich architectures such as the BTA lay the foundation for the level of traffic which will be found on the highway. To efficiently carry the information to and from the home, a digital compression system will provide the ideal vehicle. When coupled with advanced graphics and processing technology from the PC industry in a client / server architecture, basic applications will be transformed to sophisticated multimedia environments.

The in-home component of the network will turn out to be the most critical element. With the modular design of the CompuVerter, a portion of the capital investment decision can be postponed until real market data on consumer demands can be assessed. The need for an intelligent in-home component of the network is more critical than ever in an interactive world due to the need to ensure flexibility, security, and growth. The CompuVerter provides such capability in a cost effective approach without limiting the potential for future applications. It also enables the entry into interactivity in an analog world and the growth in the digital world of the future.

INTEGRATING NEW TELECOMMUNICATION SERVICES INTO THE BROADBAND NETWORK

Andy Paff, President
Executive Vice President, Strategic Planning and Technology
ANTEC Corporation
8101 East Prentice Avenue, Suite 210
Englewood, CO 80111
Phone: 1-800-FIBER-ME
FAX: (303) 694-0127

ABSTRACT

The cable television industry is just beginning the process of interconnecting disparate headends within a geographic region. The primary economic drivers relate to the existing core business and tend to be "broadcast" in technical configuration. This paper assumes that Synchronous Optical Network (SONET) standards are deployed as the transmission backbone network. The focus will be on the integration of likely new services into this digital platform. Three general areas will be addressed:

1) Telephone and related applications
2) Entertainment video (new applications)
3) Multimedia

The regional network will provide the primary core for this examination, although some discussion of integration into the RF residential network will be discussed.

INTRODUCTION

The cable television network architecture has been driven by the on-going requirements to deliver new services. Until recently, this evolution has been restricted to the residential network and the need to broadcast more NTSC channels in a point-to-multipoint configuration. The future seems dominated by opportunities involving:

- Interactivity
- Interoperability
- Broadband digital transmission

For the cable television operator, the implications associated with these new opportunities are enormous. While positioning for the future is essential, today's core business still revolves primarily around broadcast video entertainment services. The challenge is to leverage today's current business while positioning the network to be capable of handling future telecommunications requirements. This task has been made more difficult due to the reduction in expected core business cash flow through recent government regulation; however, the urgency continues to be elevated by the growing group of pending competitive threats to the cable television industry.

As fiber optics dramatically changed the residential network from a "tree and branch" structure to a fiber/coaxial "star-bus" architecture, the perception grew that cable television's broadband platform would provide the most cost effective and flexible solution for the future. Cable operators in 1994 are defining "foundation" elements of the network, such as passband, node size and migration strategies, while more complex issues (e.g. modulation format, set-top interoperability) are reviewed. It is critical to define these foundation elements (physical layer) in order to move the process forward. In similar fashion, the regional interconnect must define common elements to ensure future migration into non-broadcast services.

SONET AS THE REGIONAL INTERCONNECT FOUNDATION PLATFORM

It is beyond the scope of this paper to review how SONET will be technically configured to provide cost effective video transport for the cable television business today. There are several papers available for review on the subject.[1]. The critical issue for this discussion is that the regional interconnect should be designed today with the full expectation of being the transport layer for a variety of future applications. Interoperability between networks and each network element should be constant themes in planning the digital transmission platform. SONET standards represent several important advantages:

- SONET is a well defined, open standard assuring interoperability between vendors and networks;

- SONET intrinsically provides critical features such as ring configuration, drop and insert, drop and pass, etc.;

- SONET is the defined physical transport layer for Asynchronous Transfer Mode (ATM), which will become the interoperability standard for wide area network (WAN) and multimedia applications;

- SONET will facilitate interconnection to the public switched telephone network for new applications such as competitive access and interactive residential services;

- SONET provides a defined operational support protocol (TL1) and the potential to integrate network operations and support systems.

There are many challenges related to interoperability as the regional network develops. Cable operators will generally be looking for greater efficiency through network integration and consolidated operations. Proprietary systems in the regional interconnect or the residential network may offer short term capital savings but will quickly become impossible to manage as discrete elements. Many new applications will become virtually impossible to offer cost effectively through a patchwork network.

What are these opportunities and how will they be integrated into the SONET platform? Following the "convergence" model, this paper will examine three areas:

- Telephone
- Entertainment Video
- Multimedia

APPLICATIONS

Telephone

SONET was initially developed to operate within the public switched telephone network (PTSN). Telephone applications will grow out of the embedded base of equipment based on the North American Digital Hierarchy Standards (DS-N), in addition to SONET. Figure 1 shows the basic granularity of each.

Virtual tributary (VT) channels will be used to map DS-1 to SONET. The VT channels at various rates will be important for "mapping" digital applications to the SONET frame at optical carriers, OC-1 or OC-3. Equipment to digitally convert individual voice channels and multiplex the digital equivalent (DS-0) to higher level is well established in the PTSN; the PTSN also has a sophisticated switching infrastructure in place.

The cable operator, seeking incremental revenue opportunities, will utilize the regional SONET platform as leverage to extend the scope of the "telephone" opportunities. The two most often discussed applications are commercial business applications (competitive access) and residential telephony (voice and video).

Commercial Applications

Competitive access is currently a $300 million dollar business annually.[2] It is limited primarily by individual state regulation and economics. Generally, competitive access involves bringing business customers directly to their long distance carriers, instead of utilizing the local exchange carrier (i.e., the local telephone company). The economic issues center around the cost to serve the majority of the access universe. Only about seven percent of the total access revenue universe involves large businesses while most of the access fees come from residential customers and smaller businesses, a much larger market. But the cost to reach these small customers is difficult to justify in light of the revenue generated. The cable operator will enjoy a competitive advantage if the SONET backbone network is in place and justified through core business applications (Figure 2).

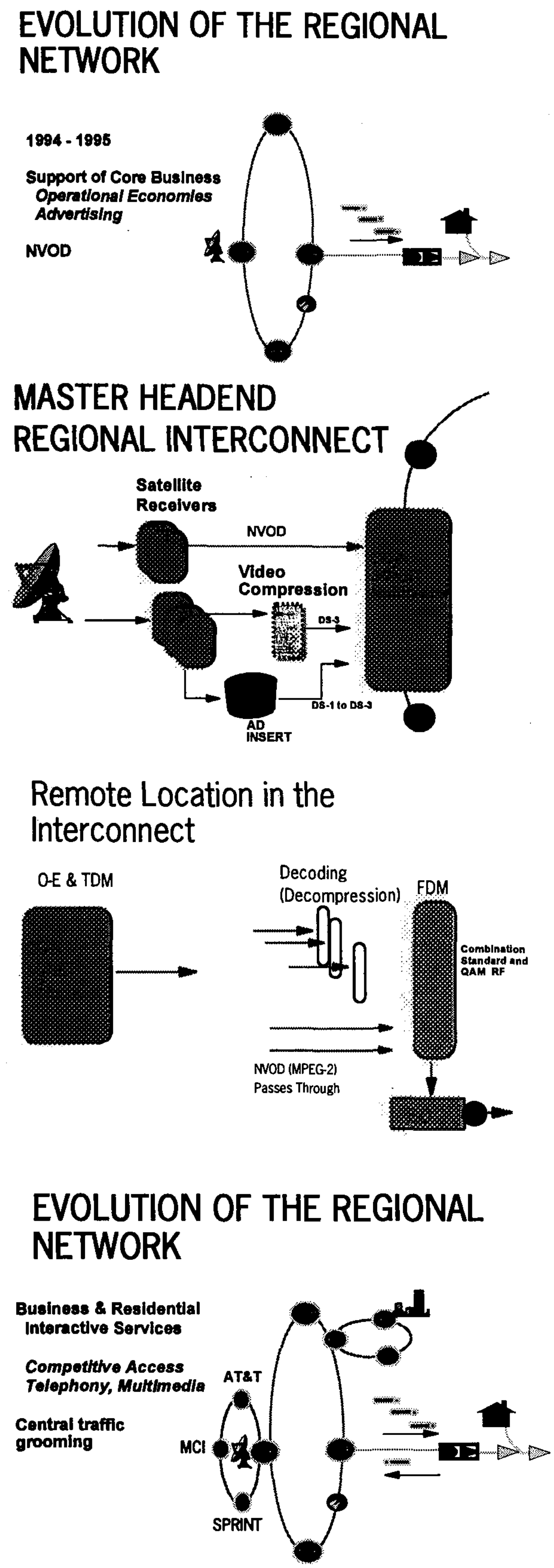


Figure 2

The marginal cost to serve smaller customers will be lower and, thus, the universe of potential customers will be greater. Even more importantly, digital switching and grooming equipment, when needed, will be deployed across the entire region. Some access providers have made significant inroads into providing switched services, as opposed to "dedicated routing" of circuits from the customer to the long distance carrier. This is controlled by regulation, but the general trend toward deregulation is clear.

The SONET regional backbone will facilitate economies of scale by spreading the costs of the switch, or any large centralized signal processing component, across the entire base. For example, the Class 5 switch is a multi-million dollar investment, a difficult piece of equipment to cost justify by a single cable system. Standard time division multiplexing (TDM) hardware will be used to process the signals on and off the SONET backbone in the remote locations (Figure 3).

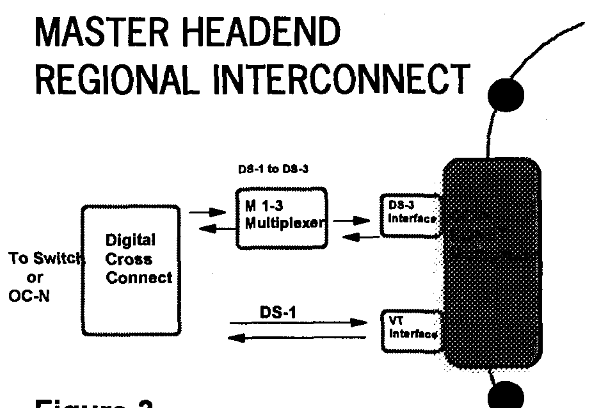


Figure 3

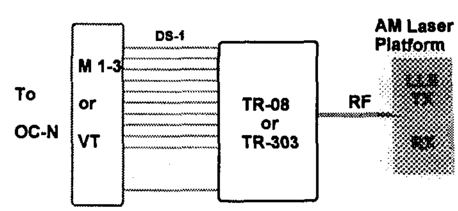
As traffic on the network increases, concentration equipment, such as digital crossconnects will be utilized.[3] Eventually, remote switching modules can be employed at critical remote sites to manage traffic more efficiently. The critical theme will be to use processing hardware only when transmission efficiencies justify the capital expense. This is an elegant model for incremental, revenue-justified growth.

Residential

The same infrastructure used for commercial customers will extend to residential telephone customers; however, the aggregation of residential telephone circuits through the RF fiber/coax network is not within the scope of this paper. Several systems are available for this purpose and most have the ability to offer multiple line, dynamic digital bandwidth (N x 64) and a range of features that will facilitate not only voice telephone but a variety of new services to the home.

The function of the residential system will be to aggregate circuits at the headend and present a standard interface to the next level in the network. This will typically be the DS-1 rate (Figure 4).

Headend Hardware for Residential Telephone



Headend Hardware for Residential Telephone

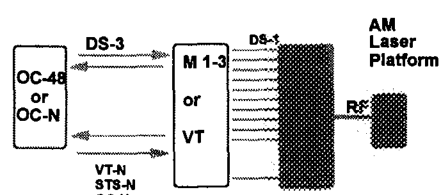


Figure 4

Figure 4 demonstrates circuit roll up DS-1 to DS-3 to SONET (STS-1 to STS-3c to OC-3). The SONET backbone network will back haul traffic to the master headend for further grooming. This may involve cross connections to a dedicated path such as an inter-exchange carrier or local exchange carrier. As in the "commercial" example above, switching may also be a possibility at the master location.

Using the SONET regional infrastructure for traditional telephone services is straight forward. Dynamic bandwidth applications, such as video telephone, will pose little additional technical burden on the transport network. The rest of the public network, especially the imbedded base of switches and asynchronous equipment may represent a bottleneck outside the regional inter-connect. Evolving PSTN standards such as TR-303, which defines many of the switch interface and dynamic bandwidth protocols, will take time. The evolutionary pace for the cable operator will not be governed by an imbedded base in the regional interconnect as long as the correct fundamental elements are put in place from the outset.

Entertainment Video

The specific focus of this paper will be on the delivery of near video-on-demand (NVOD) and video-on-demand (VOD) services through the SONET regional network. Assuming that MPEG-2 is received at the master location, the function of the regional network is to transport that information to the remote locations. Since NVOD implies broadcast services, each headend will receive the same information and will broadcast either in a digital format or an analog scrambled format to the subscriber through the RF fiber/coax network. Regardless of the RF transport, the implications for the regional network remain the same. The satellite transponder will deliver 30 Mbps to 45 Mbps to the master headend. At this time, MPEG-2 compatibility with ATM is undefined. This is not a severe problem until VOD and interactive requirements emerge. The data stream could be mapped to the SONET frame directly, although this would be inefficient. The 30 Mbps rate would need to be "bit stuffed" to match the payload size of the STS-1 frame. The more likely solution is to utilize SONET virtual tributaries to "container-ize" the transponder output and then map that output to the SONET frame. The granularity requirements suggest VT-6 (6 Mbps virtual tributaries) should be optimal. Once the digital information is retrieved at the remote site, two possibilities exist:

1) Digital RF modulation and transport to a digital set-top at the subscriber's location.

2) Decompression of the video; reassembling of the baseband signal, including scrambling information and BTSC audio; RF modulation to an analog set top at the subscriber's home.

The key point is that digital set-tops are not a prerequisite for implementation of NVOD services if regional distribution is efficient. The MPEG-2 compression standards should enable between 10 and 15 NTSC programs per STS-1 frame. The expensive encoding component of the MPEG-2 process will not be duplicated in the region. Satellite reception, SONET mapping and decompression at either the remote headend or the subscriber's home will provide a very cost effective initial deployment strategy.

Multimedia (High Speed Data/LAN Interconnect)

Data communications in the regional interconnect will develop initially from point to point applications (e.g. user location to inter-exchange carrier) or point to multipoint service (e.g. broadcast teleconferencing or application like NVOD and advertising insertion as described above). Switching, bridging and routing functions will grow out of this "provisioned circuit" business.

The end user applications will drive the evolution of the multimedia network configuration. The business LAN environment has already begun the process of moving from a mainframe, or centric, environment to a distributed, or network-centric, environment. This direction has been driven by the application, decision-support vs. production[4], as well as the rapid development of computing hardware. For years, software programmers have been told that processing power will double every two years and the relative cost will decline by half in that same time frame. With such an equation to work with, it is no wonder that software processing, storage and memory requirements have increased exponentially. Software complexity usually translates into user simplicity and/or increased utility.

High-speed data applications will be driven from two directions:

1) Wide area network (WAN) connection of local area networks (LAN) will likely involve interactive ATM transport through the SONET regional platform and the RF broadband residential network.

2) Applications driven by the imbedded telephone infrastructure will emphasize North American Digital Standards such as DS-0, DS-1, N x 64, etc. This will involve new applications such as video teleconferencing and high speed data communications to the home, in addition to more traditional voice and data applications.

The existing public switched telephone network is designed around the North American Digital Standards as described in (2) above. There is still concern over the impacts of latency in ATM voice transport, particularly through the public switched telephone network that has been optimized for time division multiplexing and analog technology. It is likely that new applications requiring bandwidth below 1 Mbps transmission will evolve from the DS-0 and N x 64 world.

ATM holds great promise for the future and will likely become the wide area network transport standard. SONET provides a well defined physical transport layer at the DS-3 and STS-3c levels. ATM granularity will reach the DS-1 level soon and the "promise" is that ATM will ultimately support N x 64 and voice applications. This would represent a smooth transition for the SONET regional interconnect, given the physical transport definition between ATM and SONET.

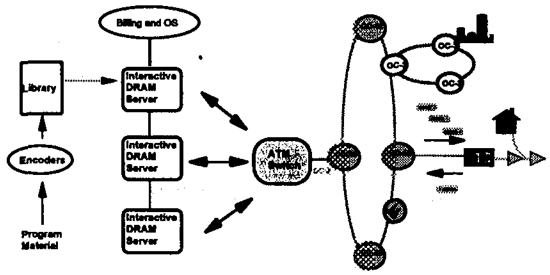
ATM would enable the network provider to offer packet services and thus move into the "retail" side of the competitive transport business. Instead of one 155 Mbps circuit for one customer, many customers and locations would share a single 155 Mbps ATM connection, paying for bandwidth as it is utilized.

ATM is still not a complete standard. Virtual connection oriented ATM, the ability to create short term circuits for ATM traffic between two or more locations, is not yet available as a final standard. Other issues, such as latency, need to be resolved for isochronous telecommunications. However, ATM could be very important in the long run for network connectivity and dynamic digital bandwidth services.

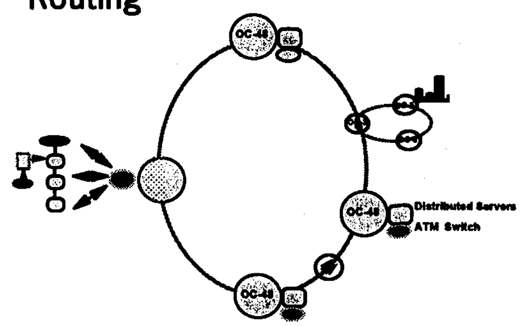
The regional interconnect based on SONET standards will provide cost effective multimedia transport for ATM and the North American Digital Standards protocol. Other bridging, routing and gateway functions between dissimilar and similar LANs can be accomplished. The transport network provider may want to define either an ATM, SONET, or NADS interface to the network. It would be extremely difficult to manage various network bridging hardware for the many local area network configurations (e.g. TCP\IP, SNA, Ethernet, etc.).

The scope of multimedia services for the cable television operator will likely include content services in addition to transport (conduit) services. File servers may be employed initially for "broadcast" applications such as advertising insertion and NVOD. The same platform will support interactive services such as games and catalog shopping, two of the more discussed applications. Figure 5 demonstrates a decentralized approach to file servers. This has the advantage of reducing network traffic in the regional infrastructure and facilitating more truly interactive services. The tradeoff between distributed file servers and regional network bandwidth may reflect more of a migration than a design conflict. Meanwhile, reduced memory and processing costs tend to favor the distributed approach sooner, rather than later.

Centralized Server Overlay



Distributed Server and ATM Routing



Remote Headend Processing in Distributed Server Architecture

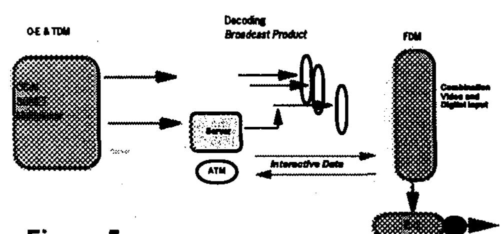


Figure 5

REGIONAL INTERCONNECT TRANSPARENCY

Transparency above the SONET regional interconnect will involve public switched telephone network (PSTN) standards and interface through the inter-exchange carriers' local "point of presence" and the local exchange carriers' central office. The PTSN interface will be defined in terms of evolving SONET and switch standards, such as TR-303, which define parameters of dynamic bandwidth in an N x 64 environment, among other things. This is reasonably straightforward, assuming a SONET regional platform.

Transparency below the regional interconnect involves seamless interconnection with the residential RF network, the fiber optic/coaxial hybrid that is today's cable television system. The time when proprietary formats could be tolerated has passed for the cable television industry. The need to adapt to evolving transmission standards becomes even more critical as high level elements, such as file servers and inter-network connectivity, become issues. Two digital formats will need to be supported at the residential network level initially:

- N x 64 -- DS-0
- ATM

RF delivery of digital signals will require further definition in terms of modulation format, contention rules, and mapping. In general, the fewer the variables within the network, the better for network efficiency and cost. The industry must work toward physical and data link layer definition (e.g., ATM, SONET, 16 VSB modulation) that will emphasize element strengths. For example, one 6 MHz carrier containing 40 Mbps of ATM cells, could represent several applications. Bandwidth would be dynamically assigned between applications or within the same application, such as the video compression rate based on program content. Working with selected standards will provider volume leverage for key network components.

CONCLUSION

By adhering to open architecture standards in the regional interconnect, new services will require incremental investments that will be proportionate with the new opportunities. This will bring many applications into focus that traditionally would not have been deemed financially viable. This is important, as most of the revenue from telecommunications today comes from small users, not large business. The marginal cost to serve small, more numerous customers must be low enough to make the business worthwhile. The cable industry has some significant bandwidth advantages in the residential network. The regional interconnect will leverage that advantage further while providing gateway access to virtually any service that is available on the public network.

---

[1] C.J. McGrath, "Digital Technology for CATV Networks," AT&T Bell Laboratories, 1991.

Andy Paff, "Applications and the Evolving Regional Interconnect," Society of Cable Television Engineers Emerging Technologies Conference, 1993.

Andy Paff, "Evolution of the "SONET Regional Interconnect in the Cable Television Environment," SCTE Emerging Technologies Conference, 1994.

Stephen D. Dukes, "Photonics for Cable Television Design," Cable Television Laboratories, Inc., 1992.

[2] The Yankee Group, "CAP Market Update: Year of Transition", White Paper, Vol 7, No. 2, February, 1992

[3] Scott Nelson, "Moving Toward the Full Service Digital Network," Communications Technology, July 1993

[4] The McData Link, "How to Avoid Gateway Chaos in Multiprotocol Networks", Internal publication, McData, 1991

INTEROPERABILITY ON THE INFORMATION SUPERHIGHWAY: THE CONTINUING SAGA

H. Allen Ecker
Chief Technical Officer
Scientific-Atlanta, Inc.

I. INTRODUCTION

Most of the participants who plan to build or use the "Information Superhighway" agree that **INTEROPERABILITY** is one of the most critical success factors. However, as in most cases in which agreement on broad general concepts is universal, we find that "the devil is in the details". It is hard to argue with the concept that Interoperability is required to have multiple equipment and system providers and to have access by the consumer to multiple programming and other service providers. However, the key questions are how to define Interoperability in the complex network structure of the "Information Superhighway" and how to implement Interoperability in the intensely competitive and highly political market.

History tells us that the only way to resolve Interoperability issues such as we find on the "Information Superhighway" is for all the players to understand accurately the technical and economic implications in sufficient detail so that rational business decisions can be made. In the past year the Communications Industry has made significant progress in understanding the technical details in many segments of the "Information Superhighway". Although progress has been made in understanding the economic details, uncertainty still exists both in the area of cost of installation and operation and in the area of derived revenue from the variety of potential programs and services. The large scale trials that are being implemented in 1994 by Time Warner, US West and others will address some of the unanswered technical and economic questions.

From a technical perspective it is useful to consider the Interoperability issues in the following categories:

1. Digital video compression
2. Digital audio compression
3. Digital system multiplexing and transport
4. Modulation and error correction
5. Security and conditional access
6. Network operating system

Of course, there are multiple sub-layers in each one of the above categories that must be addressed in detail to achieve Interoperability. This paper attempts to address important sub-layers in each of the major categories in which agreement has been reached on details and to identify several critical Interoperability details on which the industry must still reach agreement to have a truly Interoperable "Information Superhighway".

II. DIGITAL VIDEO COMPRESSION

The International Standards Organization (ISO) Moving Pictures Experts Group (MPEG) has led the way in defining standards for digital video compression that will allow Interoperability in this important part of the "Information Superhighway". It seems to be almost universally accepted now that the MPEG-2 digital video compression syntax will become the Interoperability standard for digital transmission both in the United States and in the rest of the world.

The MPEG-2 standard has been designed with inputs from the leading digital video compression experts around the world. It makes available a "tool kit" of digital video compression techniques that are the most powerful that can be practically implemented in real time video compression encoding today. MPEG-2 custom Integrated Circuit (IC) decoder chips exist today that can handle all of the techniques in the Main Profile Main Level of the MPEG-2 standard. The video encoding strategy can be selected from the MPEG "tool kit" according to specific applications and programming needs with the decoders capable of handling whatever choices are made in encoding.

There has been some confusion and perhaps misunderstandings on how MPEG-2 maintains Interoperability while allowing options for the user on the prediction modes for motion compensation. This dilemma is usually centered on the advantages and disadvantages of using what are called B-Frames in predicting vectors for motion compensation.

For reference MPEG-2 allows a choice in the prediction mode for motion compensation in the encoding strategy. The major choices are the following:

1. I-Frame only: In this mode there is no prediction of current frames either from previous or future frames. The compression process is based entirely on techniques that operate within each individual frame. I-Frame only modes are used when frame cut editing and multiple generation recording are required in studio and production applications.
2. P-Frame prediction: In this mode a current frame is predicted from motion compensation vectors that are derived from a previous frame only. Field based prediction is possible on a macro-block basis in this mode. However, two motion vectors, one for each field, are required.
3. P-Frame dual prime: In this mode prediction is done on the odd and even fields in a frame separately using information from the fields in the previous frame. However, only one motion vector with a small differential motion vector is transmitted for each block in the pair of fields and an algorithm is used to adjust the motion vector for the difference in line structure and time of occurrence for the odd and even fields.
4. B-Frames prediction: In this mode both a previous frame and a future frame are used to predict the current frame. Tests in the MPEG process have shown that this predictive mode is the most accurate of all those considered. However, it requires more memory in the decoder for an additional frame store than the other modes.

A video compression decoder designed to handle the MPEG-2 main profile can recognize and decode any of the predictive modes summarized above. However, these modes can not be used simultaneously on the same frame. MPEG testing has shown that for a given data rate the picture quality ranking and memory requirements are as shown in Figure 1.

For movies B-Frame prediction has a dramatic advantage over other prediction modes. Movies are completely framed based source material and field based prediction modes are not applicable. For programming services that are primarily movies B-Frame prediction can maintain good picture quality at significantly lower data rates than all other prediction modes.

Although there may be some differences in subjective evaluations of the degree of difference in picture quality among the four modes, experts in video compression rank them in the order given in Figure 1. Also, it is universally accepted that to implement B-Frame prediction for broadcast and studio quality resolution the incremental frame store will require an additional 8 Mbits of DRAM in the decoder. For lower resolutions less additional memory is required. No additional memory is required for horizontal resolution of 352 lines or less.

The difficult forecast is to determine the cost of additional DRAM in a decoder in future years. Today the volume unit for DRAM memory is 4 Mbits, therefore, two of the 4 Mbit units are required for P-Frame prediction either standard or dual prime. Four 4 Mbit units are required for B-Frame full resolution prediction. Even in very high volume production an 8 Mbit difference in memory can add $16 to the material costs of the decoder today.

We know that the cost of memory will come down in the future as the Semiconductor Industry goes to smaller geometry and larger memory unit sizes. The unit size for memory and the cost is driven by the Computer Industry today. Predictions are that the Computer Industry will skip the 8 Mbit unit and go directly to a 16 Mbit unit as the next memory unit size for volume production. However, the exact date that 16 Mbit DRAM will be available in volume with the necessary bus structure and low cost is very difficult to predict exactly. 16 Mbit DRAM unit volume production is estimated by some to peak in 1997 with a significant reduction in cost per Mbit. Bus structure and memory access time are critical parameters for DRAM used for video frame stores.

Today, typically four 4 Mbit parts each with a 16 bit bus are used to achieve the 16 Mbit requirement. They are typically accessed simultaneously (i.e. 64 bit bus) to achieve the required bandwidth. In order to use a single 16 Mbit part with a 16 bit data bus, a factor of 4 improvement in access time is required. Memory suppliers are starting to produce "Synchronous DRAM" which has the required access time. Parity in "cost/bit" between the "Standard 4 Mbit" parts and "Synchronous 16 Mbit" parts is expected in mid 1995.

The flexibility of the MPEG-2 syntax and the universality of the MPEG-2 decoder chips allow network operators to select the encoding techniques that best fit their applications and cost requirements. Although MPEG-2 supports all of the prediction modes described above, an operator could limit the memory in the decoder initially to hold down initial decoder costs. In the long term most likely all decoders will have 16 Mbits of memory as memory costs come down with time and the volume unit for memory becomes 16 Mbits.

III. DIGITAL AUDIO COMPRESSION

All of the digital audio compression techniques use the same fundamental principles of allocating the available bits in a fixed rate data stream according to the frequency content and dynamic range of the audio source material being compressed. All of the algorithms use a psychoacoustic model to set priorities dynamically for bit allocations in frequency and amplitude segments of the audio spectrum. In recent years, the two major contenders for digital audio approaches have been the MUSICAM group led by Philips and the Dolby Laboratories.

The audio compression algorithms chosen for the MPEG-1 standard for mono and stereo audio signals are based on the basic MUSICAM algorithm. It is this MPEG-1 stereo that has already been adopted internationally and is the core of the audio part of ISO/IEC 11172-3 audio standard. The proven high quality MPEG-1 stereo can provide the stereo pair for use for MPEG-2 video applications. Equipment is in production around the world using the ISO MPEG-1 audio standard for digitally compressed stereo.

Surround sound using stereo transmission channels can be provided for MPEG-2 video by the use of two channel matrix coded multichannel audio such as Dolby Laboratories Pro Logic® surround in combination with the MPEG-1 independent stereo coding. The MPEG-1 MUSICAM compression at compressed data rates of 256 kb/s and 384 kb/s for a stereo pair has sufficient transparency margin that it is both Pro Logic® compatible and capable of multiple cascades without causing audible coding artifacts.

In the recent completed formal listening tests conducted for MPEG, none of the 5 channel coders tested achieved transparency for all of the audio test selections at the tested bit rates. Although none of the coders achieved transparency, the performance of the PAC coder of AT&T and the AC3 coder of Dolby Laboratories indicated that higher quality audio could be achieved with techniques not backward compatible to MPEG-1 stereo. The desire to provide the best possible audio quality in multichannel audio within the MPEG standard has opened the door for a new multichannel extension in

MPEG-2 that is not backward compatible. This new mode will be pursued as a supplement to the necessary backward compatible modes. A call for new submissions has been issued by MPEG.

All of the multichannel coders are experiencing improvements in performance as further optimization is occurring and new encoding strategies are being tested. The standard has been constructed to allow these improvements to be made in the encoders without changing the transport and syntax that are specified in the standard.

The 5 channel digital audio compression standard for MPEG-2 has become a controversial issue with both technical and political disputes. Because the processing "engines" for all approaches in contention are very similar, it may be possible to develop a decoder chip set that will handle multiple approaches using a common processor and memory. The programs to execute the different approaches could be stored in separate ROMs.

IV. SYSTEM MULTIPLEX AND TRANSPORT

Outstanding progress has been made by the MPEG-2 System Sub-Committee in defining the required multiplex and transport approach for transmission of digitally compressed signals. Some of the most critical issues to resolve were the following:

1. Packet based approaches vs frame based approaches
2. Fixed length vs variable length packets and packet lengths
3. Number of bits available for overhead
4. How to map to MPEG into the Asynchronous Transfer Mode (ATM)
5. Program multiplexes imbedded in the system multiplex
6. Defining appropriate "hooks" for conditional access and encryption

The diagram in Figure 2 defines the basic packet and multiplex structure for which the MPEG-2 System Sub-Committee has reached consensus. A fixed packet of 188 bytes with a payload of 184 bytes has been designated. The key parts of the 4 byte header are the sync word and the prefix. The prefix includes error indicators, transport priority, packet identification and other indicator and control bits. The MPEG-2 system has been designed with an adaptation field for information that may be required for specific applications. Also, a single MPEG-2 transport packet can be fit into the payload of 4 ATM cells.

Key features of the MPEG-2 transport include the following:

1. Transport stream is independent of transmission data link to allow both terrestrial and satellite applications.
2. Synchronization of program service data (video, audio, etc.) is handled by MPEG-2 systems layer and is not dependent on transmission link timing.
3. Error protection requirements can be matched to transmission medium where errors actually occur.
4. MPEG-2 transport packets have error detection functions built in (priority bit, packet-error-indicator and packet continuity counter).
5. MPEG-2 systems packet-based transport structure allows for simple remultiplexing operations to be performed on data received over non-synchronized transmission data links.
6. MPEG-2 adaptation field allows flexibility in the handling of multiple conditional access streams, private data and other optional features.

Custom ICs for the MPEG-2 multiplex and transport are in the final stages of design and should be available for initial use in products by June of this year.

V. MODULATION AND ERROR CORRECTION

A. Satellite Applications

The Communications Industry both in the United States and in the rest of the world has adopted Quadrature Phase Shift Key

(QPSK) modulation for applications using digital video compression over satellite. The noise characteristics of the transmission path and the characteristics of amplifiers in satellite transponders make QPSK an optimum compromise between data rate capacity in a given transponder and ruggeddness at practical threshold levels at the receiving terminal.

Concatenated Viterbi and Reed-Solomon error correction coding is used with interleaving to reduce the effect of burst errors. Trade-offs can be made between the amount of the total digital data stream used for error correction, the maximum information data rate and the receiver threshold. Systems have been designed with rate 3/4 Viterbi coding (1/4 total data rate used for error correction) and with rate 7/8 Viterbi coding (1/8 of the total data rate used for error correction coding). At rate 3/4 the information rate in a typical 36 MHz C-band transponder is 27 Mb/s. At rate 7/8 the information rate in a comparable C-band transponder is 38 Mb/s. All things being equal the threshold at the receiving site is approximately 2 dB lower for rate 3/4 than for rate 7/8. This difference in threshold can be significant or insignificant depending on the power of the satellite used and the antenna gain and LNA noise temperature at the receiving site.

B. Broadband Cable Applications

The transmission path characteristics over broadband cable in the forward direction and the requirement to maximize information data rate in a given bandwidth led to a choice of multi level AM for modulation with only Reed-Solomon error correction.

Quadrature Amplitude Modulation (QAM) and Vestigial Side Band Amplitude Modulation (VSBAM) are the multi level AM techniques currently under consideration. Both QAM and VSBAM are well known techniques that have been utilized in numerous applications for many years. Scientific-Atlanta uses 4 VSBAM to modulate digital audio in the horizontal blanking interval in the B-MAC System. Multi level QAM has been used in many applications over the years including digital microwave radio and cable modems.

At a CableLab's seminar on digital modulation for broadband cable in December 1993 all of the presenters except one supported QAM. Many concluded that QAM had both performance and cost advantages over other modulation approaches for digital applications over broadband cable. The other presenter, Zenith Corporation, advocated VSBAM over other approaches. Zenith has implemented an 8 VSB system for trials to determine the broadcast standard for digital HDTV and has demonstrated a 16 VSB system over broadband cable.

The United States digital HDTV Grand Alliance has selected 8 VSB as the broadcast modulation for digital HDTV based on tests conducted of Zenith's implementation of 8 VSB and other implementations of 32 QAM. The tests show that Zenith's implementation of VSB was superior to the implementations of QAM. One reason for the test results is that Zenith used 8 VSB with a much higher overhead than the 32 QAM implementation. Theoretically 8 VSB has the same data rate as 64 QAM. See Figure 3 for comparison of rates for VSB and QAM. Zenith used the difference in rates between 8 VSB and 32 QAM for error correction overhead. A second reason for the test results is that Zenith incorporated a comb filter to reject rf carriers and color sub-carriers for NTSC to minimize co-channel interference. Co-channel interference is an issue in over-the-air broadcast and not in broadband cable applications.

Head-to-head tests of optimized implementations of 8 VSB vs 64 QAM and 16 VSB vs 256 QAM over broadband cable have not been conducted by independent testing groups. Many tests and simulations to date suggest that 64 QAM and 256 QAM have performance advantages on cable over Zenith's implementation of 8 VSB and 16 VSB for the following reasons:

1. Zenith's implementation requires a pilot signal with each channel to accomplish carrier recovery in the receiver. This pilot signal uses some

of the total energy available and has a significant effect in a multi channel broadband applications.

2. The Zenith implementation requires a very sharp filter with an alpha of 11%. In the QAM applications larger alphas are used which should improve the performance in a multi channel environment and reduce the cost of filters in the decoders.

Calculations show that the proposed VSB approach will have approximately 2 dB poorer Signal-to-Noise (S/N) in a broadband cable plant than the proposed QAM approaches. As shown in Figure 4, overshoot produced from Gibb's phenomenon from sharply truncated channels will require 1.2 dB more "back off" to eliminate amplitude clipping. This difference in overshoot results from the difference in alpha: the excess bandwidth ratio. The energy used in the pilot carrier in the VSB approach reduces S/N by another approximately 0.7 dB.

The European Digital Video Broadcast (DVB) group formerly called the European launching group has selected 64 QAM as the standard for digital video modulation in Europe. This decision was made after extensive simulation and testing by the members of the group. Although the European DVB has agreed on standards for the details of the Viterbi, interleaving and Reed-Solomon codes, complete agreement has not been reached in the United States on details.

Statements advocating compatibility between the modulation technique for over-the- air broadcast for digital HDTV and digital video compression modulation over broadband cable often fail to mention the real issues. The modulation scheme for digital HDTV is 8 VSB with a specialized trellis code for error correction that provides an information rate that is essentially no higher than 4 VSB or 16 QAM with error correction needed for cable. The information rate for the HDTV signal is only approximately 19 Mb/s in the 6 MHz channel. This information rate would be totally unacceptable in a broadband cable environment. Therefore, the 8 VSB implementation used for digital HDTV would not be acceptable in broadband cable. If a different implementation of VSB were used for cable than for digital HDTV then direct compatibility would not occur even though an 8 VSB demod with some extra circuitry will do 16 VSB. The issues of performance in the broadband cable plant and cost are much more important than a digital HDTV set with a dual mode VSB demodulator. Initial deployments by Scientific-Atlanta and General Instrument for applications over broadband cable will use QAM.

Scientific-Atlanta has tested 64 QAM in the fiber-to-the-serving-area network in the Time Warner system in Orlando with an IC implementation for demodulation and equalization. The decoder utilizes a three chip set: an equalization IC, a demodulation IC and a control IC. The full custom chip set is capable of either 64 QAM or 256 QAM operation. The test results show this 64 QAM implementation operates within 1 dB of theory and yields the desired Bit Error Rates (BER) with the appropriate Reed-Solomon Forward Error Correction (FEC).

The 256 QAM mode has been tested in the laboratory back-to-back with the custom IC implementation. These tests also show operation within 1 dB of theory. Field tests of the 256 QAM mode in an actual FSA Network will be made in the coming weeks.

VI. SECURITY AND CONDITIONAL ACCESS

It is usually instructive to consider security and conditional access in two major parts: (1) the encryption algorithm used to encrypt program content and (2) the conditional access data stream to distribute keys or information to generate keys. Usually the encryption keys used to encrypt the program content material are called seeds and are changed very rapidly to achieve temporal security. Temporal security can be thought of as changing a key so often that a pirate can not practically discover the nature of the key before it has been changed at the next time cycle. Often these seed keys are changed with time cycles of less than one second.

If a common seed and a common encryption algorithm is used to encrypt all program content, multiple conditional access data streams can be used to distribute the seeds or the necessary information to generate the seeds. The seeds in the conditional access data stream can be encrypted with different keys and different encryption algorithms from those used for the program content. Different program providers can use their own proprietary conditional access data stream approach or could agree on common conditional access data streams. Special provisions would have to be taken in any system with multiple conditional access data streams to prevent pirates who might break one conditional access data stream to create clones that could be used universally for all conditional access data streams.

The current approach being proposed by the European DVB is to have a common encryption algorithm for program content but to allow multiple proprietary conditional access data streams. There is strong political pressure in Europe to have multiple proprietary conditional access data streams to allow programmers to protect and control their own subscriber base without automatically giving other programmers access.

One practical approach to Interoperability for security and conditional access is the digital equivalent of the analog system we use today in the United States. There is a defacto standard over satellite and then each local cable operator selects from one of multiple vendor choices for his local security and conditional access system in his cable system. The local cable operator could likewise in a digital system decrypt the signals received over satellite and have a different local security and conditional access for both digital and analog signals in his system. This approach of different independent security and conditional access systems also reduces the possibility of a wide spread catastrophic pirate break of a common system. A clone in one local system would not be transferable as a clone in another system with a different security and conditional access system.

VII. NETWORK OPERATING SYSTEM

Even if we resolve all the issues in the previous areas that are required for Interoperability, we would not have Interoperable equipment or networks unless the industry develops an Interoperable network operating system. An operating system is essentially a collection of software programs in the headend and in the processor in the Home Communications Terminal (HCT) that lets a particular service provider or programmer deliver his services and programs to all subscribers regardless of whose HCT or headend equipment is in his system. The analog in the computer world is an operating system such as MS DOS in which any application software written to the MS DOS interfaces and specifications will operate on a given PC using that same version of MS DOS.

The part of the operating system that is in the HCT must take high level commands from a given application software program and convert them to software drivers for the various hardware and software functions in the HCT. For example, if a specific application requires digital data from a specific channel, the operating system must interpret the request for data and give the commands within the HCT to set the tuner to the correct frequency and to extract the data from a specific multiplex within that frequency channel. Also, if conditional access is involved, the operating system must check that the appropriate conditional access authorizations have been granted.

Many companies are working diligently to develop operating systems in hope that they will become the Microsoft of the "Information Superhighway". However, one of he major challenges for the operating system that will achieve practical implementation is to achieve the required performance without pushing the processing speed and memory size in the HCT beyond an acceptable cost point. The price of the HCT must be consistent with the anticipated revenue it generates. Another part of this trade-off is to determine how fundamental the Application Program Interface (API) can be to keep processing and memory cost low without inhibiting the wide spread development of creative third party applications.

Interoperability will be best served if the hardware and software designers of the products and systems in the Information Superhighway adhere to the ISO's Open System Interface (OSI) seven layer model for the system design structure. Interface specifications will be necessary at critical interfaces within the products and at defined interfaces in the system. Use of the seven layer OSI model will help facilitate the development of these interface specifications for Interoperability.

VIII. CONCLUDING REMARKS

The industry has come a long way toward Interoperability in the last two years. But, it is obvious that we still have a long way to go. I believe that action in at least the following four areas will ultimately produce Interoperability:

1. Strong industry standards groups such as MPEG-2
2. Industry consensus such as occurred for QPSK modulation on satellite
3. Industry pressure for multiple suppliers
4. Government and regulatory pressures for Interoperability. For example, current FCC regulations on video dial tone for Bell Operating Companies (BOCs).

I believe that Interoperability is inevitable because both the cable MSOs and telco BOCs will demand it. It may be a painful process and there may be costly false starts. I am convinced, however, that the only way to realize the vision of the "Information Superhighway" and the rewards it can bring to the Communications Industry and the consumer is to push for Interoperability as quickly as possible.

Prediction Mode	Picture Quality Ranking	Frame Store Memory Required
B-Frame	1	16 Mbits
P-Frame Dual Prime	2	8 Mbits*
P-Frame Standard	3	8 Mbits
I-Frame Only	4	N/A

*Requires memory with faster access than other modes

Figure 1 Comparison Of Prediction Modes For Picture Quality And Memory Requirements

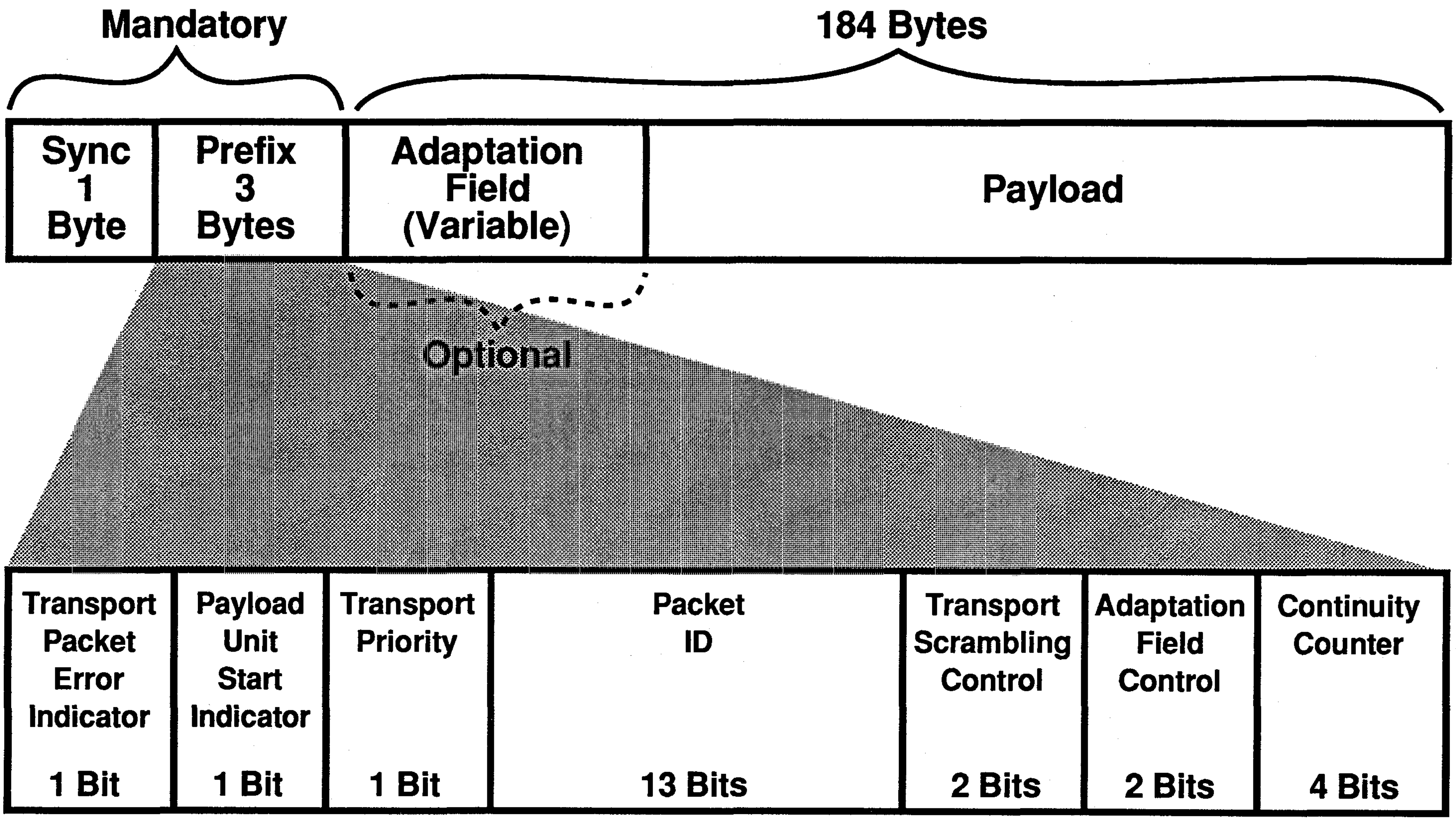


Figure 2 MPEG-2 Transport Packet Format

Scientific Atlanta
M0394-18

QAM	Bits per Hz Theory	Bits per Hz Practical	VSB-AM
16 QAM	4	3.3 - 3.5	4 VSB
64 QAM	6	5.0 - 5.2	8 VSB
256 QAM	8	6.7 - 7.0	16 VSB

Bit rates in TV channel (6 MHz)

8 VSB or 64 QAM	16 VSB or 256 QAM
~ 30 Mb/s	~ 40 Mb/s

Figure 3 What Bit Rates Are Possible?

M0394-20

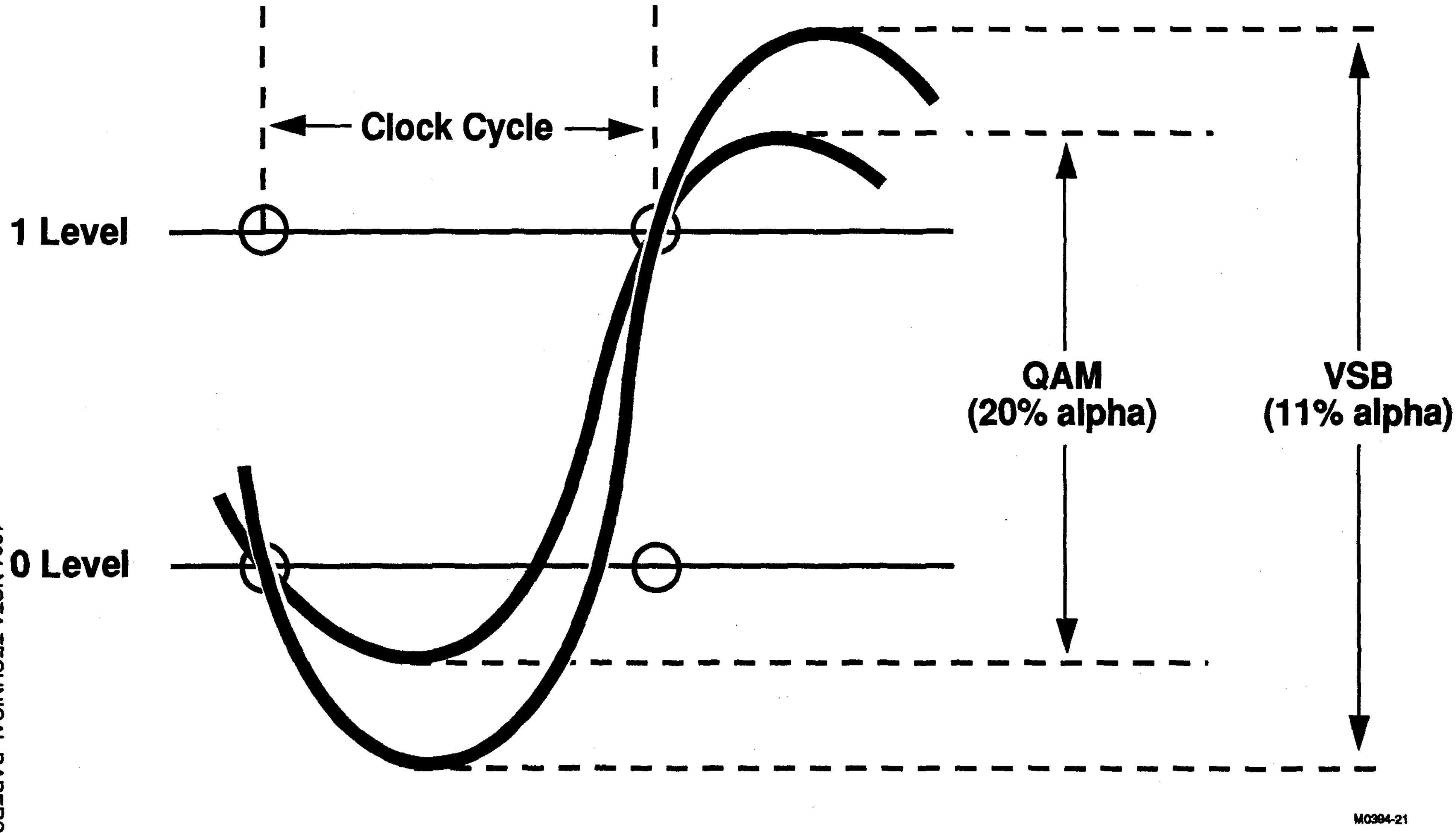


Figure 4 Overshoot for VSB (11% alpha) vs. QAM (20% alpha) with resulting 1.2 dB difference in peak power

ISSUES IN HANDLING CABLE SIGNALS WITHIN DIGITAL OPTICAL INTERCONNECT NETWORKS

James O. Farmer
Electronic System Products
An Antec Corporation Business

Edward J. Callahan
Antec Corporation

INTRODUCTION

As cable systems seek to add telephony services and interconnect headends for operational efficiencies, they will migrate toward use of digital fiber optics for the interconnections. This is true even of systems which expect to continue to deliver analog programming to the home for the immediate future. When developing a digital optical interconnection system, one may choose either proprietary systems, or may select compatible standard systems available from a number of manufacturers. Here-in the case is made for the latter approach, in which standard systems, originally developed for telephony applications, are used to transport video and ancillary signals from one headend location to another.

Interface issues include not only video, but also audio services, addressable data, control and new data services. The versatility of the network is enhanced by integrating cable television's needs with standard multiplexing systems. Multiplexing is explained, and the hierarchies in common usage are introduced. Finally, a short space is devoted to describing some of the quality issues in hybrid analog/digital networks.

INTERCONNECTING HEADENDS

Headend interconnection is becoming popular today to allow efficiencies of operation, more controllable advertising insertion and improved reliability. Figure 1 shows how headends may be connected. Two architectures are shown. Figure 1 (A) shows a single master headend connecting with three hubs, or sub headends. Fiber cables interconnecting the headends are routed two ways, so if one path is interrupted, for example by a car hitting a pole, the other path can be used. Figure 1 (B) shows a modified concept in which the headends are connected in rings. Signals travel around the ring in both clockwise and counterclockwise directions, passing through each headend on the way to the next. In this case two master headends are used. Signals are routed completely around the ring in two directions. As in (A), if one path is interrupted, the path around

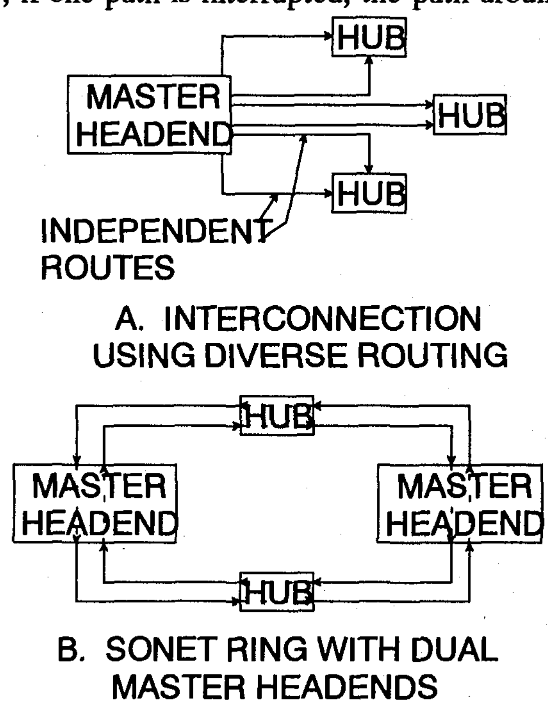


Figure 1. Options in Interconnection

the ring in the other direction is available to carry signals. Further, if something happens at one headend, there is a second master headend to take over, so that service is interrupted, if at all, to only one node.

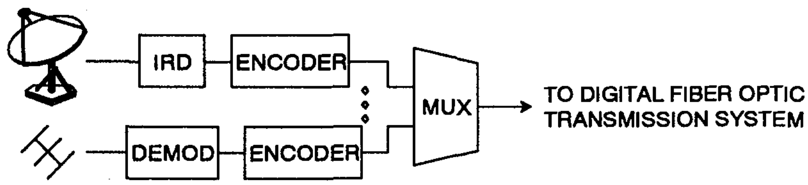


Figure 2. Simplified Master Headend

The preferred method used to transmit the signals is the SONET Synchronous Optical Network. It is a fiber optic based system with digital transmission of all signals. Figure 2 shows the basic conditioning of a video signal at the master headend. Signals are converted to baseband and supplied to *encoders*, which convert to digital and compress the signal. At this time MPEG compression is not being used due to the cost of the encoders, though this may change in time. The digital signals are multiplexed (described below) onto a single data stream for transmission throughout the network. At each node, the signals are passed through with provision to drop and add signals if necessary. The signals are *demultiplexed* and converted to NTSC for modulation and conventional transmission to subscribers.

WHY DIGITAL SIGNAL TRANSMISSION?

Within the network, there are excellent reasons for transmitting TV signals digitally. Digital transmission can be less costly *when applied intelligently*. Because fiber transmitters for baseband digital signals don't have to control distortion (be linear), they are potentially lower in cost. Receivers are also potentially less costly. Of course realization of the cost savings will depend on volume, which cable television may drive in the future.

Another reason for digital transmission is that, within limits, no video degradation occurs during transmission or switching.[i] Of course we can and do pick up distortion in the process of digitizing (and compressing) a signal. That distortion is a trade-off in how few bits we use to represent the signal, the quality of the signal and our ingenuity in compressing the video.

In the future the industry is likely to begin switching signals more, as we subdivide systems into fiber nodes to which we can route different signals and as we provide more customized cable features and transmission path back-up. These will require more switching and routing of signals. These operations can be done much better if the signals are in digital format, since we don't have to worry about signal degradation.

Video file servers, big brothers to the file servers used in local area networks (LANs) today, will, in future years, become commonly used for storage of programs. The programs will then be transmitted to subscribers, either on a subscription basis or on demand. File servers store video digitally, so in many cases it will make sense to transmit the signals in digital form, at least part way to the subscriber. The trick will be to balance the cost of transmission against the cost of converting the signal back to NTSC.

Once we have a signal digitized and compressed, the next effort is route the signal to a given subscriber, along with other signals. Simultaneously other signals will be routed to other subscribers. The process of combining all of the signals for transmission is the subject of this paper. We call this combining process "multiplexing."

INTRODUCTION TO DIGITAL MULTIPLEXING

As we move into a world in which video is carried in digital form, we encounter the term *multiplexing*. This is another way of saying that we are combining signals. In the cable television world, we are accustomed to combining RF signals in a *combiner*, which puts many signals together on a single cable. In traditional cable television, these signals are put together using a technique called *frequency division multiplexing* (FDM). In FDM, we put several signals together on one cable by assigning each its unique band of frequencies. The same idea is practiced in the digital world, except instead of employing FDM we normally use *time division multiplexing* (TDM). TDM is nothing more than a different way of putting signals together on a single cable. Rather than putting each signal on its own unique RF carrier **frequency** and multiplexing by combining the signals on a cable, we assign each signal to its own unique **time** slot, switching from one signal to another at the appropriate time.

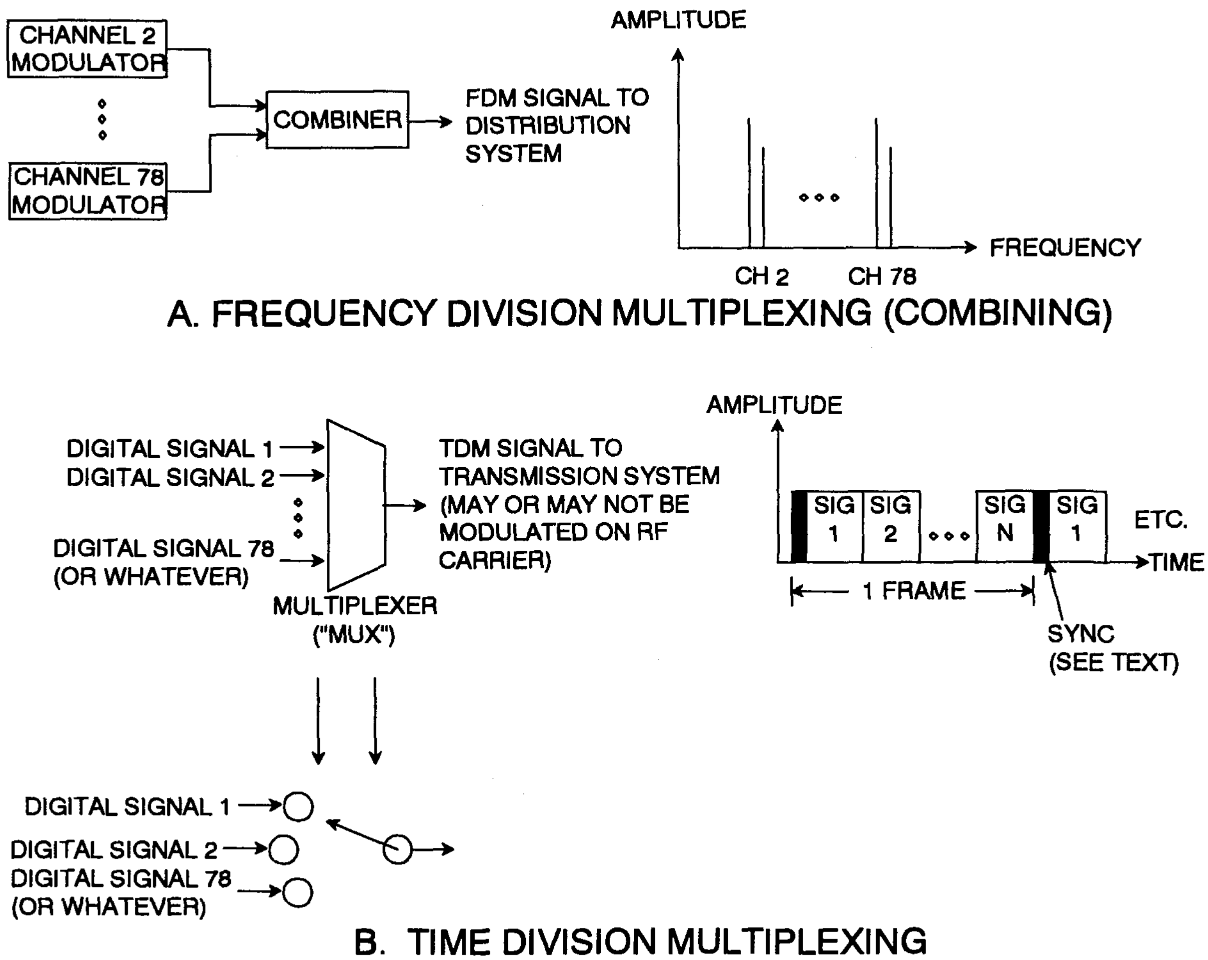


Figure 3. Concept of Multiplexing

Figure 3 compares time and frequency division multiplexing. In (A) we are doing frequency division multiplexing at a headend, by sending each signal (on its own frequency) to a combiner. The spectrum diagram shows what the resulting signal looks like. We have the familiar diagram of each channel in its place on the cable. In order to recover one of the signals, we tune it with circuits that can differentiate one frequency band from another.

In contrast, (B) shows time division multiplexing, as practiced in the types of systems discussed here. Each digital signal (a series of ones and zeros coming one after the other) is applied to a multiplexer (commonly called a "mux"), which serves an analogous function to the combiner in a conventional headend. The amplitude vs. time diagram shows the result. A short segment (one or more bits) of Signal 1 is transmitted, followed by a short segment of Signal 2 and so on. Notice that, as in all digital systems, everything has the same amplitude: amplitude data is carried as a series of bits which represent the numerical amplitude. In order to recover the individual signals, we apply the data stream to a demultiplexer ("demux"), which takes the signal back apart, into its individual data streams. The demux does this by switching Signal 1 to its "bin," then switching Signal 2 to its bin and so on. In order to let the demux know when a particular signal is being sent, we must add synchronization (analogous to vertical and horizontal sync in TV) to the data stream. The sync is shown in the diagram. Depending on the system, sync may be part of each data stream, or as shown here, may be added to the composite data stream.

In order to get all of the incoming digital signals into the same data stream, we must speed up the data in the composite stream. (Think of a river having several tributaries: streams that feed it. As more water is added to the river, it must either flow faster, wider or deeper.) For example, suppose we had 24 streams of 64 Kb/s (kilobits per second - often informally called, simply, "kilobits"). The data rate required to transmit all of the data streams on one cable is at least 24 X 64 Kb/s = 1.536 Mb/s. The actual data rate may be higher still if sync is added external to the data, as in Figure 3. For example, we may add enough sync to require an additional eight Kb of data to be transmitted each second. In this case, the resultant data stream would be 24 X 64 Kb/s + 8 Kb/s = 1.544 Mb/s. By the way, these are not arbitrary numbers: we shall encounter them again.

SPECTRUM

The spectrum of a digital signal extends to at least one half the bit rate of the data stream: if we transmitted a one followed by a zero, we would have one cycle of a square wave. In some cases, we want to filter the signal so we have nothing above the minimum frequency required to get data through. In other cases, we would handle the signal as a square wave, with a spectrum much wider than required.

Figure 4 shows a sample sequence of a baseband digital signal and the resulting spectrum. In (A) we see a sequence of bits that starts out with the sequence 1010, followed by a sequence 11001100, then a random sequence. The 1010 sequence represents the highest frequency the system will be called on to transmit. As shown, two bits, a one followed by a zero (or vice-versa), make up one cycle of a periodic waveform. Thus, we have an inherent "efficiency" of two bits per Hertz with this form of data transmission.[ii] In (B) we show the spectrum of this square wave (1010...). The fundamental is shown as a solid line at one half the bit rate. Since we are dealing with a square wave we have harmonics of this signal, and these are shown as dotted lines at two and three times the fundamental frequency. (Granted, a square wave doesn't have even harmonics, but we will have them by the time we get differing duty cycles caused by different data patterns.)

The sequence 11001100... has a rate one half that of the 1010... sequence, so the fundamental of its waveform is one half the fundamental of the 1010... sequence. This is shown as the long-dashed dashed line in (B). The harmonics are not shown. As we go to waveforms having longer sequences of ones and zeros we have even lower frequency components. If we have sequences of random lengths of ones and zeros, such as suggested at the right of (A), the spectrum is filled in from the one half bit rate point of (B) down to nearly zero frequency. Harmonics exist, and may or may not be removed, depending on the following transmission system. This is illustrated in (C), which shows the frequencies occupied by the fundamental and harmonics of the random data waveform. If we didn't do something to limit the number of ones or zeros that could occur in succession, the fundamental would occupy frequencies down to zero Hz. In most systems, this would cause problems with clock recovery and modulation, so something is added to the data to prevent excessively long strings of ones or zeros. This limits the low frequency end of the spectrum, as shown in (C).

The limiting of the longest sequence of ones or zeros is normally done by exclusively OR-ing the data stream with a pseudo-random data sequence known to both the transmitter and receiver. This process, unfortunately, is called "scrambling." It has nothing to do with hiding a signal from an unauthorized viewer (the definition of scrambling traditionally used in our industry).

Of course a multiplexed digital signal can be modulated onto an RF carrier, and this is frequently done. Examples are the digital music services carried on many cable systems today. In both systems the left and right channels, along with certain data, are muxed onto a single RF carrier. In one of the systems, the RF carrier consists of only the left and right channels and data, for one stereo pair. The signals

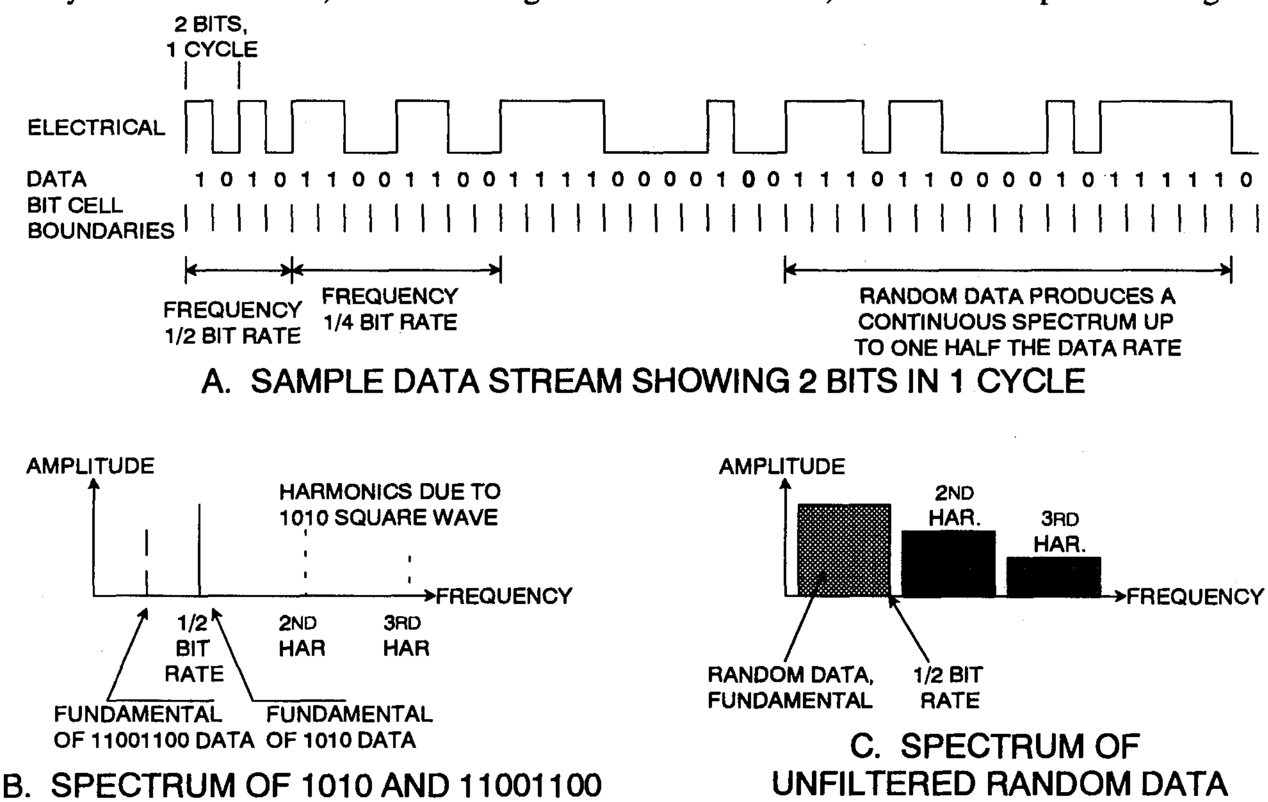


Figure 4. Spectrum of a Baseband Digital Signal

for various programs are then combined (frequency division muxed) onto the cable, in channels about 600 KHz wide. In the other system, some additional TDM is done, in that five stereo pair are TDM'd and modulated onto a single carrier about three MHz wide. The arguments concerning the merits of the two approaches are far beyond the scope of this paper.

COMPARISON OF FREQUENCY AND TIME DIVISION MULTIPLEXING

It is quite instructive to consider further the difference between a frequency multiplexed system and one that is time division multiplexed. This will help illustrate the advantages of handling baseband digital signals where practical. Figure 5 shows a system for transmitting analog or digital data using modulation and frequency division multiplexing in (A), and using digital baseband transmission in (B). (C) shows waveforms useful in understanding (A) and (B). In (A) we have taken in digital data at the left. A sample waveform is shown in (C), as circle X. This is the way data would look coming out of any normal digital process, such as a codec that converts video to compressed digital data. The spectrum is shown in (A), as explained above. In order to reduce the transmitted spectrum as much as possible, the data is supplied to a low pass filter having a cut-off nominally at one half the bit rate, to

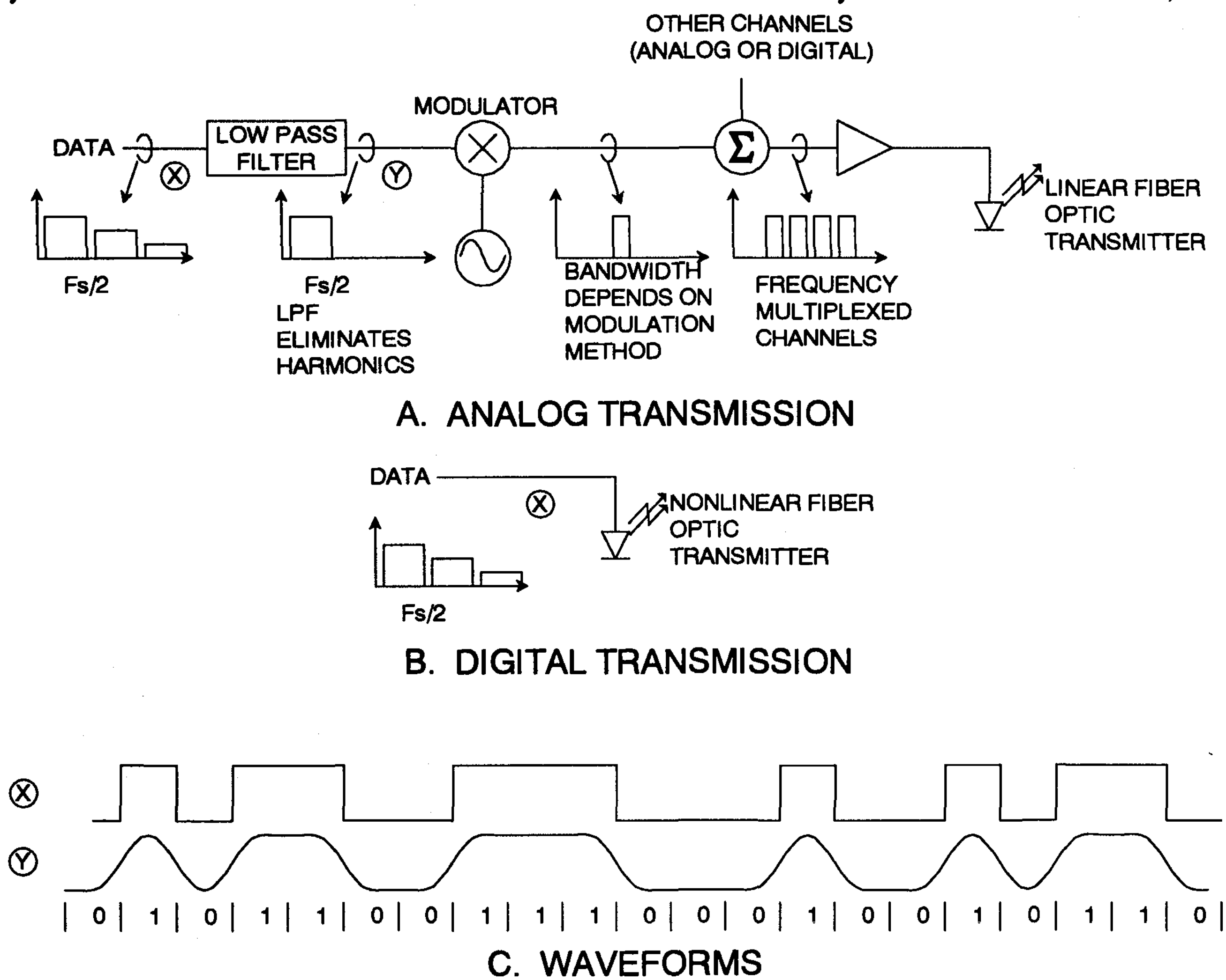


Figure 5. Linear vs. Non Linear Transmission

eliminate harmonics. If transmitted, the harmonics would require power, and would spread out the bandwidth required for transmission, without adding any useful intelligence to the signal. The spectrum at the output of the filter is shown. The resulting waveform is shown as circle Y.

From the low pass filter, the signal is applied to a modulator, which places it on a suitable frequency for carriage on the cable system. At the output of the modulator, we have a spectrum at some RF frequency, depending on the frequency of the modulator. The occupied bandwidth of the signal is a function of the low pass filter, and also of the type of modulation used: higher orders of modulation take up less spectrum, but are more expensive to decode and are more susceptible to transmission errors.

The output of the modulator corresponds to the output of a conventional cable television modulator, which is supplied to the headend combiner. In (A) Σ (sigma) represents the combiner, which combines this signal with other digital and analog signals. The composite spectrum is shown at the output of the combiner, and is amplified and supplied to a laser transmitter in this example. So far, we have a conventional cable television system with a mix of analog and digital signals. Because of the multiplicity of signals, the fiber optic transmitter must be linear. That is, it must handle any instantaneous signal amplitude in the correct proportion, a requirement to which we are accustomed in cable television. The transmitter is relatively expensive due to this need for linearity.

By way of contrast, section (B) of the figure shows what is required for time division multiplexed digital transmission. The waveform of circle X is applied directly to the laser. Of course, the data rate is much faster than is the data rate of a signal in (A). Rates up to 2.4 Gb/s are in use today. Note that the laser transmitter only needs to be on or off, according to whether a one or zero is being transmitted. This allows a transmitter which potentially costs less than a linear transmitter (as in (A)) having similar information capacity. Of course the cost of getting data into and out of the digital format is high, and this must be considered in developing a system. Fast multiplexed data may not be a suitable format for transmission to subscribers, due to the recovery cost, but is suitable for headend interconnect or for use anywhere in the network where costs are shared among a number of subscribers. (We may see some applications for data multiplexed at lower rates and transmitted directly to consumers, though.)

WHY MULTIPLEXING STANDARDS

We need a standard way of muxing in the TDM world. Granted, there are manufacturer-specific (non-standard) muxing schemes in use today, and they work well within the confines of what they were designed to do. However, they have the disadvantage that one can only obtain equipment from that specific manufacturer. If he doesn't have the particular function you need or if he leaves the business, you are stuck. If you want to interchange programming with another cable system (in a regional headend interconnect for example) and that system uses different equipment, you will have to bear the cost of some sort of conversion, if you can make the conversion at all. Maintenance is a problem because personnel must become familiar with specific techniques that won't apply if the vendor is changed. A proprietary multiplexing system will not allow a system to add telephony services easily. Finally, we can expect limited cost reductions because the vendor has no competition once an initial purchase is made.

The telephone industry has had these same problems and has evolved a standard way of handling data. We may as well take advantage of what they have developed. Not that the telephone industry has implemented their data standard perfectly: only now are standards being implemented that will let a North American signal interconnect gracefully with a European signal. Further, despite efforts to the contrary, just because a piece of equipment conforms to a particular standard (such as DS3), that doesn't mean that one manufacturer's piece of hardware will interconnect with another. However, the telephone industry had a big head start on the cable television industry in solving these problems, so we may as well take advantage of what they have developed. That way, we don't have to reinvent the wheel ourselves, and if we want to interconnect with them in the future (a situation almost guaranteed by recent events), we will be ahead of the game if we use the same standards.

Finally, the cable television industry is likely to begin handling telephone signals in the next few years, with networks we are installing now. In order to interface with other telephone vendors, we will have to use the same set of standards that they use.

In this spirit, we shall attempt here-in to describe how multiplexing is done in the telephone industry, and relate it to our needs in cable.

The multiplexing standards allow for various levels of interconnect, mostly differentiated by data rate. In order to facilitate the development of standardized equipment, only certain data rates are allowed. These can be combined in prescribed manners. The process of defining the particular ways data can be combined is called a "hierarchy." Consider relevance to the commonly understood definition of the word: "a group of persons or things arranged in order of rank, grade, class, etc."[iii] In this case, the *group* is a set of standard data rates, and the *rank* is the way in which they are ordered. There are at least two hierarchies involved, an older asynchronous digital hierarchy (ADH) and a newer synchronous digital hierarchy (SDH). The SDH in North America is the still evolving Synchronous Optical Network (SONET). The words "asynchronous" and "synchronous" refer to timing and synchronization between individual data streams. While important, this is not a topic that can be covered within the present scope.

In connection with a discussion of the hierarchies, we shall introduce a number of relevant terms and the data rates associated with them. Everything builds on what came before, though the manner in which this is done is a bit inconsistent for historical reasons. We shall not go into the history of how things came to be as they are, though a paper[iv] published not long ago contains some interesting insights into the topic, as well as other useful information.

ASYNCHRONOUS DIGITAL HIERARCHY

The older asynchronous hierarchy developed during the 1960s, as telephone companies attempted to digitize voice signals for efficient switching and transmission. A suitable data rate for carrying voice traffic is 64 Kb/s.[v] Thus, everything is developed around this data rate. Figure 6 shows the hierarchy that exists. The voice signal to be transmitted is digitized in a device called a "codec" (for coder-decoder: in the telephone industry everything is symmetrical, and where a signal must be coded, a related signal must be decoded). The output of the codec (really the coder portion) is a 64 Kb/s data stream, a *DS0* (spoken "DS zero") data stream. It is muxed with 23 other 64 Kb/s data streams, plus eight Kb/s of sync, into a 1.544 Mb/s data stream called a *DS1* signal. (Notice the data rates used here, compared with the example of multiplexing given earlier.)

The reader may recognize this as the so-called T1 rate that has been used in earlier stand-alone cable television data applications. Technically, calling this T1 is not correct. The

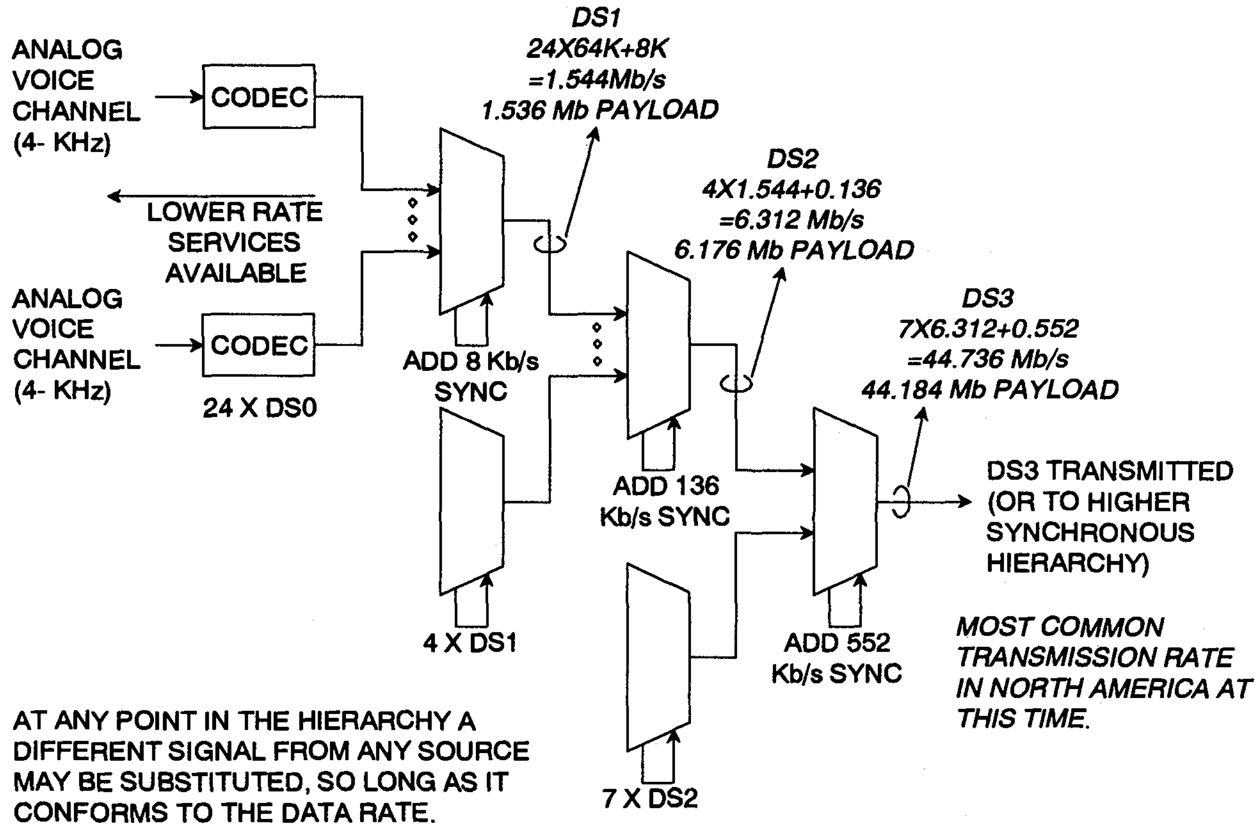


Figure 6. Asynchronous Digital Hierarchy

term T1 is reserved for a DS1 data stream carried on twisted pair (conventional telephone) cable. The definition of the data stream, which is what cable television has used before, is contained in the DS1 specification. T1 refers to a particular type of modulation used to get the data onto a twisted pair.

Note that data rates lower then 64 Kb/s (the DS0 rate) exist and can be muxed into a data stream.

Four DS1 data streams can be muxed into a DS2 data stream, and seven DS2s can be muxed into a DS3, which operates at 44.736 Mb/s. At each stage in the multiplexing, sync is added as overhead, as illustrated in Figure 6 above. The overhead is necessary in order to recover the data, but does not carry data itself. This is one of those necessary evils of life: we have to tolerate the overhead, even though it doesn't help us by carrying useful (to the user) data. Later we shall see hierarchies that depend on imbedded sync and don't utilize added overhead.

Figure 6 shows the computation of the amount of payload (useful data) that exists at each point in the hierarchy. The payload numbers shown consider sync from lower levels (to the left), to be payload, not overhead. Whether or not this is true depends on the application. If the application is voice traffic, where everything is transmitted in 64 Kb/s channels, then the sync overhead from previous levels of the hierarchy is clearly still overhead, not payload. However, if the data from a previous level is something else, it may all be useful payload data. A simple computation from information in Figure 6 shows that a DS3 channel bears up to 672 voice channels of 64 Kb/s each, or a net payload of voice data, of 43.008 Mb/s. The difference between this and the 44.184 Mb/s payload suggested in the figure, is the overhead added at the DS1 and DS2 levels.

To get a feel for transmitting video, consider that the professional D1[vi] data rate, a method of recording and transmitting video used in some broadcast and production studios, is 140 Mb/s. This is uncompressed video, with no attempt to minimize the bandwidth. Compression systems that reduce the data rate to permit a video channel to fit in a DS3 data stream, have been used for specialized transmission of professional video for a while. The motivation to squeeze the video into one DS3 is that this is currently the most commonly available data rate interface in North America.[vii] This year equipment has come on the market for cable television use, that puts two NTSC signals in one DS3. Each signal includes three audio channels (stereo and SAP) and an 18 Kb/s data stream. The application is headend interconnect.

Depending on what one believes MPEG compression will ultimately deliver, one may be able to put up to about ten compressed NTSC signals in one DS3, though the hardware to do so is not quite available yet. A disadvantage of MPEG compression systems for network use is that MPEG is an asymmetrical process: the cost of compression is very high, while the cost of decompression is low. This works in our favor for transmission to the subscriber, but works against us in the network.

Signals may be muxed into the hierarchy at any point. For example, at the DS2 level, one could mux together some combination of digital audio, voice and data signals as required, so long as each existed at a data rate supported by the commercially available multiplexers. Out of band addressable data from a headend computer to set top converters is normally transmitted at either 9.6 or 19.2 Kb/s, depending on the system. These rates are standard interface rates and can be interfaced to the multiplexer at any level where a suitable bandwidth channel is available.

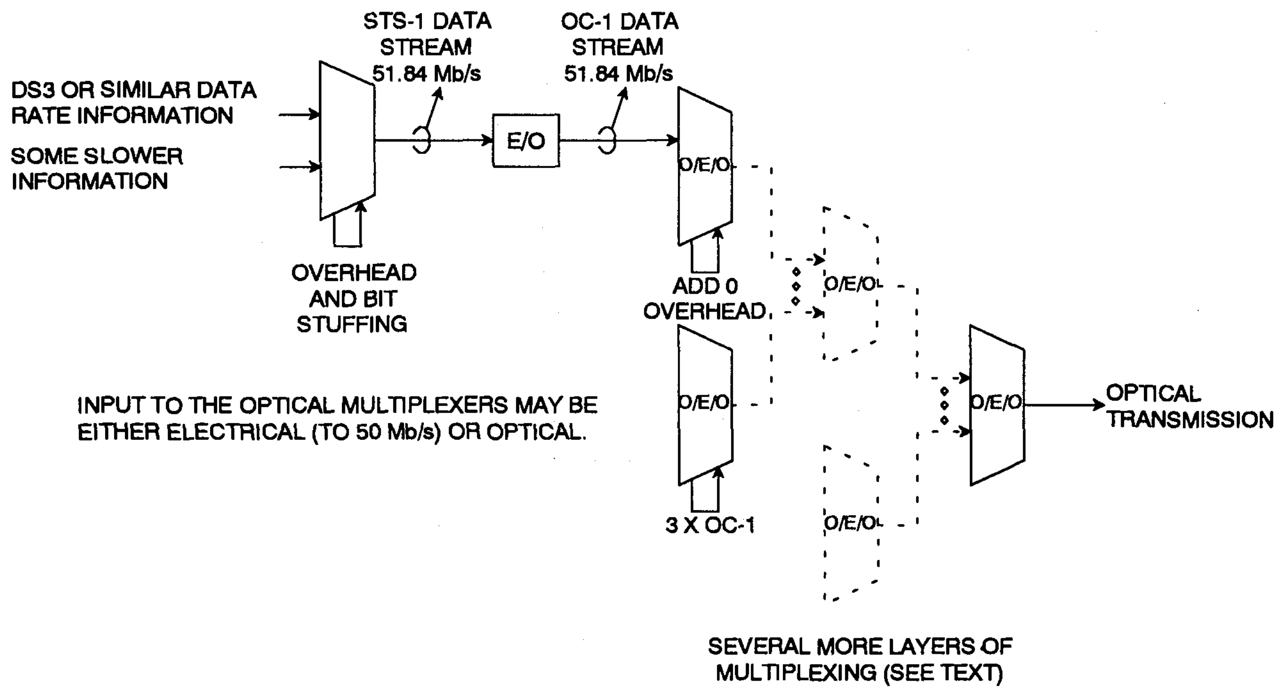


Figure 7. Synchronous Digital Hierarchy

On the other hand, digital audio systems don't operate at standard data rates, so some sort of converter must be supplied by the manufacturer or a third party, to raise the data rate to the next highest standard so it can be muxed in. The data rate is raised by "stuffing" bits. That is, extra, or "junk" bits are added to the data stream to bring the total rate up to the next standard rate that can be muxed. These extra bits are thrown away at the receive end.

SYNCHRONOUS DIGITAL HIERARCHY

The newer synchronous digital hierarchy (SDH) is currently in an advanced state of development and in initial stages of deployment. It is fiber optic based. A significant motivation for SDH is that we cease "burning" bandwidth for synchronization: the sync information already in the data is used for synchronization, so all we have to add is payload. The synchronous hierarchy begins at the DS3 level. Figure 7 shows this hierarchy.

Data at the DS3 rate (or lower) is supplied to a multiplexer whose output is at the synchronous base rate of 51.84 Mb/s. This rate is chosen to allow multiplexing in either a DS3 rate or the European equivalent. This is the point at which an attempt is made to reconcile the North American and European rate standards. (Full compatibility is not attained, however, until the next level of multiplexing.) The electrical output at 51.84 Mb/s is known as an STS-1 data stream. In the past, this signal has not been utilized for external interface with multiplexers, though one multiplexer vendor has announced product availability about mid 1994. A more common interface point is after the data stream has been converted to optical form in the electronic to optical conversion process shown. From this point on in the multiplexing hierarchy the signal interface is optical, even within the confines of one headend or other center at which multiplexing is done ("office" in telephone terminology). Multiplexers at later points in the hierarchy include conversion of the input signals from optical to electronic form, internally muxing the signals in electronic form, then re-conversion to optical on the output. Optical interfaces are much easier to handle, even in the space of one room, than are electronic interfaces. This fact is painfully apparent to anyone who has spent late nights fighting ground loop problems!

The hierarchy is a bit more consistent in the synchronous world than it was in the asynchronous. At each level, three channels of the previous level are muxed. All synchronization is built into the data streams from the DS3 level or equivalent, and this synchronization is interpreted by the multiplexers and demultiplexers. Consequently no external synchronization is needed, and no additional overhead is added in the process of muxing to higher levels.

From the OC-1 level, three data streams are muxed to form an OC-3 stream at three times 51.84 Mb/s, or 155.52 Mb/s. The process continues in groups of three to the maximum currently defined rate of OC-48, 2.48832 Gb/s. This rate can actually be carried on a single fiber optic cable today. To compute the data rate at any level of multiplexing, multiply the data rate at OC-1, 51.84 Mb/s, by the level of multiplexing. For example, the OC-48 data rate is 48 times 51.84 Mb/s. All of these frequencies are exact and are shown with no rounding error.

REGIONAL INTERCONNECTS

Now that we see the why and the how of multiplexing signals, let us examine the use of digital SONET rings to accomplish cable television headend regional interconnects. There are several reasons why one or more cable operations might want to interconnect headends.

1. Ad insertion. A regional interconnect allows ad insertion equipment to be centralized at the master headend site, with regionalized ads inserted in the appropriate channels for transmission to all other hubs on the ring. Localized ads specific to one or more hubs on the ring could be stored at the master hub site and then

transmitted over the ring to specific hub locations at the appropriate time.

2. Headend interconnect and/or consolidation. Off air satellite-delivered signals can be received at the master headend, encoded as necessary, then transmitted on the ring to all other hubs. Depending on the situation in a particular market, it might be wise to designate one of the other hubs on the ring as a back-up headend site with appropriate off-air and satellite reception capabilities along with encoding equipment necessary to interface those backup signals with the digital ring.

Satellite delivered digitally compressed video signals (such as MPEG-2) for delivery to a digital set-top box at a subscriber location will be interfaced with the ring at the master headend with devices called "variable rate multiplexers". As their name implies, those devices can flexibly package various lower speed, non-standard rate data streams, into virtual containers (called "virtual tributaries") which fit nicely into higher speed , standard rate data streams. The input to a variable rate multiplexer connects to a line card tailored to the application. The output of this line card conforms to the appropriate higher level standard rate, making it east to fit into a standard hierarchical structure.

Other digital sources which operate at non-standard data rates, such as digital music services, electronic program guides, and some downloadable game channels, can also be handled with variable rate multiplexers.

At the remote hub sites, video sources which were digitized, compressed, and multiplexed into higher rate data streams at the master headend, are now demultiplexed, decompressed, and converted back to analog format, and are then modulated to the appropriate cable channel.

Compressed digital video services which had been satellite delivered to the master (or backup) headend, are received from the high speed digital ring, demultiplexed to the correct rate, and then modulated onto an RF carrier for transmission to the subscriber location. Other digital data sources (digital music services, game channels, electronic program guides, etc.) are received, demultiplexed, and presented to an appropriate digital modulator for transmission over the broadband hybrid fiber/coax cable television system to the subscriber.

MULTIPLEXING INTO THE STANDARD

Equipment is available to multiplex in all of the standard data rates used today. These start at low rates we know from the EIA 232 (formerly RS-232) standard: 1200 b/s, 2.4 Kb/s and so on through at least 56 Kb/s. Any 64 Kb/s data stream can be multiplexed into a higher level data stream using off the shelf equipment, as can any higher level of multiplexing shown in the figures above. 1.544 Mb/s (DS1) data is commonly encountered. The 10 Mb/s data rate of Ethernet local area networks (LANs) can be accommodated with commercially available muxing equipment. We mentioned above that DS3 data (44.736 Mb/s) is the most common interface currently available in North America. Any signal at any of these data rates may be multiplexed into a higher level using equipment available from a number of manufacturers.

The table shows the interfaces that are currently available for OC-48 multiplexers. Also shown are some lower data rate levels not generally supported in OC-48 multiplexers. These rates may be multiplexed in at lower asynchronous rates, before one gets to the optical multiplex levels.

To our knowledge the data rates used in addressable converter systems are standard, so one can get addressing data from one point in a network to another with no difficulty. Of

Table 1. Interfacing Optical Multiplexers

INTERFACE	COMMENTS
DS0, DS1	Not likely to be interfaced directly in OC48 mux due to cost. Can be interfaced at lower level.
DS3	44.736 Mb/s. Most common North American interface today.
STS-1	51.84 Mb/s. Electrical equivalent of OC-1 optical rate.
STS-3	155.52 Mb/s Electrical.
STS-3c	155.52 Mb/s. Not likely due to cost, interface issues. (The "c" indicates a concatenated signal - made up of lower level signals transported as a single signal.)
OC-3, 3c	155.52 Mb/s Optical.
OC-12, -12c	622.08 Mb/s Optical.
OC-24	Uncommon.

course, there may be issues regarding how headends are logically configured. One data stream inserted at a master site will serve all hubs identically.

The situation is not as clear today concerning some of the non standard data rates used in cable television today. Various signals in use in headends today include 3.3 Mb/s and 5.33 Mb/s. These non standard rates were adopted for good reasons, but they cannot be accommodated with existing multiplexers. As cable television interests begin to use the standard multiplexing system, one assumes that equipment will become available to accommodate these data rates. One company has announced such. These data rates can be accommodated by adding extra bits to raise the rate to one of the standards that can be handled with off the shelf multiplexing equipment. For example, if a data stream operates at 5.33 Mb/s, an extra 982 Kb/s can be added to the data stream to produce a DS2 payload. That is, a DS2 data stream has a total data rate (payload plus sync added at the DS2 level) of 6.312 Mb/s. The difference between this and 5.33 Mb/s is the data rate that needs to be added, or "stuffed." The extra bits may be used for another data stream of less than 982 Kb/s rate, or if the data "bandwidth" is not needed, random data bits can be inserted.

WHAT TO EXPECT FROM DIGITAL TRANSMISSION

Arguably, this is off the subject a bit, but due to confusion we wish to briefly address the concept of quality of a digital signal in a cable television environment. This is an area subject to considerable mis-interpretation today. Figure 8 illustrates qualitatively the difference between digital and analog transmission. We plot noise level (roughly equivalent to carrier to noise ratio) vs. video quality, measured by any convenient metric. Analog transmission quality vs. noise level is shown as a heavier curve. It follows the well-known rule of gradual degradation with increasing noise. At very low noise levels the quality of the video is excellent. As the noise level increases, for a while a viewer cannot perceive the noise, so the picture quality appears to remain unchanged. At some noise level (perhaps 50 dB C/N) the picture quality begins to degrade with increasing noise. The degradation is gradual at first but as the noise gets worse the picture quality degrades until at some moderately arbitrary level we say the picture is unusable.

The lighter lines describe two different scenarios for digital video. One scenario is without error correction to the digital signal, and the other is for error correction. Error

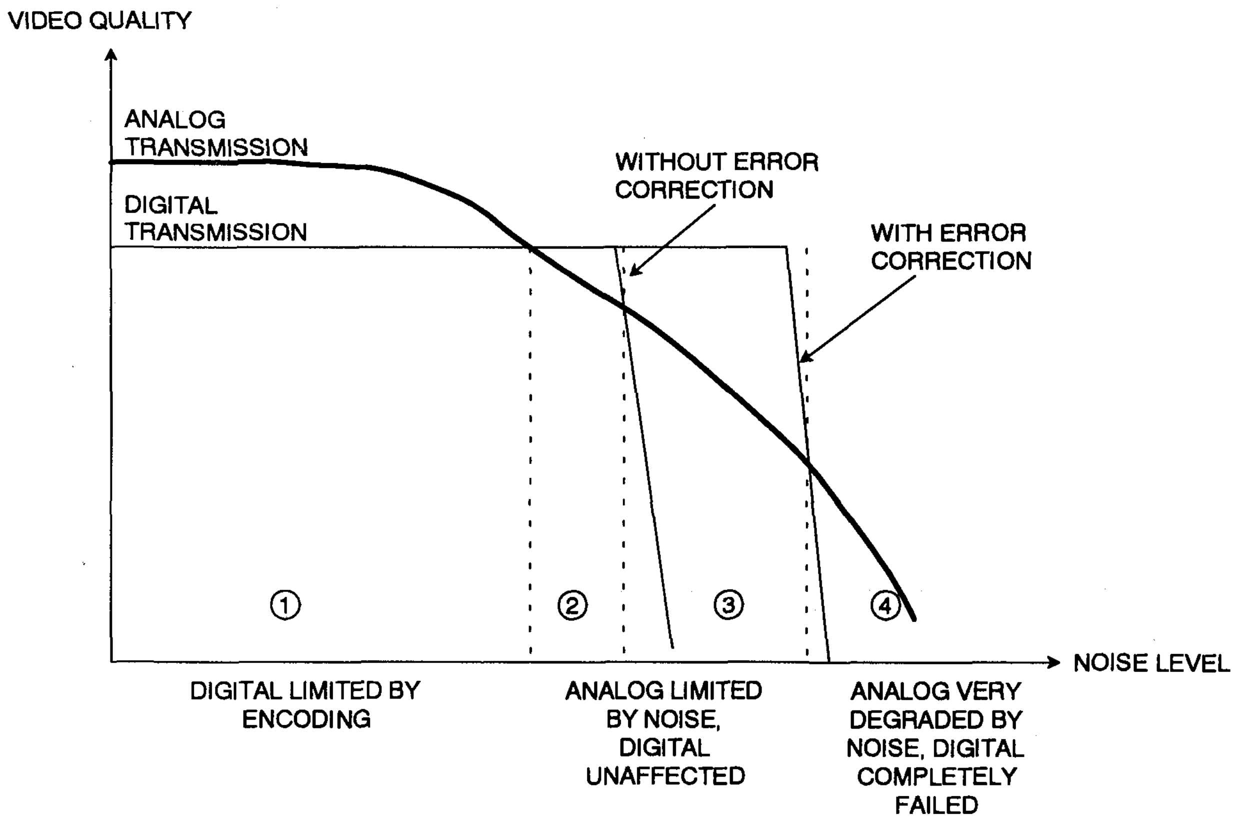


Figure 8. Degradation of Analog vs. Digital Transmission

correction is a way of transmitting extra bits with the digital signal, such that if some bits are received in error, they can be identified and corrected. Without error correction digital transmission would hold few if any advantages over analog transmission.

In order to compare analog with digital transmission, we have divided the graph into four regions identified by circled numbers. In region 1 the analog signal is better than the digital signal. This is because we have inevitably lost something when we digitized the signal, an added step. In some cases, one can argue that the digital picture is better than the analog, but this will only be true for non compressed signals that were never in NTSC format before being digitized. Compressed pictures will certainly be degraded over the corresponding analog signal, by definition (compression is a lossy process, in which some information is removed). The trick in compression is to remove information from the picture in such a way that the removal is undetectable by the viewer. In region 1, both the analog and digital pictures would be considered excellent.

In region 2 noise is beginning to affect the analog picture noticeably. The digital picture, with or without error correction, is not affected by the noise, because the noise is not yet sufficient to cause errors in recovering the bits that make up the digital signal. Above a certain noise level the non error corrected signal degrades fairly quickly, going from "perfect" to unusable with just a few decibels increase in noise (region 3). The error corrected signal in region 3 is surviving with no noticeable degradation because error correction is masking the transmission errors. The analog signal is continuing to degrade significantly.

Finally in region 4 the noise level is so great that even the error corrected signal fails. When it fails, it does so very quickly, going from a near perfect picture to nothing with very little increase in noise. The analog signal is very degraded but may remain recognizable.

The exact shape of the curve and the relative position of the analog and digital curves is a function of a lot of variables including the type of system through which the signals have passed, the type of digitization and compression used and the error correction used. If a digital transmission link is cascaded with an analog link, as in a regional interconnect done with SONET digital transmission between headends, followed by analog distribution, the signal to noise ratio of the digital link will roughly add to the carrier to noise ratio of the analog portion. The reason the addition is not exact is that the digital noise is composed exclusively of quantizing noise related to the number of bits, the compression algorithm used. This noise probably does not have the same spectral characteristic and peak to RMS value that thermal noise has.

Most certainly the signal to noise contribution of the digital link will NOT add to the carrier to noise contribution of the analog link. This can be seen from Figure 8, which shows the digital link signal to noise ratio (video quality) not changing as link noise increases in the way noise adds in the analog world. The signal to noise ratio measured at the originating point is the same as the signal to noise ratio at the receiving point.

We have been a bit sloppy in mixing baseband signal to noise and RF carrier to noise above. If the baseband signal to noise ratio is measured using the CCIR unified weighting network, using the proper definition of "signal," then it is numerically within a couple of tenths of a decibel of the RF carrier to noise ratio as defined by the NCTA. This is the way many instruments make signal to noise measurements today. However, it is a source of considerable confusion, and addition of noise on combined digital and analog links must be done very carefully.

ACKNOWLEDGMENTS

K. Lynch and R. Reynard contributed valuable research material. T. Engdahl contributed a vast amount of information concerning telephone practices. M. Dionne was very rough on the manuscript, but her suggestions improved the paper immeasurably.

END NOTES

[i] There are limits to what we can do with the signal: digital transmission is generally characterized by an all or nothing situation. If the signal is above a threshold, the signal out is just as good as the signal in. Below that threshold, the signal is lost completely.

[ii] In some transmission systems we would have two transitions per bit, but this is usually done when economy is at a higher premium than is bandwidth efficiency.

[iii] *Webster's New World Dictionary of the American Language*, World Publishing Company.

[iv] McGrath, C. J., *Digital Delivery Technology for CATV Networks*, Technical Papers of the SCTE Fiber Optics Conference, 1991.

[v] In telephone terminology, the voice transmission quality needed is called "toll quality." As described in the McGrath paper, achieving this quality at 64 KB/s required compression and expansion respectively at the transmit and receive ends of a circuit. Compare this rate for transmitting voice with the approximately 704 KB/s data rate to transmit one channel of CD quality audio!

[vi] Perhaps the use of the term "D1" here is confusing: D1 stands for the *first digital* video recording standard to be used. It has nothing to do with the DS1 multiplexing level of telephone company usage.

[vii] The corresponding European rate is 2.048 Mb/s, known as E1.

Interoperability Requirements For Interactive Cable

John Richardson
Dave Williams
Robert Adams
Intel Corporation

Abstract

The term 'interoperability' denotes a concept quite familiar to the computer and communications industries. Interoperability is the concept of systems working together for the purpose of transparent exchange of information. It is a very difficult thing to accomplish and is full of both technical problems and political pitfalls. The computer and communications industries have learned, with great pain, that interoperability must be addressed at several different levels to be effective. The Cable industry is about to come face to face with these issues as it tries to implement a broadband network to the home.

This paper will explore the essential elements of interoperability and outline some of the essential requirements to stimulate debate in the cable industry. It presents some of Intel's findings in developing systems for cable that are open to interconnect with other systems and will cover those areas considered most crucial to the achievement of interoperability in interactive cable.

1. WHAT IS "INTEROPERABILITY"?

Interoperability is one of those terms where "everybody understands what it means but nobody can define it." It is often defined by example. A remote is said to be interoperable with a TV if it can control the TV. This is a rather narrow interpretation. If the remote and the TV were designed together, then interoperability is not too interesting. However, if the same remote can control many TVs, then interoperability becomes a market factor and interoperability becomes interesting. This broader interpretation of interoperability has two requirements:

- (At least) Two devices must communicate, and
- Those devices must have been independently designed.

Perhaps this terminology from the computer world below will help with the distinction:

- *Portable* — typically applied to software that can be moved to another environment and it will work. Binary portability means that the software can be moved to any system and it will work without modification. Source portability means that the original program needs to be recompiled in the new environment for the software to work, but that no source changes are required.
- *Connectivity* — this term is applied to multiple systems that are designed to work together (for example, a TV and remote from the same manufacturer)
- *Coexistence* — groups of systems that provide connectivity, but do not interoperate (for example, 2 TVs and 2 remotes where each remote only works with one TV)
- *Interchangeability* — systems that have identical functionality and interfaces
- *Interoperability* — multiple systems that were independently designed to work together

The cable industry is striving for *interoperable* systems — not necessarily limited to *interchangeable* systems. For example, a set-top converter might offer an electronic program guide. A PC connected to the same cable system that was able to also read the electronic program guide would be interoperable with the set-top, though not interchangeable.

Interoperability is required when there could be incompatible solutions for the same problem. Interoperability is achieved through standard interfaces.

Many of today's standard interfaces are actually a compatibility layer that translates between one interface and another. A *compatibility layer* is software that translates between one interface and another. In the "real world", a language translator serves this function. Using a translator, neither the speaker or the listener need to change. This is important in the computer world, because it supports diversity, and diversity allows for innovation. Interoperability should not require that two solutions be identical, just that a developer can easily take advantage of either solution.

Complete interoperability could lead to one or more of the following:

- The ability to move cable box from one house to another
- The ability to have settops and computers share the same network
- Consumer choice about what to devices plug into the cable network — but any device will work

2. DISCUSSION FRAMEWORK

To provide a framework for the discussion below, we must agree on some basic concepts. These concepts have been the source of much debate in the cable and computer industries. I do not claim to have *the* answer. Instead, I will simply provide working definitions relevant to the discussion of interoperability.

It's not just video any more — the cable system is evolving from the distribution of entertainment video into a full service network. This network will include today's video, along with digital video, computer data, telephony and digital multimedia services.

Full service is not just a spiffy video selection system — data in the cable plant will be used for far more than video on demand and electronic program guides. Data will be the basis for a whole host of rich interactive services. Often cited examples include shopping, tele-commuting, personalized news, personal messaging, and interactive games.

Tomorrow's applications won't look like today's computer interfaces — This interactive, full service network will be competing for the consumer's leisure time. Its face will be competitive with today's TV -- fast moving, engaging, with sound, action, and rich colors.

It's the Services, stupid — Data connectivity is not just pushing bytes. Consumers are not interested in just connectivity, they are interested in services. They want to be able to do new and interesting things. Consumers will need to be isolated from the details of network implementation and management. They want to believe that they are interacting personally with the service of their choice.

Services will come from far and wide — while many services will be provided at the head-end, many more will come from regional or national distribution centers. The full service network will need to connect to these remote service centers.

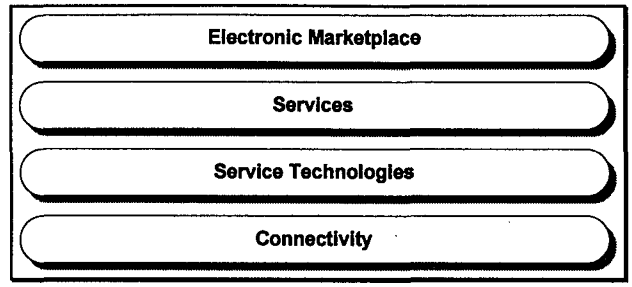
New billing models will be possible — by providing computer-controlled connectivity, control and usage information will be available with more detail and accuracy than every before. This will enable billing systems based any combination of subscription, time, per-use, resource usage, application activity (e.g., when you use a certain part of the system), electronic transactions (e.g., buying the right to view a photo), and, perhaps, supplemented by advertising.

3. THE LAYERS OF INTEROPERABILITY

These new applications will need to interoperate at all levels. If one layer is missing, the whole capability falls apart. For example, you can pick up the phone and call anywhere in the world. However, that doesn't mean that you'll be able to carry on a conversation with whoever picks up the phone.

In this section we will build interoperability from the bottom up. We will start with the most fundamental layer (electrical) and work all the way to the most complete (interoperable applications). In a later section we will provide an overview of the relevant standards for each layer.

Connecting services together is a complex task. At Intel, we use the model below to help organize this connectivity so that we can focus on key problems independently. This model is based on standards and experience from both the cable and computer industries. It is not meant to be a rigorous treatment of networking, but instead a framework for discussion. Those readers familiar with computer networking will recognize this as a collapsed version of the ISO stack. Those familiar with cable TV standards will see that the model is largely built on top of existing industry standards and practices.



3.1. Connectivity

3.1.1. Electrical

Today's cable TV provides a basic framework for looking at electrical connectivity — electrical levels and timings are specified and frequency spectrum allocation is determined. For example, in the US, downstream bandwidth is allocated in 6 MHz units. Depending on the individual cable plant, downstream bandwidth may be available anywhere in the 50 MHz to 750 MHz range. In most cable plants that are 2-way capable, the upstream bandwidth is limited to the 5 - 35 MHz range.

It should be noted that individual cable operators are the ones who actually assign the specific channels to various services. Standards may specify the available options, but cable operators will determine the service mix based on their particular market and installed technology.

3.1.2. Modulation

The next step is deciding how to represent binary data in these channels. This is the modulation scheme.

Modulation provides a method for sending binary data over an analog medium. For example, a very simple modulation scheme might specify that an 1800 Hz signal for 10 mS represents a zero and a 2400 Hz signal for 10 mS represents a one. Using this scheme, one could send 100 bits in a second.

Today's modulation schemes are far more sophisticated, utilizing amplitude, frequency, and even phase to communicate data. Modulation scheme selection is based on carrying capacity, immunity to noise, and cost to implement. Current modulation schemes include QAM (quadrature amplitude modulation), VSB (vestigial side band), QPSK (quadrature phase shift keying), and FSK (frequency shift keying). Each has its strengths and weaknesses. Currently VSB and QAM are used primarily for downstream transmission, while QPSK is used for upstream.

Once we have a method to send streams of bits down the wire, the next step is selecting a coding scheme that allows the receiving unit to tell data from idle noise. A coding scheme typically involves specifying a unique pattern that begins a bunch of data and a similar pattern that signifies the end of the data. For example, one system specifies that each data burst begins with 24 bits of alternating zeros and ones.

Many of the systems in use today also select a coding scheme that includes some error correcting capabilities. For example, an error correcting scheme may use 6 bits to code a 4 bit value. If any one of the 6 bits has an error, the receiver can detect that an error has occurred and, if the error is only one bit gone wrong, can decide what the actual data *should* have been. Error correction schemes that allow the receiver to determine the correct data, even in the presence of small errors, are called *forward error correction (FEC)*. FEC is particularly important for time-critical information, since the receiver doesn't have time to say, "I'm sorry, Mr. Sender, I didn't get that last bit of data, could you please send it again?" The science of error correcting codes comes largely from the telecommunications industries and is the subject of a vast amount of research.

ABCDEFGHIJKL → ABC DEF GHI JKL

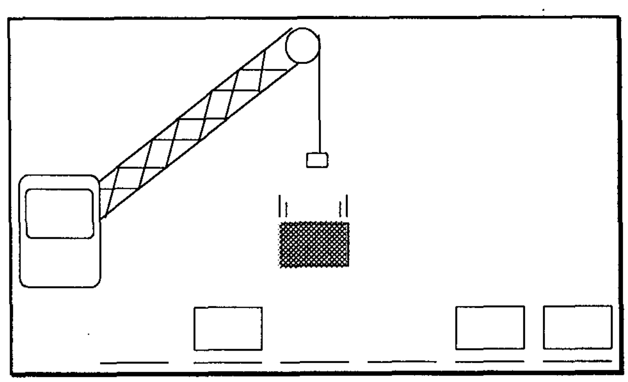
Rules are also set up to define how many characters can come "in a row". These are called packets and the rules specify the largest number of bits that will come across in a group. Special codes are allocated to mark the beginning and end of a packet.

At this point we have defined enough to send packets of data.

3.1.3. Media Access Control

The next step is to figure out how to share the data channel. Because spectrum is a precious resource, we must be efficient in its utilization. This is even more important in sending upstream data, since all consumers must share the scarce 30 MHz of bandwidth. This means all consumers using all services, including pay-per-view television, telephony, wireless communication, and computer data.

There are a number of standards that define media access control. For example, Ethernet sets rules that say anybody who wants to send something should listen first (to make sure nobody else is in the middle of sending) and, when the line is clear, go ahead and send a packet. Of course, there are complex rules about what happens if two devices start to send at the same time.



In the world of cable, access control is a much larger issue for the upstream connection than it is for downstream. For downstream there is only one unit "talking" -- the device at the head-end. For upstream, however, every home could potentially have data to send.

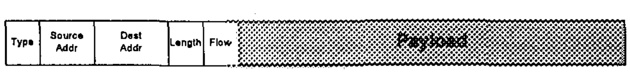
The upstream problem is aggravated by the fact that individual homes in the cable plant can't "hear" all the other homes. This is because of the architecture of the coax plant and amplifiers. Suffice it to say that this makes using the scheme that ethernet uses more difficult (though not impossible).

Sharing upstream cable connections is an area where no standards exist. There are many solutions that have been built (or at least tried in the lab), but none is acknowledged as clearly the best. For the industry to achieve interoperability, upstream media access control will be a major obstacle.

To continue building our interoperability layers, we now have the ability to put a packet onto the network.

3.1.4. Packet Layout

The next step is to decide what the data in the packet means. Packets are defined to have some *header* information that is used to deliver the packet to the right place at the right time, and the rest of the packet is known as the *payload*. Part of the header is reserved for addressing information. This provides a way for an individual network component to determine whether the packet is of interest or not. In data networking, packets include information about who sent the data and who is supposed to receive it. These are referred to as source and destination addresses.



Certain types of data (broadcast video, for example) don't need both a source and destination address. The destination is "anybody who want's to see the video [and is authorized to do so]". The receiver selects the desired stream based on the source address only (e.g., "this is HBO").

This layer has many incompatible standards. There are a number of packet layouts defined by the cable industry for digital video transmission (e.g., MPEG). The computer industry has standards such as ethernet, token ring, ATM, IP, IPX, SNA.

3.1.5. Layering — Division of Labor

Layering is used as a way to isolate the details of connectivity from the software that takes advantage of it. In the computer world, a PC might see a file server as just another disk drive. Any program that wants to get information just reads and writes files as if they were local. However, the next layer down worries about *redirecting* those file access requests to a file server running on another machine on the network. The redirector is only responsible for intercepting the file requests and saying to the next layer down, "please send this request to the file server". The next layer down is charged with making sure the request gets there, retransmitting the request if the file server takes too long and didn't appear to hear, and making sure that the response from the file server has the valid data.

So, above we have created three layers:

- application (thinking it's reading and writing local files),
- redirector (pretending it's a local disk and sending requests to the file server), and
- network (making sure information is reliably sent between the client PC and the file server).

Notice the division of labor — the application doesn't care how its requests get carried out, it just wants the data, while the network layer doesn't care what information it needs to transfer, it just cares that the information gets to the right place and that it's accurate. This is a powerful concept and very important to networking. Each layer has a single job to do. It relies on the support of the layers below to perform their job, but the basic job of a layer is to provide services to the layers above.

Let us return to our packet example from the last section. As information is passed *down* the networking stack, each layer treats the request from the layer above as data (payload), adds its own control information around the data, and passes the request to the layer below. At the bottom of the stack, the physical layer gets a bunch of bits that it needs to send out on the wire. On the receiving end, the reverse happens. Each layer of the stack looks at its control information, makes sure that it does its job, and passes the payload to the layer above.

So, for example, TCP is responsible for the reliable delivery of information to the other end of the connection. IP has the responsibility of taking a single packet and asking that it be sent over the network. IP is refereed to as a datagram service — it will put the packet on the wire but it won't guarantee that it gets there. So, TCP must pass a packet to IP, and mark it so that the TCP layer of the stack on the other end of the connection gets that packet. The receiving TCP layer gets the packet and sends an acknowledgment to the sender saying that the packet was successfully received.

If there is a problem and the packet doesn't make it, the sender's TCP will recognize that "it's been too long — maybe the receiver didn't get my data, so I'll send it again." This is how TCP makes sure that it can reliably deliver data — it has the receiver send an acknowledgment back for each packet. Note that the IP layer doesn't care what TCP does. It gets a packet and does its best to send it but doesn't really worry about it if the packet doesn't make it. TCP, on the other hand, doesn't worry about how IP delivers the packet. It just passes the packet to the IP layer and counts on IP doing the rest. If IP doesn't get the packet through in a "reasonable" amount of time, then TCP will try again.

It should also be noted that IP doesn't really care what medium the packet is sent over. If the layer below IP uses ethernet, token ring, or ATM, the packet will still get to the receiving end. In fact, if the packet is sent part way over ethernet and then the rest of the way over ATM, IP still won't care. The actual method for packet delivery is the responsibility of the layers below IP.

This is a good example of how the computer industry has used layers to hide differences. The upper layer formulates a packet and asks the next layer down to deliver it. It is the responsibility of the next layer to worry about actually how to deliver the packet. The upper layer only cares that its counterpart on the receiving system gets it.

To resume our narrative, we now have achieved the ability to send a packet to anybody on the network. This provides basic connectivity. We can now start to "do something"!

3.1.6. Protocols

Protocols are conventions for interaction. We have so far achieved the ability to send a packet to somebody on the network. What will they do with it? How will they know what it means? How can we be sure they got it? That's where protocols some in.

Let us invent a protocol that will allow you to ask your bank for a list of the last 5 checks you wrote. In our protocol, the bank's computer hangs out waiting for a request from you (or anybody). When you formulate your request, you need to tell the bank computer what you are asking for (last 5 checks, please), and what your account number is. (We'll ignore the security implications for now.) You write this information into a packet, address is to the bank's computer, and ask the next layer down to deliver it.

Source = You	**Dest = Bank**	**Request Last 5 Checks**	**Account number**

The bank's computer gets your request, looks up the information on its database, and creates a response. The response is formatted so that there are 5 entries, each one has a 5 digit check number and a 10 digit amount. The bank's computer gets your address from your request and sends the response.

Source = Bank	**Dest = You**	**Check 1**	**Check 2**	**Check 3**	**Check 4**	**Check 5**

One problem we might have is that the connection is noisy. Suppose your response got slightly garbled and it says you wrote a check for $1,000,000. To deal with that problem, we will modify our protocol. We'll have the bank computer put the 5 checks in, as before, but it will also add a 6th entry that is the total of the 5 checks. That way when you receive the packet, you can add up the 5 check amounts, compare their sum to the bank's total, and see if they match. If they don't, you know you got a bad response and you just ask the bank to try again.

Source = Bank	**Dest = You**	**Check 1**	**Check 2**	**Check 3**	**Check 4**	**Check 5**	**Total**

Why is this important to interoperability? Well, interoperability requires that both the sender and receiver be following the same protocol. The sender needs to know how to make requests of the receiver and the receiver needs to know to listen for those requests. They each need to agree on the meaning of the data that is sent. It would be pretty chaotic if the bank thought the first character of the packet was the type of request while the sender thought that was where to put the account number.

3.1.7. Gateways

Often systems are designed under different circumstances and build on top of different standards. To allow these different systems to interoperate, *gateways* are often used. A gateway is basically a translator. To the sender, a gateway appears to be a receiver of the matching standard. However, the gateway takes the original sender's request and reformulates it to another standard and forwards it on to a recipient. The recipient thinks that a system matching *its* standard has made the request and responds appropriately. The gateway translates the response into the form appropriate for the original sender and sends it back.

Gateways are interesting in that they require no changes to either the client or the server — they both think that they are interacting with a "like" counterpart. It is not always possible to build a gateway to translate between two standards, since the basic operations of different standards may not always have matching counterparts. However, since similar standards often perform similar functions, it is usually possible to map the majority of the capability through a gateway.

Gateway systems are often used to connect dissimilar database management systems. In computer networking, this functionality is often performed by a bridge or a router — both devices that take input in one format and output that same information in a different format. The distinction between bridges, routers, and gateways is beyond the scope of this paper.

The interested reader should consult a text on computer networking for more details.

3.2. Service Technologies

3.2.1. Making this a business

Cable operators are in the *business* of delivering services. This means that they need to deal with all aspects of running that business. All of the problems below must be addressed — whether with custom solutions or common technologies. These technologies include:

- *Access Control* — customer management, enabling and disabling services for customers

- *Security* — dealing with the issues of privacy, theft of service, theft of content, perhaps through technologies such as encryption

- *Metering & Billing* — metering is the measurement of usage, billing is using that information to bill the customer, along with all the other functions that come along with billing (accounts receivable, ...).

- *Management* — making sure the network stays up, dealing with capacity issues, solving problems, planning for expansion, etc.

All of these are "behind the scenes" activities. However, they still have serious implications for interoperability. Access Control, for example, will play a key role in interoperability. If the consumer's system doesn't have the appropriate capability to participate in the access control protocol, then it will not be interoperable.

3.3. Services

From the consumer's point of view, they are connected to services. The fact that they are using a network-connected device — whether it's a set-top or a PC, is less important. Each service does something for the user and many of the services are interconnected. The consumer doesn't necessarily distinguish the lines between services — the user wants something done and asks "the system" to do it. How it happens is only relevant to the builders of the system, not the consumer.

3.3.1. Directory

The directory provides the most basic service to the user. It provides information about "what's out there". This is where the user learns about all the other services that are available.

From a software point of view, we should be careful to distinguish between the application that presents information to the user, and the underlying service that provides the information. The user could be presented with the perspective that they are running "their portal to the wide world of cable services". In fact, they are running a local application that queries the directory service and presents what it finds to the user.

The directory will contain not only service information for presentation to the consumer, it will also contain enough information to enable the network access device to find the server for that service, and how to talk to it (the protocol).

3.3.2. Individual Applications

Each service the user can activate will have its own way of interacting with the user. However, there are a number of capabilities that will be similar for many applications. These capabilities will be supported by common technologies. The basic idea is that the problem is solved once and the different application developers can take advantage of it. This is not a requirement — each developer will have the option to build their own solutions.

From an interoperability perspective common technologies will mean service will behave in a similar fashion. For example, if each service that charged the user money asked for confirmation in a different way, the user could quickly become confused. A common (interoperable) technology to manage the user's account would solve that problem.

3.4. Building an electronic marketplace

As new and richer services become available, the cable industry will move toward creation of an electronic marketplace. Home shopping will be the likely fore-runner, but it will by no means be the only participant. There are many potential players in this arena, from the cable operators themselves to traditional retailers, electronic retailers, and even today's financial industry.

Support for the electronic marketplace will require the technologies of an electronic "trading floor". These could include:

- *Electronic transactions* — the ability for a service to ask to have a customer pay an amount. This must include audit trails, customer verification of the amount, and so on.

- *Authentication* — ways for the customer to "prove" they are who they say they are (something similar to a PIN), for the service provider to "prove" they are legitimate (and not some hacker trying to steal money), and so on. This activity has many parallels in today's credit card world.

- *Account Management* — the ability for the consumer to control their expenditures. This includes setting limits, managing sub-accounts (the kids), verifying purchases, and funds transfer.

- *Marketplace* — some way for consumers to find vendors. This could be operated by the cable operator or, perhaps, by outside companies. This could be far more than "yellow pages". These gathering places could include interactive advertising, support test drives, and provide vendor "presence" with customized identities. Different marketplaces could set up reputations for having different types of vendors, and so on. The technology to provide this "world" will instantiate the electronic marketplace.

Interoperability is absolutely critical for this environment. Today's credit card companies will likely play a role here, but existing standards are not sufficient to deal with all of the problems and opportunities of electronic transactions.

4. SHARING INFORMATION

The top level of interoperability is sharing application information. Consider the ability for the shopping application to interact with the personal finance application, automatically verifying funds from the appropriate account and allocating the purchase to the correct budget category. Or, perhaps, providing your personal likeness to the shopping application so you could see what that sweater would look like on YOU.

The problems posed by this level of information sharing are large. They will require evolution of the applications and services before interoperability can be achieved. The evolution of interoperability might start with established services that have natural synergy working together to provide direct connectivity. Once experience is gained with specialized connectivity, generalized interoperability can follow.

5. STANDARDS

This section provides a short list of some of the key standards mentioned above. This list is not all inclusive; please forgive me if I miss an important protocol or standard. I include this list mainly to highlight the fact that interoperability and standards require cooperation at many levels.

As described above, connectivity is a many-layered problem. Standards are applicable at each layer. In fact, if the layering model is done correctly, the standards at one layer will have no impact on the layers above or below. In the computer world, this is illustrated by the fact that the TCP/IP standard does not change whether it uses ethernet, token ring, or ATM to provide the packet transport.

Interoperability is achieved through standard interfaces. Standardization can be as simple as two people agreeing to do things the same way. It can also be an arduous process involving entire industries. It's an evolutionary process.

The concept of a full service network is one of the offspring of convergence. It is still in its infancy. Many companies are working hard to provide solutions that work. They will worry later about interoperability.

There are many standards being brought in from other industries. Where possible, those will be used to save time and money. However, many of the key problems are not yet solved.

The important thing to remember is that standardization *will* happen in this area. If it doesn't, the converging industries are doomed to the same result as the tower of Babel.

Key Standards for Each Layer

Modulation	QAM, VSB, QPSK, FSK, ...
Media Access	Ethernet, X.25, ATM, MPEG, proprietary, ...
Reliable Connection	TCP, SPX, ...
Computer Addressing	IP, IPX, ATM, ...
User Addressing	DNS, X.400, phone numbers, postal service,
File redirection	Windows for Workgroups, Netware, Network File Service (NFS), Vines, ...
Access control	Cable equipment vendor proprietary, network OS vendor proprietary, ...
Directory	Proprietary, X.500, ...
Database	ODBC, DBMS vendor proprietary, object database interfaces, ...
Electronic Mail	X.400, SMTP, MAPI, VIM, CMC
Transaction Processing	DCE, Tuxedo, Encina, ...
Billing and Electronic Transactions	Proprietary

6. HOW DO WE ACHIEVE INTEROPERABILITY?

- Use existing standards wherever possible (don't reinvent the wheel)
- Provide gateway capabilities where standards differ
- Provide layered software that isolates the impact of differences
- Work together in industry forums to drive toward common interfaces

7. CONCLUSION

Interoperability is not a simple matter of electrical connectivity. It must happen at all layers — from connectivity through services. Existing standards will provide a strong foundation, but much work remains to be done, especially at the services and service support layers.

Delivery of rich, interactive services is new for everyone. There is lots of experience in related areas from all of the contributing industries but nobody has experience with delivery of these new services on a large scale.

Convergence has spawned a new industry by combining the skills and products of the cable, computer, communication, and content industries. Interoperability will be required if this fledgling industry is to survive.

LONG TERM COMPATIBILITY

Vito Brugliera
Zenith Electronics Corporation
Glenview, Illinois

Abstract

Today's television receivers and VCRs tune the cable spectrum, however some consumers experience problems when connected to cable systems. The set-top decoder, when connected to a television receiver, reduces the functionality of a television receiver to a monitor, disabling most of the television convenience features. Despite the availability of other alternatives, for cable operators the technology of choice for tuning access, subscriber denial and control is the set-top addressable decoder. As television technology becomes more sophisticated the loss of these features creates frustration and anger with consumers. Congress responded to this frustration, as well as perceptions about cable operators, and the Cable Act of 1992 resulted. We shall review the history, causes, legislation and what can be done long term to alleviate this problem.

Cable History

Cable television originated in the rural areas of Pennsylvania, where mountains interfered with reception of terrestrial TV signals. In 1948, an enterprising appliance retailer built an antenna tower on a mountaintop, ran cable to his appliance store, and set a TV in his window. Residents nearby purchased TV sets and required cable from his store to their homes for reception.

Today, over 98% of television households are passed by cable and 65% of them are connected to cable.

FCC involvement in cable began in 1965, requiring cable systems to carry all local television stations. The FCC also prohibited duplication of local signals by importing signals from another city, and importation of distant signals into the top 100 markets. This latter prohibition was rescinded in 1972.

The 1975 launch of SATCOM I into a geostationary orbit offered a cost effective way for programmers to distribute software to cable operators. Later that year HBO initiated a premium pay movie service that was distributed nationwide to cable operators. Cable has evolved from simple delivery of terrestrial signals to remote areas to the dominant provider for television viewing. Today, there are over 90 different program providers offering choices for every taste.

Technology Developments

Tuning Access

The rapid expansion of cable programming choices quickly exceeded the tuning capacity of television receivers then on the market. Television tuners initially tuned only the VHF band, then after the FCC mandated all-channel tuning, UHF. These terrestrial channels occupy discontinuous portions of the RF spectrum. Cable operators placed additional programming on channels in the gaps of the VHF spectrum, effectively adding channel capacity. Operators used RF block converters to translate these channels into portions of the RF spectrum that television receivers could tune.

These block converters served several purposes: additional programming was available despite television tuner limitations; reception of programming was denied to unauthorized viewers; DPU (direct-pick-up) problems were reduced when co-channel interference was present.

However, television tuner technology was evolving as well. Electronic tuners replaced mechanical tuners and the tuning range of tuners was easily extended to cover the "holes" in the VHF and UHF spectrum. Thus was born the "cable compatible" television receiver, to which the cable service could be connected directly without the apparent need for a set-top box.

Signal Denial

Cable-compatible tuning solved one problem, but created another problem. The extended tuning range of receivers, without the need for block converters, eliminated the means of denying signals to unauthorized viewers. It was necessary to use other techniques for signal denial. One method was the use of "traps" (RF filters) to prevent desired signals from entering the home. Traps are designed to filter out a single channel or a band of channels.

Traps are physically installed between the pole and the user's home. This was satisfactory for a time. However, as programming choices continued to expand, consumer interest became more volatile. This required frequent truck runs to physically change traps and an increasing operational expense for cable systems.

Control

The operational expense of physically changing traps became the impetus for a different technology. What was needed was technology that allowed control by remote authorization. The solution was the addressable converter. The basic technique attenuates the synch signal, confusing the television receiver signal processing circuits, which are unable to separate the synch signal from video. The resulting video is scrambled and not viewable.

The set-top cable hardware technology employed for control and denial is not designed to any industry standard. Each vendor has proprietary technology and there is little interchangeability among suppliers. This lack of a single standard allows innovation and serves to inhibit piracy because of no single target.

Control is achieved by sending a parallel signal, out-of-band or in-band, containing information which allows restoration of the synch signal. Using this technique, individual channels can be scrambled or unscrambled remotely from the system headend. Operationally, this was a success and initially met with user acceptance.

Television Developments

At first, cable customers welcomed the set-top addressable converters. The set-tops brought new programming choices and convenience: remote control tuning, volume control and mute. In the 1970's and early 1980's remote control was available only in high end television receivers. These features represented retail increments of $100 to $150, the perceived value was high and the box was welcome at a small monthly rental fee.

However, the consumer electronics industry is dynamic, innovative and continually driving down costs with new technology. Remote control, cable compatible tuning, other features such as picture-in-picture, originally in high end models, were driven down into the product line, with smaller and smaller retail cost increments. By the 1990's remote control was already in 90% of 13 inch receivers.

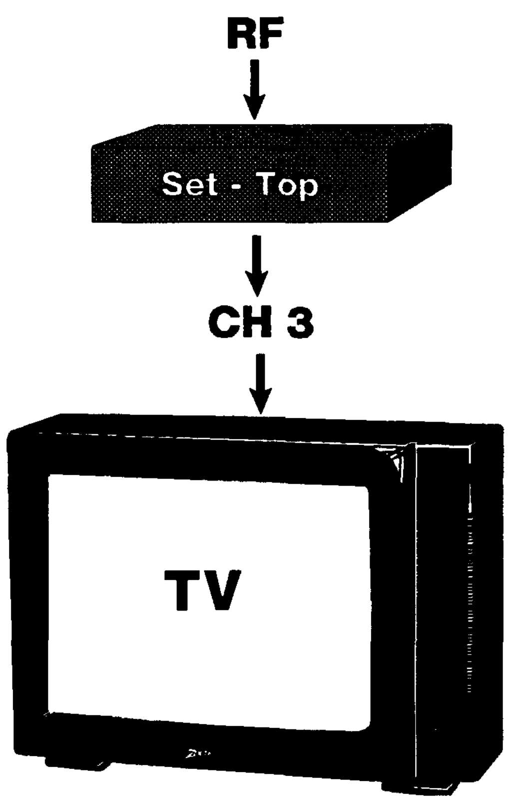


Figure 1

The rapid rise of VCRs increased the resentment against the set-top converter, especially when watching and recording two different programs. Traps, which physically removed channels, brought only in the clear signals into the home, were transparent to users and did not interfere with these features. The set-top converter did. When a set-top was placed between the cable and the television receiver, the receiver essentially was reduced to a channel 3 monitor. As television receivers became more featured the set-top began to lose its welcome. See Figure 1.

Regulatory Consequences

Despite the decreased use of set-top converters brought about by a decline in premium pay subscriptions and a greater preponderance of basic only customers user compatibility problems persisted. Cable vendors and operators made attempts to alleviate the user compatibility problems, but perceived progress was slow. By 1992 many of these user issues: -- loss of features, remote control rental, VCR problems and frequent rate increases—became political issues. The result was legislation reregulating the cable industry that survived a presidential veto. The Cable Act of 1992 not only reregulated cable rates—ultimately reducing rates by 17% and cutting industry revenues by $ 3 billion—but Congress also mandated that the FCC promulgate rules to resolve consumer cable and equipment compatibility issues. The cable and consumer electronics industries, under the auspices of the NCTA (National Cable Television Association) and the EIA (Electronic Industries Association) formed an

Advisory Group consisting of technical, administrative and marketing personnel from both industries. The Advisory Group, along with other commenters, assisted the FCC by recommending solutions for solving the compatibility problems. In particular, the Advisory Group recommended a set-back decoder interface allowing use of the television tuner and providing for communication between receiver and set-back.

The Cable Act of 1992, in order to preserve for consumers the full functionality of their deluxe TV and VCR equipment, calls for improving compatibility between consumer electronics and cable hardware while maintaining the cable companies' control over signal access and security. This section of the legislation is focusing attention on the considerable work done by both the consumer electronics and cable industries in recent years, in pursuit of solutions to these very problems. Predictably, each industry leaned toward solutions which put the major burden of performance on the other, but the Congressional mandate and FCC ruling have spurred progress.

Preferred Consumer Electronics Solutions

Solutions such as Interdiction and Broadband Descrambling promise clear, broadband signals being delivered to the consumer's video equipment, permitting full use of all the functions and features of that equipment. In this environment, the consumer electronics industry could continue to make and sell the highly-featured products that many consumers demand (and the products which generate better margins for a profit-starved Industry).

The cable industry cites high cost and limited security as substantial shortcomings of these approaches. Interdiction schemes add the cost of hardware even for subscribers not buying premium services, and the clear signals transmitted on the distribution system are especially vulnerable to theft in multifamily housing. Broadband descrambling reduces the number of security options available to the cable operator, and fails to block the audio on scrambled channels...a condition unacceptable to many local authorities. And in addition to these limitations in today's analog environment, neither approach will work in a future environment which includes compressed digital signals.

Preferred Cable Solutions

The cable industry desires that TVs and VCRs be made more truly "cable-ready," that is, having more resistance to interference, more channel capacity, and a setback device "interface" such as the ANSI Standard 563 "Multiport." The consensus is that the "Multiport" interface as originally defined is now no longer adequate because of changes in both industries. A new decoder interface has been proposed that offers users full enjoyment of all television receiver features, while cable signal access and security would be handled through a "set-back" (rather than "set-top") box.

The consumer electronics industry points out that these design changes, if required for all TVs and VCRs, substantially increases the cost of these products and places the burden of higher prices on all consumers, whether or not they subscribe to the cable services which drive the need for the enhancements. In addition, the sorry track record of the consumer electronics industry in implementing price increases suggests it may not be possible to pass along even some of the increased costs. And finally, this approach does not improve the fortunes of the owners of the 180 million color TVs and 100 million VCRs already in service.

An Interim Solution

The new decoder interface offers a solution in keeping with the intent of the Cable Act: improvement in compatibility with consideration of the costs and benefits to consumers and with consideration of the control and security needs of cable operators:

- Establishes a new "cable-ready" specification for TVs and VCRs, incorporating an IF interface port, communication bus between decoder and television receiver and related tuner and performance improvements.

- Permit manufacturers to offer VCRs and additional TV models of this design at their option, as the marketplace.

- Cable operators make the appropriate interface decoders available to buyers of these "cable-ready" products, and offer those subscribers a reduction in their monthly rate.

"Cable-Ready"

Television receivers and VCRs labeled as cable-ready will have the following improvements as shown in Figure 2.

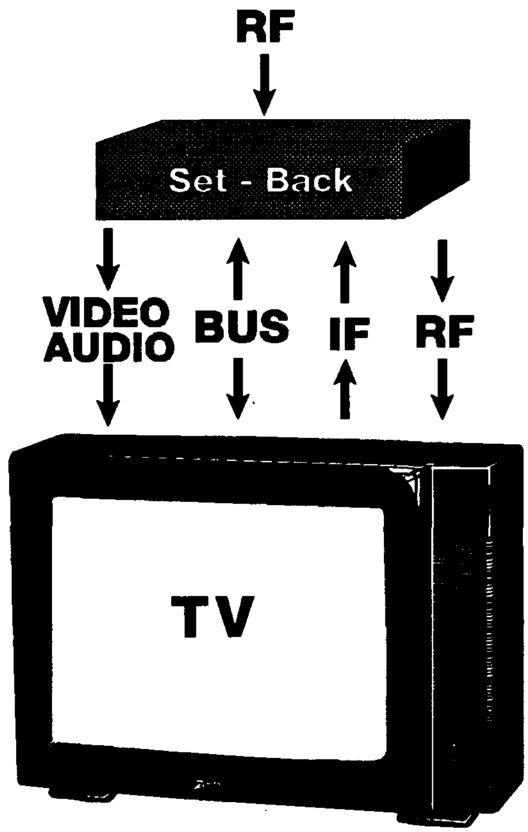


Figure 2.

- *Direct Pickup.* DPU are addressed by adding more shielding to the chassis and tuner and RF input(s).

- *Tuner.* To handle future advances in cable service such as digital compression, tuners with flat response and low phase noise oscillators will be needed.

- *Communication Bus.* The decoder Interface permits communication between the set back decoder and television receiver to provide tuning control by the decoder when needed. The television receiver retains full functionality of special features, while program access, descrambling and decompression are performed by the outboard interface box.

- *IF Interface.* The unfiltered IF signal is fed to the set-back decoder for processing. This signal passes present analog scrambling signals and will pass digital compressed NTSC signals to a digital set-back decoder in the future.

Interface Decoders

The set-back decoder unit, which would interface with the port on the "cable-ready" product, would not include a tuner, display, keyboard or remote control as a conventional set-top cable box does today. Such decoders could be priced to the cable operator as much as $40 less than the $100-$120 per unit invested today.

Cable-Ready Advantages

The following are positive factors which provide cable operators justification for reducing the monthly rate for subscribers of new "cable-ready" product:

- *Less Capital Invested.* Applying the "rule-of-thumb" that $1/month of cash flow amortizes $50 invested, the lower-cost box may allow the operator to pass along savings of nearly $1/month to the subscriber.

- *Less "Churn."* The improved compatibility will keep the current subscriber happier and hooked up, and attract new subs. Each sub is a $1500-$2900 equity consideration, as cable systems are valued these days.

- *Higher Revenue per Subscriber.* The greater ease of use may stimulate increased Premium and PPV revenues.

- *Perceived value.* Subscribers could envision a "payback" of the premium paid for the TV or VCR in 18 months or so...thus making the selling of the higher-priced cable ready product much easier.

Long Term Solution

Digital Migration

Analog scrambling is a fragmented technology with brand-specific scrambling methods. It is anticipated that analog technology - and scrambling - will be replaced by digital technology. Digital technology holds promise for a long term solution through the adoption of a national digital standard.

The ATV proponents, joined in the Grand Alliance, have adopted as the ATV standard a system consisting of MPEG2 for compression and VSB modulation for transmission. For ATV transmission via broadcast, 8VSB trellis coded, and for cable, 16VSB.

There is no standard for SDD (standard definition digital or compressed digital NTSC). MPEG2, in some variations, is assumed to be the NTSC compression standard. There are two different transmission technologies under consideration for SDD transmission on cable: 64QAM and 16VSB.

64QAM, as embodied in brand specific proprietary technology, offers only a migration path from set-top to set-back decoder. Even in a digital world with HDTV television receivers, with 64 QAM the major portion of signal processing would remain outside the television receiver. Dual mode reception of both 16 VSB and 64 QAM is necessary if the signal processing is to be within the receiver. That accommodation forces the consumer to bear additional cost because of the additional complexity and hardware to accommodate both technologies.

Adopting 16VSB as the transmission standard for both HDTV and SDD signals on cable, would allow a migration path for eventual incorporation of SDD signal processing within the receiver. Initially digital decoding, demultiplexing and decompression would be accomplished in the SDD set-back decoder. See Figure 3. Eventually, decoding, demultiplexing and decompression would migrate to the receiver.

These functions would be incorporated within receiver circuits common to both SDD and HDTV signals. All that would remain external to the receiver would be conditional access and security functions. This would significantly reduce complexity because receiver tuners would be optimized for only one transmission technology.

The advantages of this to consumers and operators are significant:

- For consumers, compatibility would be assured
- and for cable operators a major capital expense transferred to subscribers.

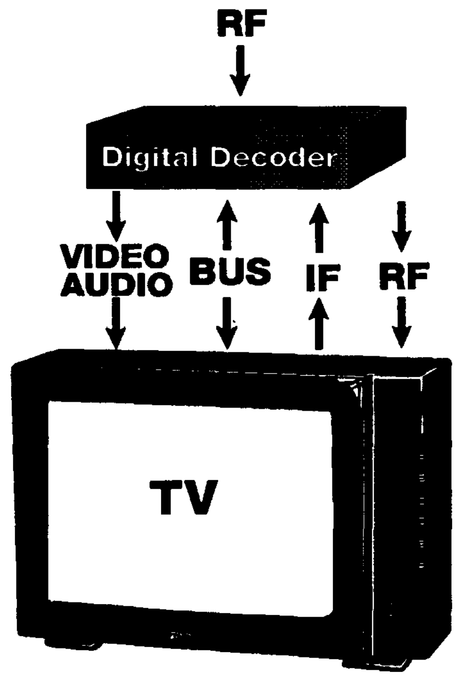


Figure 3.

Conditional access and security functionality would be available at much lower cost and replacement (if needed in a digital environment) less onerous. See Figure 4.

Another advantage for a common digital technology is the opportunity to provide access for other applications that are proposed for the "information highway". The decoder-television receiver command language is being written in a susbset of CEBus. This allows the interface to connect with other peripheral devices to provide new applications and services. A benefit for having these applications occur within the television receiver is that access to the display allows higher resolution graphics than would be available with additional processing in the set-back or set-top device. This is especially important with menus and other graphical interfaces required for making consumer choices easier

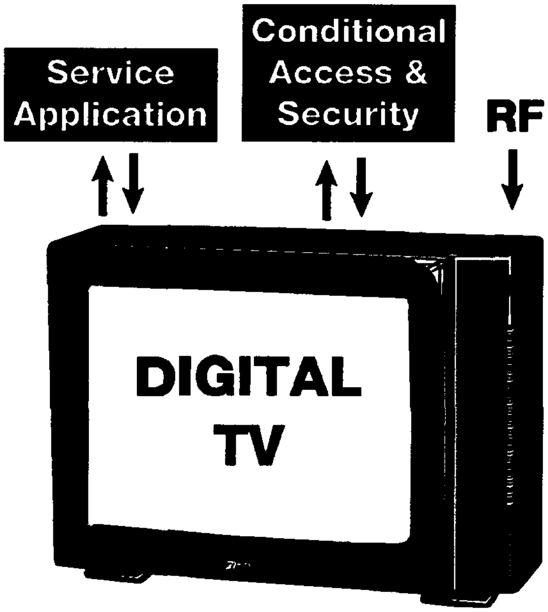


Figure 4.

Conclusion

There has been concern expressed that interface-port solutions fail to address the problems consumers are experiencing today. There are technically feasible enhancements to set-top cable converters which would improve, although not make perfect, the compatibility of cable with the "installed base" of TVs and VCRs.

For example, an active splitter and automatic A-B switch built into the set-top terminal can permit pass-through of all channels when the decoder is inactive, and thus enable the subscriber to use all the special features of his consumer electronics equipment except when watching a premium channel.

However, the IF interface port is a better long-term solution. It is capable of handling analog and digital signals and making the security and control functions of the cable operator transparent to the consumer. Thus permitting the consumer to enjoy full functionality of his TV and/or VCR. And it represents a system architecture capable of handling the compressed digital signals of both NTSC and HDTV.

A common transmission standard allows the decoder interface a ready migration path from analog to both SDD and HDTV signals. For the consumer it offers the promise of true compatibility

Measurement of Differential Gain, Differential Phase, and Chrominance to Luminance Delay in Cable TV

Rex Bullinger
Hewlett-Packard Company

Abstract

Beginning in mid-1995 cable TV operators must measure the quality of the color signals they are delivering to subscribers. The measurements chosen to judge color quality are differential gain, differential phase, and chrominance-to-luminance delay.

Local insertion of full-field signals allow the operator to test the performance of the system but requires program interruption. Local vertical interval test signal (VITS) insertion allows in-service testing without program interruption or picture impairment.

Another in-service approach utilizes programmer supplied VITS. In addition to lower cost, this is an end-to-end test which more closely links the test result to the actual subscriber picture quality. This approach, however, risks the signal not being present when needed or that the signal, if present, is already impaired enough to call its value into question.

INTRODUCTION

A great deal has been written on the 3 "color" tests selected by the FCC to be performed by cable TV operators[1,2,3,4,5,6]. Rather than repeat work already accomplished in describing how to do the measurements, this paper will look at some of the issues surrounding getting set up to make these measurements. Minimizing disruption of service will be a high priority.

The author is a relative neophyte in making these measurements so the approach of this paper will be to review what has been written and apply that knowledge to getting started doing it the first time. The intended audience are other people in the cable TV community who are also neophytes in these matters. Also, with the concept in mind that you can't know where you're going if you don't know where you've been, we'll take a look at some of the history involved.

Legal Requirements

As of June 30, 1995, the color rules will apply. Chroma delay shall be within 170 nanoseconds, differential gain shall not exceed +/-20%, and differential phase shall not exceed +/-10 degrees. Generally, the number of channels that must be tested are 4 plus 1 for every 100 MHz of spectrum use. (See the rules for the specific requirements[7].) However, since a consultant or FCC auditor can check any channel, and since all channels must comply, few operators will feel comfortable unless all channels are tested. Thereafter, these tests must be done once every 3 years.

The FCC requires these tests of the headend related signal processing equipment only. Since the trunk and distribution system components are all broadband, their contribution to color signal impairment is minimal. Not having to test outdoors in the system should be good news especially during bad weather.

A LOOK AT THE TEST SIGNALS

The differential gain and differential phase measurements are done using a modulated stairstep signal (Figure 1). The modulation on each step is the color frequency at an amplitude of

+/-20 IRE on each step. What is being measured are changes in amplitude and phase of that color signal as the luminance signal is stepped through its range of operation. Ideally there would be no changes at all in the chroma amplitude and phase as the luminance staircase goes from black to white.

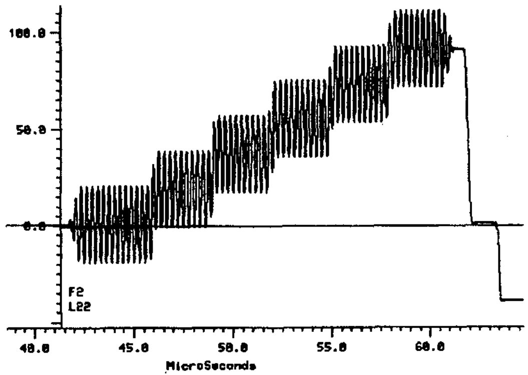


Figure 1. Modulated Stairstep Signal (NTC)

The third test, chrominance-luminance delay inequality (CLDI, also called relative chroma time or RCT) uses a different test signal called variously the modulated 12.5T $\sin^2$ pulse, or chrominance pulse. This cleverly designed signal consists of both luminance (Fig. 2) and chrominance (Fig. 3) components which occur simultaneously. Figure 4 illustrates the 12.5T pulse as it appears after the low and high frequency components are added together. As the two components which are at different frequencies travel with the TV signal, any speed differences encountered because of being at different frequencies become readily visible through predictable distortions of the 12.5T $\sin^2$ pulse. Generally, the higher chroma frequency will encounter more delay than the luminance signal so the test is probably named for this. However, negative delay, where the chroma component arrives first is not uncommon.

These test waveforms, the modulated stairstep, and the 12.5T $\sin^2$ pulse, have both been incorporated into 2 common test signals called the FCC Composite (Fig. 5) and the NTC-7 (Fig. 6, hereafter just NTC) Composite signals*. The 12.5T $\sin^2$ pulse is identical in both signals, but the modulated stairstep is slightly different. In the FCC Composite signal the 5th stairstep has a DC level such that the color signal at peak excursion is exactly 100 IRE. However, in the NTC version the DC level of the 5th step is already at 90 IRE so the color signal's peaks extend to 110 IRE. This has implications for the cable TV operator that will be explored later. Compare Figures 5 and 6. Figure 1 is the NTC modulated stairstep in more detail.

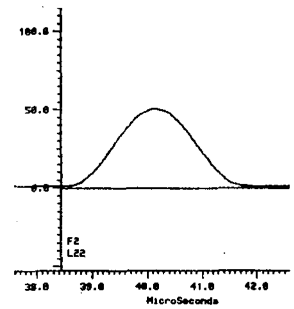


Fig 2. Low Freq Portion

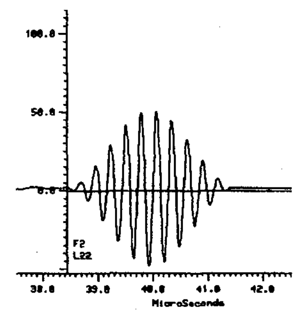


Fig 3. Hi Freq Portion

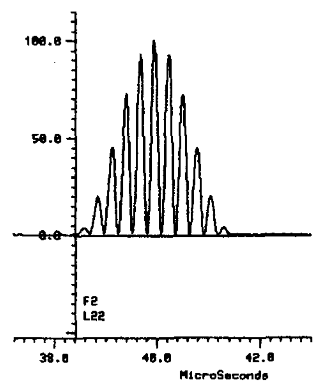


Figure 4. 12.5T Chrominance Pulse

The reason there are two such similar test signals seems related to differences in their application. The NTC is intended for network video transmission to affiliate broadcasters and the FCC is for the terrestrial broadcast environment. The NTC Composite signal is explicitly intended for

* Note that it is easy to confuse the name "Composite" signal with that of the NTC-7 "Combination" signal which is quite different. Fortunately, there is no "FCC Combination" signal.

testing "video facilities leased by the major television networks from the Bell System[8]." For this purpose, Johnston[9] writes that the NTC members preferred having the line bar occur first because it gave a more valid reading on line-time distortion. In addition, the author suspects the excursion of the 5th stairstep's modulation in the NTC Composite signal to 110 IRE is good for testing marginal network capability but would put broadcast transmitters under unduly harsh treatment.

The FCC Composite signal, on the other hand, seems intended for terrestrial broadcast. Indeed, it's called the "Composite Radiated Signal" in the NAB Engineering Handbook[10].

Other differences between the FCC and NTC Composite signals are the order of occurrence of the test signal elements. In the FCC Composite signal the modulated stairstep comes first, while in the NTC Composite signal it comes last. The 12.5T $\sin^2$ pulse occurs in the middle of the signal for both, but the timing of the center of the pulse after the leading edge of the sync pulse is 39.2 uS for the FCC Composite and 37 uS for NTC Composite signal. In general, these timing differences should not matter to the cable TV operator. However, the amplitude differences of the 5th modulated stairstep does matter. This issue will be covered in a later section.

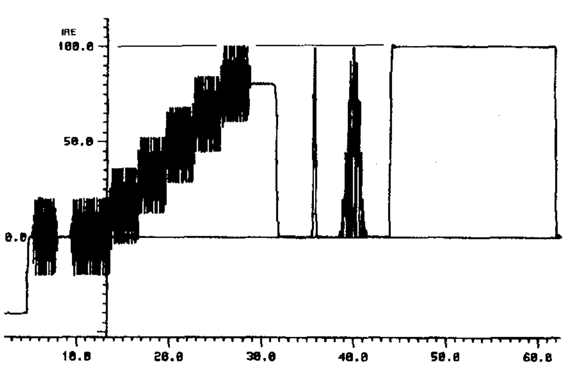


Figure 5. FCC Composite Signal

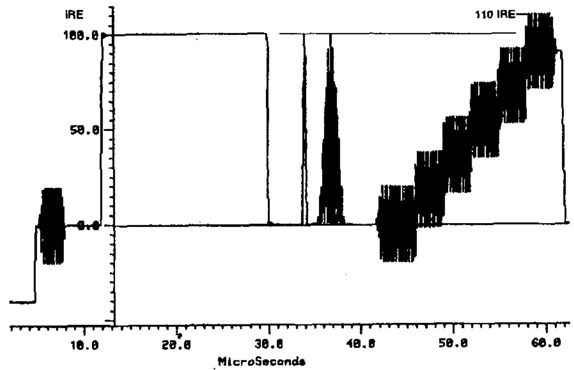


Figure 6. NTC-7 Composite Signal

Both Composite test signals also include two other features called the bar and the 2T pulse. The 2T pulse is the short pulse occurring just before the 12.5T $\sin^2$ pulse. It is used for measuring short-time waveform distortion. The bar is a relatively long period of time with the signal at the white level. Both of these features are used in other video tests not specifically called for by the FCC, however the bar is useful in cable TV for measuring depth of modulation.

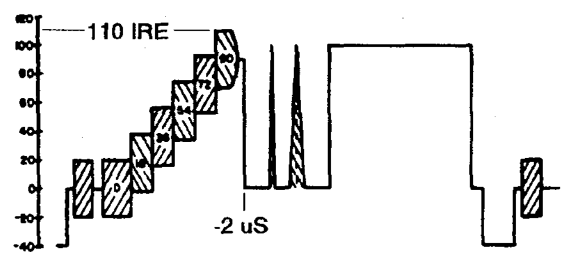


Figure 7. STOC Composite Signal

Figure 7 illustrates a now rare but not completely absent test signal developed by the Satellite Technical Operational Committee and reported on in the early 1970's[11]. This reference says its purpose was to "test microwave radio relay systems". Another reference[12] says it was for "testing satellite NTSC transmission". It is strikingly similar to the FCC Composite signal and can be easily mistaken for it. The differences, however, between the STOC and FCC Composite signals are the chroma peak amplitude in the 5th stairstep, and a 2 uS shift forward in time of the rest of the signal components after the 5th stairstep.

The main reasons to bring this up here are that the chroma level on the 5th stairstep, like the NTC Composite, reaches 110 IRE. The other reason is that the author recently saw this signal, or something very similar to it, on an off-air signal in Arkansas. As will be discussed below, chroma at 110 IRE can lead to overmodulation and clipping, thus compromising the differential gain and phase tests. And the fact it has been seen means it's lurking out there threatening to confuse the situation.

A Bit of Test Signal History

Taylor[13] writes that in the 1950's and 1960's TV network video feeds carried by the Bell System to TV broadcasters sometimes became impaired. Attempts to locate the source of the impairment could lead to finger-pointing between Bell System and TV broadcaster personnel. To help resolve this problem the National Transmission Committee (NTC) was formed. This committee consisted of engineers from ABC, CBS, NBC, the Public Broadcasting Service, and AT&T and produced the well known NTC-7 Engineering Report[14].

Jenkins[15] wrote that in 1966 the NTC was considering using both fields of lines 18 and 19 for Vertical-Interval Test Signals (VITS). Field 1 of line 19 was to carry a "sine-squared signal" defined as a 2T pulse and bar (no 12.5T pulse yet). Field 2 was to carry a modulated stairstep. These two signals were later combined with the 12.5T pulse onto a single line.

Rhodes[16] wrote that in 1965 P. Wolf introduced the 20T modulated $\sin^2$ pulse in Munich, Germany. Rhodes then proposed that a 12.5T pulse was more appropriate for NTSC. He also developed a number of nomographs for determining CLDI from the amplitudes of the distortion of the pulse baseline.

USING THE COMPOSITE TEST SIGNAL

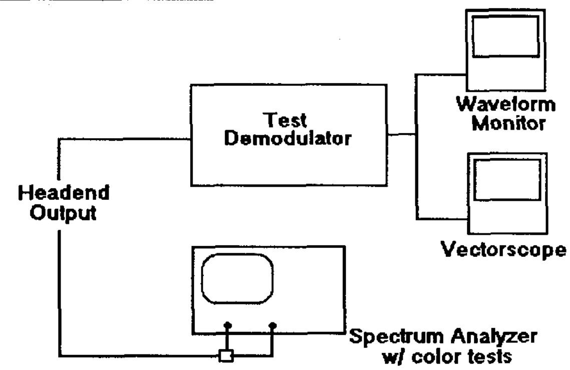


Fig. 8 Measurement Setups

The 3 color tests can be measured at the output of the final headend combiner using a directional coupler. They can also be obtained from the input test point of the first trunk amp.

The traditional equipment needed is a precision demodulator, waveform monitor or TV triggered oscilloscope for the differential tests, and a vector scope for the CLDI test. The above pieces of test equipment are available in a wide range of performance and costs. Alternatively, a new spectrum analyzer has recently come to market which combines these 3 color tests with the traditional cable TV test capabilities in one portable unit.

VITS Testing

VITS, as opposed to full field, testing allows measurements to be made on the equipment in service. This minimizes service disruption and when programmer supplied VITS are used, the result is an indication of the actual signal quality delivered to the subscriber.

Substitution Full Field Testing

Another option for doing the color tests is to substitute temporary headend signal processing equipment into the signal path of the channel to be tested. The equipment substituted for would include channel filters, processors, modulators, and any other components in the signal path that would have narrowband enough amplitude and/or phase characteristics to affect the color measurements. As with trunk and distribution equipment, headend switching and combining equipment should not affect the color tests.

This method would limit service interruption to the time to change cable connections. This would allow bench testing using a full-field test signal. Other advantages of this method are not needing a VITS inserter in order to maintain service, and because of full-field testing, having the ability to test over the full range of average picture level (APL) as recommended in the NAB Engineering Handbook.

The actual processes for making these measurements has been well explained by other authors. The purpose of this paper is to explore the in-service options available to the cable operator for obtaining the Composite test signal and analyze the merits of each option.

Satellite & Local Origination Channels

The first point at which the cable TV operator has any control over a satellite or local origination signal is as it is delivered as baseband either from a satellite receiver or from another source such as a VTR or text generator. If a Composite signal is present in the signal it may or may not be usable. If not, a VITS inserter can be placed in the baseband video path at this point.

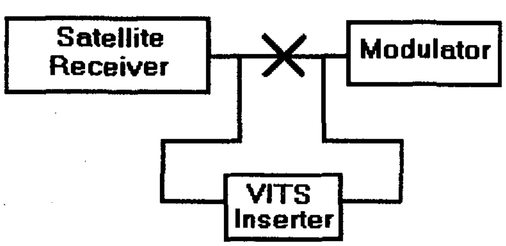


Fig. 9 Satellite Channel VITS Insertion

Recently a VITS survey was done for the NCTA of 118 satellite services by Reed-Nickerson[17]. He found that 33% had the FCC Composite signal, 33% carried the NTC-7 Composite signal, and 40% had no Composite signal at all. (There were some which carried both Composite signals.)

It is attractive to use the satellite programmer supplied test signal because it is convenient and does a complete end-to-end test. If the result is within test limits, not only are the FCC requirements met, but the cable operator can have greater confidence in the final quality of the product at the subscriber terminal (ignoring the effects of a converter). A potential inconvenience, though, is that the test signal can be interrupted or changed because of programmer or local signal switching. Local ad insertion is a prime example.

Since initial indications are that satellite programmer provided Composite VITS can be a usable, if not always continuous, signal source for the cable color tests, one hopes the 40% not carrying the Composite VITS will diminish.

Off-Air Channels

Channels using processors present a bit more of a challenge. While broadcasters are more likely than other sources to have a Composite signal, VITS derived from off-air channels tend to be more impaired than when delivered by satellite. The author speculates this mostly results from impairments in broadcast transmitters, multiplexers, and/or effects of multi-path. This increases the likelihood that VITS will have to be inserted. But the Composite VITS will have to be inserted at RF at the input to the processor in order to do the test properly. This requires the use of a test modulator in addition to the VITS signal generator. See Figure 10.

To insert VITS into an RF signal it must first be demodulated, VITS inserted, then the signal is re-modulated to RF and connected to the processor input. For this case the demodulator does not need to be precision. It merely needs to supply a good enough signal to carry the inserted VITS and keep the channel in service during the period of the test. An agile demodulator is the best choice because the audio path is most easily managed. In a pinch, a VCR or TV with a baseband video output could be used. However, the audio would most likely be available as left and right and not a composite audio signal at 600 ohms.

Modulators

To the author's knowledge, one cannot buy a "precision" video modulator specifically designed for test. Precision demodulators for test are available, but not their reverse function. Therefore the operator must make do with buying the one with the best specifications then verifying its performance with a precision demodulator.

However, at least a couple of cautions must be observed. Bowick[18] reports that many modulators in cable TV use may not have CLDI precorrection circuitry. If the measurement assumes precorrection is present, the results can be off by 170 nS. Since the FCC requires this precorrection of broadcasters, it almost certainly is assuming that precorrection will be present in the signal delivered to cable subscribers. It may be hard to justify use of modulators without it.

What we are considering here is defining a modulator to be used for test. But what about headends now populated with modulators without precorrection? This issue and how it relates to the concept that the cable TV operator is not to change CLDI by more than 170 nS, nor is the operator charged with fixing a bad signal, is beyond the scope of this paper.

Another modulator issue is peak clipping. Some modulators used in cable TV limit modulation to 90 - 95% to avoid overmodulation. Recall that the NTC and STOC 5th stairsteps have chroma amplitudes at 110 IRE. Even when depth of modulation is correctly set to 87.5%, this 5th stairstep could cause the modulator with modulation limiting to reduce the chroma amplitude. This will significantly affect the differential gain measurement. The Recommended Practices document of the NCTA[19] addresses this by stating "if the modulator under test contains a clipping circuit (90% modulation) ignore the fifth step in a 5 step signal". Using the FCC Composite test signal and making sure depth of modulation is properly set minimizes the effect of this problem.

A last caution is that if using a precision test demodulator, the test modulator must have a stable enough carrier for the precision demodulator to be able to phase lock to it. It has been seen that at least one state-of-the-art agile modulator is not stable enough for a top-of-the-line precision demodulator to phase lock to it. As noted below, phase lock is needed for synchronous detection which is required for some tests. At this time, this author has no data on the potential extent or seriousness of this issue.

In summary, when choosing a modulator to be used as a reference test modulator, chroma-luminance precorrection must be present, white clipping would ideally not be functioning, and it would have a stable carrier output.

Demodulators

Synchronous versus envelope detection is an issue that deserves attention. Rhodes[20] writes that synchronous detection can be used for all video measurements as long as ICPM and static phase error are known to be negligible on the signal to be tested. Excessive incidental carrier phase modulation (ICPM) can make differential phase measurements read artificially high with true synchronous detection. In this case envelope detection may help. Static phase error can affect

CLDI, however, it still must be measured using synchronous detection.

Rhodes goes on to say that depth of modulation when measured at baseband with synchronous detection can be strongly affected by ICPM. He suggests envelope detection here but cautions that envelope detectors are usually quite nonlinear at low (white) carrier levels. This author suggests that since depth of modulation is such an important setup parameter for the color tests, that it be measured at RF using a spectrum analyzer. This relatively narrow band technique uses envelope detection and a calibrated linear scale to avoid problems with ICPM and linearity.

In summary, when making color measurements with a test demodulator, it must be capable of synchronous detection which must be used for all tests except for differential phase and depth of modulation which may use envelope detection if the signal under test is impaired by ICPM. Additionally, the test demodulator must be capable of phase locking to the test visual carrier signal.

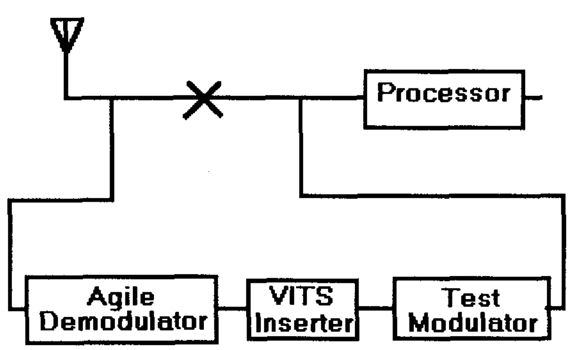


Fig. 10 Off-Air VITS Insertion

CONCLUSIONS

The intent of this paper was to identify the assumed few subtle issues that might arise when doing the color tests in the cable TV environment. Unfortunately, a search of the available literature uncovered more subtleties than anticipated. It is hoped that most if not all of the issues raised here will prove to be minor in practice. Until experience proves otherwise, though, here is a summary of some of the things to watch out for in no particular order.

1. Depth of modulation has to be set to 87.5%. Broadcasters shoot for 85-90%. The NCTA Recommended Practices suggests 80-84% to ensure against overmodulation. This lower modulation will also guard against white clipping due to high chroma levels in the 5th stairstep of some Composite test signals.

2. Differences in average picture level can affect test results. The NAB Engineering Handbook recommends measuring at 10, 50, & 90% APL. The effects of different APL's on test results, however, may be more of an issue for broadcast transmitters than for cable TV modulators.

3. Some cable TV modulators may not have chroma-luma delay precorrection.

4. Johnston[21] warns that "many signal generators actually produce triangular or ramp shapes for signals which should have sine squared shapes". This he says could compromise measurement accuracy. This author has never seen a 12.5T chroma pulse that has straight sides, but it's worth being on the lookout for.

5. The chroma level must be sufficient for a reliable differential gain and phase result. Chroma levels in off-air programmer VITS appears particularly problematic.

6. The test demodulator must use synchronous detection and be capable of phase locking to the test signal.

7. If using programmer supplied Composite VITS, programmer and/or local signal switching, such as ad inserts, can change or eliminate the test signal at inconvenient times.

[1] James O.Farmer and Blair Schodowski. "Measuring and evaluating video signals in the headend--Part 1 & 2." Communications Technology, August 1992, p. 20; September 1992, p. 28.

[2] John Vartanian. "Tech rereg: Differential gain, differential phase, chrominance-to-luminance delay inequality." ibid.

[3] Mark J. Everett. "Some FCC compliance testing suggestions." Communications Technology. October 1993, p. 28.

[4] Jeff Noah. "Color signals: FCC-required measurements." Communications Technology. November 1992, p. 34.

[5] "Color Measurements." Recommended Practices (Washington, DC.: National Cable Television Assn. 2nd ed. October 1993), sec. 1.H.

[6] Margaret Craig. "Television Measurements--NTSC Systems." Tektronix. 25W-7049-1, July 1990. pp. 25-29 & 47-55.

[7] Code of Federal Regulations, Title 47, Chapter I, Part 76, "Cable Television Service".

[8] NTC Report No. 7--Video Facility Testing Technical Performance Objectives. (Network Transmission Committee of the Video Transmission Engineering Advisory Committee, June 1975, rev. January 1976)

[9] Warner Johnston. "Examining Composite Test Signals." TV Technology. September 1991, p. 30.

[10] E. B. Crutchfield et al., eds., Engineering Handbook (Washington, D.C.: National Association of Broadcasters, 7th ed., 1985), p. 3.5-140.

[11] Charles W. Rhodes, "An Automated Vertical-Interval Test-Signal Monitoring System for NTSC", IEEE Transactions on Broadcasting, Vol. BC-19, No. 2, June 1973.

[12] Charles W. Rhodes, "Measuring Distortions in the Television Signal", Tektronix, 1972.

[13] Archer S. Taylor. "NTC-7." CED. July 1992, p. 70.

[14] NTC Report No. 7, p. 7.

[15] S. C. Jenkins (of AT&T). "The 'VITS' Program for Intercity Television Network Testing." HP JOURNAL, Feb. 1966, 17, No. 6, p. 12.

[16] Charles W. Rhodes, "The 12.5T Modulated Sine-Squared Pulse for NTSC", IEEE TRANSACTIONS ON BROADCASTING, BC-18, No. 1, March 1972, pp. 8-17.

[17] Linc Reed-Nickerson, "VITS Survey", NCTA Engineering Committee, January 1994.

[18] Chris Bowick, "Delay predistortion", CED, January 1990, p. 16.

[19] "Color Measurements." NCTA.

[20] Charlie Rhodes, "Testing and Using Synchronous Demodulators", Tektronix, Application Note No. 28, 20W-4177-2, Feb. 1987

[21] Warner Johnston, ibid.

MODULATION CONSIDERATIONS FOR DIGITAL TELEVISION

Joseph B. Glaab
General Instrument
Hatboro, PA 19040

Abstract

The modulation technique selected for the Advanced Television standard may not necessarily be optimum for the Cable environment. In fact there may be technical and financial reasons for having alternate modulation techniques for each application.

COFDM Modulation

Different modulation methods are being proposed and tested in various transmission environments. The developers of the European Advanced Television System for High Definition Television (HDTV) are investigating a method which sends hundreds of low bit rate carriers in the normal 8 MHz wide television channel. Each of these carriers is modulated by data using a higher complexity modulation technique such as 16 Quadrature Amplitude Modulation (16 QAM). In experimental systems built so far, the carriers are generally spaced at a frequency close to the 15 KHz Horizontal Line rate of the systems. This is a Frequency Division Multiplex (FDM) system. Since the carriers are very close together, the sideband energy overlaps the adjacent carriers to some extent. In order to minimize the inter-symbol interference (ISI), the adjacent carriers are in quadrature with each other. Quadrature in geometric terms is Orthogonal: Thus the name OFDM. Unfortunately, the phase modulation part of QAM information sometimes looks like the orthogonal relationship of the adjacent carriers and data is subject to corruption. In order to minimize this problem, the data bits on alternate carriers are delayed in transmission and the system uses heavy Forward Error Correction Coding (FEC). The result is COFDM; Coded Orthogonal Frequency Division Multiplex. In an 8 MHz RF Channel, about 512 carriers spaced at 15.625 KHz would fit exactly. If each of those carriers were modulated at 50 KHz using a 16 QAM the total data rate would be 26 MBits/s. If roughly ¼ of those bits were for FEC, the true data rate would be 19.5 MBits/s. The United States TV channels are only 6 MHz wide and the TV Horizontal line rate is about 15.750 KHz. This would permit about 380 channels, and following the same logic, a true data rate of about 14.5 MBits/s. This does not compare favorably with the present 19 MBits/s supplied by both 8 VSB and 32 QAM. There are proposed implementations which use up to 1856 carriers spaced every 2.93 KHz and 64 QAM modulation.[1] A system of this complexity would really stress the phase noise performance of the recovery circuitry. Rather than build hundreds, or thousands, of transmitters and receivers, the signals are generated by supplying the instantaneous value of the digital symbol for all the 500 or more sub-channels to a Digital Signal Processor (DSP) chip. This DSP does an Inverse Fast Fourier Transform (IFFT) and generates the frequency spectrum associated with the data. When this is transmitted, any delay distortions (ghosts) theoretically are minimized in effect because the ghosts are predicted to be shorter that the length of the symbols being transmitted. At the receiver site, a Fast Fourier Transform (FFT), is used to recover the individual data bits prior to being reconstructed for use. This technique has tested well, in spontaneous tests, in the European environment where one channel is reused by low power transmitters in adjacent cities. The European plan is to use one channel countrywide for the same program. Interference from adjacent cities appears as a ghost and is therefore eliminated. Another advantage of this system is that the IFFT can be generated with the frequencies at and around the normal picture and sound carriers of a normal Amplitude Modulated Vestigial Sideband television eliminated. This removes one interference method in which the Digital signal gets into the normal television picture creating a low frequency beat. The reason that the beat could be seen is that the transmitted bits are multi-microseconds long, and therefore look like a relatively low frequency if considered as a noise source. As a side benefit, this method of transmission eliminates the need for an adaptive equalizer. The trade-off does not seem to present any compelling reason to select this modulation and transmission method for the United States where the transmission constraints are different. The U.S. uses transmission frequencies which are not available in adjacent cities because of other reasons. One of the reasons is that the television sets already in homes have selectivity and radiation limitations which restrict using those frequencies in a normal way.

While COFDM seems to work well in the particular broadcast environment where it has been tested, it is not expected to be ready for use until about the year 2000. Another limitation is potentially attempting to do some sort of a double rate transmission for the more benign Cable TV environment. The European market is still planning to use 64 QAM for Cable TV distribution.

QAM Modulation

Two generic modulation techniques were being considered for the U.S. HDTV standard. Both have their advantages and disadvantages for the environments in which they must operate.

While The Grand Alliance, a consortium of HDTV proponents, has decided to recommend 8 VSB over 32 QAM for the terrestrial broadcast and 16 VSB over 256 QAM for Double Rate (carriage of two HDTV broadcast channels on one Cable TV channel)[2], we will discuss the merits of both techniques.

Quadrature Modulation can be considered to be the individual modulation of the in-phase and quadrature (90 degree phase shift) carrier phases. In the simplest level of modulation, the in phase carrier can signify either a zero or a one. Likewise the quadrature carrier can also signify a zero or a one. These values are established by either turning the carriers on or off at the sample time. The resultant carrier is the sum of the two amplitude and phases and has four possible values. These are defined as zero, if both bits being sent are zero, to any combination of the two bits. The result of this special case modulation is that there are four possible states of carrier phase and amplitude and the modulation method is called Quadrature Phase Shift Keying (QPSK). In a similar way, each carrier phase could transmit two bits of information by sending an instantaneous value as generated by a Digital to Analog Converter. This would give amplitudes of 0, 1, 2, and 3 to each phase of the carrier. The combination of two phases creates 16 instantaneous phase and amplitude possibilities, giving what we call 16

Quadrature Amplitude Modulation (QAM). Higher orders of linear modulation are easily derived. Three bits per phase yields 8 possible levels per phase with 64 possible amplitude-phase points. Four bits yield 16 values and 256 QAM locations. It can be seen that going from QPSK to 16 QAM doubles the data rate and going to 256 QAM doubles it again. The complexity of going to 256 QAM from 16 QAM is significant.

QAM signals occupy frequency spectrum based on the number of symbols transmitted every second. A 6 MHz channel can support about 5 Mega-Symbols per second since the signal looks similar to a double sideband, suppressed carrier modulated signal. This means that the total data bit rate for a 16 QAM signal would be 4 bits per symbol times 5 MS/sec or 20 MBits/sec. 64 QAM yields about 30 MBits/sec and 256 QAM, 40 MBits/sec.

In QAM modulated signals, echoes and multipath (ghosts) in the transmission path can modify the received signal in both the net resultant amplitude and phase. Because of the short symbol time, 200 ns in the case of a 5 MS/sec system, broadcast length echoes of up to 24 or so microseconds requires an adaptive equalizer of 256 taps. Two taps are commonly used per symbol time. This gives 100 ns per tap correction resolution. For more information on adaptive equalization, see ref 3. The important issue here is that cable systems usually don't have strong and long echoes at the same time. This means that cable system transmission is relatively benign in this respect.

VSB Modulation

The Modulation method recommended by the Grand Alliance for broadcast HDTV is 8 Vestigial Sideband (8 VSB). This technique could be considered to be a single phase carrier, amplitude modulated by the value of the digital data. The actual signal is implemented by using signal shaping and a quadrature modulator to cancel the second sideband. Ideally only one sideband would be sent, but this is not practical in real systems, and a small part of the second sideband is sent along with the desired sideband. A small amount of carrier is added back to the signal to aid in locking the receiving circuitry. Because of the vestigial sideband, the carrier must be displaced somewhat into the channel. Even with this displacement, the symbol rate can be double that of a system with the carrier centered in the channel. With the depressed carrier and 6 MHz channel bandwidth limitation, about 10 MegaSymbols per second may be sent. This means that if a two level bit was to be sent, 10 MBits/sec could be transmitted. Divide the amplitude into 4 levels (4 VSB) and you get 20 MBits/sec. Doubling that again to 16 VSB gets to 40 MBits/sec. The problem is that for 16 VSB, the resolution of the sampling must be as fine as for 256 QAM. That makes is highly unlikely to be acceptable for broadcast HDTV. An advantage of this system is that since all the information is in phase with the carrier, it is not necessary to have as complex an adaptive equalizer.

The modulation methods being considered for broadcast HDTV are more complex than straight VSB or QAM. In the VSB system, a number of unique operations are included which improve the performance of the system while making some other operational aspects more complex. The VSB signal implementation has a repetitive sync period to improve synchronization of the electronic circuits. With the incorporation of a sync signal, the adaptive equalizer can adapt to optimize the sync waveshape, and by implication, the data symbols. The down side to this is that the sync only appears every 65 micro-seconds, compared to that of the blind adaptive equalization of QAM which can fine tune values at the symbol rate. The Grand Alliance HDTV will probably use Blind Equalization with the 8 VSB.

With the strict timing involved with the relationship of sync to data, the makers of Studio Transmitter Links (STL) used to carry data from Production Studios to the transmitter are having problems. This is because the addition of HDTV to the present STL links is difficult. The FCC has not given any additional STL bandwidth when they assigned HDTV broadcast channels. The plan now is to multiplex the digital HDTV signal with a digitized NTSC signal on the STL, and the precise timing requirements are causing concern.

The good overall performance of VSB requires close matching of the transmit and receive filtering. In order to accommodate the transmitter non-flatness, a receiver was incorporated into the tested transmitter so any error created by the transmitter was sent back to the transmitter equalizer filter to pre-distort the transmitted signal. There is also come concern that filtering associated with frequency conversion may present problems.

8 VSB is what was 4 VSB with one bit per symbol added in a scheme called trellis coding. 16 VSB does not use trellis coding. Trellis coding is used in both VSB and QAM systems and works by making it impossible to get to new signal locations from the closest Euclidean location. This means that the likelihood of some symbols is not possible from some previous symbols. This technique uses a more complex constellation and transmits more bits overall while improving the system equivalent carrier to noise by 1 to 3 db. Three db is the equivalent of doubling the transmitted power. The 32 QAM system tested uses a trellis code which gives a constellation which is not rectangular, but rather looks like a 6 by 6 constellation with the corners removed. For more information on coding for transmission, see reference 4.

Discussion of Issues

For reference, some of the data which was used to decide the modulation method of choice are listed below[2].

Modulation Format	32 QAM	8 VSB	256 QAM	16 VSB
C/N (dB)	14.8	14.8	29.3	27.6
dB Min Tap Isolation	14.0	20.7	24.0	30.0
dB Peak-Avg Carrier Pwr	5.8	6.4	6.4	6.5
Fiber Optic Clipping				
Depth of Mod	>6.3%	6.2%	3.8%	4.7%
C/total Dist	<26.9 dB	26.4 dB	48.6 dB	38.9 dB
CSO dB	<42.9	42.2	79.5	57.2
CTB dB	<36.6	35.3	60.2	53.8

The results shown above were made using equipment with performance about as close to theoretical as can be achieved with present technology. Using technology and techniques common to Consumer Product design, some relaxing of performance may be expected. Tap isolation is a particular case in point. While static changes are compensated by the adaptive equalizer, it can be reasonably assumed that customers on adjacent taps could cause each others' Digital receiver to stumble when they "surfed" channels. From the chart above, up to 30 dB minimum isolation was required for glitch free reception. Considering that millions of taps are installed which have only 20 or 25 dB isolation; even with drop line loss, some problems may develop especially on lower frequency channels. Some protection can be achieved by using isolation amplifiers, which would need low band bypass for return path applications, thus adding complexity. Fiber optic clipping is another area of concern. The higher order digital modulation formats have a low depth of modulation tolerance by modern standards.

Both systems described above would give a high enough net bit rate to conveniently allow simple transcoding of two broadcast HDTV programs into one CATV channel. The Forward Error Correction (FEC) ideally would be different for cable than for over the air broadcast. Broadcast FEC coding is a combination of Reed Solomon and the aforementioned Trellis Coding. The Cable environment might manage with only a strong Reed Solomon, however a more economical combination may be to use a simpler Reed Solomon in conjunction with a simple Trellis code.

As can be seen from the preceding, it may not be possible to have a simple scheme which would double the number of Broadcast HDTV channels in a CATV channel.

Conclusions

COFDM seems to best fit the channel reuse concept being proposed in Europe. It appears that either 64 QAM or 8 VSB will work on cable with few problems. Both of the higher order, 256 QAM and 16 VSB, modulation formats present risk in their ability to work well in all cable systems.

For non-technical reasons, cable may find that is beneficial to select one modulation scheme over the others.

It is also reasonable to believe that an inexpensive digital demodulator could be developed that would automatically reconfigure itself so as to receive and process both QAM and VSB signals.

References:

1 - COFDM Report, Summary of European Trip report. General Technology Assessment & Recommendations COFDM Report, Feb 24, 1994 ACATS - Transmission Expert Group

2 - ACATS Technical Subcommittee Meeting of Feb 24, 1994 Handout; Henderson/James

3 - An overview of Adaptive Equalization Communications Engineering & Design, Oct 1992 J Waltrich

4 - Error-Control Techniques for Digital Communication, John Wiley & Sons, New York, 1985 A Michelson & A Levesque

- Theory & Practice of Error Control Codes, Addison-Wesley Pub Co Reading, Mass, 1983 R Blahut

- Analysis of Recurrent Codes, IEEE Trans. Information Theory, IT-9 (1963) A Wyner & R Ash

- Application of Error Control Techniques to Digital Transmission via CATV Networks 1992 Technical Papers J Griffin

Additional Readings of Value:

From Northern Telecom Meeting, IBC92 Amsterdam.

- Digital Terrestrial Television Development in the Spectre Project A G Mason & N K Lodge

- VADIS: Digital Television at under 10 MBits/s K McCann

- Digital Television Broadcasting with High Spectral Efficiency R Monnier, J B Rault & T de Couasnon

Mugged on the Information Superhighway: Security Problems and Solutions

by
Robert Adams
of
Intel Corporation

Abstract

With the information superhighway speeding toward us, how will we protect our secrets? Suppose you use your credit card to order a pizza. How will we keep your neighbor from seeing the number? How will we keep the pizza house from using your credit card for a month's worth of pepperoni? Security, privacy, and encryption are key elements of the emerging cable-based broadband networks.

This paper describes the infrastructure necessary to provide security as part of the underlying infrastructure for broadband networking. It will provide a brief tutorial on the key elements of security, with a focus on the implications in a broadband, cable-based network and participation in a global information network.

Data Networking Comes to Cable

or It's 5pm and Do You Know Where Your Data Is?

Connecting a computer to the cable system is different then connecting a television. And the computers, whether embedded in settop devices or general purpose personal computers, will bring with them a multitude of new applications and new ways to steal.

Computers in the home will be talking at high speeds over the cable plant to services. The services will provide information, entertainment, and commerce. But it will be valuable information, paid for entertainment, and monetary commerce. There must be some way to protect the data.

When connecting computers to servers over the cable system, the first impulse is to use existing computer networking systems. This solves some problems (routing, interconnection, and cost) but the design of computer networks does not include private and personal data.

Existing Computer Networks to the Home

or Send Lawyers, Guns, and Money

Existing computer networks rely on physical security, trust, and passwords for its data security. These methods have their strengths and weaknesses.

Physical security is using walls and locks works for many businesses using computers. If the computer network is all in one building and the server is in a locked room, data security is easy. Outside connections (phones with modems) can be the weak link and we've all heard stories of computer breakins.

Agreements with all who have access to the data (see "physical security" above) tie legal and moral bands on a very weak link in any security system -- the people. These agerrments name all users *trusted users* of the computers. The trust creates needed accountability because only very specialized computer systems have audit trails to remember who did what, when.

The lowly *password* is the first line of defense for nearly all computer systems (and thus networks). Passwords fail to protect all too often because of the people who use them. When picking something easy to remember, most people choose something that's easy for others to guess -- first name, spouse's name, last name backwards, etc.

These have been sufficient because computer networks were developed and used in small or single business environments (where physical security was sufficient) or by the academic, research community (where trust and passwords were enough and, besides, who wants the overhead of audit trials and accountings?).

Today, the "Information Superhighway" does not exchange money, protect private conversations, or tell you if you are communicating with who you think you are. If you tap into the Internet today, you will find plain text mail messages and adhoc administration.

Existing computer networking does not solve the problems of security over the cable plant and we are challenged to address an area of legal entanglements, unknown assailants, and complex problems -- send lawyers, guns and money.

Security Problems

Let's take the data network outside the building. Assume you have a server that people need to communicate with and anyone else can snoop on the bits that are sent between the server and a client. This is how networking to your server over the cable plant would be -- the RF goes into everyone's house and business.

Sending data between the client and the server as "plain text" is certainly out of the question. Anyone listening to the coax cable could receive and look at what I sent -- whether it be my VISA card number or a letter to Aunt Minny.

I could modify the data I'm sending to the server by some algorithm. The server would have to know to use the inverse of the algorithm to create the plain message. Anyone listening to the conversation could not know what was being sent. Well, not until they found out or figured out the algorithm.

A better way would be to use a data modifying (or data "encryption") algorithm that has a little part that is easy to change and is not generally known. This is called a "keyed encryption algorithm" -- the encryption algorithm includes a small part (called a *key*) that I can choose. So, even if the listener knows the encryption algorithm, unless he has the key, the plain message can not be extracted.

I can now send data to the server by encrypting it by the algorithm that uses my key. The server has the decrypting algorithm and the proper decrypting key so it can understand what I sent. The client and the server have a *shared secret* that they use to send data between them. Anyone listening could not receive the data even if they knew the encryption algorithm because they wouldn't have the keys.

How did the client and server institute this shared secret (the keys)? They couldn't send the information over the network un-encrypted because everyone listening would know it also. They could send it through the mail or by messenger. Should there be one shared secret or many? Should I use the same key whenever I talk to this server or would I need a new key every time we exchanged data? Would there be different keys for all the servers I communicate with or would they all use the same? The creation and juggling of all these keys requires some *key management*.

In addition to encrypting the data between my client and the server, there is the problem if the server knowing if it is really me talking to it. For instance, if I send a VISA number, how does the server know it's my VISA number and not a stolen number? Existing manual systems require a PIN or a signature to verify identity. So, we need some sort of *authentication* -- the ability to authenticate the identity of a data sender or receiver.

Authentication identifies a person before something is done, but what about leaving proof that something was done. How do you "sign" an electronic document so that it could be proven in court that it was indeed you that signed it? Thus there is the need for *certificates* that can be used to identify a person or an association (this person signing this document).

Bulk Encryption
or How To Keep People Out of Your Diary

Bulk encryption refers to algorithms that encrypt blocks of data. They are used to encrypt a message or a stream of data. The encryption algorithm usually includes a key -- the sender knows the encryption key and the receiver has the decryption key. So, for Mary to send a message to Bob, Mary and Bob somehow set up their keys. Mary then encrypts the message with the encryption key and sends it to Bob. Bob decrypts the message with the decryption key. Anybody with the decryption key can decrypt the message so the secrecy of the key is very important.

The most popular encryption algorithm is DES ("Data Encryption Standard") that was defined and endorsed by the US Government in 1977. DES is a "secret key, symmetric" cryptosystem meaning the sender and receiver share a secret key and that the key is the same for encryption and decryption. Researchers have tried to break DES for more then a decade and no one has been successful, so it is felt that DES is reasonably secure.

Skipjack is a new algorithm that is the bulk encryption algorithm chosen for the U.S. government's Capstone project[1]. Skipjack is new and unknown (the algorithm is classified) but it could be the next de facto standard the same way that DES became one -- the U.S. government mandates it as their standard.

Beyond these two algorithms, there are thousands of algorithms that can be and are used for bulk encryption that range from fancy, classified algorithms to Captain Midnight decoder rings.

Public Key Encryption

Public-key cryptographic systems are a relativily new invention (1976) that adds new capabilities to encryption. In this system, each person gets a pair of keys -- one the public key and the other a private key. The public key is published for anyone to see and the private key is kept secret. The encryption algorithm is asymmetric: data encrypted with the public key can be decrypted with the private key.

Say Mary wants to send a message to Bob using public-key encryption. Mary would encrypt the message with Bob's public key (kept in her address book or looked up in some directory). The message would be sent to Bob who could decrypt it with his private key. Anyone could send encrypted messages to Bob and the messages could only be read by him.

The advantage over shared secret cryptosystems is that Mary and Bob didn't have to arrange the shared, secret key in advance. In this public-key example, there was no key exchanged. This has definate advantages for uses like email where the relationship with the receiver might be transitory.

Public-key systems can be used for authentication -- Mary would "sign" the message with her *digital signature*. The digital signature includes the message encrypted with Mary's private key plus other identifying information[2]. Mary would append the digital signature to the message to Bob and then encrypt the message with Bob's public key. Bob, after he decrypted the message, can use Mary's public key to verify that only Mary could have created the signature.

---

[1]Capstone is the is the U.S. government's long term project to develop a set of standards for publicly-available cryptography. There are four major components: a bulk data encryption algorithm, a key exchange protocol, a digital signature algorithm, and a hash function.

[2]To make the size of digital signatures managable, the whole message is not encrypted but a hash function is used to generate a few hundred bit code that represents the document. Several hash codes exist that create unique codes for nearly any document.

One disadvantage of public-key systems is the speed of the algorithms. Because it must be difficult to compute the private key from the public key, the keys are very long (500, 1000, or more bits). This makes encryption and decryption slow. Thus, public-key cryptosystems are usually used for authentication and for key exchange and a bulk encryption algorithm is used for the majority of the data.

The predominate public-key system in use today is RSA, invented in 1977 by Ron Rivest, Adi Shamir, and Leonard Adleman (the last name initials gives the "RSA" name). This system is based on large primes and the difficulty in factoring same. Several companies and agencies exist to create and certify public and private keys for individuals.

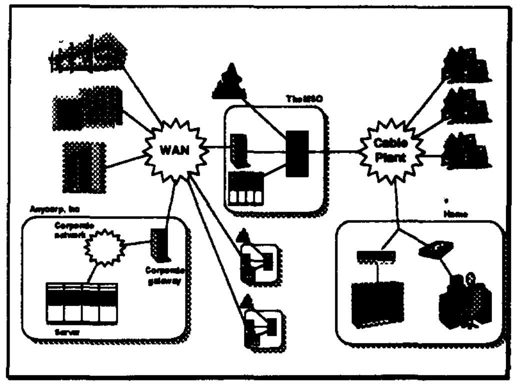
The U.S. government's Capstone Project has chosen an algorithm named DSS ("Digital Signature Standard") for creating digital signatures. It is not as general as RSA and cannot be used for key exchange, for example.

Secure Telecommuting

or "Humm, I wonder what my competition is up to."

Let's look at an application: telecommuting.

In the picture below, Anycorp has employees scattered over the metropolitan area that can telecommute to the companies' offices. Like most metropolitan areas, there are several MSOs serving the region so the company is connected through some wide area data network ("WAN") to the different MSOs. The MSOs provide the data connection from the WAN to the computers in the home.



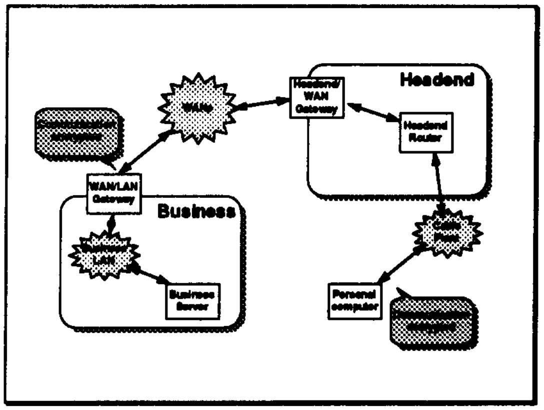
There are many companies in the area that have telecommuting employees so there are multiple companies connected to the WAN and employees from competing companies connected to one MSO. It's obvious that data from one telecommuter will get multiplexed with data from another telecommuter as the data moves between the companies and the employees at home.Additionally, inside each home, competitors' data will be on the cable.

The multiplicity of companies and employees and MSOs adds an interoperability requirement. For instance, if the person in the home is a consultant, she might need to telecommute to two different companies. This means that any security system must be compatible and, if there is specialized hardware to do security, that hardware must be either compatible or everywhere or the person in the home will need different hardware for everyone they're communicating with.

With corporate data moving through various networks and even into competitor's homes, physical security, trust, and passwords won't protect the data.

The way an employee is connected to the company can vary: phone lines (for calling in while on the road), wireless (when using their PIM), and cable (when connected at home). Each of these has its advantages and disadvantages but, to provide useful connectivity, all must be supported and, ideally, the security solution for one should work for all. The security solution cannot just work for connecting a computer to a cable modem.

Any company would make sure that any data leaving the company is encrypted until it gets to the employee as shown in the following simplified drawing.



In a simple system, the employee trys to connect to the company and is authenticated (this could be as simple as a password). All communication thereafter would be encrypted by a shared secret key. A twist on this would be to make the shared secret key the password. In this case, the employee would encrypt all communication to the company with the password and the company could decrypt it because it had the password in its user databases.

For this simple system the MSO is not involved in the security aspects of the communication and thus does not incur extra hardware, administration, or liability. Even if the MSO is selling the "service" of telecommuting, the implementation is mere connectivity between the employee and the company.

Secure Correspondence
or electronic envelopes

The shared password worked in the above telecommuting example because it is reasonable that the employee has to be "registered" with the company before he can connect. This wouldn't work, though, if you wanted to send secure mail to a friend or a business associate. It would be unreasonable to have to arrange a password with everyone you wanted to send mail to.

A public-key system works well for electronic mail. The sender does not have to pre-arrange a shared secret (e.g., a password) with the recipient. The message is encrypted with the recipients public key and sent.

So, while passwords (or fixed, shared secrets) can be used for pre-arranged connections, they are not suitable for the multitude of connections that can be made over a large community or for relationships that are transitory.

Money and Contracts
or digital money can move faster than paper money

If you always went into the same stores, they'd know you and know to accept your checks. If there's only one store available on the cable network, you can have a password to that store. If there are many stores, you need something better then passwords.

When you do buy something and exchange money, merchandise and warrenties, both side of the transaction will want more then just a forgable note saying it was done (digital bits are easily changed). It's not that important when buying a toaster but someday it will be possible to buy houses. Transactions need more then just un-snoopability.

There are also two types of monetary transactions: transactions with a third party (e.g., existing plastic money where someone authroizes and records the transaction) and between only two parties (e.g., paper money). Some people like the latter because of the anonymity. It's easier to implement the former, though.

Digital signatures allow someone to send an unforgable message to a merchant to purchase and item and to exchange money. These can be considered mini-contracts. Also, communication with a credit vendor (a plastic money provider) can be authenticated and transmitted without fear of being overheard.

Digital signatures and the related public-key cryptosystems enable much of the commerce the new interconnected world is supposed to make happen.

Conclusions

Existing computer network systems solve most connectivity problems but they don't address security, privacy or accountability. When the computer network is "up on the poles", the security systems that have worked in the past don't work any more.

Privacy, security, and accountability are required for communication over the cable system. Even if it only supports video-centric services (e.g., video on demand), the requirements of commerce can't allow breakends and forgeries.

With greater interconnectivity, where the real power of having computers at both ends of the connection creates compelling environments, the security needs also include inter-operability and flexibility. A fixed, position dependent, subscriber dependent solution does not create the base for the growth we all envision.

Glossary

authentication - is the process of verifying that the person communicating with is the person expected. In the physical world, an example is showing your driver's license to use a check.

bulk encryption - encrypting large amounts of data.

Capstone - the name of a U.S. government project to define and standardize a bulk encryption scheme, a digital signature standard, and a key exchange protocol for government agencies.

certificate - in public-key systems, they are digital documents that attest to the binding of a public key to an individual. Certificates rely on (contain) the certificates of the attesting agency. This creates a certification chain up to some well trusted authority.

Clipper - the silicon chip that implements the Skipjack bulk encryption algorithm. Part of the Capstone project.

digital signature - data that is unique to the sender and un-forgable. This is usually implemented with a public-key cryptosystem.

key - a "seed" for an encryption or decryption algorithm. The algorithm can be widely known but, without the keys, encrypted data cannot be decrypted.

key management - the processes and protocols used to distribute and store encryption and decryption keys.

password - usually a work or phrase that is the shared secret between a client and a server.

physical security - using physical barriers and monitoring to create

shared secret - some information that the sender and receiver of data knows that no one else knows. This is usually the keys that go with some encryption algorithm.

Skipjack - the bulk encryption algorithm chosen for the U.S. government's Capstone Project.

trusted third party - two people can rely on another agency to distribute and keep communication keys secret. This third party generates keys that are used to communicate. For instance, the Kerberos system has a server on a computer network that will authenticate users on the network and hand out session encryption keys.

MULTIMEDIA DELIVERY DEVICE for FIBER/COAXIAL HYBRID NETWORKS

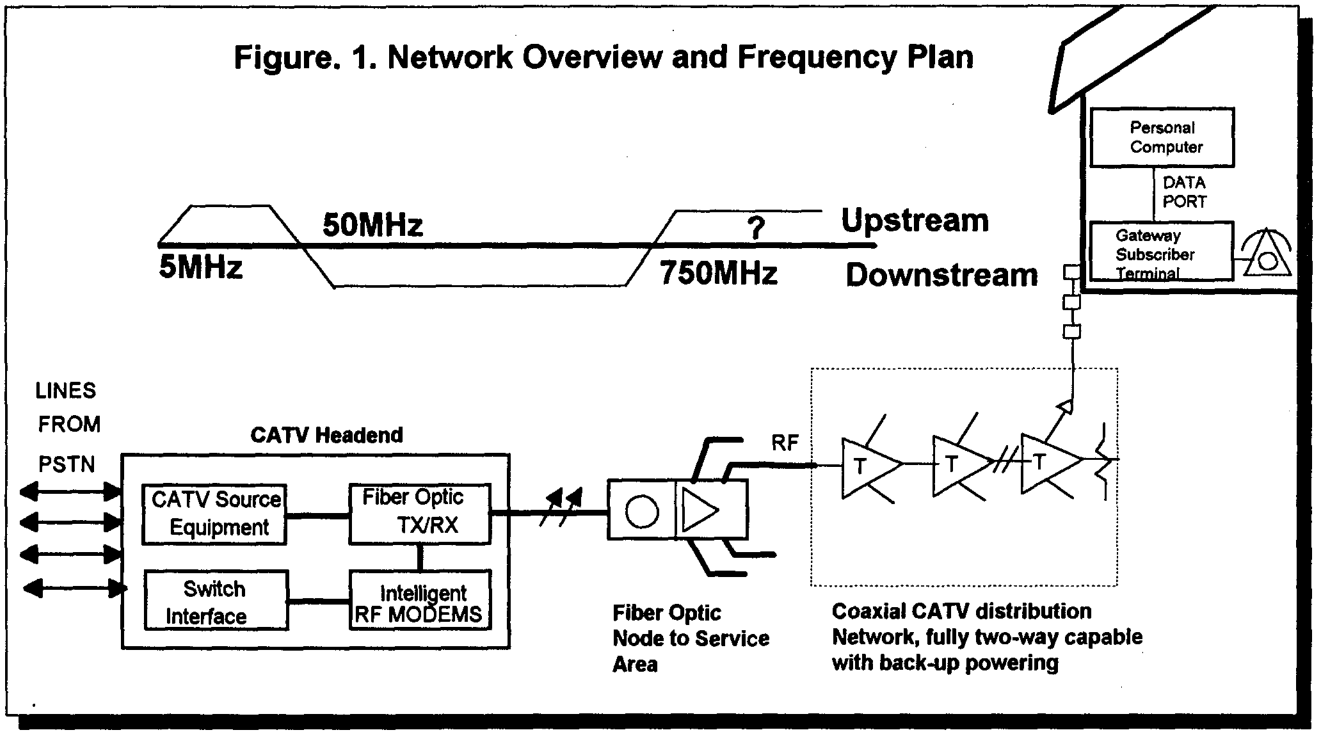
by
Aravanan Gurusami, Jeffery Cox, and Mark Chapman
Philips Broadband Networks Inc.
100 Fairgrounds Dr.
Manlius, New York 13104

Abstract : *The fiber/coaxial network which is currently used for carrying entertainment video in mostly analog format is likely to carry other types of traffic in the near future. The types of traffic that will traverse the super highway fall in three broad categories, namely, (i) one way for traditional broadcast video delivery (ii) asymmetric two-way for video on demand services where minimum traffic exists for subscriber initiated controls and (iii) symmetric two-way services like voice, video-telephone, and computer network inter-connections. This paper focuses on the design aspects of a multimedia delivery device with symmetric delivery capacity of Nx64KBits/S for subscribers connected to the CATV network. The performance objectives for circuit switched and packet switched networks for both connection oriented and connectionless types are discussed with acceptable compromises required for a low-cost implementation of a Broadband Communication Gateway (BCG) system. This system provides Nx64KBit/S capability with appropriate interfaces for voice and data delivery and proper telephone network interfaces for universal switching by the Central Office (CO).*

1.0 Introduction : The Fiber/Coaxial hybrid network is emerging as a leading contender for the communication highway of the future. While high bandwidth services on an Asynchronous Transfer Mode (ATM) backbone are a distinct possibility, the upstream bandwidth of many CATV networks in existence today accommodates only modest services like telephony and data applications that use 64KBit/S, or multiples thereof on a full time basis. This paper describes a BCG system, including the architectural aspects of headend interfaces; internal architecture of the device; and application areas for voice telephony, symmetric data transport and Compact Disk Interactive (CD-i) interface for interactive entertainment. The paper concludes with the author's view of future enhancements of the gateway product for the highly competitive consumer market.

2.0. System Architecture of BCG

The BCG system architecture consists of the CATV network Distribution System, headend telephone network (PSTN) to CATV distribution bridge, and the gateway transceiver device at the subscriber's home. Figure1 illustrates the system architecture of the BCG and a frequency plan model for a full service network in which the symmetric services use the conventional CATV return band from 5 to 40MHz for upstream communications. The system is designed to be frequency agile to approximately 800MHz. The BCG functions over the CATV Fiber to the service area topology where the node splits will serve from 500 to 2000 subscribers. The bandwidth efficiency of the BCG allows 400 to 500 full-time 64KBits/S circuits to function in the available upstream bandwidth of a conventional CATV system.

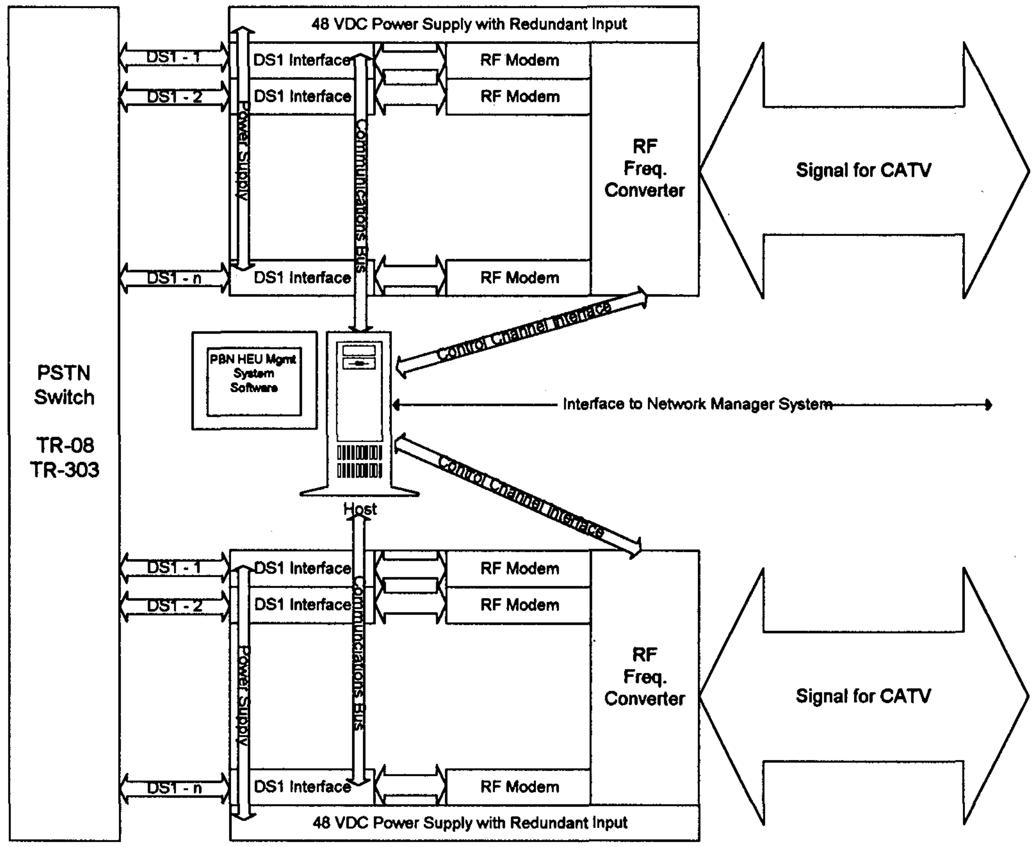


2.1. Bridging PSTN and CATV at the Headend

Modern telephone systems use a Time Division Multiplex (TDM) structure for transmission and switching. The protocols and switching methods have evolved into well established standards and require compatibility for seamless interconnects. The CATV distribution system uses an equally mature frequency division multiplexing for the backbone of its distribution system, and it is unlikely to change in the near future. For transporting TDM structure efficiently on a CATV network, the issue of differential distances of subscribers encountered in CATV distribution has to be resolved. Figure. 2 illustrates the PSTN to CATV distribution interface system used in the Philips BCG product family.

The switching interface to the PSTN complies with industry standard TR-08 and TR-303 requirements. Each subscriber is assigned a dedicated time slot from the switch; thereby no concentration is provided in the cable transportation system. The RF MODEMS provide necessary RF transportation for CATV interfaces. The MODEMS are intelligent to handle the differential distances encountered in a CATV distribution architecture and work together with the subscriber premise device to handle burst traffic in the upstream path. Bandwidth efficiency is realized using multilevel modulation techniques that are appropriate for a CATV environment; frequency agility provides flexibility and capability to avoid "trouble spots" in frequency associated with typical return systems. Hot insertion of electronics and redundant power supply functions can be provided to reduce down-time. A Management System provides network monitoring primarily at DS1s and configuration control of the DS0s. Many of the control functions are performed using an out-of-band channel without interfering with the operation of telephony channels. The system controller is based on an industry standard open platform with Graphical User Interface (GUI) making multi-vendor, multi-application management possible from a single management console. The RF signals from the MODEM and control channel connect to the CATV system for distribution through fiber optic transmitters and receivers.

Figure 2. PSTN to CATV Distribution Interface



2.2 Subscriber Terminal Unit

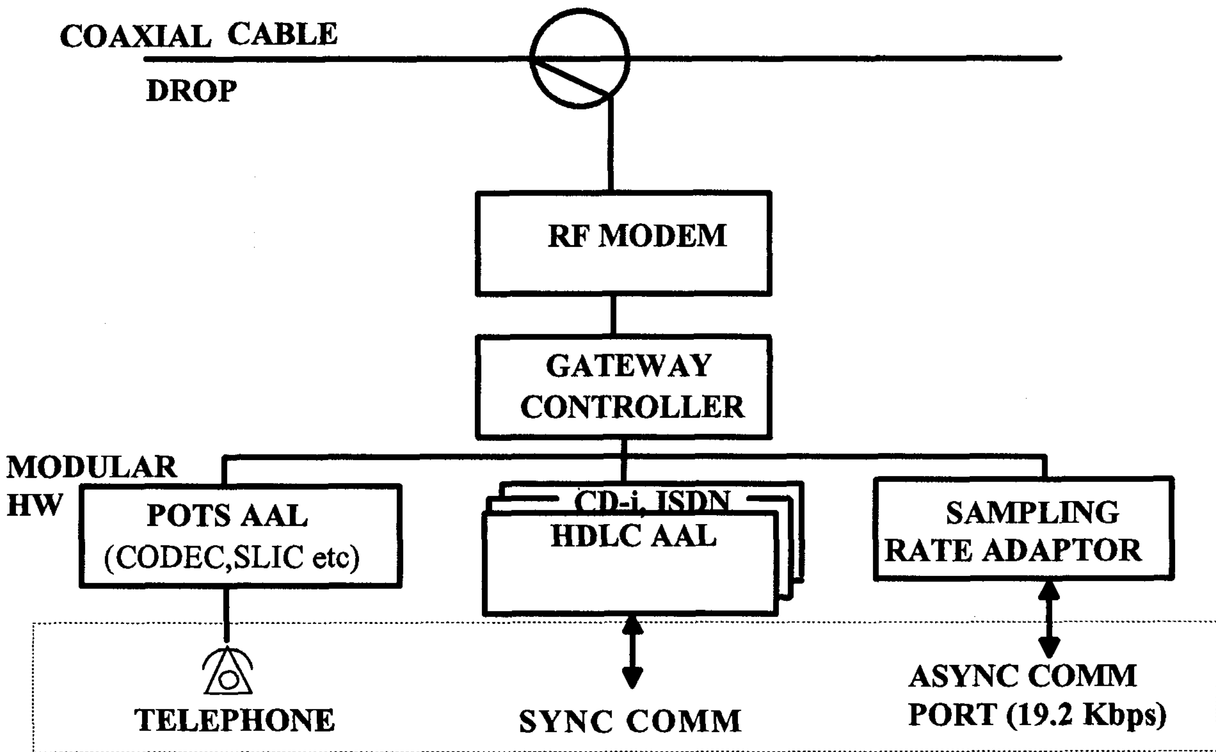
The internal architecture of the BCG subscriber terminal unit (SU) is shown in figure 3. The SU provides the physical and electrical connection between broadband distribution network and subscriber baseband connections. It provides Modulation and Demodulation, Analog to Digital and Digital to Analog conversions, Data Port protocol and rate adaptation for transmission and reception over the broadband CATV medium.

Features provided by the SU are:

1. BORSCHT Functions,
2. Packet-oriented protocol supporting Nx64Kbs channels,
3. Battery Backed Power Supply option,
4. Adaptive ranging for differential distance management on CATV medium,
5. Plug and Play - No setup requirement,
6. Data port for multimedia applications, and
7. Modular structure to accept various interface adaptation devices.

The microprocessor and the MODEM are capable of handling payloads at data rates in

Figure 3. BCG Subscriber Terminal Unit Architecture



Nx64KBits/S format up to a maximum of 2.048Mbits/S. The size of the payload for an application delivered to the Application Adaptation Layer (AAL) devices by the gateway controller is variable in 64KBits/S steps to a maximum of 1.536 MBits/S for ANSI requirements and 2.048MBits/S for CCITT requirements. Different AALs provide the necessary interface for different applications. For example, the POTS AAL is a CODEC for a voice application and the sampling AAL can be a CCITT V110 rate adapter for a simple serial port interface at 19.2KBits/S asynchronous port. The following sections discuss the AALs, specifically in their connectivity and isochronous voice application, a packet switched AAL using High Level Data Link Controller (HDLC) for connection oriented and connectionless services, and a rate sampled 19.2KBits/s interface for personal computers.

3.0 AAL for Voice Application

A typical CODEC for voice applications samples an analog line at 8KHz and provides 8 bits of data per sample after companding using standard companding methods like A-law or μ-law. So in the downstream direction the CODEC payload is delivered directly to the CODEC. In the upstream direction, several bytes are collected over a period of time from the CODEC and burst-out in the proper upstream time slot. The burst packet transmission in the upstream direction is used to conserve bandwidth and reduce hardware requirements at the headend. However, the data collection introduces some time delay for transmission of the packet; therefore, the packet size for a connection oriented delay sensitive service should be kept minimal to reduce the impairment due to talker and listener echoes. For POTS applications, a Subscriber Line Interface Circuit (SLIC) is necessary to do the BORSCHT functions; however, for a computer multimedia application, a CODEC with sufficient drive to activate a low power headset is adequate.

3.1 AAL for Synchronous Data Port

For data communication, some networks use packet-switched technology, in which blocks of data called packets are switched from a source to a destination. Source and destination can be terminals, computers, printers, or other types of data communication devices. In a packet -switched technology, it is possible to share the same distribution and transmission facilities in connection-oriented (where call-setup is required before data exchange) or connectionless (no logical connection is setup prior to data exchange) modes. Synchronous protocols based on HDLC and other protocols like X.25, LAPB, LAPD etc. that evolved from HDLC provide standards for call setup and data exchange methods from the data stream provided by the gateway controller. A typical AAL provides an ISDN-B service extracting two bearer channels (2B) at 64KBits/S each and one data channel (D) at 16KBits/S derived from three 64KBits/S time-slots provided by the gateway controller. The ISDN-B service provides a low-cost high-performance connectivity solution to residential cable customers for high-speed links computer links or video conferencing applications. Similarly, a higher speed service like H0 (384 KBits/S) and H11 (1.536 MBits/S) can be provided by using the proper AALs from the gateway controller.

3.2 AAL for Interactive Services

The proliferation of interactive devices such as the CD-i system provides an additional application for two-way capacity on CATV systems. Many of the current interactive systems are based on some form of CD-ROM storage. The CD can be readily replaced by a communications channel, effectively placing the CD drive at some remote location. A CD retrieves data at about 1.4 MBits/S, nearly filling one T1 channel. Applications can then be run transparently. There is no difference to the user between local CD applications and remote, server based, applications. Of course, the whole point of this is to provide a truly interactive system. Getting the commands from the user back to the server with a minimum delay is critical to the interactivity of the system. The return delay needs to be limited to just a few milliseconds. In fact, the system we are describing adds just a couple of milliseconds of delay to the return data. The user will not be able to perceive the difference between a local application and an application running on a server miles away.

As in all interactive applications on broadband networks, the return capacity is a critical resource. In the networked CD-i application, using the basic BCG technology, we are not making very effective use of the return capacity. Certainly we do not need 64KBits/S of return capacity to be able to control the interactive applications. By using smaller packets, we can accommodate more users in a given amount of return bandwidth. As the packets get smaller, however, the transmission system gets less efficient. We can add users, but the overall data capacity of the return channel is reduced. For example, if we can support 24 users at 64KBits/S, we can possibly support 36 users at 32 KBits/S by sending smaller packets. We can also share the bandwidth by sending packets less frequently. This adds delay to the system and so we need to use this option carefully. By using both of these techniques, we can comfortably support 64 users in a single 1.5 MHz return channel. Each user is allocated 16KBits/S, with the system adding about 3 msec of delay due to the return channel protocol.

3.3 AAL for Asynchronous Data Port

Today, most homes that are equipped with personal computers use telephone MODEMS to communicate to bulletin board services and other data exchanges via the serial ports. These serial ports usually handle about 19.2KBaud and in some cases are capable of going as high as 115KBaud . It is desirable to offer these services using a CATV system with the connectivity of the telephone network without compromising the telephone voice connection. Also, a direct digital port eliminates the need for traditional voice MODEMS, thus reducing the cost of the service provided by the CATV versus the twisted-pair telephone network. There are many rate sampling methods like CCITT V.110 that provide standards for transporting data up to 19.2KBits/S through asynchronous lines onto a 64KBits/S synchronous highway. Direct rate sampling with start bit timing corrections are available from some component vendors like Mitel to provide low-cost methods of rate sampling for asynchronous lines. The modest 19.2KBits/S data rate compares favourably with available telephone MODEMS that have to resort to data compression to achieve higher speeds and occupy the telephone line. Most personal computers can work at these speeds with the standard serial ports and some may require upgrades to communication boards even at these modest rates. Higher data rates can be achieved by providing more time slots and custom or parallel printer port drivers on the PC.

4.0 Conclusion

The world is undergoing some of the most revolutionary changes in interactive communications since the advent of telephones. The fiber/coaxial hybrid network is in place for the "last mile" to integrate the communication needs of new services like multimedia. This paper gave an overview of Nx64KBits/S transport on a CATV network with full PSTN connectivity. The service can be extended to higher data rates to provide multimedia services like video conferencing, interactive games, and high-speed ISDN networks. Many of these services can be provided today on the CATV network for a typical home user if commercial justification and usefulness can be demonstrated. The BCG starts with a telephone MODEM sized device that simultaneously transports 64KBits/S voice and 64KBits/S data for a full two-way symmetric service. The product line will be extended to offer more services like interactive video and work-at-home applications, while simultaneously transporting voice. The product will migrate to the side of the house to provide lifeline telephone service and ISDN connectivity when regulations permit commercial success of such a network.

REFERENCES :

1. Charles A. Merk "An Economic Evaluation of Broadband Distribution Networks to the Home" 1991 SPIE Conference

2. Charles A. Merk et.al. "Integrated CATV/Telco Network for Switched Broadband Services" 1992 SPIE Conference

3. William Stallings "Networking Standards : A Guide to OSI, ISDN, LAN, and MAN Standards" 1993 Addisson-Wesley

4. Harold C. Folds " Data Communications Standards" Edition III, McGraw-Hill, 1982

5. Gino Caira et.al. "CATV Bandwidth Assessment (A Practical Approach to System Design)", 1993 NCTA Technical Papers

6. Stephen Dukes " Next Generation Cable Network Architecture" 1993 NCTA Technical Papers

Network Management System Advantages and Implementation

Guy Sucharczuk

Harmonic Lightwaves Inc.

The emerging technologies that place greater and greater demands on our present day CATV Networks must also provide easier and more cost-effective ways to manage and control them. As CATV companies derive more of their income from PPV and other transactional services, network reliability grows from merely important to imperative. Historically, Network Management Systems (NMS) were not cost effective and provided no control of the CATV system. Fortunately as fiber moves closer to the home newer architecture's and Network Management Strategies become much more viable. In the new networks based on new architectures, Network Management Strategies become an integral part of the overall system design. By using sophisticated Network Management software and hardware in these new system architectures, network down-time can be virtually eliminated and system maintenance costs dramatically reduced. This paper will discuss the implementations and advantages of state of the art CATV Network Management Systems.

Network Management System Advantages

NMS systems are cost effective ways for allowing CATV systems to reach high performance goals and achieve optimization of valuable resources. As the revenue stream generated by transactional services becomes an increasingly larger portion of a systems revenue generation reliability becomes extremely important. People will not tolerate outages is services such as pay per view, alternative access, and telephony. Since the system operator systems will be unable to bill the customer during network down time, revenue becomes significantly effected if the network is unreliable. Additionally, data and telephony customers will not utilize the CATV network if it is unavailable due to repeated system failures, thus further affecting revenues. It must achieve the same level of accessibility as today's telephone network.

The other key advantage of NMS systems is resource optimization. By doing predictive analysis and using redundant equipment in the correct architectures network down time can be virtually eliminated. In addition by utilizing the above mentioned strategies system maintenance costs can be greatly reduced. The NMS system provides the user with non intrusive visibility into the network, thus providing the user with the ability to predict when a network element is going to fail. The NMS systems enables the user to allocate the correct resource when a failure does occur. By knowing which element failed, the user can either re-route the flow of information through the network or dispatch the correct resources to deal with the fault. This allows the system operator to maximize the effectiveness of system resources.

Determining A Network Management Strategy

In order to achieve the above mentioned gains, it is imperative that the user plan, design and implement a NMS strategy correctly. The hardest part of implementing a cohesive cost effective NMS system is deciding on the NMS strategy. The NMS strategy can be cut into two distinct sections.

The first section is the network elements that the user wants to monitor. The user must decide what are the important network elements that need to be monitored and how much effort to allocate to the task. There are several key parameters that can be used to determine the importance of a element with respect to what effect it has on the network as a whole. If the element under consideration effects a large portion of the network it is important to manage it. For example a fiber optic transmitter that feeds up to 5000-10000 homes node is probably a good candidate for network management. A receiver that feeds a 500 home node is also a good candidate. Equipment with controllable functions are obvious candidates especially those which provide system operators redundancy is a good candidate, because it can provide the system operator with a means of preventing system outages. Power supplies are another good example of what to monitor because they can have a large effect on the network. There are two more key parameters that are not so obvious. The first is Mean Time Between Failure of the element. If an element under consideration is prone to severe environmental conditions, it might justify the extra cost to be network managed. The final parameter that should be considered is the typical failure mode of the element. If the element fails gracefully over time then the element might warrant being considered so that the system operator can predictavely remove the unit from use before a hard failure, thus causing a service interruption.

In addition to managing network equipment there is additional network elements that should be considered in the NMS strategy. This is equipment that was designed specifically to help monitor the network. For example equipment that allows the user to monitor discrete digital and analog points that can examine air conditioning systems, temperatures, open doors , etc. Another useful NMS system related piece of equipment is an end of line monitor. This piece of equipment allows the user to examine the RF signal almost at the customer premise. These two pieces of equipment must also be examined while forming the NMS strategy . A typical example of how to judge whether a network element warrants monitoring is illustrated below. Let's look at an amplifier. Typically amplifiers do not have the ability to reroute the network flow. Amplifiers don't tend to fail gracefully. Amplifiers are being removed as fiber is being moved closer to the home, so the amplifier may not be the most optimal piece of equipment to monitor. An alternative strategy that would be more cost effective might be to have a backup link and an end of line monitoring trasnsponder that tells the user the condition of the service after the cascade. This optimization has two distinct advantages. The biggest is that the customer realized no service interruption because the system operator is providing him service through another route. The second big advantage is that regardless of how many amplifiers fail the end of line monitor will tell the user the status of the service. So, in a sense the system operator is monitoring all the amplifiers in this cascade with one piece of equipment. In addition as the user removes amplifiers the investment made doesn't become obsolete, also the user may be able to use this equipment to compile FCC proof of performance data. Once the user has determined which elements to monitor they must then choose the right equipment to enable them to implement their NMS strategy. This task is not as easy as it sounds. Most major CATV manufactures equipment claim to be status monitoring compatible, but what does that mean. The important criteria is does the network element provide enough monitoring and control points to be useful. For example a fiber optic transmitter that doesn't tell me the optical output on the NMS system is of little use. The fact that a unit failed will be instantly identifiable by the number of telephone calls from the subscribers. What's more important is the fact that the optical transmitters power is dropping and it can be determined in advance of a hard failure. In order for a network element to be truly network manageable it must be designed to be network managed from it's conception. Trying to network manage an element that wasn't designed to be managed provides minimal rewards at best. If a unit was not designed with all of it's essential parameters and controls to be microprocessor controlled, then the network management system will be unable to look into the internals of the element to effectively manage it.

The second section that must be examined is the application element. This is the software portion of the NMS system. It is as important as choosing the right elements for fulfilling the NMS strategy. It must be flexible enough to provide the user with all the functions that are necessary. But it shouldn't be overly flexible so that its to difficult to use. The ability to monitor and control the parameters that the network elements provide is the highest concern. The next parameter that the user should examine is ease of use. If the system is not plug and play people will not use it. This is where a point and click user interface becomes extremely important. The user should examine the NMS application element with respect to a typical customer service representative not a systems engineer. It should allow the systems engineer the flexibility to view all the important parameters of each element, while still providing a system overview at all times. Data logging of the network is an extremely important feature to be considered.

There are two type of logging that must be provided. The first type of logging is based on change of state. This would include open housing alarms, A/B switch information, and hard digital alarms. The second type of logging would include analog performance that should be used to perform analysis on element history and performance. Ideally the user would have the ability to set the interval of data acquisition. For example he may decide to log the performance of his elements at 12:00 noon because that's when his equipment runs the hottest or at 5:00am when it's at it's coldest. The ability to remotely monitor the network using readily available dial-up modems is a large consideration. Rather than mimicking the screen and keyboard of the system's computer, the remote software should be able to run independently and not effect the main terminal directly. The ability of the manufacturer to examine the system via the PSTN is another important consideration in determining the NMS strategy. This allows for multiple people to be examining the network and look at different elements simultaneously. Security is especially important if the user is planning to use the remote function. Multiple levels of passwords should be considered to fulfill this requirement.

While the application software is the part of the NMS system the user interacts with everyday. It's the network elements that will dictate how successful the NMS strategy becomes. The NMS system will be a constantly changing system that requires a good base in order to grow with the future. If the above NMS strategies are implemented the user will able to realize great gains in the networks reliability and management.

Noise Accumulation in CATV Distribution Systems Employing 1550-nm Externally Modulated Transmitters

D. Atlas and R. Pidgeon
Scientific-Atlanta, Inc.

Abstract
In this paper we present a method to determine the SNR and CNR performance of 1550-nm CATV distribution systems formed by optical amplifier-fiber cascades. The distribution of multiple AM CATV channels over long fiber spans is degraded by the presence of Rayleigh backscatter-induced low-frequency interferometric noise. When the laser source is modulated externally the low-frequency interferometric noise is mixed and translated around the AM carriers. The scattering-induced noise sets a limit on the link CNR as a function of fiber characteristics, fiber length, and source linewidth. Due to the narrow linewidth of solid-state and single-frequency semiconductor laser sources, the standard CNR measurement can give erroneous results because of the narrow-band translated noise. In this analysis we compared NCTA CNR and baseband video SNR performance of a CATV distribution system employed in a hub-ring architecture.

INTRODUCTION

With the advent of Erbium-Doped Fiber Amplifiers (EDFA's) employed as power and/or in-line amplifiers combined with externally-modulated 1550-nm transmitters, long fiber spans can be deployed to replace coax lines in the CATV network infrastructures. Furthermore, operation at the low fiber loss window at 1550 nm, the use of chirp-free external modulators, and the commercial availability of high output power EDFA's has made it possible to deploy longer fibers to broadcast CATV channels from the headend to the local service areas such as to hubs, fiber star feeders, and distribution nodes.

Performance of fiber-optic transmission systems can be seriously degraded by multiple reflections from discrete reflection points [1]-[5] and by Rayleigh backscattering within the fiber [5]-[11]. Although multiple reflections along the fiber can be suppressed by careful system design such as by using obliquely polished connectors, fusion splices and high return loss passive components, suppression of Rayleigh backscatter generated interferometric noise poses some fundamental limitations. In addition, employing in-line optical amplifiers at the hub sites to extend system reach could generate even more severe impairments when no isolators are used [8],[9],[11].

Noise is also inherent in an optically-amplified transmission system due to the spontaneous emission and amplification of photons in the active fiber of the amplifier. The signal-to-noise performance of the amplifier is characterized by its noise figure and is a function of a number of external parameters including amplifier gain, input signal power, and wavelength. The noise figure can be quantified in the optical domain by measurement of output signal power and amplified spontaneous emission (ASE) spectral power and density. The amplifier noise figure is also measured in the electrical domain, and this method is probably more applicable for AM CATV applications. In addition to ASE noise, interferometric noise can be generated due to reflections within the optical amplifier. The noise mechanism is the same as for discrete reflections and Rayleigh double-backscatter in the fiber link except that doubly-reflected signals in the amplifier may be amplified in each direction by the gain medium. Thus, it is particularly important to minimize the reflection of signals within and through the amplifier [11]. In this paper we present a method to determine the CNR and SNR performance of a CATV distribution network employing optical amplifier-fiber cascades in a counter rotating ring architecture [12].

NOISE FIGURE MODEL OF AMPLIFIER-FIBER CASCADE AM SYSTEM

Figure 1 shows a schematic of a CATV distribution system employing N optical amplifiers in an amplifier-fiber cascade. The CATV system noise factor can be expressed as follows:

$$F(f)=F_a+F_{sh}+F_{rin}+F_{th}+F_b(f) \qquad (1)$$

where F_a, F_{sh}, F_{rin}, F_{th}, and F_b are the noise factor of the accumulated optical amplifier noise, shot noise, laser relative intensity noise (RIN), receiver thermal noise and accumulated fiber

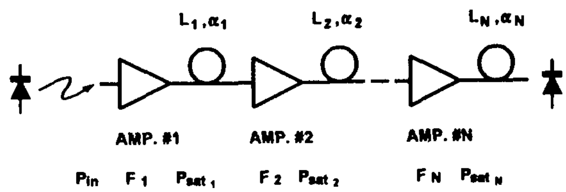


Fig. 1. Block diagram of N optical amplifier-fiber cascades.

Rayleigh double-backscatter noise, respectively. The total noise factor of N cascaded optical amplifiers is defined as

$$F_a=F_1+Pin\cdot\sum_{k=2}^{N}\frac{F_k}{\alpha_{k-1}\cdot Psat_{k-1}} \qquad (2)$$

where Pin is the input optical power at the first optical amplifier, F_k is the noise factor of the k^{th} amplifier, $Psat_k$ is the saturated output power of the k^{th} amplifier and α_k is the fiber and splitting loss after the k^{th} amplifier. The shot noise factor is given by

$$F_{sh}=\frac{Pin}{\alpha_N\cdot Psat_N}\cdot\frac{1}{\eta} \qquad (3)$$

where η is the photodetector quantum efficiency. The thermal noise factor is given by

$$F_{th}=\left(\frac{1}{\alpha_N\cdot Psat_N}\right)^2\cdot\frac{\langle i_{th}^2\rangle}{\left(\frac{\eta\cdot q}{h\cdot\nu}\right)^2}\cdot\frac{Pin}{2\cdot h\cdot\nu} \approx\left(\frac{1}{\alpha_N\cdot Psat_N}\right)^2\langle i_{th}^2\rangle\cdot\frac{Pin}{2\cdot h\cdot\nu} \qquad (4)$$

where $\langle i_{th}^2\rangle$ is the receiver equivalent current noise density. The laser RIN noise factor is expressed as

$$F_{rin}=RIN\cdot\frac{Pin}{2\cdot h\cdot\nu} \qquad (5)$$

where h is Planck's constant and ν is the optical frequency of laser emission. The total fiber Rayleigh backscatter noise factor is defined as

$$F_b(f)=RIN_b(f)\cdot\frac{Pin}{2\cdot h\cdot\nu} \qquad (6)$$

where $RIN_b(f)$ is the fiber Rayleigh double-backscatter relative intensity noise (including the low-frequency and translated interferometric noise). The Rayleigh double-backscatter RIN is expressed as

$$RIN_b(f)=\kappa\frac{4}{\pi}\left(\frac{\alpha_s S}{2\alpha_t}\right)^2\left[2\alpha_t\sum_{k=1}^{N}L_k-N+\sum_{k=1}^{N}e^{-2\alpha_t L_k}\right]\left(\frac{\Delta f}{\Delta f^2+f^2}+\frac{m^2}{2}\sum_{j=1}^{N_c}\frac{\Delta f}{\Delta f^2+(f-f_j)^2}\right) \qquad (7)$$

where κ is the depolarization factor, α_s is the Rayleigh backscatter coefficient, S is the fiber capture factor, α_t is the fiber absorption coefficient, m is the optical modulation index and N_c is the total number of channels. Inspection of expression (7) indicates that Rayleigh double-backscatter RIN has a non-flat noise floor, and therefore, a non-flat noise factor.

AM-VSB VIDEO CNR AND SNR DEFINITION

The AM-VSB CATV fiber optic distribution system employing optical amplifiers is deteriorated by shot noise, thermal noise, laser relative intensity noise (RIN), amplifier spontaneous noise and Rayleigh signal-backscatter beat noise. The video signal quality due to broadband noise is defined by NCTA RF CNR [13] and baseband video SNR measurements [14]-[16]. Assuming all the noise terms except the signal-backscatter noise term have flat power spectral densities the theoretical NCTA CNR at channel f_j is defined as

$$CNR=\frac{Pin}{2h\nu}\cdot\frac{m^2}{2}\cdot\frac{1}{F(f_j)\cdot B_n} \tag{8}$$

where $B_n=4$ MHz is the NCTA noise bandwidth (we have assumed a brick-wall shape bandpass filter). Note that F(f) (the non-flat noise factor due to Rayleigh backscatter RIN) in the CNR definition of expression (8) is determined with the carrier at frequency f_j turned-off (i.e., $m=0$ at f_j). The baseband video filter used in SNR measurements consists of cascaded 10 kHz high-pass, 4.2 MHz bandwidth low-pass, and unified weighting filters. The theoretical baseband video SNR is defined as

$$SNR=\frac{Pin}{2h\nu}\cdot\frac{(m\cdot A_m)^2}{4}\cdot\frac{1}{\int_{f_l}^{f_c}\left(V(f+f_j)+V(-f+f_j)\right)|H_c(f)|^2} \tag{9}$$

where A_m is the blanking-to-peak white video signal level in percentage, $V(f)=F(f)\cdot H_v(f)$ is the total noise factor filtered (weighted) by the vestigial sideband demodulator, $H_v(f)$ is the vestigial sideband demodulator filter transfer function, $H_c(f)$ is the CCIR 567 baseband unified weighting filter transfer function, and $f_l=10$ kHz and $f_c=4.2$ MHz are the video low-pass filter lower and upper frequency limit, respectively. Theoretical SNR calculation with unified video weighting filter yields a 0.6 dB improvement over CNR calculation when the noise power spectral density within the channel bandwidth is flat. Our experimental measurements yielded 0.4 dB to 0.8 dB improvement of SNR over CNR (with minimal translated noise around the AM carrier) which is consistent with the theoretical 0.6 dB improvement. If the noise power spectral density (PSD) is not flat, theoretical CNR and SNR calculations will yield different results, indicating that RF CNR measurements can not be faithfully employed.

1550 nm EXPERIMENTAL CATV DISTRIBUTION SYSTEM

Figure 2 shows the experimental set-up [17] used to measure the NCTA RF CNR and baseband video SNR of our 78 channel AM-VSB CATV distribution system. Inspection of Fig. 2 reveals that the transmitter consists of a laser source, a

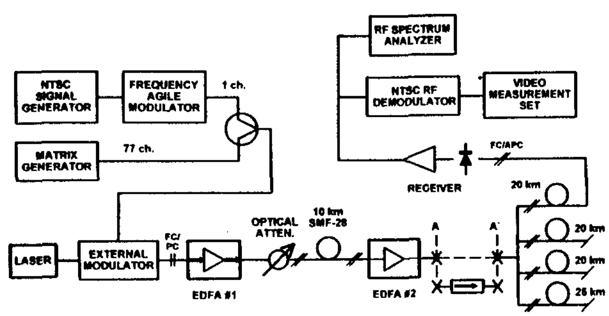


Fig. 2. Experimental set-up of a 78 channel AM-VSB CATV distribution system. Section A-A' denotes the position of the in-line amplifier output isolator.

linearized external modulator and an EDFA (with internal input and output isolators) used as a power amplifier. The laser source is a 1554 nm wavelength 700 kHz linewidth SL-MQW DFB laser chip packaged in an internally isolated module. We used a single channel agile AM carrier modulated by a NTSC video signal (luminance) combined with 77 unmodulated AM-VSB channels using a RF directional coupler. The power amplifier has a saturated output power of about $P_{o1}=13$ dBm at $P_{i1}=2.8$ dBm and a noise figure of about 5.5 dB. The saturated output power of the in-line amplifier is about $P_{o2}=12$ dBm at $P_{i2}=3$ dBm and has a noise figure of about 4.4 dB (measured with the output optical isolator placed in section A-A' in Fig. 2). The optical attenuator at the transmitter output is adjusted to maintain the in-line amplifier input power from -3 to 3 dBm. The received optical power after 20 km of single-mode fiber from port 1 is kept at about -0.9 dBm. The optical receiver equivalent input current noise density is less than 8 pA/√Hz with an optical return loss greater than 51 dB. The agile carrier is tuned either to channel 2 or to channel 29 for CNR and SNR measurements. For NCTA CNR measurements, the NTSC video signal at the transmitter is turned-off (only the unmodulated carrier is transmitted through the system) and the RF output from the receiver is connected to a RF spectrum analyzer. The AM carrier peak power level is first measured and then the channel under test is turned-off to measure the spot frequency noise power level [13]. This noise power measurement procedure is adopted to minimize ambiguities in measuring the non-flat translated noise spectra. For baseband video SNR measurements, the NTSC video signal at the transmitter is turned-on and the RF output connected to a Tektronics 1450-1 television demodulator operating in the synchronous detection mode. The baseband signal from the demodulator is sent to a Tektronics VM 700-A video measurement set (video analyzer). Within the video analyzer the baseband signal and noise pass through a 10 kHz high-pass filter, 4.2 MHz bandwidth low-pass filter and unified weighting filter. Finally, the noise at the filter output is integrated and compared to a nominal 100 IRE unit video signal (luminance) level.

Figures 3(a) and 3(b) show the low frequency and translated RF spectra of the low end channels with and without an in-line amplifier output isolator, respectively. Inspection of Fig. 3 indicates that the low frequency noise spectral density is suppressed as much as 7 dB when input power is increased from -3 dBm to 3 dBm. When an in-line amplifier output isolator is employed the signal-backscatter spectral density is suppressed by about 5 dB as compared to the spectral density without output isolator. The audio carrier which is about 20 dB lower in magnitude and 4.5 MHz away from channel 2, is generated by the agile modulator (shown in Figs. 3(a) and

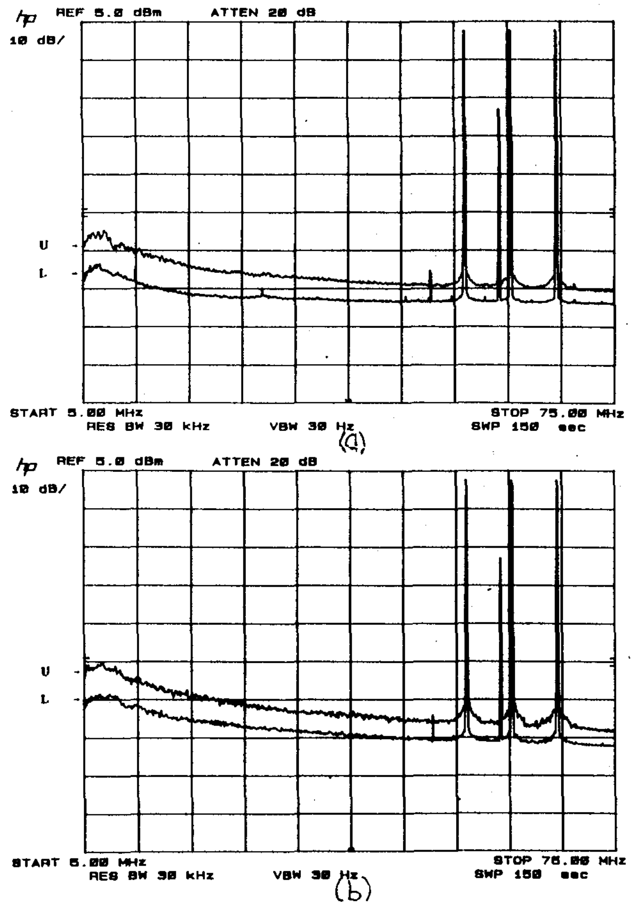


Fig. 3. RF spectrum of low-end channels:(a) with; and (b) without isolators. Upper (U) and lower (L) traces indicate the spectra at P_{i2}=-3 dBm and 3 dBm, respectively.

3(b)). Without an output isolator both the SNR and the CNR are degraded due to doubly amplified signal-backscatter beat noise; note that the lower frequency channels are the most degraded by the tail of the low-frequency and by the translated noise spectral density. The spot frequency noise spectral density for the CNR measurement is determined with the carrier turned-off. The SNR and CNR without output isolator is 44.1 dB and 44.8 dB at P_{i2}=3 dBm, respectively. At P_{i2}=-3 dBm the SNR and CNR is measured to be 39.7 dB and 40.5 dB. With an output isolator in place the SNR measurement is about 0.3 dB lower than the CNR measurements indicating that 0.9 dB degradation is due to

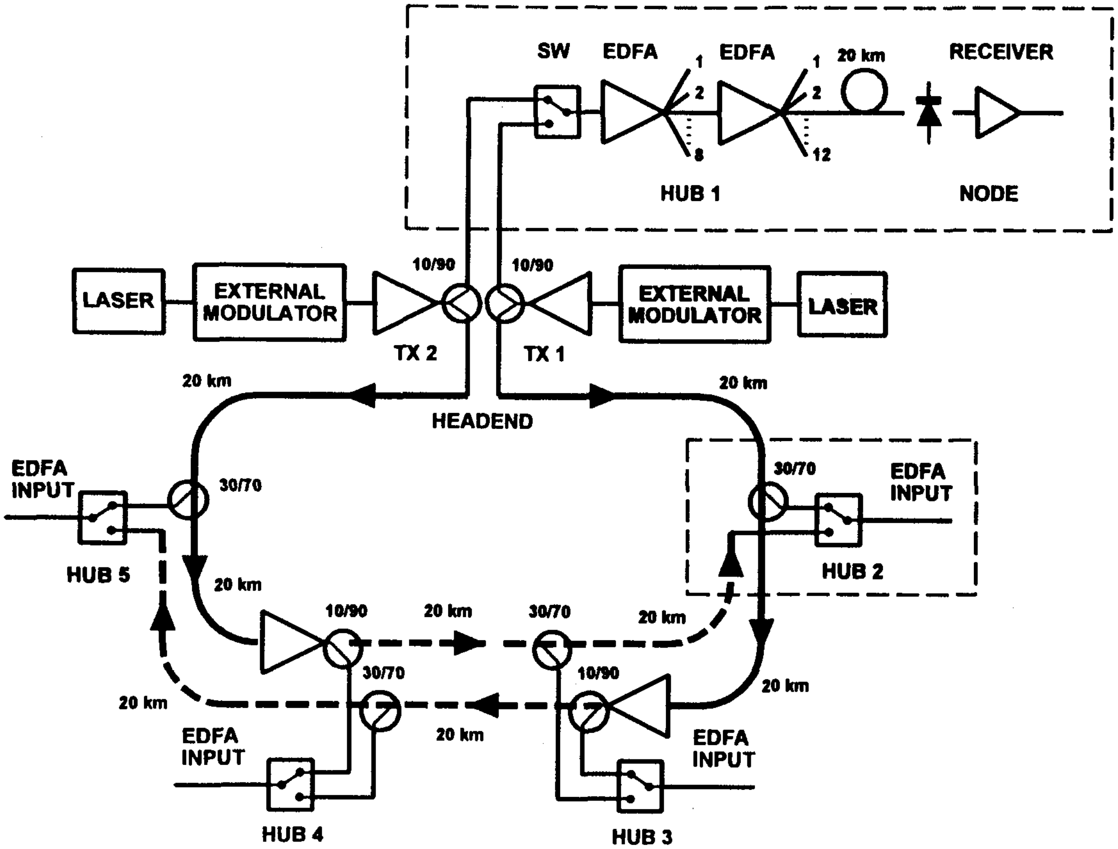


Fig. 4. Counter rotating CATV hub-ring optical network. Solid line indicates primary and dotted line indicates redundant fiber transmission path.

translated noise (with flat noise the SNR is about 0.6 dB higher than CNR). The SNR at $P_{i2}=3$ dBm is 49 dB and at $P_{i2}=-3$ dBm is 44.6 dB. Finally, we measured the Composite Second-Order (CSO) and Composite Triple-Beat (CTB) distortion of our AM-VSB CATV distribution system. For input powers ranging from -3 dBm to 3 dBm the CTB at 307.25 MHz is better than 60.5 dB; no CTB degradation is seen without the output isolator. The CSO at 55.25 MHz is greater than 64 dB and 65 dB with and without the output isolator, respectively. The CSO at 547.25 MHz is measured to be about 1 dB worse than the CSO measured at the low-end channel.

1550 nm CATV HUB-RING NETWORK PERFORMANCE ANALYSIS

Figure 4 shows a schematic of a counter rotating five-hub ring optical network. The optical ring network consists of one headend and five hub sites (one hub site is co-located with the headend). The headend consists of two externally modulated optically amplified transmitters. The input power at the transmitter power amplifier is about 6 dBm. The saturated output power of the power and in-line EDFAs are assumed to be 16 dBm with a noise figure of 4 dB. To ensure network survivability, the AM-VSB CATV channels from the headend are routed clockwise through hub 2 to hub 5 and counter-clockwise through hub 5 to hub 2 by transmitter 1 and 2, respectively. The distance between hub sites is 20 km, therefore, total distance from the headend transmitter 1 (or transmitter 2) to hub 5 (or to hub 2) is 80 km. In our analysis broadcast channels are distributed and trunked by two cascaded in-line amplifiers (see Fig. 4) to 96

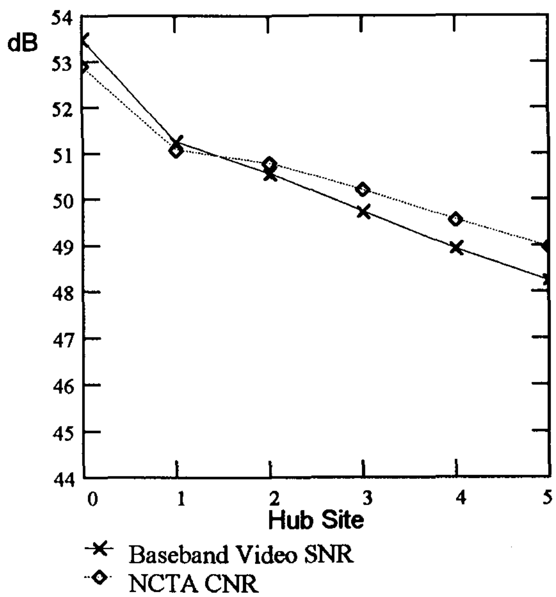


Fig. 5. NCTA CNR and baseband video SNR at channel 2 of the CATV distribution ring network for 100 kHz laser linewidth.

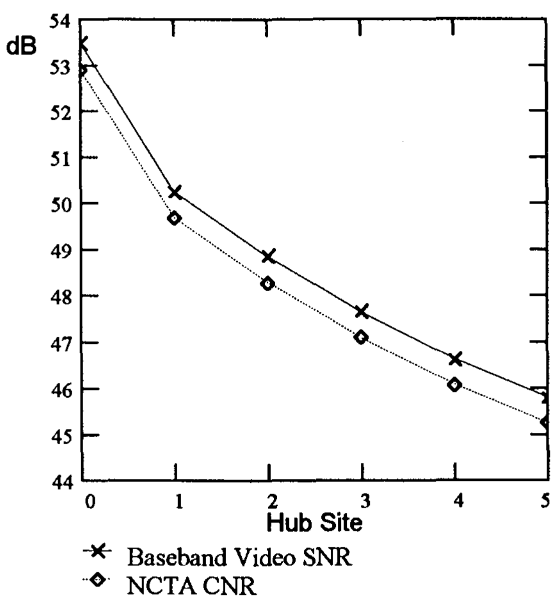


Fig. 6. NCTA CNR and baseband video SNR at channel 2 of the ring network for 5 MHz laser linewidth.

nodes (8X12) which are 20 km away from the hub sites. Our model assumes a 78 AM-VSB channel transmission with a 3 % optical modulation depth per channel.

Using expression (1) in (8) and (9) we compared the NCTA CNR and baseband video SNR at channel 2 for 100 kHz and 5 MHz laser linewidths. Figures 5 and 6 show the baseband video SNR and NCTA CNR at 100 kHz and 5 MHz laser linewidth, respectively. The SNR and CNR at hub site 0 in Figs. 5 and 6 are determined at the output of the headend (output of transmitter 1 and 2). The received optical power at hub sites 1, 2, 3, 4 and 5 for AM carriers transmitted by transmitter 1 (rotating clockwise on the hub-ring optical network) is 6 dBm, 5.3 dBm, 6 dBm, 5.3 dBm and 4 dBm. The SNR and CNR at the headend output is 53.5 dB and 52.9 dB, respectively. At the node of hub site 1 the SNR and CNR is 51.3 dB and 51.1 dB for 100 kHz linewidth, and 50.3 dB and 49.7 dB for 5 MHz linewidth, respectively. At the node of hub site 5 the SNR and CNR is 48.2 dB and 49 dB for 100 kHz linewidth, and 45.8 dB and 45.2 dB for 5 MHz linewidth, respectively. Further inspection of Fig. 5 and 6 reveal that the system performance at 100 kHz laser linewidth is degraded mostly by translated noise and at 5 MHz laser linewidth the system is degraded by the low-frequency interferometric noise.

CONCLUSIONS

An experimental 78 channel AM-VSB CATV distribution system is constructed employing two EDFA's simulating headend and hub sites and we compared NCTA CNR and baseband video SNR measurements using a 700 kHz linewidth externally modulated 1550 nm DFB transmitter. With no output isolators within the in-line amplifier we experimentally and theoretically determined that even for laser linewidths as low as 700 kHz (100 kHz with our theoretical calculations) the CNR and SNR performance is degraded by Rayleigh backscatter induced low-frequency and translated noise. Therefore, both input and output isolators are essential within in-line amplifiers employed for AM-VSB distribution. We presented a simplified method to determine the SNR and CNR performance of 1550 nm CATV distribution systems formed by

N number of amplifier-fiber cascades. Using our expression we evaluated the performance of a counter rotating CATV hub-ring optical network [12].

REFERENCES

[1] M. M. Choy et. al., "Interferometric conversion of laser phase noise to intensity by single-mode fiber-optic components," Electron. Lett., vol. 23, no. 21, pp. 1151-1152, October 1987.

[2] J. L. Gimlett, N. K. Cheung, "Effects of phase-to-intensity noise conversion by multiple reflections on gigabit-per-second DFB laser transmission systems," J. Lightwave Technol. vol. 7, no. 1, pp. 888-895, June 1989.

[3] W. I. Way et. al., "Multiple-reflection-induced intensity noise studies in a lightwave system for multichannel AM-VSB television signal distribution," Photon. Technol. Lett., vol. 2, no. 5, pp. 360-362, May 1990.

[4] T. E. Darcie, G. E. Bodeep, A. A. M. Saleh, "Fiber-reflection-induced impairments in lightwave AM-VSB CATV systems," J. Lightwave Technol., vol. 9, no. 8, pp. 991-995, August 1991.

[5] A. F. Judy, "Intensity noise from fiber Rayleigh backscatter and mechanical splices," Proc. ECOC'89, Gothenburg, Sweden, 1989, paper TuP-11.

[6] H. Yoshinaga, K. Kikushima, E. Yoneda, "Influence of reflected light on Erbium-Doped fiber amplifiers for optical AM video signal transmission systems," J. Lightwave Technol, vol. 10, no. 8, pp. 1132-1136, August 1992.

[7] A. F. Judy, "Generation of interference intensity noise from fiber Rayleigh backscatter and discrete reflections," Proc. OFC'91, San Jose, CA, February 1991, paper WL4, p. 110.

[8] J. L. Gimlett et al., "Observation of equivalent Rayleigh scattering mirrors in lightwave systems with optical amplifiers," Photon. Technol. Lett., vol. 3, no. 3, pp. 211-213, March 1990.

[9] R. K. Staubli, P. Gysel, R. U. Hofstetter, "Power penalties due to multiple Rayleigh backscattering in coherent transmission systems using in-line optical amplifiers," Photon. Technol. Lett., vol. 2, no. 12, pp. 872-874, December 1990.

[10] R. K. Staubli, P. Gysel, "Crosstalk penalties due to coherent Rayleigh noise in bidirectional optical communication systems," J. Lightwave Technol., vol. 9, no. 3, pp. 375-380, March 1991.

[11] J. C. van der Plaats, F. W. Willems, "RIN increase caused by amplified-signal redirected Rayleigh scattering in erbium-doped fibers," Proc. OFC'94, San Jose, CA, February 1994, paper WM6, pp. 158-160.

[12] A. R. Johnson, D. Raskin, "Thinking seriously about 1550 nm technology for CATV," Tech. Digest SCTE'94, pp. 143-154, 1994.

[13] *NCTA Recommended Practices For Measurements On Cable Television Systems*, Second Edition, 1989, National Cable Television Association (NCTA) Science and Technology Department.

[14] *CCIR Recommendation 567*, 1978.

[15] "Electrical Performance Standards for Television Relay Facilities," EIA Standard RS-250-B, September 1976.

[16] T. M. Straus, "The relationship between the NCTA, EIA and CCIR definitions of signal-to-noise ratio," Proc. NCTA'74, pp. 58-63, 1974.

[17] D. Atlas, R. Pidgeon, F. Little, submitted to J. Lightwave Technol., 1994.

Performance of a 256-QAM Demodulator/Equalizer in a Cable Environment

K. Laudel, E. Tsui, J. Harp, A. Chun, and J. Robinson

Applied Signal Technology, Inc.

Abstract

The premise of this paper is to show that 256-QAM is a viable modulation type for the cable television environment. Performance results for various cable system impairments such as AM hum, FM phase noise, residual FM, SNR, and microreflections are presented for an actual prototype 256-QAM Equalizer/ Demodulator that verify this premise. The prototype demodulator has also undergone extensive testing at the Advanced Television Test Center (ATTC) in support of the cable portion of the HDTV testing (January 1994). All prototype testing results conclude that, with a well conceived architecture and design, 256-QAM is an eminently practical format for the transmission and reception of high-data-rate cable television signals. QAM and VSB modulations are compared. There are advantages in the use of 256-QAM over 16-VSB in areas involving carrier and phase tracking and blind equalization which are analyzed here and also supported by results obtained in carrier offset, residual FM, phase noise, and channel change acquisition time testing [8]. Cable microreflection environments are modeled and equalizer length trades are presented that indicate that 16-32 taps (possibly 64 taps in some cases) are generally required to overcome the effects of cable microreflections.

Introduction

The advanced methods of digital compression have brought new visions of interactive television which will soon be realized in practice. Although some cannot fathom the use of hundreds of new television channels, the broadcaster's vision is that as video-on-demand grows in popularity due to sports events, home shopping, etc., the search for more capacity will continue. This paper will present results of a QAM system, e.g., 256-QAM, that will increase the capacity by 33% over 64-QAM. Section 1.0 of this paper presents an overview of what is theoretically possible and what is or can be practical. Section 2.0 compares QAM with VSB which has recently been recommended for HDTV broadcast application. Section 3.0 presents results of extensive simulations and laboratory testing of a 256-QAM prototype modem for digital television over cable. Finally, results are summarized in Section 4.0.

1.0 SIGNAL DESIGN TRADES

1.1 Introduction

Over the years, many of the advances in modem design have been brought on by the desire to increase capacity over telephone lines. Many of the characteristics of the telephone line are similar to cable channels (strictly band-limited channels with amplitude and delay distortion) and thus, recent techniques that have increased the capacity of the phone lines from 9.6 to 14.4 kbps up to the mid twenties of kbps can be adopted. There are however, some significant practical differences as well. One of the obvious differences is that phone lines are point-to-point and therefore the channel between the transmitter and receiver can be sensed and adjusted prior to transmission. This is the basis of attaining channel capacity via tailoring the transmission spectrum to be highest where there is highest SNR, etc., and is made practical via Tomlinson and Harashima precoding techniques (see Reference [1] for further details) and multi-carrier techniques ([2]). Unfortunately, in the cable channel, the situation is point-to-multipoint and thus the channel can be significantly different from consumer to consumer depending on the neighborhood cable layout and the consumer's own premises cabling. Even in the situation of full video-ondemand, the subscriber would share a 6 MHz digital carrier with 5 to 10 other subscribers and thus the 6 MHz carrier could not be "tailored" to any particular household. Thus, as has been shown in a number of prior NCTA papers [3], [10], the requirement for adaptive equalization for each consumer is necessary.

The cable signal needs to achieve the most capacity for a given (reasonable) cost. In a sense, this can be considered a "cost-limited" channel situation. If operator revenue can be considered to increase linearly with number of channels transmitted (or data rate per 6 MHz carrier), it can be stated as a general rule that each increase of 2 bits/Hz requires 6 dB more SNR and thus one more bit of precision in the demodulator and equalizer. Thus, for example, a multiplier may have to be increased from 9 x 9 bits to 10 x 10 bits or a rough increase of 100/81 = 24% increase in die area/complexity. This is roughly linear with the 33% increase in capacity stated earlier for 256-QAM versus 64-QAM, and thus the semiconductor costs will be roughly proportional to the increased revenue from the operator.

Thus, what limits us now? Of course the complexity of the set-top box is not solely governed by the demodulator subsystem but also by the tuner, filters, quality of splitters, fiber, etc. (see Figure 1-1) which would tend to degrade the higher capacity signal and would also have to be included in the cost equation if they would have to be modified. As a result, it would be cost-effective if the signal design could work with minimal modifications to the entire cable infrastructure. It is for this reason that we have added a system cost-limited constraint on the signal design.

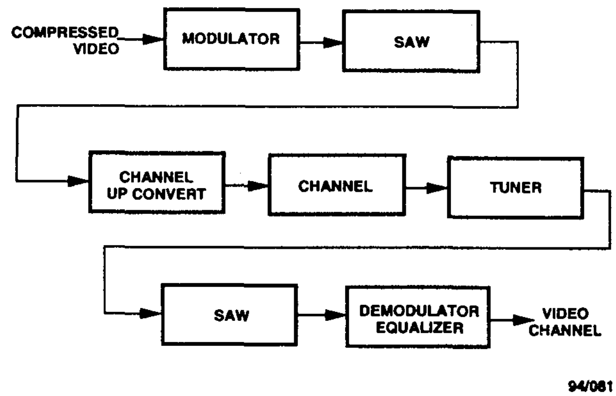


Figure 1-1. System Overview

1.2 BER versus SNR Bounds

Considering the theoretical bounds, it has been shown that for a strictly band-limited, high SNR, Gaussian noise/interference channel there is a 9 dB difference between the BER curves of an ideal Shannon system and uncoded MxM QAM [1, 4]. This 9 dB bound includes the knowledge of the subscriber channel and some shaping gain [4]. Practical coding schemes (e.g., Trellis, pre-coding, and shaping) have achieved as much as 7 dB gain [4]. Thus, for example, for 64-QAM, the required SNR (measured in the baud rate bandwidth) for 10^{-6} BER (using proper signal design techniques) can be as low as 19.6 dB and for 256-QAM, 25.6 dB, and 512-QAM, 28.6 dB. Thus, in an ideal band-limited cable channel with only amplitude/delay channel distortions, the above BER vs. SNR bounds hold.

The available SNR in reasonable worst-case cable environments have been presented in [5] as 30 dB SNR and in [6] as 32 dB SNR. Accounting for the fact of the inevitable losses in SNR due to consumer-caused cable wiring and losses due to other cable impairments (e.g., splits without pre-amplifications, ingress, etc.) a safe number might be 26 to 27 dB SNR with 3-6 dB of margin. Thus, in this environment the 256-QAM system appears to be a reasonable capacity-approaching compromise for the transmitted signal due to its requirement for only 25.6 dB SNR.

It should be noted that if constellation expanding trellis coding is applied to the signal, the actual signal may be a 512-QAM although the bps/Hz is still that of 256-QAM. As an example, a practical realization of a coded 256-QAM signal would be a 256-QAM signal with Reed-Solomon encoding which requires 27-28 dB SNR for 10^{-6} BER and results in an additional baud rate overhead of approximately 10%. (Note: Trellis coding would not result in any additional overhead.)

Finally, the above results depend on the demodulator circuits being properly designed to substantially remove (without additional significant SNR degradation) the effects of upconverter, tuner, etc., phase noise, AM hum, etc. This is very important since many cable environments are <u>not</u> SNR-limited but rather phase-noise limited.

1.3 Baud Rate Trades

It is interesting now to determine the maximum baud rate that can be obtained with the coded 256-QAM system with a given level of complexity of the demodulator. The 6 MHz channel is essentially band-limited by the combined effects of the transmit and receiver SAW filters. The maximum baud rate as a function of demodulator complexity will now be addressed with the constraint of the low-cost SAW filters.

Since the allocated channel bandwidth for each digital TV signal is 6 MHz, it is theoretically possible to operate a system at a symbol rate approaching this value. However, practical limitations arise due to:

1. the achievable frequency selectivity of input band-select filters,
2. the achievable SNR over the channel, and
3. the amount of digital processing and coding which can practically — from both a technical and economic standpoint — be applied to achieve a desirable level of performance (i.e., BER).

This technical and economic trade-off is illustrated in the family of curves shown in Figure 1-2. The details of these curves may change slightly depending upon the initial assumptions, but the character of the trade-off remains. Shown in this figure are curves of coded demodulation performance loss (at a BER of 10^{-6}) versus symbol rate for 256-QAM modulation. In these simulations, a typical SAW filter model has been used which exhibits 40 dB of rejection outside of the 6 MHz channel bandwidth. This channel-select filter model, derived from measurements on an actual device used in this application, is shown in Figure 1-3. Also assumed in these curves is the use of a particular Reed-Solomon code (t = 10) which can achieve roughly 4.5 dB of coding gain at a BER of 10^{-6}. (The vertical scale shows the loss relative to <u>uncoded</u> performance.) A fractional, T/2, equalizer has been assumed, and the curves show the performance loss for different length equalizer structures.

These curves illustrate that, as the symbol rate is forced higher for a given channel, in order to maintain a given level of performance (say 10^{-6} BER), a higher channel SNR must be achieved <u>and</u> a more powerful equalizer must be used. Both requirements lead to higher system costs.

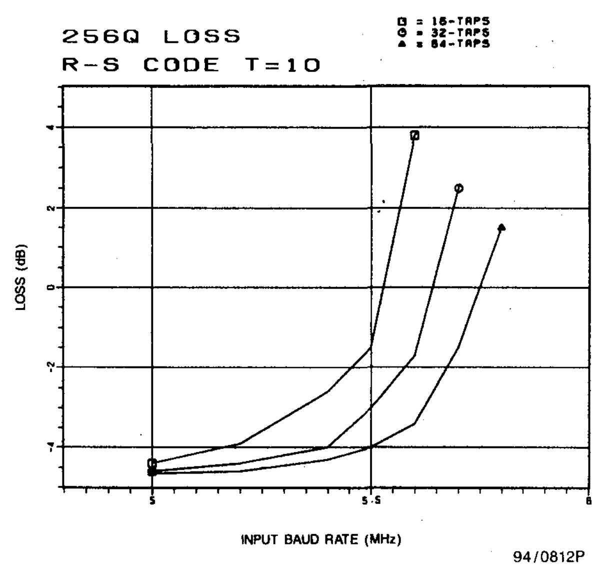


Figure 1-2. 256-QAM Loss versus Baud Rate for 6 MHz Channel and Reed-Solomon (t=10) Coding

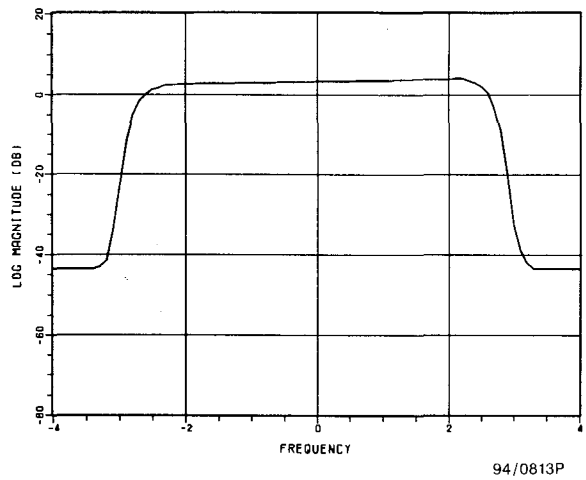


Figure 1-3. Model of Combined Transmitter/ Receiver SAW Channel-Select Filtering

2.0 QAM AND VSB TRANSMISSION TECHNIQUES

2.1 Introduction

Two transmission methods, QAM and VSB, are the principle candidates for use with digital television systems. Both make efficient use of the available transmission bandwidth and may be implemented using highly integrated digital architectures. Though the transmission and reception techniques for the two are quite different, the actual transmitted waveforms are similar. Both are generated using similar quadrature modulation techniques (VSB can also be generated by direct filtering at IF). Both methods require the same amount of transmission bandwidth for a given data rate and their spectra are very similar in appearance.

The transmitted spectrum of VSB is distinguished from that of QAM by the presence of a small pilot tone at the carrier frequency. VSB demodulation requires a coherent carrier reference in order to reconstitute the original double-sideband signal. QAM demodulation, on the other hand, is able to recover the coherent carrier from the quadrature waveforms and so needs no such reference. The waveform equalization of the VSB signal is less robust than that of QAM, due to the way in which the carrier pilot must be used in VSB reception. A VSB demodulator must acquire the carrier pilot, via a relatively narrow tracking loop, prior to any equalization. A QAM demodulator, on the other hand, can perform blind equalization in the presence of carrier offsets and final equalizer convergence takes place after the carrier is removed via a relatively wide bandwidth tracking loop. VSB transmissions also frequently include periodic "training sequences" (strings of known bit patterns) to assist in adapting the equalizer to the channel.

For the same transmission data rates, these two modulations exhibit the same theoretical performance in the presence of additive white Gaussian noise (AWGN). Hence, bit-error rates versus E_b/N_0 (bit-energy-to-noise-density ratio) are the same for 8-VSB and 64-QAM as they are also for 16-VSB and 256-QAM. Figure 2-1 shows these two performance characteristics. These characteristics do not account for the additional VSB power due to the pilot tone. The pilot will decrease the VSB SNR by a few tenths of a dB, as noted later. The equalization length (complexity) requirements for these two, in the presence of multi-path (echo) distortion, are virtually the same, as is noted later.

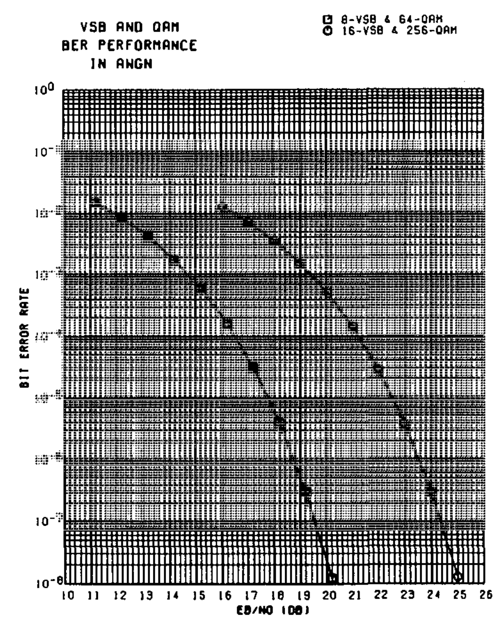


Figure 2-1. Ideal BER Performance for QAM and VSB

2.2 Generation of QAM and VSB Waveforms

In this section, the general methods for generating the two candidate waveforms are discussed. In order to make this discussion relevant to digital television transmission we will use in these discussions 64-QAM and 8-VSB signal types each operating at a gross bit rate of 32.4 Mbps — a nominal rate for digital TV transmission. With appropriate baseband filtering, either of these signals will fit within the 6 MHz channel allocation. (These same discussions can also apply to 256-QAM and 16-VSB — the number of amplitude levels and gross bit rates increase by a factor of 2.)

The block diagram of a typical 64-QAM generator is shown in Figure 2-2. The input bit stream (at 32.4 Mbps) is first multiplexed into two parallel streams of 16.2 Mbps. These two independent streams will be impressed upon the quadrature components of the QAM waveform. Each stream passes through an 8-level (3-bit) amplitude encoder. The rate of amplitude modulation at the encoder output is (16.2)/3 or 5.4 MHz. Since each stream carries 3 bits of information per level, the combined streams will convey 6 bits of data per transmitted symbol. The output data rate is then 32.4 Mbps, identical to the input rate. The amplitude encoded streams are passed through identical baseband filters in order to limit the transmitted spectrum to 6 MHz. The normal filtering used in this application has a square root, raised-cosine (SRRC) frequency response. A shaping factor of 10% will result in a transmitted bandwidth of 1.1x5.4 MHz or 5.94 MHz, just within the 6 MHz limit.

The filtered waveforms then are used to amplitude modulate two quadrature tones centered at the transmission center frequency. (The quadrature tones may actually be at an IF which is subsequently up-converted to the final transmission frequency.) The transmitted spectrum from this process is shown in Figure 2-3, with an SNR of 25 dB.

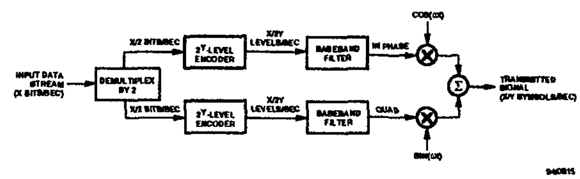


Figure 2-2. General QAM Generator and Modulator

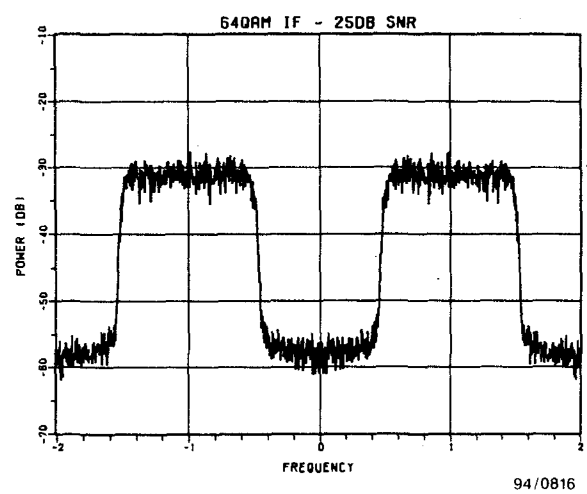


Figure 2-3. Spectrum of Transmitted 64-QAM Signal

The block diagram of the comparable 8-VSB generator is shown in Figure 2-4. In this method the input stream is first encoded into 8 levels (3 bits per level) resulting in an output symbol rate of (32.4)/3 or 10.8 MHz. This stream is next passed through a complex baseband filter producing two filtered output streams — a filtered version of the input stream and a phase-shifted (quadrature) version. The complex frequency response of this filter has a similar shape to that used in the QAM generator — in this example we will use the same 10%-shaped, SRRC response — with the filter center frequency offset from 0 Hz by one-half of the symbol rate. A small pilot tone at 0 Hz (carrier frequency) is also inserted at this point. The two streams are each carrying data at the full symbol rate of 10.8 MHz. The streams then amplitude modulate quadrature tones at the transmission frequency, which are summed together to produce the output waveform. Though the quadrature components carry full-rate data, the unique relationship between the two data streams results in a transmitted spectrum which is made up of one sideband of the quadrature components, a small vestige of the other sideband passed by the filter, and a small pilot tone at the carrier frequency. Figure 2-5 shows the transmitted spectrum of this signal. The pilot is approximately 11 dB below the signal power [6] and the SNR is 25 dB on the data portion of the signal. The pilot adds 0.3-0.4 dB to the total power. The required transmission bandwidth for 8-VSB is the same as that for 64-QAM.

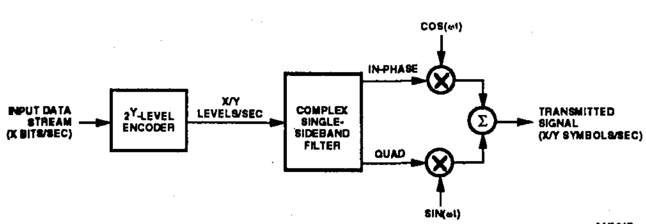


Figure 2-4. General VSB Generator and Modulator

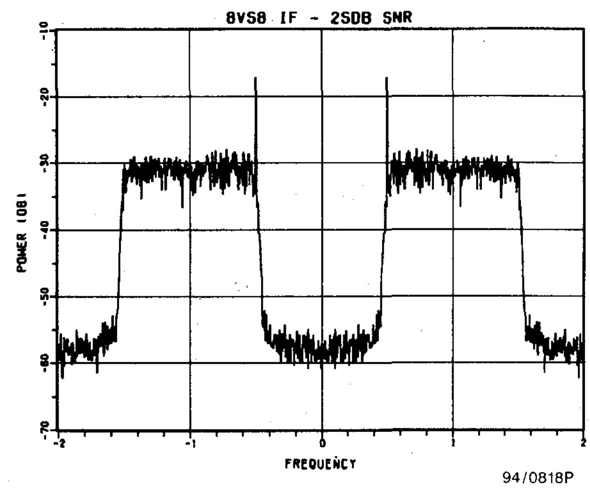


Figure 2-5. Spectrum of Transmitted 8-VSB Signal

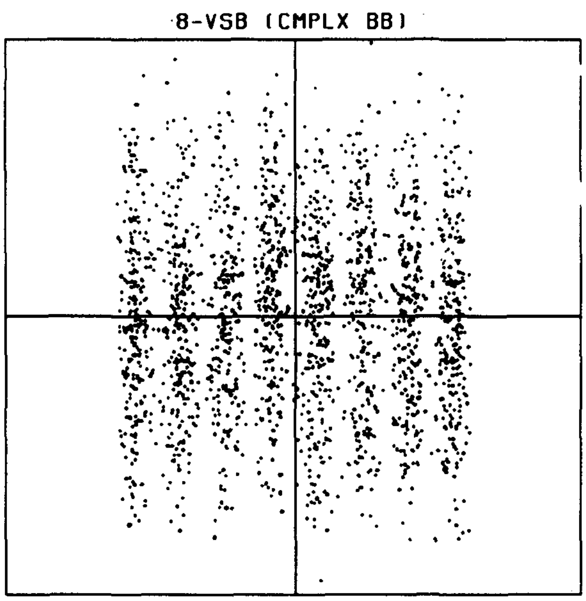
2.3 Comparison of QAM and VSB Transmissions

The two techniques as discussed have very much in common. In Section 2.1 it was shown that for the same data rates and same AWGN channel the two have identical error rate performances ignoring the VSB power associated with the pilot. The baseband filtering used on each is very similar and can have identical frequency responses as shown in Section 2.2. Both transmitted waveforms consist of a carrier with quadrature amplitude modulation (for QAM the quadrature modulating data is independent, for VSB the data is identical but is a phase-shifted version of itself in quadrature). The required transmission bandwidths are the same for both modulations; any channel distortion will be observed to have a similar effect on the spectra of both signals.

The equalization complexity requirements are identical for the two modulations. QAM requires a complex structure (equivalent to 4 parallel real filters) while VSB requires only a single structure. However, for equivalent performance the equalizers must have the same duration impulse response. The VSB structure must run at twice the clock rate of the QAM, since VSB has twice the symbol rate. Therefore the VSB equalizer must have twice as many taps to achieve the same impulse response duration. For comparable equalization, the same number of multiply-accumulates per second are required of each structure.

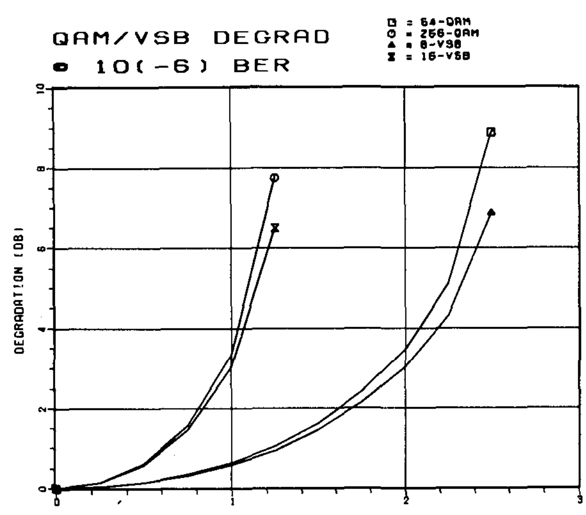
Differences in the two techniques reside generally in the processing required for demodulation and in particular in the requirement for VSB to recover the carrier prior to equalization. A strong reliance is placed upon accurate carrier recovery in order to achieve high performance from VSB demodulation. Unlike QAM, the VSB equalizer cannot work in the presence of a non-zero carrier frequency, and small carrier phase errors or biases can cause performance degradations in the equalized signal. The use of training sequences in the data enable the demodulator to compensate for carrier phase errors and to assist the equalizer in initial adaptation. Figure 2-6 shows a complex baseband VSB constellation with distortion due to the transmitter baseband filtering. Since the received constellation possesses no angular symmetry it cannot be used for carrier acquisition; the carrier offset frequency is recovered from the received pilot. Reference [6] presents an architecture for performing some carrier phase correction following the equalizer; this, naturally, adds to the demodulator complexity.

QAM demodulation performance is also highly sensitive to received carrier phase. In fact, the performance loss resulting from carrier phase errors is virtually the same for both modulation types. Figure 2-7 shows the performance losses for QAM and VSB modulations resulting from carrier phase errors. However, the QAM demodulator is able to deal more effectively with system-induced carrier frequency and phase offsets because of the way the carrier is recovered in the demodulation process. The equalizer can begin the adaptation process in the presence of carrier frequency offsets. Figure 2-8 shows a noiseless 64-QAM constellation with distortion due to baseband filtering. The 4-quadrant symmetry of the 64-QAM constellation allows blind adaptation techniques to be applied during signal acquisition. Full carrier phase information is retained in the data at the equalizer output where noise and distortion has been largely removed from the signal. This allows more efficient carrier tracking to be used and thereby to very effectively track out carrier phase jitter. In addition, the complex equalizer is able to correct for any phase biases introduced in the tuner/demodulator hardware.



94/0821P

Figure 2-6. 8-VSB Quadrature (Complex) Baseband Constellation



94/0822P

Figure 2-7. QAM and VSB Performance Degradation from Carrier Phase Errors

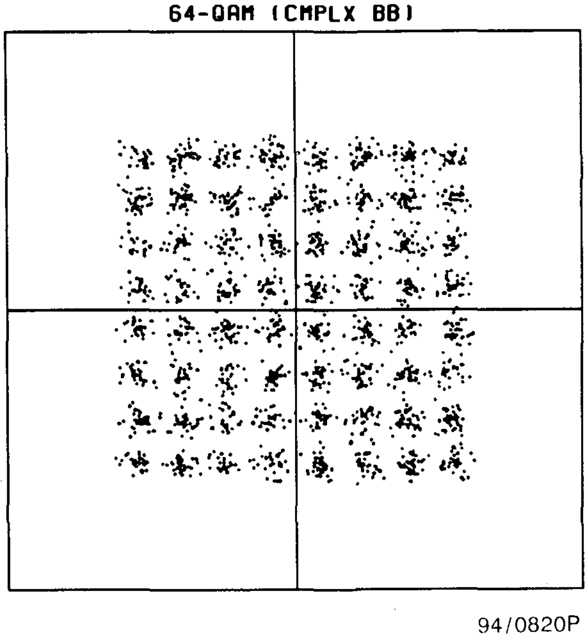


Figure 2-8. 64-QAM Quadrature (Complex) Baseband Constellation

3.0 CABLE IMPAIRMENTS

There are many impairments which need consideration when designing a QAM equalizer/demodulator for Cable TV. A brief description of some of these impairments is listed below.

3.1 Random Noise

Random or Gaussian noise is primarily added into the system via the trunk, bridge, and line extender amplifiers in a cable system. The figure of merit for these amplifiers is the noise figure, which is defined as the total measured output noise level of the amplifier minus the sum of the thermal noise and the amplifier gain. A cascade of identical trunk amplifiers with gain G dB is typically set up so that the cable between them attenuates the signal by –G dB. In this manner the noise level is increased solely by each amplifier's noise figure contribution. The noise figure of a cascade of identical amplifiers is then given as:

$$NF_n = NF_1 + 10 \log n$$

Thus, for example, every time the number of cascaded amplifiers is doubled, the noise level is increased by 3 dB. Increasing the random noise on a QAM signal has the effect of making each constellation point look more 'cloudy' or 'fuzzy.' Eventually, the amount of noise will begin to cause constellation points to cross the decision regions between points, causing errors or an increase in Bit Error Rate.

3.2 Composite Second Order (CSO) and Composite Triple Beat (CTB)

Of these two, CTB is typically the most problematic. Because of nonlinear characteristics in the trunk, bridger and line extender amplifier transfer characteristics, some mixing of the various video carriers occurs. Distortion products of the form F1 ± F2 ± F3 are called triple beats. F1, F2, and F3 are video modulated carriers. Thus, the more channels (or the more bandwidth) the system has, the more triple beats your system will incur. Triple Beats from different video carriers may fall at the same spectral location, adding with different phases and building upon one another to cause Composite Triple Beats. Second Order distortions are caused by the same mechanism, however push-pull amplifiers have overcome this problem, making it a non-issue. A very comprehensive study of the CTB/CSO problem was published in Reference [5]. The conclusion was that the CTB/CSO interference would cause little degradation (less than 1 dB) as long as the digital carriers were transmitted 8-10 dB below the analog carriers.

3.3 Microreflections or Multipath

Microreflections are caused by impedance mismatches in cable systems. If the terminating impedance is different than the characteristic impedance of the line, a portion of the incident wave is reflected back towards the source. The portion of the incident wave reflected is dependant upon the difference between the terminating impedance and the characteristic impedance, measured in return loss. If the incident signal ray arrives at the receive end accompanied by a delayed, attenuated version of itself the receiver must be robust enough to eliminate this distortion. Reflections can cause scalloping in the signal spectrum which the adaptive equalizer must equalize out by building a filter whose frequency response enhances the notches in the input spectrum, thus creating the inverse of the channel frequency response. Reflections between subscribers on separate taps are typically not a problem since the tap to tap isolation is usually around 30 dB. An example of a scenario where microreflections start to become a problem might be when the signal on one tap is connected to a splitter with low port to port isolation and sent to a separate TV with a relatively short run of cable. In this case, subscriber 1 may be watching a particular channel in one location while a viewer 2 may change to this same channel. In this case, the equalizer at subscriber 1 must respond to the change in channel conditions (i.e., level of reflection from the viewer 2 TV input port) that occurred during the step in return loss that occurred when the channel was changed. In Sections 3.8 through 3.10, results of actual and simulated equalizers on realistic microreflection environments are presented.

3.4 AM Hum

The trunk, bridger and line extender amplifiers are usually powered by a 60V, 60 Hz quasi-rectangular AC supply. This quasi-rectangular waveform can amplitude modulate the video signal within the amplifier. Although a single amplifier does not contribute significantly to hum, a long cascade of amplifiers can cause a buildup of hum to the point where it can appreciably distort the video signal. Other sources of hum can come from a wide variety of sources at the subscriber site. For example, if the cable carrying the video is in close proximity to fluorescent lights, a significant amount of AM hum can be imposed on the signal.

AM hum affects a QAM signal by elongating the constellation points in the radial direction. The elongation is proportional to the amount of AM hum present. As the elongation extends closer to the constellation point decision regions, an increase in the BER occurs.

3.5 Phase Noise

Phase noise can be thought of as random phase fluctuations imposed on the video carrier. In Cable TV systems, phase noise is imposed on the carrier by the LOs in upconverters, AML links, head-end modulators, and set top converter tuners. Phase noise forces the constellation points of a QAM signal to traverse a circular arc whose length is proportional to the amount of phase noise present. As the arc length extends closer to the constellation point decision regions, an increase in the BER occurs.

3.6 Residual FM

Low cost power supplies in set top converters which supply VCOs are the prime contributors to residual 120 Hz FM. Other residual FM sources include upconverters and modulators, however these are usually less significant contributors. The effect of the residual FM is to cause the signal frequency sweep between ±Fmax (the maximum frequency deviation) at 120 Hz and harmonics rates. The amount of deviation, Fmax, varies with tuners and can be from several kHz up to 50-60 kHz or more for the most inexpensive tuners.

3.7 Lab Tests

A series of in-house tests were performed on the Applied Signal Technology 256-QAM demodulator prototype to determine the viability of using this design as the basis for a VLSI implementation for the Cable TV industry. The tests were selected from a CableLabs report [7]. Only a subset of the tests were performed, which characterized the demodulator in the presence of:

1. Random Noise Interference
2. 120 Hz AM Hum
3. Residual FM
4. Residual FM and Noise
5. Channel Change
6. CW Interference

The results were preliminary in the sense that Applied Signal Technology intended to duplicate these tests at CableLabs in the near future. The following tests were not performed.

1. Composite Second Order (CW) Interference
2. Composite Third Order (CW) Interference
3. Composite Third Order (ATV)
4. CTB (CW) and Noise

These tests would be characterized later at the Advanced Television Test Center [8]. The QAM demodulator employed a 64-tap equalizer throughout the tests.

3.7.1 Random Noise Interference

3.7.1.1 Summary of Test Method

An HP3708 Noise and Interference Test Set was used to add Gaussian noise to the signal. The noise level was varied until an uncoded BER of approximately 10^{-4} was obtained (accounting for the effects of a simple error coding scheme). The noise power in a 5 MHz bandwidth was recorded.

3.7.1.2 Test Results

(Note that no coding gain has been applied. These are raw BER measurements.)

256-QAM: SNR = 31.4 dB
(~1.4 dB loss from theory)

3.7.2 AM Hum

3.7.2.1 Summary of Test Method

The HP 8782A Vector Signal Generator was used to produce the 256-QAM signal. The generator AM port was driven by a signal generator. The QAM signal was amplitude modulated using both a triangular and squarewave 120 Hz signal. The level of the modulation was varied to obtain a BER of approximately 10^{-4}.

3.7.2.2 Test Results

Hum Modulation:
256-QAM: 5.9%
(triangular modulation source)
4.2%
(squarewave modulation source)

3.7.3 Residual FM

3.7.3.1 Summary of Test Method

The HP 8780A Vector Signal Generator was used to produce the 256-QAM signal. This generator has an FM port which was driven by a signal generator. The QAM signal was then frequency modulated using a 120 Hz sinewave. The modulation was increased to obtain a BER of approximately 10^{-4}.

3.7.3.2 Test Results

Residual FM 256-QAM: 300 kHz

This result implies that low cost tuners can be used with 256-QAM.

3.7.4 Channel Change

3.7.4.1 Summary of Test Method

This test was performed using two different methods. The first method involved changing the channel from channel 13 to channel 12 on the set top converter and filming the constellation display using a camcorder. By counting frames and knowing the frame time, one could determine how long it took for the equalizer to acquire the new channel. The second method involved using a digital storage scope and triggering with a TTL signal from the set top box which indicated a channel change. The 'Error Output' on the back of the Anritsu BER Receiver was also monitored on the scope. The time difference between start of trigger and the Error Output signal reaching 0 errors was taken to be the channel change time.

3.7.4.2 Test Results

~ 0.5 seconds (AGC, carrier, and equalizer lock)

3.7.5 FM Interference and Noise

3.7.5.1 Summary of Test Method

The residual FM deviation is varied over the course of the test. At each value of FM deviation random noise in a 5MHz bandwidth is added until a BER of 10^{-4} is achieved. The Carrier to Noise ratio (C/N) is then recorded for each level of FM impairment.

3.7.5.2 Test Results

256-QAM	
Res. FM (kHz)	C/N (dB)
10.0	31.4
20.0	31.6
37.0	31.7
100.0	32.3

64-QAM	
Res. FM (kHz)	C/N (dB)
500	28
600	30 BER=10^{-5}
700	32 BER=10^{-5}

3.7.6 CW Carrier Interference

3.7.6.1 Summary of Test Method

A CW interference carrier was inserted at 0.5 MHz offset from the center of the 6 MHz band. The level of the interferer was raised until a BER of 10^{-4} was achieved.

3.7.6.2 Test Results

256QAM: CW C/I =26.9 dB (0.5 MHz from center of band)

3.7.7 Performance With a Low-Cost Tuner

In addition to these tests, the BER performance with standard off-the-shelf consumer grade tuners was evaluated. It could be hypothesized that the integrity of the IF filtering and the phase noise of future LOs in QAM tuners would have to be improved due to the complexity of the digital signals. This translates into a higher cost for the tuner. It would be advantageous from a cost standpoint if the demodulator were robust enough to work with low cost tuners.

Applied Signal Technology selected several consumer-grade tuners to determine the relative degradations imposed on a QAM signal. The SER performance (the BER curve can be derived by dividing the SER value by eight) for one of these tuners is plotted with white gaussian noise added to the signal. The results are shown in Figure 3-1. The 256-QAM test curve represents the results of sending a 256-QAM signal through the set-top box tuner and a 256-QAM prototype demodulator. The uncoded BER results are encouraging considering the coding gains that can be realized using Forward Error Correction.

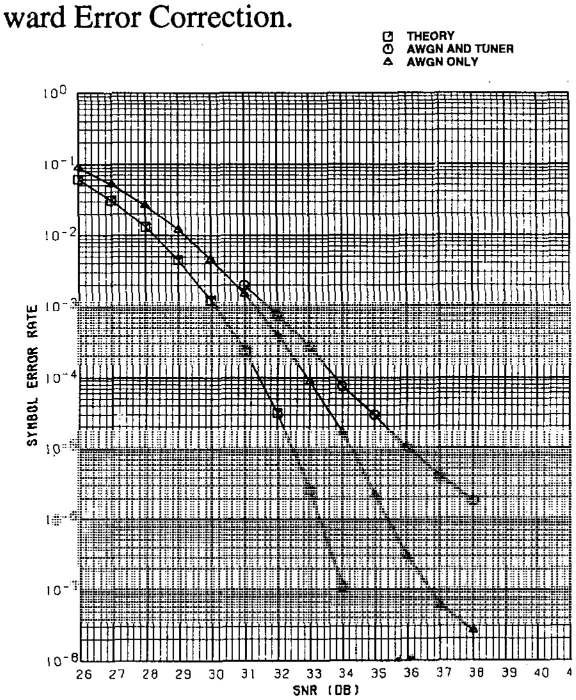


Figure 3-1. Uncoded 256-QAM Prototype Performance With and Without Low-Cost Tuner Distortion

3.8 Microreflection Degradations and Equalizer Performance

Computer simulations were performed to estimate the performance of a linear adaptive equalizer for 64- and 256-QAM for a variety of cable channels. Based on cable models of [9], microreflection models were derived for the trunk, distribution, and subscriber cable channels. In addition, models of some of the cable channels which were used in the recent CableLabs tests [8] were also evaluated.

The block diagram of the system which was simulated is shown in Figure 3-2. The baud rate was set to 5.287 Mbaud. The transmit filter which was used in the simulations was based on an actual transmit SAW filter.

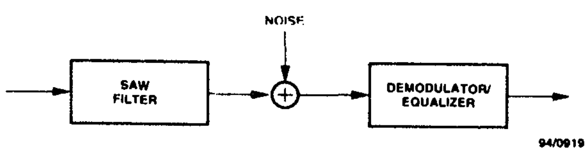


Figure 3-2. Simulated System

For each of the test channels, the optimal minimum-mean-square error equalizer coefficients were computed and the resulting error variance was determined as a function of input SNR and equalizer length. The error variance values were used to compute the symbol error rates for both 64- and 256-QAM modulations. The performance of T/2 spaced equalizers with 8, 16, 32 and 64 taps were compared to determine the minimum equalizer length required to produce performance close to ideal.

The reflections that occur within cable systems will now be evaluated. These reflections result in Inter-Symbol Interference (ISI) on the signal and can result in bit errors at the demodulator. Reflections will be evaluated at three different locations within a cable system: the Trunk environment, the Distribution environment, and the Subscriber environment. Each environment has been studied and simulated to determine the amount of equalization needed for the different cases.

3.8.1 Trunk Environment

The Trunk environment is a series of amplifiers which transmit the cable signal from a head-end or fiber node to a particular neighborhood. A typical trunk system has approximately 20 amplifiers that are spaced by 22 dB (2472 feet at 300 MHz) [10]. Figure 3-3 shows a typical microreflection scenario between amplifiers #2 and #3.

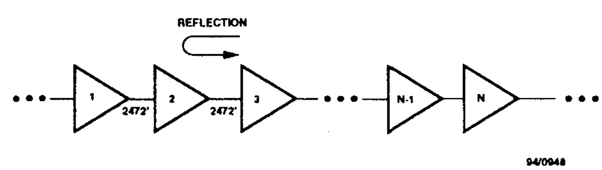


Figure 3-3. Trunk System

The power of the reflection (in dB) at the Nth stage is $10 \log (N * 10^{R_a/10}) + L_a$, where N is the number of amplifiers, R_a is the return loss of the amplifiers in dB, and L_a is the loss due to the attenuation of the line. Reflections in the trunks are characterized by relatively low amplitude reflected signals that are delayed in time on the order of microseconds.

Typical reflected powers with associated delays are:

Table 3-1. Reflected Power and Delay versus Frequency to the Trunk Channel

f (MHz)	Power of Reflection (dB)	Delay (μs)
55	–36.8	5.69
300	–63.0	5.69
450	–74.4	5.69
550	–82.8	5.69

From these results, it is clear that for the higher cable frequencies where digital signals will operate, the trunk environment results in essentially no degradation due to microreflections.

3.8.2 Distribution System

The distribution system is characterized by N equally spaced taps between line extenders, with each tap system assumed to have 5 unterminated taps. Reflections result from single reflections from nearby taps, double reflections from adjacent taps, and coupling between adjacent drops. See Figure 3-4. For analysis purposes, a software algorithm was used to generate the channel characteristics modeled in Reference [10].

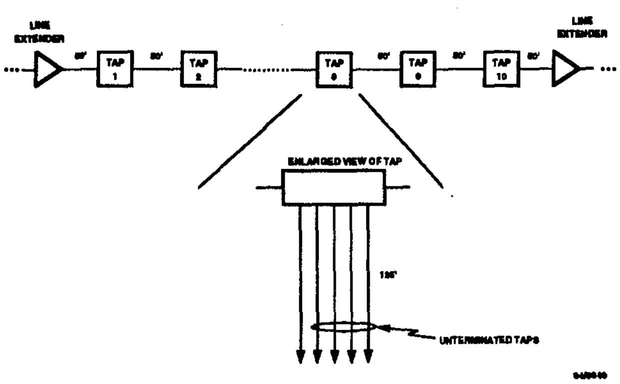


Figure 3-4. Distribution System

The assumptions placed on the channel [10] are shown in Table 3-2.

Table 3-2. Distribution Channel Parameters [10]

Number of Taps	10
Number of Drops per Tap	5
Line Loss (Taps & Drops)	0.46 dB/100ft and 1.55 dB/100ft
Delay	1.15 ns/ft
Tap Return Loss (Both sides)	20 dB
Line Extender Return Loss	16 dB
Distance between Taps	50 ft
Insertion Loss of Taps	~.8 dB
Tap Value	~21 dB
Tap to Output Isolation	25 dB
Tap to Tap Isolation	25 dB
Drop Length	125 ft
Return Loss of Drops	0 dB
Reflection of Drop	20 dB

The corresponding reflections at 55 MHz are illustrated in Figure 3-5a) and b).

The channel reflections are higher in the trunk environment than in the distribution system, but they are delayed by less time.

3.8.3 Simulation Results

Figure 3-6 illustrates the equalizer performance for the distribution system only for 64-QAM and 256-QAM, respectively. For 64-QAM, the 16-tap equalizer produces about 1 dB of degradation from ideal performance. For 256-QAM, the 16-tap equalizer suffers several dB of degradation at an uncoded symbol error rate (SER) = 1 x 10^{-6} and the 32-tap equalizer is about 1 dB from ideal. The 64-tap equalizer tracks the ideal curve with about 0.5 dB of loss.

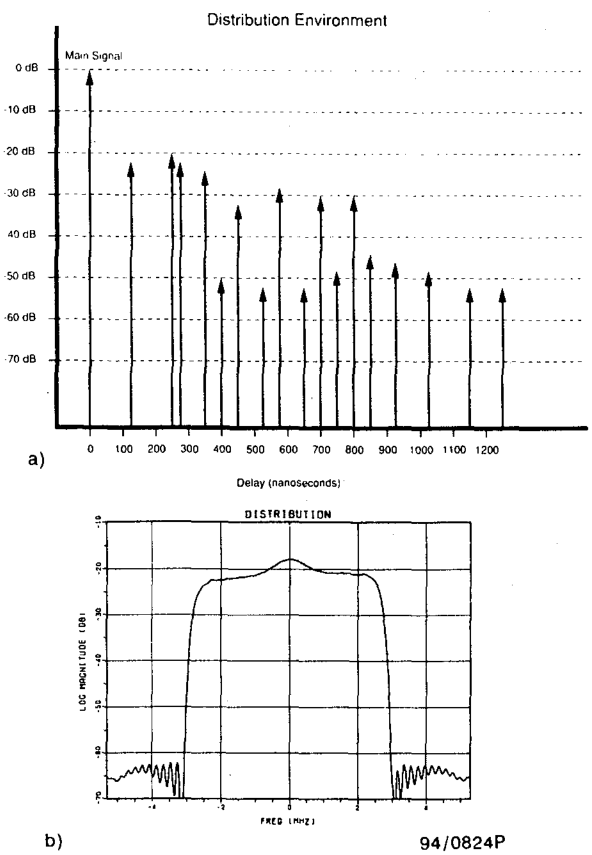


Figure 3-5. Distribution Environment Impulse and Frequency Response

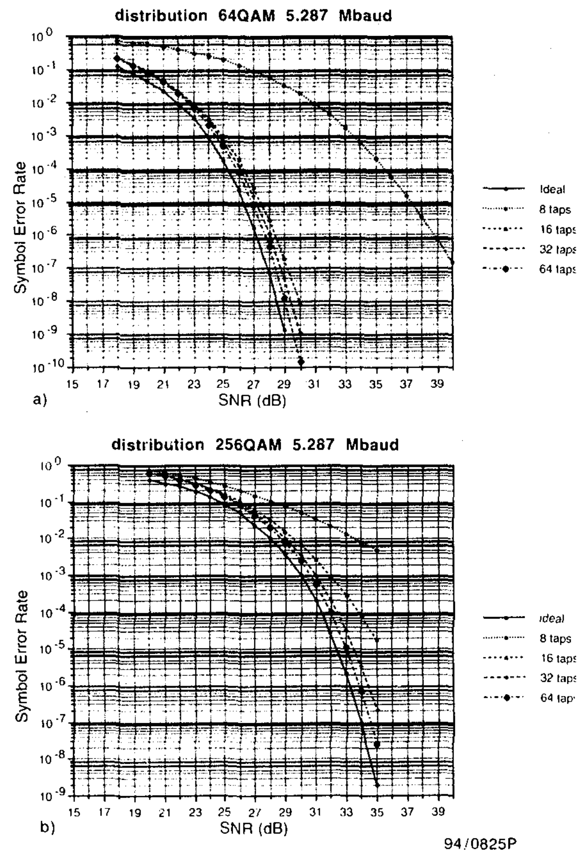


Figure 3-6. Equalizer Performance for Distribution Environments 64-QAM and 256-QAM

3.8.4 Subscriber

The subscriber environment is the reflections that take place once the cable signal enters an end-user's home. The different configurations are numerous, and a few reasonable scenarios were examined.

Two different channels were simulated. The first had splitter cable lengths of 5 feet, the length of typical cable in the living room and is illustrated in Figure 3-7. The second configuration used cable lengths of 50 feet, which is typical when TVs and VCRs, etc., are located in different rooms. The assumptions are shown in Table 3-3.

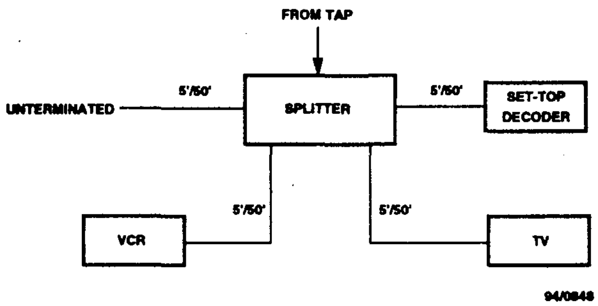


Figure 3-7. Subscriber System

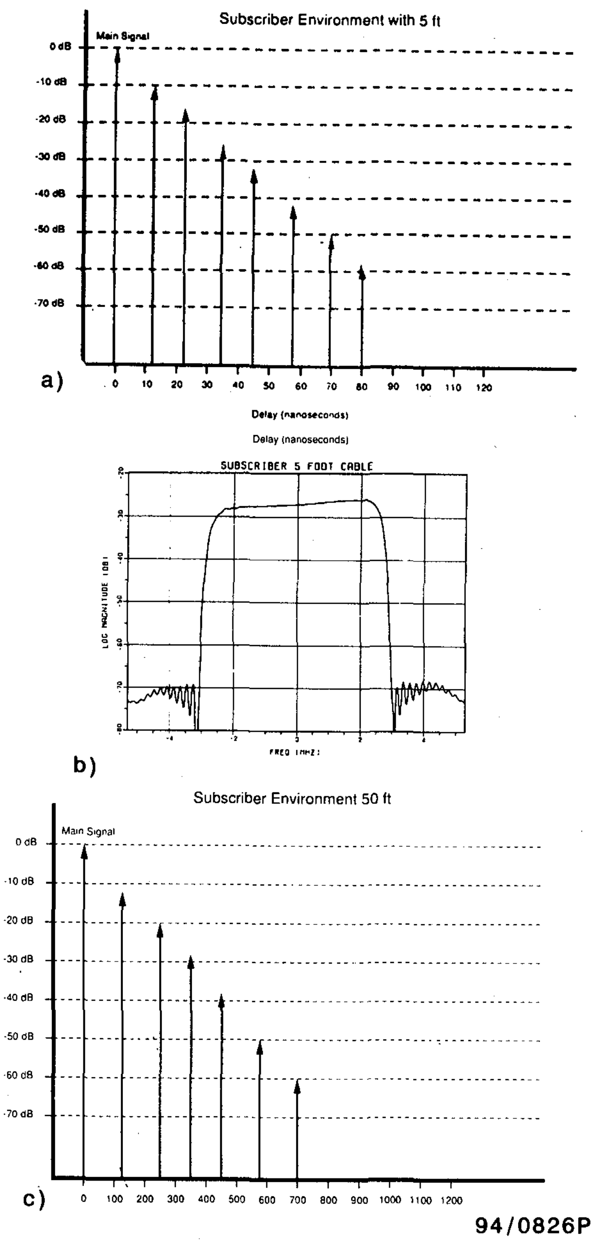


Figure 3-8. Subscriber Environment Models

Table 3-3. Subscriber Channel Parameters

Frequency	55 MHz
Number of Ports	4
Length of Cable on Ports	5 ft (case 1) 50 ft (case 2)
Termination Reflection Coefficient	0 dB
Port Reflection Coefficient	8 dB
Line Loss	1.9 dB/100ft
Delay	1.15 nS/ft
Isolation between Ports	15 dB

The corresponding 5-foot cable (case 1) reflections are shown in Figure 3-8a) and b), impulse response model and power spectra, respectively, and the 50-foot cable (case 2) impulse response model in Figure 3-8c).

Figure 3-9a) and b) illustrate the performance for the subscriber channel with 50-foot splitter cables for 64- and 256-QAM, respectively. For 64-QAM, the equalizer performance is within two dB of ideal for the 16-tap equalizer and within one dB of ideal for 32- and 64-tap equalizers. For 256-QAM, it is apparent that the 16-tap equalizer diverges from the ideal curve by a considerable amount; the 32-tap equalizer is within 1.5 dB of ideal while the 64-tap equalizer is within one dB of ideal.

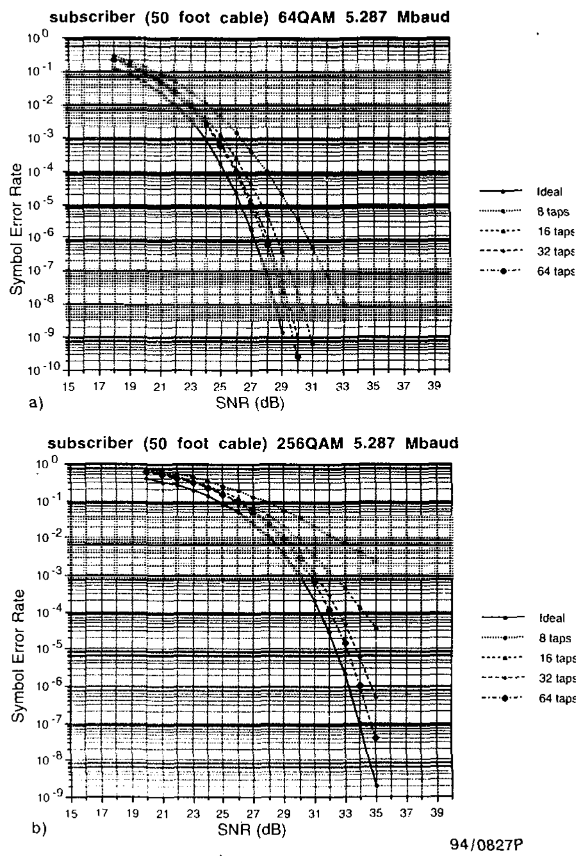


Figure 3-9. Subscriber Performance (50-Foot Cable)

Figure 3-10a) and b) illustrate the performance for the subscriber channel with 5-foot splitter cables for 64- and 256-QAM, respectively. For 64-QAM, the equalizer performance is within one dB of ideal for 16-, 32- and 64-tap equalizers. In particular, the 32- and 64-tap equalizers track the ideal curve very closely. For 256-QAM, it is apparent that the 16-tap equalizer diverges from the ideal curve by 2 dB at a SER of 10^{-6} while the 32-tap equalizer is within one dB of ideal. The 64-tap equalizer tracks the ideal curve very closely.

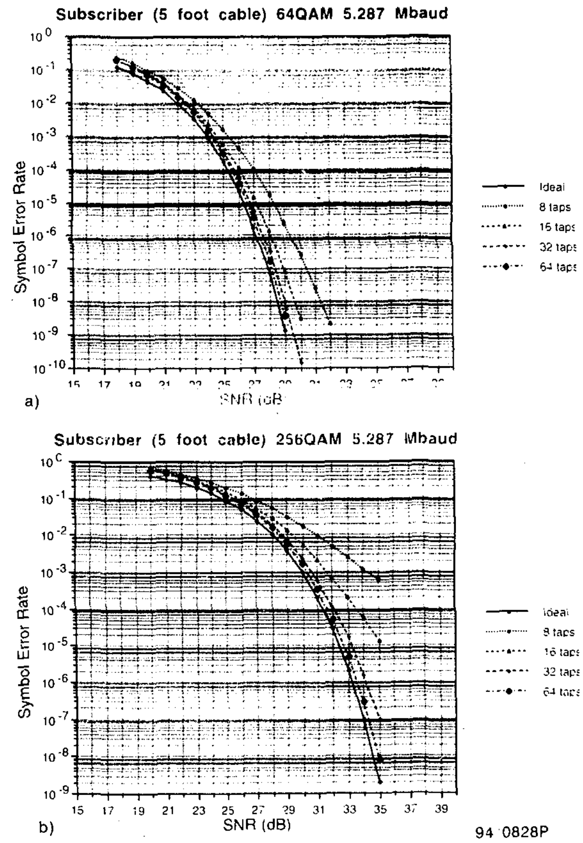


Figure 3-10. Subscriber Performance (5-Foot Cable)

3.8.5 Test Channels

Five of the channels which were used in the recent CableLabs tests [8] were also analyzed. Three channels were derived from Section 3.16 of the CableLabs tests which were used to measure the TV channel change acquisition time for the equalizer-demodulator. The parameters for CableLabs channels are shown in Table 3-4. The frequency response of these channels are illustrated in Figures 3-11 through 3-13.

Table 3-4. CableLabs 1 Channel Parameters

	Delay (ns)	Power (dB)
CableLabs 1	300	–18
CableLabs 2	2500	–20
CableLabs 3	600	–20

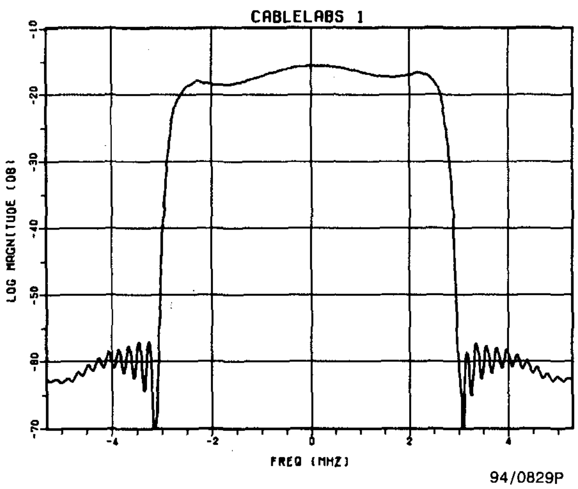


Figure 3-11. CableLabs 1 Channel Frequency Response

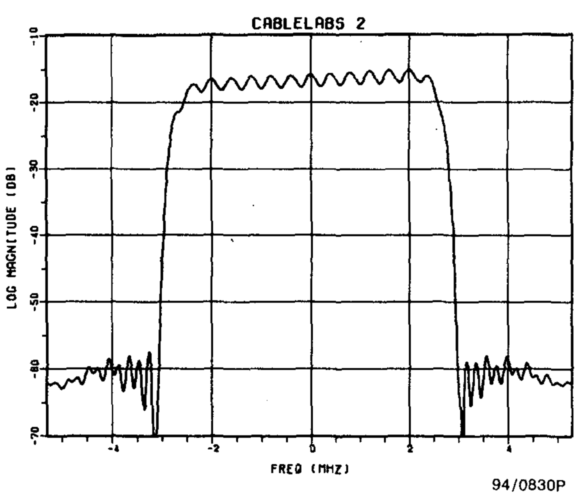


Figure 3-12. CableLabs 2 Channel Frequency Response

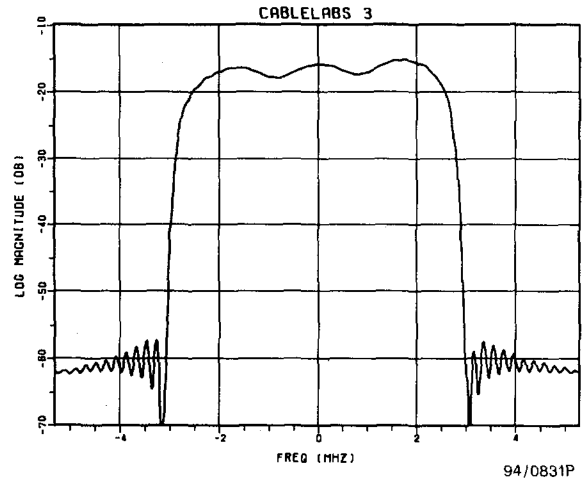


Figure 3-13. CableLabs 3 Channel Frequency Response

A strong reflection channel with reflection only –13 dBc was modeled in Figure 3-14 and a CableLabs recommended microreflection ensemble for general cable environments is shown in Figure 3-15. Parameters for the strong reflection model are given in Table 3-5 and for the microreflection ensemble in Table 3-6.

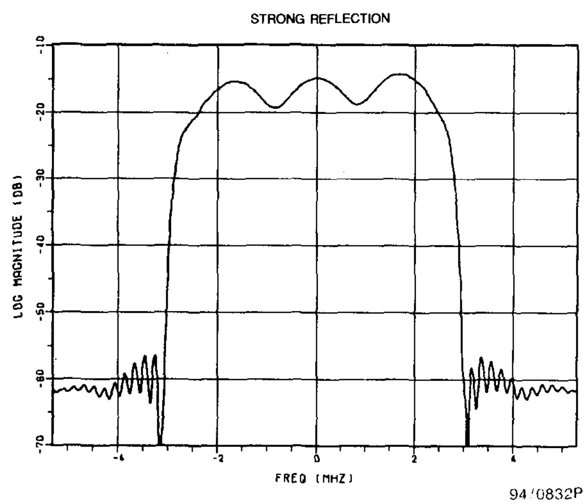


Figure 3-14. Strong Reflection Channel Frequency Response

Table 3-5. Strong Reflection Channel Parameters

Delay (ns)	Power (dB)
600	–13

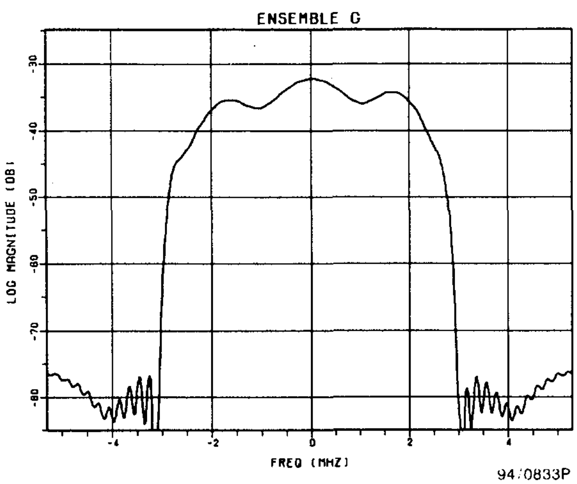


Figure 3-15. Microreflection Ensemble Channel Frequency Response

Table 3-6. Microreflection Ensemble Channel Parameters

Delay (ns)	Power (dB)
0	0
–200	–19
80	–22
150	–17
300	–22
600	–19

The SER performance vs. SNR curves for the CableLabs 1 channel for 64- and 256-QAM are shown in Figure 3-16 a) and b). It is apparent that for 64-QAM, the 16-, 32- and 64-tap equalizers produce less than 1 dB of degradation with respect to the ideal symbol error rate performance. For 256-QAM, the 16-tap equalizer produces about 2.5 dB of loss with respect to ideal and the 32-tap equalizer produces about one dB of loss. The 64-tap equalizer tracks the ideal curve fairly closely.

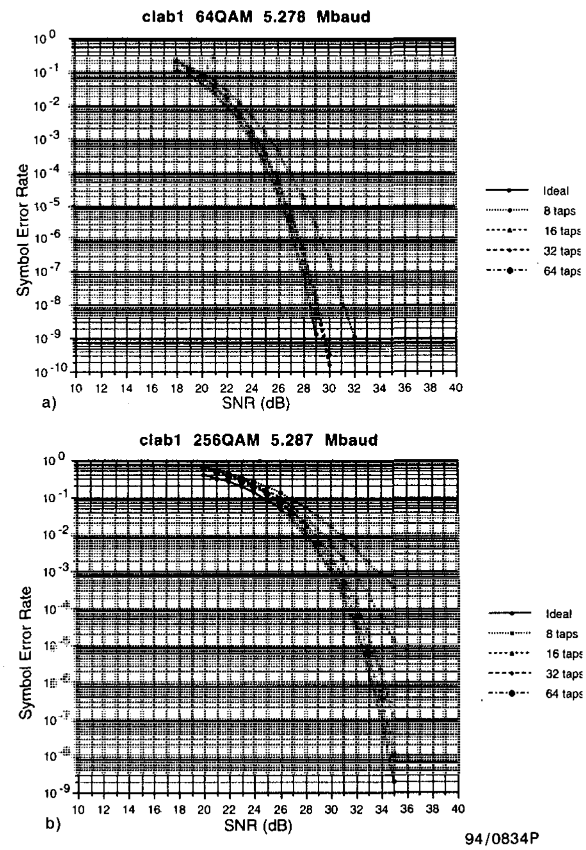


Figure 3-16. CableLabs Channel 1 Equalizer Performance a) 64-QAM b) 256-QAM

The SER performance vs. SNR curves for the CableLabs 2 channel are shown in Figure 3-17a) and b). Because of the long 2.5 microsecond delay of this channel, it is apparent that a 64-tap equalizer is required to produce performance close to ideal for 64-QAM. However, for 256-QAM the 64-tap equalizer produces about 2 dB of loss with respect to ideal; a longer equalizer or more coding gain may be required.

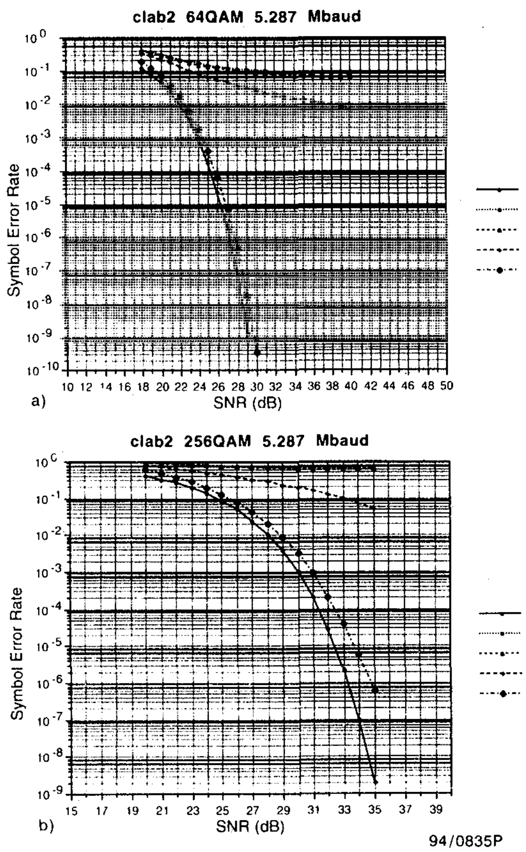


Figure 3-17. CableLabs Channel 2 Equalizer Performance a) 64-QAM b) 256-QAM

The SER performance vs. SNR curves for the CableLabs 3 channel are shown in Figure 3-18a) and b). For 64-QAM, the 16-tap equalizer diverges from the ideal curve while the 32-tap and 64-tap equalizers produce performance within one dB of ideal. For 256-QAM the 32-tap equalizer produces about one dB of loss at a SER of 10^{-6} while the 64-tap equalizer is within about 0.5 dB from ideal.

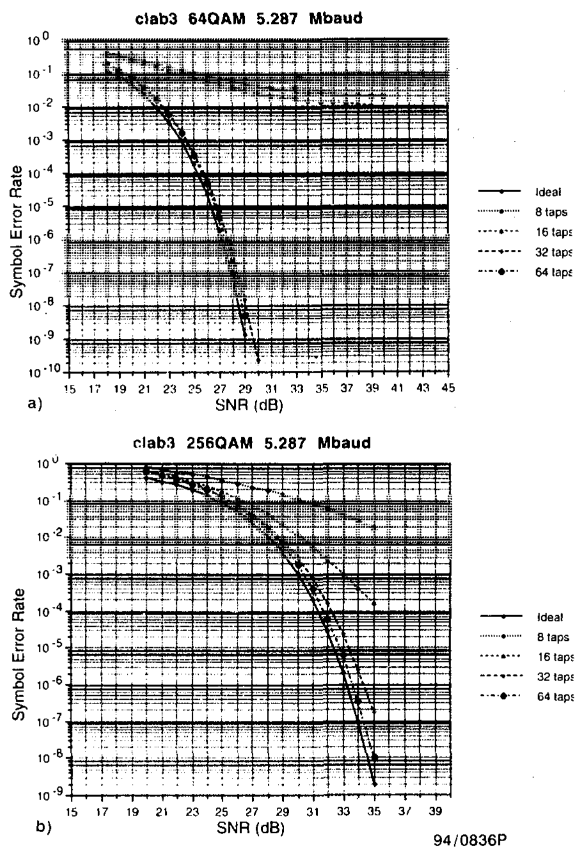


Figure 3-18. CableLabs Channel 3 Equalizer Performance a) 64-QAM b) 256-QAM

The SER performance vs. SNR curves for the strong reflection channel are shown in Figure 3-19a) and b). For 64-QAM, the 16-tap equalizer diverges from the ideal curve while the 64- and 32-tap equalizers produce performance within one dB of ideal. However, for 256-QAM the 32-tap equalizer is about 2 dB worse than ideal at a SER of 10^{-6} while the 64-tap equalizer is about one dB worse than ideal. Note that even though the echo delay for this channel is equal to that of the CableLab 3 channel, the equalizer performance with this channel is worse because the reflection is 7 dB larger.

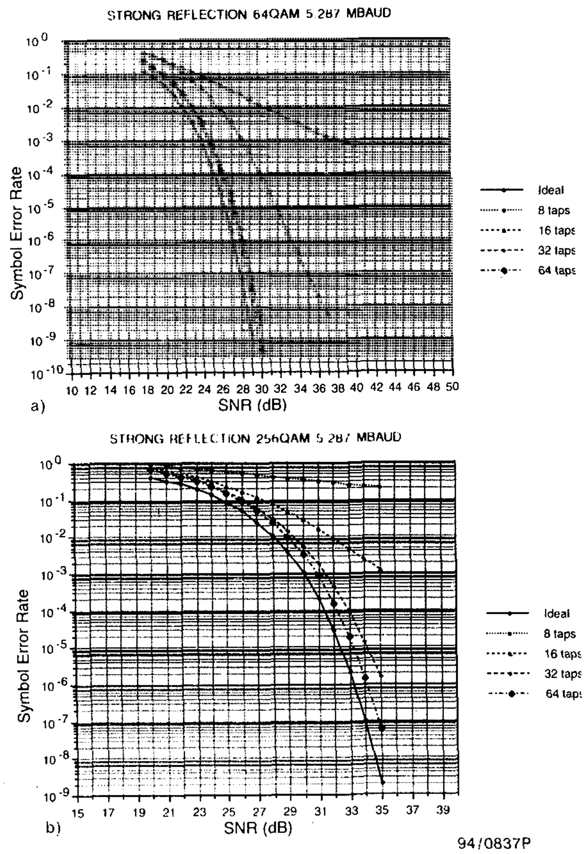


Figure 3-19. Strong Reflection Equalizer Performance a) 64-QAM b) 256-QAM

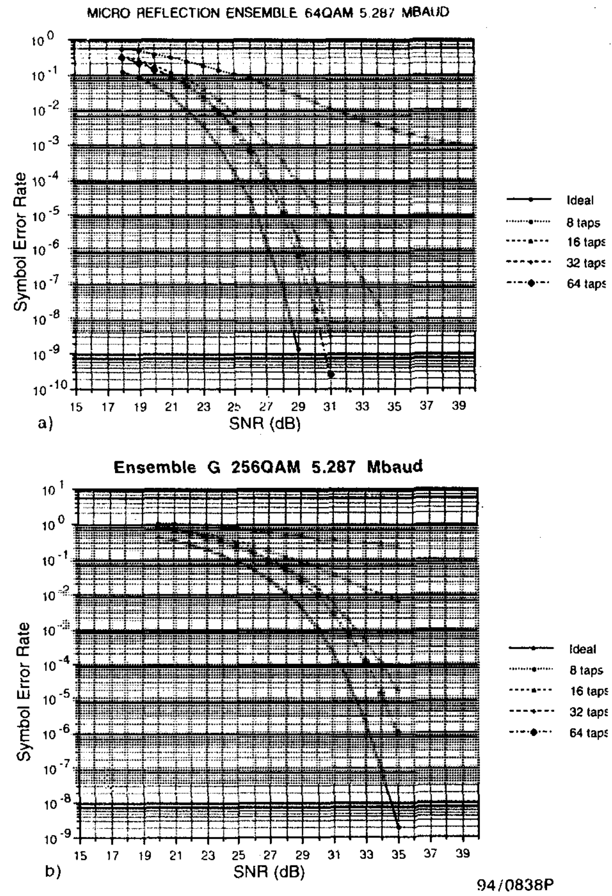


Figure 3-20. Microreflection Ensemble Equalizer Performance a) 64-QAM b) 256-QAM

The SER performance vs. SNR curves for the Microreflection Ensemble channel are shown in Figure 3-20a) and b). For 64-QAM, both the 64 and 32-tap equalizers produce about 2 dB of loss with respect to ideal. For 256-QAM the 32-tap equalizer diverges from ideal and the 64-tap equalizer is about two dB worse than ideal.

3.8.6 Summary

The performance of a linear T/2 spaced equalizer was evaluated for a variety of cable channel models for 64- and 256-QAM modulations. The simulations indicate that for a 10^{-6} uncoded BER, 64-tap equalizer is sufficient to produce performance within one dB of ideal for both modulation types for the majority of channel models. There were a couple of cases with long delays such as the CableLabs 2 and Microreflection Ensemble channels where a longer equalizer is desirable to improve performance. However, for reasonably coded systems, the equalizer lengths of 16-32 taps appear to be sufficient.

3.9 ATTC Testing

In January 1994, a series of tests were conducted by CableLabs, Inc. at the Advanced Television Test Center as part of the evaluation of HDTV transmission techniques. The testing was extensive and comprehensive and included simulations of a number of realistic cable impairments. The results of these tests serve to support the conclusions in Section 2.0 regarding the comparative performance of QAM and VSB techniques. Performance tests of 256-QAM and 16-VSB demodulators were performed in the presence of phase noise, residual FM, carrier frequency offsets, and channel switching. The QAM performance was found to be comparable to or better than the VSB. These results could be attributed to features in the QAM demodulator architecture (as mentioned in Section 2.0) such as blind equalizer acquisition and robust carrier tracking on data only.

4.0 SUMMARY AND CONCLUSIONS

The 256-QAM modulation signal has been suggested as a reasonable capacity maximizing approach when cost, SNR, and various cable environment impairments are considered. Blind QAM equalization and carrier recovery used in the prototype 256-QAM demodulator was shown conceptually and in lab tests [8] (also compared to results of [7]) to offer equal or better performance than VSB systems without the need for an equalizer training sequence or pilot tone.

A BER of 10^{-4} was achieved in the presence of 120 Hz residual FM distortion consisting of a 100 kHz peak to peak deviation and a C/N of 32 dB. For 256-QAM, the prototype has been shown to track out the phase noise and residual FM inherent in an off-the-shelf analog tuner. The 256-QAM demodulator can perform channel change acquisition in 0.5 seconds or less by virtue of acquiring blindly without the need of a training sequence. Carrier offsets on the order of hundreds of kHz have been shown to be within the pull-in range of the demodulator on a consistent basis.

It has also been illustrated that the minimum required equalizer length appears to be 16-32 taps for most multipath scenarios encountered. Two different subscriber scenarios were modeled for simulation and certain worst-case assumptions were made concerning such parameters as isolation, terminations and return loss of cable elements. The resultant channel was simulated with several different equalizer lengths. It was found that in both cases 32 taps were sufficient. A 32-tap equalizer also proved adequate in handling several microreflection models recommended by CableLabs including a microreflection ensemble which consists of five rays between 80 and 600nsec with relative levels ranging from –17 to –22 dB from the main signal. However, for 256-QAM it was shown that a longer equalizer with 64 taps performs closer to ideal (within 2 dB at 10^{-6}) for microreflections which are –20 dB and delayed by 2.5msec, while the 32 tap equalizer BER curve diverged considerably for the same case. There are many microreflection scenarios that must be considered when designing a QAM equalizer for the cable environment. Several typical scenarios have been presented with results which are achievable with a low cost demodulator implementation.

This paper has described the virtues of 256-QAM for transmission of high capacity digital television signals, and contrasted some of these virtues with VSB modulation. Clearly, if careful architecture and design techniques are employed, the simulated and laboratory data presented in this paper confirm that 256-QAM modulation is eminently usable in the cable television environment.

ACKNOWLEDGEMENTS

The authors would like to express their appreciation to Ken Snow, Joe Mitchell, and John Treichler for their support and review; also our appreciation to Mike Meschke and Larry Yamano for their efforts in the area of ATTC testing.

REFERENCES

[1] R. Price, "Nonlinearly Feedback Equalized PAM Versus Capacity for Noisy Filter Channels," Proc. International Communications Conference '72, June 1972.

[2] J.A.C. Bingham, "Multi-Carrier Modulation for Data Transmission: An Idea Whose Time Has Come," IEEE Communications Magazine, May 1990.

[3] R.S. Prodan, "Performance of Digital Transmission Techniques for Cable Television Systems," 1992 NCTA Technical Papers.

[4] G. David Forney, Jr. and M. Vedat Eyuboglu, "Combined Equalization and Coding Using Precoding," IEEE Communications Magazine, Dec. 1991.

[5] J. Hamilton and D. Stoneback, "The Effect of Digital Carriers on Analog CATV Distribution Systems," 1993 NCTA Technical Papers.

[6] R. Citta and R. Lee, "Practical Implementation of a 43 Mbit/sec. (8 bit/Hz) Digital Modem for Cable Television," 1993 NCTA Technical Papers.

[7] "Test of Zenith 16-VSB Transmission System," CableLabs, May 1993.

[8] CableLabs, Inc. Test Procedures for the FCC Advisory Committee for Advanced Television Service (ACATS).

[9] J.B. Waltrich, "64-QAM Transmission of Digital Data Over Cable and Alternate Media," 1993 NCTA Technical Papers.

[10] J.B. Waltrich, J.B. Glaab, M. Ryba, and M.A. Muller, "The Impact of Microreflections on Digital Transmission over Cable and Associated Media," 1992 NCTA Technical Papers.

TRUE VIDEO ON DEMAND VS. NEAR VIDEO ON DEMAND, STATISTICAL MODELING, COST, & PERFORMANCE TRADE-OFFS.

By
Winston Hodge, Hodge Computer Research
and
Chuck Milligan, Storage Technology Corporation

Abstract

Video On Demand (VOD) has the potential of giving individual television viewers nearly instant access to a wide range of recorded movies, video programs, games, information and other services. It is distinguished from more conventional TV viewing by a high degree of interactivity between the viewer and the material being viewed.

A perception exists in this industry today that each person interacting with their TV demands instantaneous response. This is called True Video On Demand (TVOD). As this paper will show, TVOD is extremely expensive when it provides for all services possible.

The alternative to TVOD is Near Video On Demand (NVOD). This paper will demonstrate that while NVOD is significantly less expensive to implement, an NVOD system can be designed so that its delays are not objectionable to the user for many applications. Procedures and strategies for concealing customer latency time will be described, along with the cost differential attendant to eliminating it.

Access to recorded material with zero access time is not physically possible. Fractional second access is possible, but would be very expensive for an unlimited menu of choices by an unlimited number of subscribers.

Clearly, the quantification of cost to provide service versus the latency time is of serious importance. But there is more to the implementation decision than cost. The psychological effects of waiting come into play. For example, is one second too long to wait? How about two seconds? How about two minutes? All things being equal, (which they are not), the shorter the service time the better.

This paper will provide a clear view of physically possible service times and the cost to provide those services *using advanced technology hierarchical storage.*

A model will be described which demonstrates how the system cost varies with viewer latency. This model will be applied separately and collectively to the video server, disk storage complex, *large terabyte robotic tape farms,* VOD selector switch, communications channel and viewer selection mechanism.

Block diagrams used in the systems analysis and simulation will be included, along with charts and graphs which will clarify the results of the analysis. The paper will conclude with recommendations for an economically viable system design.

Definition & Requirements

Video On Demand (VOD) trial systems in one form or another are currently being implemented. An understanding of the cost factors related to response time (i.e. viewer selection latency) will provide insight into the overall system costs.

Interactivity is much more than channel selection. It may be the simple ability of the viewer to decide **what** program he wants to watch, and **when** he wants to watch it. It might allow him to select from among several different endings to a movie thriller. It may allow him to take a simulated walk down a supermarket aisle **he** selects, ordering products from among those displayed. It could allow him to engage in a simulated trip through the solar system or a Mayan temple, making decisions about which planets to explore or which corridor to turn down, through the wonder of virtual reality. It could even allow him to engage in a simulated dog fight with another viewer through an interactive video game which could be offered.

The foregoing scenarios require progressively increasing levels of interactivity. The response times required of the system also vary widely between the applications. For example, when home shopping, the response time from advertisement to order placement is not critical but the navigation response from product to product is more significant.

The viewer may be more concerned about the time between selection and delivery of a new movie, but whether this time interval is fractions of seconds, seconds or even minutes may not be consequential.

A video arcade game or a virtual reality session requires much more rapid response--far beyond the capabilities of even a very large mainframe computer to service a large number of clients. For these applications, the interactivity will be supplied by downloading a program to a set-top box for execution. Given this fact, once more the initial response between ordering the game and its actual delivery falls into the same degree of urgency as the ordering of a movie.

Selection time is subject to the laws of physics. These laws place limits on what it is physically possible to achieve. By knowing where the limits are, and by understanding the cost of approaching these limits, one is in a position to make objective decisions on implementation approaches. This paper will enlighten the reader with the options currently available.

Strictly speaking, True Video On Demand (TVOD) requires instantaneous response, probably less than a second from the time a program request is made until the time the program is delivered. This has significant cost ramifications not only for the video server and video disk drives, but for the communications channel and other system elements not addressed in this paper.

Near Video On Demand (NVOD) requires only a reasonable and convenient response time from program selection to program delivery. This interval could range from seconds to a few minutes or in some cases even a few hours. During the interval, stock material (such as seen in theaters) or interactive advertising for food or other products to be delivered to homes, or music video interludes may be presented.

The system to be discussed will even allow a viewer to see new movies at reduced prices by selectively permitting advertising inserts in the subscriber's now less expensive pay per view movie. This scheme could allow several price levels, depending on the total number of minutes of commercials the viewer is willing to tolerate. This, in turn, would allow the service provider to offset the reduced customer billing with advertising revenues so earned.

The bottom line for the service provider should be: Which operating procedure, NVOD or TVOD, produces the largest revenue stream at what cost, ultimately providing the greatest return on investment? This paper will summarize these issues.

System Possibilities

In order to analyze TVOD vs. NVOD costs, it is necessary to understand the three prominent hardware implementation philosophies illustrated in Figures 1 through 3. The differences between approaches depends on a vendor's reliance on his installed hardware architectures, as well as his philosophy on whether a general or "tuned" solution is preferable.

In all the examples to be presented, it is assumed that the transmission system employs Asynchronous Transmission Method (ATM). This protocol utilizes data packets consisting of a five byte header and a 48-byte data field. The header describes the destination and the content of the information portion of the packet. It is further assumed that the appropriate storage solution is a 3 level hierarchy of disk and robotically managed tape libraries. The general solution uses standard operating system functions and software, and the more "tuned" solutions employ significantly more specialized software and firmware to manage the hierarchy.

For applications where the volume is not adequate to justify a custom or tuned design philosophy (such as for a small number of tests sites, or for concept validation where reduced non-recurring costs are important), the generalized solution as shown in Figure 1 may be preferable to a tuned solution. It is less expensive because it relies mostly on the procurement of off-the-shelf hardware and possibly off-the-shelf software. The generalized system can produce both TVOD and NVOD, but the cost of delivery is high.

In Figure 1, the term "mainframe" is intended to mean a general purpose processor running a "standard" operating system (e.g., a RS6000 running UNIX). Such mainframe system solutions are often more expensive than tuned solutions in production because a great deal of system hardware and software must be provided which is unnecessary for the specific application. Further, the mainframe data flow is designed for data processing, not data movement. Video applications require a great deal of data movement, with very little data processing.

The image processing (such as image compression and decompression) is usually performed by specialized hardware units. This is because affordable mainframes cannot handle the computational load required to deliver multiple video programs in real time.

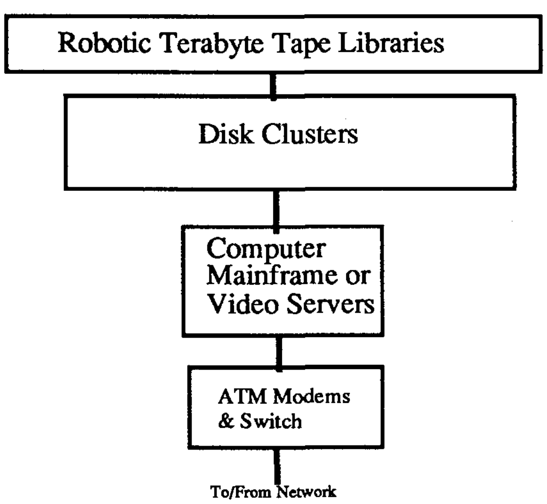


Figure 1, A Generalized Video On Demand System

When the opportunity exists to construct thousands of units for a specific application, the tuned solution is preferable because of lower cost, higher performance, superior function and just a better fit to the problem being solved.

There are various degrees of tuned systems. Some systems are very good at creating databases of still images or moving video which use general purpose operating systems, database managers, networking facilities and the like. These systems rely on small amounts of customization. They can do a good job of delivering a small number of selected videos on demand to a small customer base. As in the previous systems, they can produce either NVOD or TVOD, but the program selection is limited and the size of the client base is severely restricted when operating in TVOD mode.

These systems may be cascaded to accommodate more videos and more clients. An example of such a cascaded system is illustrated in Figure 2.

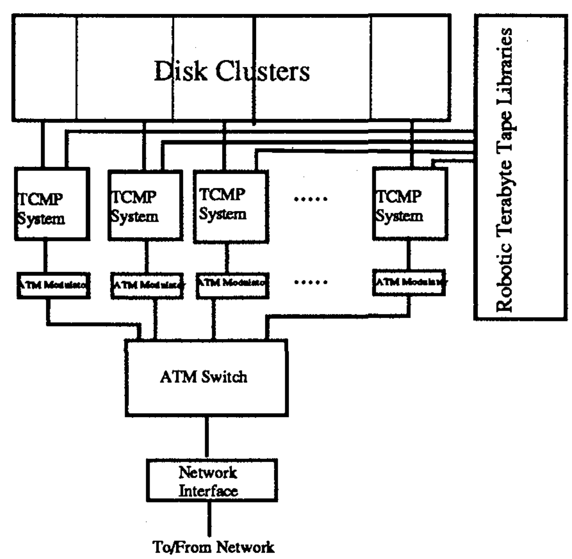


Figure 2, Cascaded Tightly Coupled Multiprocessing (TCMP) Video Server

The ultimate tuning of a video server exists when special paths are provided for moving digital video information. An architecture can be created which relaxes the throughput requirements on the computer performing the server function.

Once the server has interpreted the customer's video request, validated billing and program availability, confirmed that the requester at the customer premises is not restricted (child requesting X Rated movie), and arranged for the short term scheduling (seconds or minutes), the server computer submits the program material request and the electronic customer address to the Server Saver/ATM Switching system. Then for the balance of that transaction, the server has nothing more to do until the program is complete (for a typical movie this would be between 90 and 110 minutes). This system is shown below in figure 3.

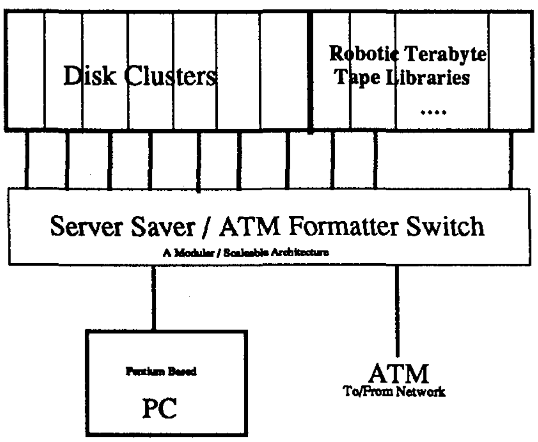


Figure 3. Composite Server Saver System / ATM Switching System

The Server Saver Sub-System permits the use of inexpensive components and simplifies data routing and manipulation while simplifying computational requirements to such an extent that a single high performance PC such as a Pentium [1] or a Power PC [2] can assume responsibility for a 500 program 10,000 subscriber system.

If a larger system is required, these systems can be cascaded to produce greater program selection for more patrons. The Server System can produce TVOD for a small number of subscribers or NVOD for a large number of subscribers, or some combination of both TVOD and NVOD. This capability is similar to that of the above systems, but **at very low relative cost.**

The Server Saver system is a simple device both architecturally and physically. It connects to a "storage farm" through multiple SCSI data paths, to the PC via one or more SCSI data paths, and to the CATV or other network through the ATM Formatter/Switch.

The Server Saver has only three types of interfaces:

(1) to/from the PC computer
(2) to/from the Storage Farm
(3) to/from the ATM network.

The Server Saver provides storage control, flow control, packet switching and an interface to the ATM network.

<u>Costs</u>

Each of the preceding systems can supply TVOD, NVOD, or some combination of NVOD and TVOD, but at substantially different costs. The cost of each of these systems varies as a function of program capacity, subscriber capacity, and the degree of responsiveness to customer requests.

It is obvious that video program capacity is a function of storage and that storage, in turn, makes up a major portion of system cost.

Each 90 to 110 minute program can require from 1 to 9 gigabytes of storage depending upon resolution requirements. Each gigabyte of disk storage will cost from $750 to $1300 at the system level, while data in robotically controlled tape systems (e.g. StorageTek Nearline offerings) will cost $7 to $10 per gigabyte of storage. There are also performance differences between disk and such tape systems. These will be discussed later in some detail.

Obviously, the more programs desired, the more storage is required, which in turn increases hardware costs.

The generalized video server systems typically cost $250,000 and upwards. Tightly coupled multiprocessing video server systems currently cost between $65,000 and $100,000 per module, each of which is capable of producing up to 25 programs concurrently.

For example, 500 channels of programming could cost (500/25 X $65,000) or $1,300,000 per video server complex, not including ATM formatting, switching or interfacing.

Each of these systems has limited capacity, requiring additional system hardware replication to yield more capacity and more responsiveness. Again, added system hardware increases system costs.

The purpose of this article is to determine for the various generic hardware approaches the costs to produce the continuum between TVOD and NVOD and how much responsiveness can an interactive TV system cost-effectively produce.

This paper will generate some approximate best case and worst case pricing for each of this trio of approaches, determine reasonable pricing intervals, and the subsequent cost relationships for TVOD and NVOD. This will facilitate the qualitative judgment as to whether, for instance, it is worth an additional $500,000 or more to give the customer a program selection response time of 1 second/minute instead of 30 seconds/ minutes.

Further, after the analysis, procedures for camouflaging program latency will be discussed.

The following spread sheet represents estimates of significant costs for each of the three prominent system architectural philosophies shown in Figures 1 to 3. While these numbers may be challenged as being tomorrow's prices, guesses or inaccurate, they do represent working approximations derived from potential vendors in this industry. It is interesting to observe that using any set of different <u>reasonable</u> numbers does not change the comparative relationship, i.e. - NVOD is much less expensive than TVOD.

This paper has alluded to video programs and threads. A thread is defined as a continuous stream of video representing one complete program, using one of the available broadcast channels. Since both tape drives and Video Friendly disks can produce data transmission rates greater than required for a single channel, it is possible to store the data in such a fashion that it can be read out multiple times in real time.

If a device is able to sustain a data rate 10 times greater than is required for normal video rates, 10 video streams or "threads" could be produced if only short duration device read interruptions occur (e.g. for turnaround at end of tape track or for head or next cylinder seeks). An alternative is for additional buffering to be used to mask longer duration read interruptions. It is possible and therefore desirable to interleave the programming material such that each thread is displaced in time.

For example, a 90 minute (1 gigabyte) video program can be structured to allow 10 threads, and would have each thread offset by 9 minutes. This can be accommodated by appropriate data structures using only one gigabyte in either tape or disk storage.

Because TVOD requires the ability to instantaneously access the first and then subsequent video frames of the program at random and arbitrary intervals, it would require that the storage device be capable of rapidly switching from one random spot to another to support even two threads, let alone a number as large as 10 or 12.

Although tape can support that many threads of NVOD, multiple thread TVOD is not feasible with tape devices, because they require seconds to move from one random spot to another.

TVOD is feasible but more expensive with disk because buffers must be included in front of each device for each thread, which substantially increases the cost per thread. More importantly, the random seeks reduce the sustainable rate of the device so that it is less efficient, and even with external buffering, can sustain significantly fewer total threads.

For example: Assume a particular disk can sustain 3 MB/sec with a maximum (because the video stream must be guaranteed) random seek time of 33 ms. If the disk is rotated at 5400 rpm, it will have approximately 33KB on a track that will spin by the head in 11 ms. (These of course are budgetary numbers, but may be adjusted for any particular device).

If a random seek is allowed at the end of each track transfer in order to switch to another thread, then the sustainable rate is:

$$3^{MB}/_{sec} \times \{(11^{ms}/_{T})/(44ms/_{(T+sk)}\}$$
$$= .75 \text{ MB/sec}$$

If a video stream requires 1.5Mb/sec (~.2 MB/sec), then the NVOD approach allows 15 threads without buffering, while the TVOD approach allows only 3 threads, <u>with buffering</u>. The buffer size, however, need not be very large, i.e.:
Since:

.2 MB/sec=>.2KB/ms

Then let buffer size for each thread $=B_T$

$$B_T=.2\text{KB/ms} \times (3 \text{ seeks of } 30\text{ms} + 2 \text{ transfers of } 11\text{ms})$$

$$B_T=.2 \times (90+22) \text{ KB}$$

$$B_T=25\text{KB}$$

Since each track must be buffered, it would adjust to 33 KB/thread => 100 KB buffer total.

This of course assumes the video friendly type of device that has no other non transfer activities to mask.

Therefore, it is clear that TVOD threads with only one (or serendipitously, a few) customers per thread will require many more disks than a NVOD with schedulable threads which allows a significantly greater number of customers per thread and a significantly greater number of threads per disk.

Furthermore, where as the TVOD approach limits tape storage to 1 thread per device (as opposed to the 2 or 3 for the disk), the NVOD approach works as well from tape as it does from disk. The systems configured below are intended to support 1,000 to 10,000 program titles and use the tape as the primary storage media. The disks are used as a buffer for the currently active programs primarily to reduce the number of passes against each tape volume for reliability purposes rather than for performance. As a matter of fact, tape performance in some instances will exceed that of disk devices in terms of the number of simultaneous threads that can be sustained. With NVOD threads scheduled in greater than 30 second increments,(e.g. 5 to 15 minutes) the delay would completely mask the initial few seconds of startup to mount the tape.

Using tape directly, or using disk as a buffer in front of the tape for most of the active programs (assuming the disk described above and that each will hold 3 to 5 gigabytes) it would be possible to have each tape or disk provide as many as 15 threads (channels) of broadcast. This could be all for one program, or split among the number of programs that could be stored on that one device (e.g. three 90 minute movies would require 3 to 4 1/2 gigabytes of storage).

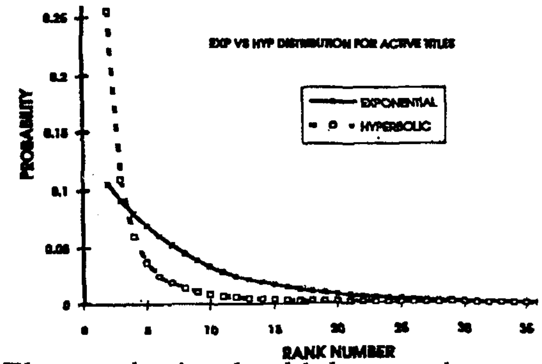
To support 200 channels of NVOD would require a minimum of 14 devices, and 500 channels would require a minimum of 34 devices.

The experience in this industry is that in any particular week there is a very small subset of programming that accounts for most of the demand. One specific example is for video rental where 97% of revenue comes from less than 25 titles. With this tight a skew, out of a population of 1,000 to 10,000 titles between 33 and 68 titles account for 99% of the demand and between 39 and 129 titles account for 99.5% of the demand. See the inset below for the details on this set of calculations.

Customer Demand vs. System Performance Limits Analysis

The given task is to identify the number of program titles necessary to satisfy a "large" proportion of the customer requests. Obviously the greater the percentage of requests one desires to satisfy, the larger the population. Also the distribution of the requests across the inventory of titles significantly affects the number requested. If the total number of titles is significantly greater than what can be simultaneously broadcast (e.g. more than an order of magnitude such as 200 channels for 2000 titles) then the true answer will generally lie between an exponential and a hyperbolic distribution. Experience has shown that the number will quite often track hyperbolic through some significant portion of the range (e.g. 95% to 99% depending on the tightness of the skew) and then drift to the exponential and then terminate at some finite number far short of where either distribution would predict.

Without knowing the actual distribution of requests to the most popular titles, it is difficult to calculate the exact number of titles that must be broadcast with any confidence. However using what little is known about the reference patterns of the video rental base (e.g. one company reports that 97% of revenues come from 20 to 25 titles) one can calculate a range and bound the problem using distributions that historically tend to fit skew problems of this sort; i.e. Binomial (to give an easy but very gross and optimistic first approximation), Exponential, and , and "Hyperbolic" (or "Pareto") probability distributions.



The emphasis should be on the use of hyperbolic distributions (with the probability density form $p(n)=A/n^k$ for $n \geq 1$). It is a convention to use the word "3-sigma" to mean the value of the tail beyond $z=\pm 3\delta$ limits for the case of a "normal" or gaussian distribution (even though the actual distribution is not normal and may not even have a "sigma". Framing the given problem between EXP and HYP limits gives the approximate value calculated here. One caveat is that historical skew distributions tend to deviate from perfect hyperbolic shapes at the high end tails (i.e. they drop faster than $1/n^k$ and this is formally called "droop"). This shortens the real use tail so that the actual expected answer should be below that calculated at the 3-sigma limit for the hyperbolic distribution.

1) **Most elementary approach (Binomial)**

p = probability of selecting "choice" movie $\approx \frac{25}{10^4} = .0025 = \frac{C}{N}$

$\mu = Np = 25$

σ binomial $\approx \sqrt{Np} \doteq 5$, $3\sigma = 15$

$\mu + 3\sigma = 40$ Titles for 99.74% of demand

Same for Poisson (same N large, p small), $\mu = \lambda, \sigma = \sqrt{\lambda} = \sqrt{\mu}$

$\mu + 3\sigma = 40 \leftarrow$

2) **Exponential approach** $\int_0^{25} \alpha e^{-\alpha x} dx = 97\% = 1 - e^{-\alpha 25}$

so $\alpha = .14$

find n ∍ tail is .26% (Gaussian tail interpetation, 2 sides at 3σ)

so CUM $= 1 - e^{-.14n} = .9974$

$e^{-.14n} = .0026$

$n = 42.5$

% Demand Satisfied	25 Titles at 97% Skew	25 Titles at 95% Skew	20 Titles at 97% Skew	20 Titles at 95% Skew
99.74%	43	50	34	40
99.50%	39	44	30	35
99.00%	33	38	26	31

Scale Invariant Distributions

3) <u>Hyperbolic Distribution</u> Assume $f(x) = \frac{A}{x^k}$, k unknown > 1

normalize (find A): $\int_1^N f(x)dx = -\frac{A}{k-1}\left(\frac{1}{N^{k-1}} - 1\right) = 100\%$

our N very large $\Rightarrow A \sim (k-1)$.

$$\text{Now CUM} = 97\% = \int_1^{25} \frac{k-1}{x^k}dx = \left(1 - \frac{1}{25^{k-1}}\right) \Rightarrow k - 1 = 1.09 \quad k \approx 2.089 \quad \left(\begin{array}{c}\text{Note:}\\ k \approx 1.931\\ \text{for 95\% skew}\end{array}\right)$$

Then for 99.749.(3σ),

$$(k-1)\ln(n) = -\ln(.26\%) = \ln(\text{"tail"}) \quad n \approx 236$$

e.g. $f(x) = \frac{1.089}{x^{2.089}}$

% Demand Satisfied	25 Titles at 97% Skew	25 Titles at 95% Skew	20 Titles at 97% Skew	20 Titles at 95% Skew
99.74%	236	599	161	385
99.50%	129	296	93	200
99.00%	68	107	51	100

Max values: reality less due to "droop".

If the 25 titles were placed on shared disks at 12 threads each, and the rest of the programming spread with a few on tape drives, the 200 channels could be supported by 12 disk drives and 12 tape drives. The 500 channels would require about 24 disk drives and 16 tape drives. The following cost analysis is on the basis of about 20 disks at 10 threads per disk plus 16 tape drives.

Individual disks and RAID (Redundant Array of Independent Disks) systems have different performance characteristics, so the numbers derived for individual disks and RAID systems is different. However, even when using the more expensive RAID technology, only a few TVOD threads can be produced.

		1 thread/disk	10 thread/disk	ATM Encoder/	1 Thread/disk	10 Thread/disk
COSTS for 25 Thread Video System		Disk Cost	Disk Cost	ATM Switch	system cost	system cost
MP Server Only	$65,000	$25,000	$2,500	$6,250	$96,250	$73,750
Server Saver	$30,000					
Server Saver+Pentium	$40,000	$25,000	$2,500	$6,250	$71,250	$48,750
Mainframe	$250,000	$25,000	$2,500	$6,250	$281,250	$258,750
COSTS for100 Thread Video System						
MP Server Only	$260,000	$100,000	$10,000	$25,000	$385,000	$295,000
Server Saver	$120,000					
Server Saver+Pentium	$130,000	$100,000	$10,000	$25,000	$255,000	$165,000
Mainframe	$500,000	$100,000	$10,000	$25,000	$625,000	$535,000
Cost for 250 Thread Video System						
MP Server Only	$650,000	$250,000	$25,000	$62,500	$962,500	$737,500
Server Saver	$232,500					
Server Saver+Pentium	$242,500	$250,000	$25,000	$62,500	$555,000	$330,000
Mainframe	$1,250,000	$250,000	$25,000	$62,500	$1,562,500	$1,337,500
Cost for 500 Thread Video System						
MP Server Only	$1,300,000	$500,000	$50,000	$125,000$1	925,000	$1,475,000
Server Saver	$480,000					
Server Saver+Pentium	$490,000	$500,000	$50,000	$125,000$1	115,000	$665,000
Mainframe	$2,500,000	$500,000	$50,000	$125,000$3	125,000	$2,675,000
Cost for 1000 Thread Video System						
MP Server Only	$3,250,000	$1,000,000	$100,000	$250,000$4	500,000	$3,600,000
Server Saver	$930,000					
Server Saver+Pentium	$940,000	$1,000,000	$100,000	$250,000$2	190,000	$1,290,000
Mainframe	$5,000,000	$1,000,000	$100,000	$250,000$6	250,000	$5,350,000

Figure 4 - Assumptions Comparing the Three VOD Video Server Architectures

Total# Thread System	MP Server Only System	Svr+Srvr Saver	Mainframe
25	$96,250	$71,250	$281,250
100	$385,000	$255,000	$525,000
250	$962,500	$555,000	$1,562,500
500	$1,925,000	$1,115,000	$3,125,000
1000	$4,500,000	$2,190,000	$6,250,000
	10 threads / disk		
25	$73,750	$48,750	$258,750
100	$295,000	$165,000	$535,000
250	$737,500	$330,000	$1,337,500
500	$1,475,000	$665,000	$2,675,000
1000	$3,600,000	$1,290,000	$5,350,000
	Cost comparison of 1 Thread vs 10 thread systems		
25	1.305084746	1.461538462	1.08695652
100	1.305084746	1.545454545	1.1682243
250	1.305084746	1.681818182	1.1682243
500	1.305084746	1.676691729	1.1682243
1000	1.25	1.697674419	1.1682243

Figure 5 - Table representing completed video server costs for Multiprocessor System, Server Saver System, and Mainframe System. The last 5 rows of numbers represents cost improvement multipliers per thread.

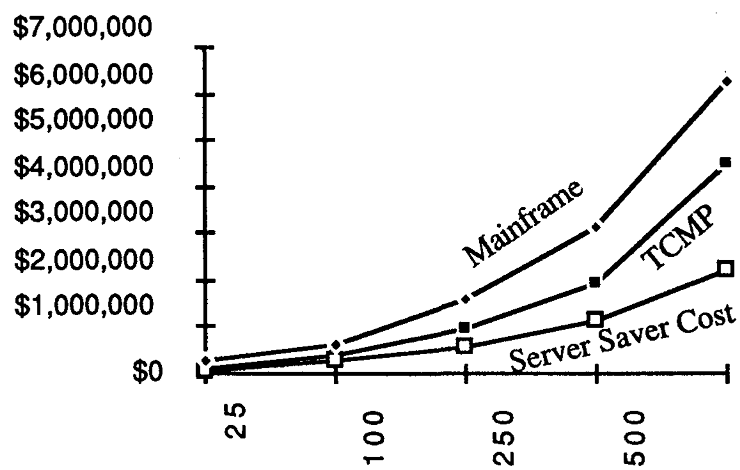


Figure 6 - Chart depicting relative system costs for each of the 3 candidate TVOD video server system implementations.

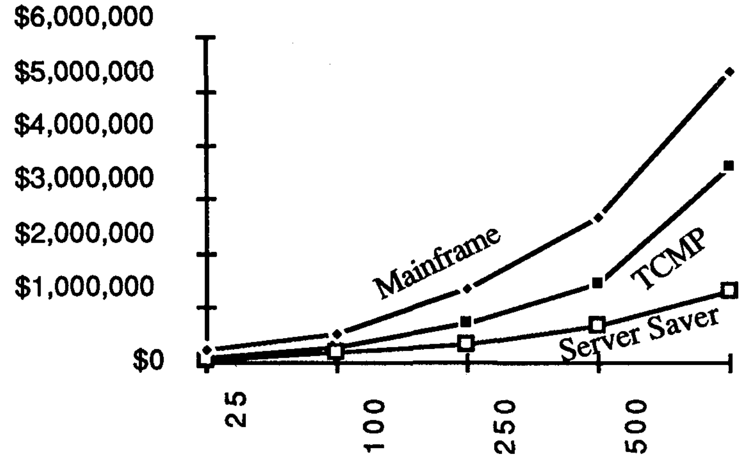


Figure 7 - This chart represents NVOD cost per program thread for each of the 3 candidate systems assuming 10 threads are available from each storage device simultaneously. Depending upon desired video quality and device performance, these numbers can change, but their relationships remain the same.

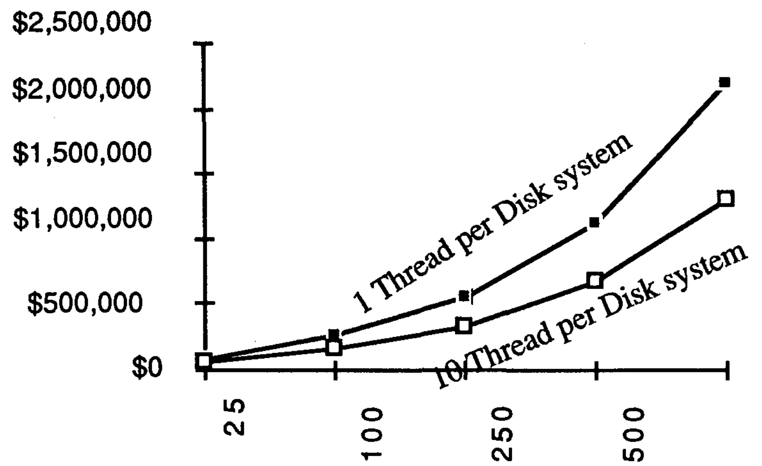


Figure 8 - This chart illustrates the cost of the Server Saver application. The upper curve represents system cost when only 1 thread per storage device (TVOD) is provided and the second curve represents system cost for a 10 thread per disk system is implemented.

The chart depicted in Figure 9 illustrates the cost savings as a percentage savings using the Server Saver System Architecture for 1 thread per storage device giving TVOD and 10 threads per storage device rendering Unlimited Capacity NVOD with a response time of 10 minutes.

When the system program capacity is 20 units, NVOD can be produced for about 68% of the cost of TVOD while systems above 250 programs flatten out such that NVOD costs less than 60% of TVOD systems as depicted in Figure 9.

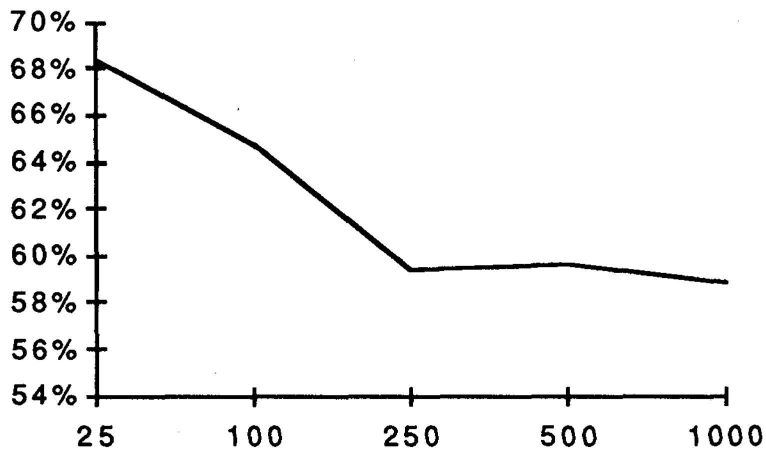


Figure 9 - The above chart depicts cost savings of NVOD system over TVOD system for Server Saver style architectures.

Figure 9 demonstrates how the cost per thread is reduced as the number of threads is increased. The vertical axis represents the cost relationship between the server saver system with one thread per storage device (TVOD) and the same server system with 10 threads per storage device NVOD). Ten threads per storage device implies that for a 90 minute movie, 10 equal space start times can exist providing a new start time for the movie every 90/10 or 9 minutes.

The horizontal axis represents the number of threads (channels) available to subscribers. The multiple thread system assumes that the disk storage system is video friendly.

Unlike standard drives, Video Friendly drives are designed to provide a worst case data rate that will assure highly predictable delivery of data so that discontinuities in the audio/video data stream will not exist.

Figure 10 illustrates the savings that multiple thread disks (NVOD) can have on each of the candidate architectures versus single thread (TVOD). NVOD produces more programming at less cost per program than TVOD.

Furthermore, since NVOD has an interval of time during which subscribers can request a program, NVOD can accommodate unlimited subscribers without requiring a subscriber to tune in late. Therefore, NVOD can produce substantially more revenue.

This paper is also intended to determine the cost consequences of employing tuned solutions to the TVOD

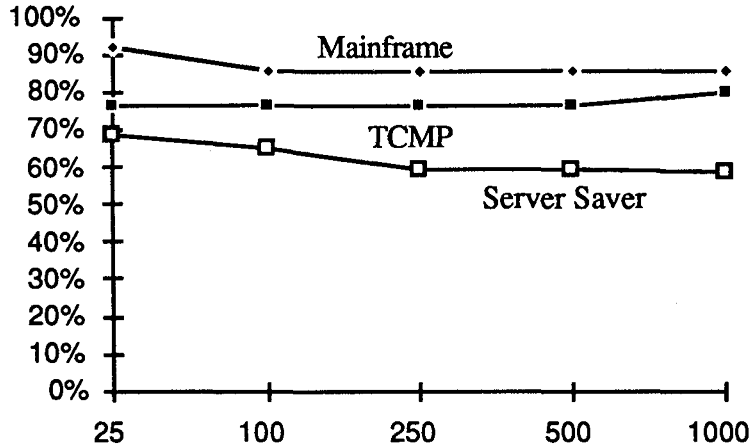


Figure 10 - This chart depicts the same server saver information as figure 9, but it includes the related information for the Tightly Coupled Multiprocessor Application and Mainframe Application

application versus general purpose solutions, or partially tuned solutions. Figure 11 illustrates that the tuned solution (i.e. the Server Saver architecture) with 200 or more threads will cost about 50% as much as the partially tuned solution (Tightly Coupled Multi-Processor) and about 25% as much as the general purpose (mainframe approach)!

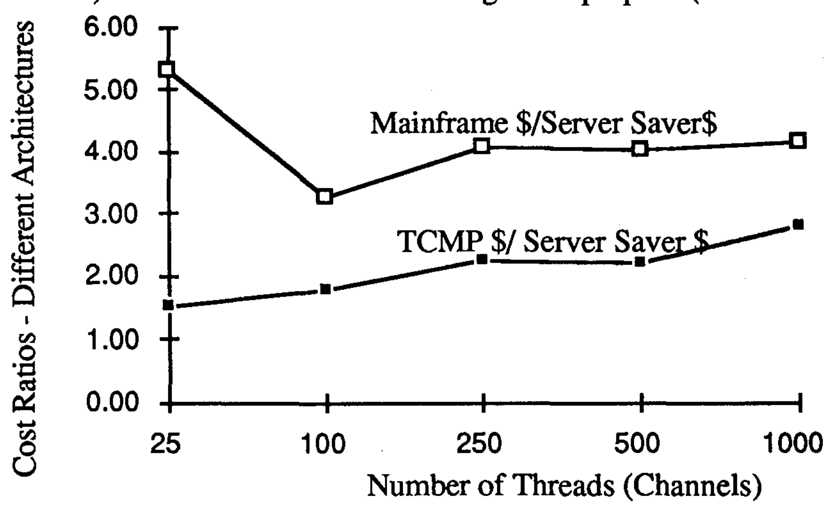


Figure 11 - This chart depicts NVOD based system costs normalized to the Server Saver Architectural approach.

The crucial element to facilitate both TVOD and NVOD is smooth, high bandwidth, uninterrupted device transfer capability which we will refer to as "Video Friendly". Interruptions in data will produce interrupted video unless extensive and costly video buffering is provided. Interrupted video of course is unacceptable.

Figure 12 shows Transfer time vs. wall clock time for four representative vendor disk drives.

The horizontal line at 33 ms represents the threshold of intolerable disk access times.

It should be observed that only one vendor drive achieves this requirement

(Micropolis) and one approaches this requirement.

Why use Video Friendly Devices

Video friendly devices are able to cost-effectively produce multiple threads of smooth program video without the requirement for external video buffering which requires extensive amounts of video RAM. Therefore, video friendly drives are a significant component in reducing system cost.

Why use Server Saver Style Architecture

The Server Saver architecture represents a highly tuned VOD application, not a generalized solution. It is the most cost effective tool to solve the VOD problem. It provides significant advantages, including cost / performance, system flexibility, simplicity, high uptime and low maintenance, as discussed in the author's previous referenced VOD articles (1). It can produce TVOD, NVOD and combinations of TVOD and NVOD.

Also, NVOD systems can produce unlimited customer showings per movie (unlike TVOD systems).

Why use NVOD instead of TVOD

NVOD systems can require approximately 1/2 the hardware cost to produce 10 times the video flow as do TVOD systems. Therefore, NVOD systems are the highly preferred economic approach. NVOD has been shown to cost substantially less to implement than TVOD and has the ability to support unlimited clients. NVOD can be tuned by the system operator to produce waiting intervals other than discussed in this paper.

Perhaps an average 3 minute wait for the program is too long, even though that time is used for information on upcoming attractions, to sell food to be delivered to the home, to sell other services, or to merely provide a music-video interlude, or some combination of these.

The system operator can reduce the NVOD interval by 50% while increasing his hardware costs substantially less than 50%, thus moving closer to the TVOD model. This procedure can be repeated as often as desired to further reduce viewer latency time.

Studies in one TVOD vs. NVOD trial by a hotel pay per view TV operator indicated no increased revenue stream for the TVOD application, only added cost to provide the function to the same number of clients.

One could make a career of looking at numerous other variations of data in the spread sheet and graphing and plotting them. It seems obvious to the authors if an operator is decided on a TVOD system, he can use the Server Saver technology and video friendly disks. If he desires the economies of NVOD, he can also cost effectively employ the Server Saver technology.

If the operator is unsure of whether he wants TVOD or NVOD, he can use the Server Saver technology and provide both styles of programming to his clients. Statistics collected from the real world will probably tell the real story.

What is the Impact of VOD on CATV delivered ATM

The basic non-cascaded Server Saver supports a 500 thread (or channel) system. The industry seems to support the idea of employing 50 MHz to 500 MHz for conventional analog TV and 500Hz up to 1000 MHz for digital interactive TV, while leaving 5MHz to 50 MHz for reverse channel communications.

If this is the case, then it is expected that as many as 500 streams or channels of digital interactive video could be placed in the upper CATV frequency band. If it were desired to support more program sources or threads, a different delivery system (such as fiber optic cables) might need to be in place.

Since fiber optic cable would only go to a city section, block or curb, costly ATM switches would be required to move the proper packets from one transmission facility to another. This leads to the hotly debated question: Does a city require more than 500 channels of interactive TV and if so, how much more will it cost to provide them?

NVOD will not require as many channels for transmission as TVOD to support the same number of viewers; hence provides a great deal of relief from the expenses required to provide the infrastructure to support the greater number required by TVOD.

Filling in the Viewer Latency Time

The following strategy is proposed as a means of preventing the viewer from becoming frustrated at the delay between the time he makes his selection and the time it is actually delivered.

Assume a maximum viewer latency time of 10 minutes. A number of pre-packaged "mini-programs" may be prepared. They could be binary divisions of the 10 minute maximum time to be filled if a viewer requested a program only 1 second after the previous start time.

Thus, there could be one of several ten minute cartoons, five minutes of coming attractions, two and one half minutes of news headlines, one minute and 45 seconds worth of public service announcements, 50 seconds of helpful hints, 25 seconds of quotable quotes, 12 seconds of inspirational messages, and up to 12 seconds of a warning that the feature is about to begin. Using various combinations of the above, any amount of time up to the maximum latency time may be filled with entertainment. When it is determined what the delay will be, the viewer could be advised of the time remaining before the next feature

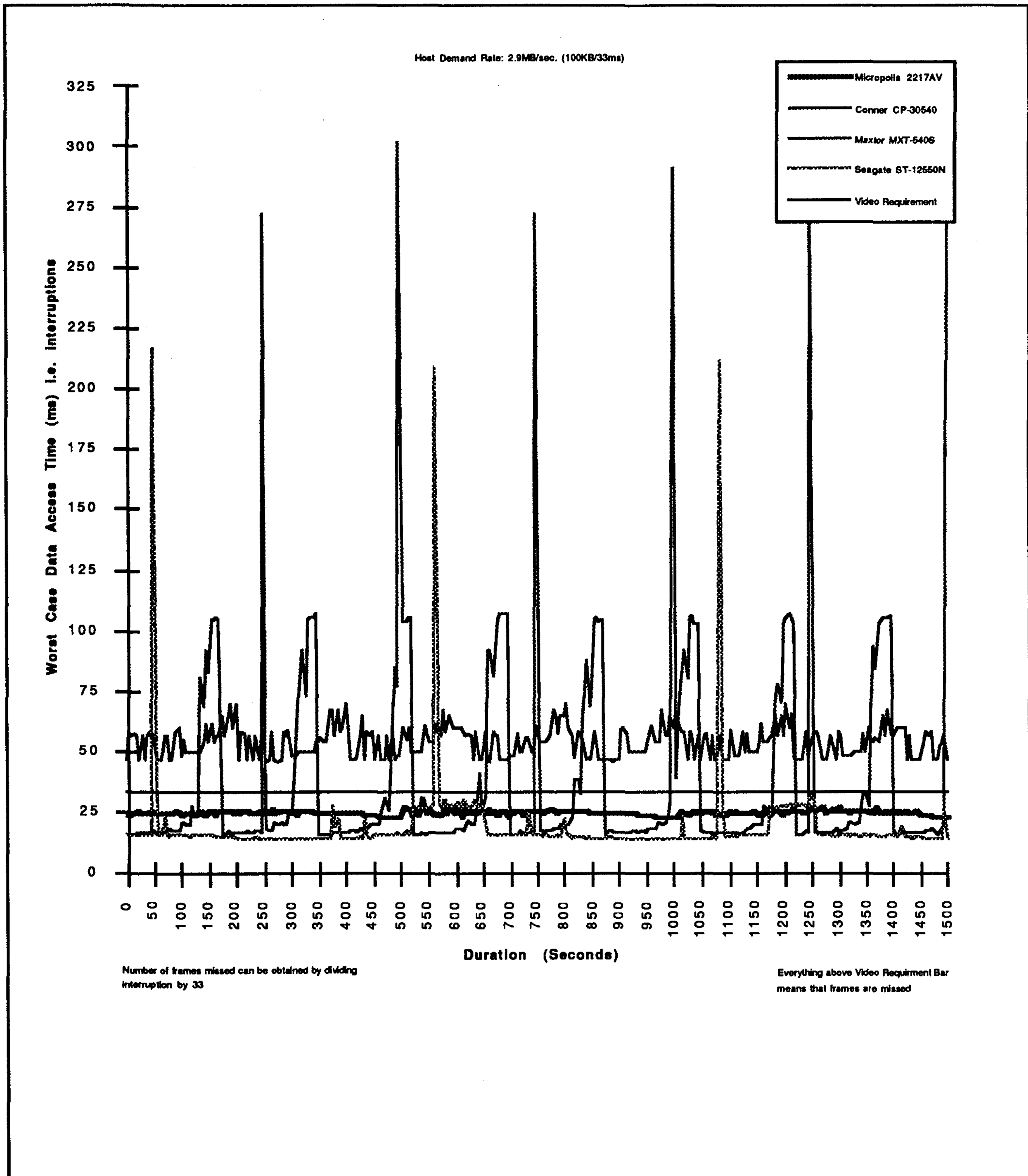


Figure 12. Transfer Size vs. Time. Video Friendly disk drives provide almost a linear relationship between transfer length and transfer size. The data in this graph includes command overhead, and is measured with the demand rate of 2.9 MB/s. This figure is presented courtesy of Micropolis Corporation.

starts, and given a menu from which he could select his own "fill in" entertainment.

Conclusions:

The ultimate goal of interactive TV is to provide the subscriber nearly instantaneous access to the programming of his choice. While this goal is attainable at very high cost, for a very limited number of subscribers, the authors do not believe it to be economically feasible to provide this type of service to the number of interactive TV subscribers projected over the next five years by leading industry market researchers.

NVOD offers a reasonable compromise between the ideal (zero viewer latency time)and an acceptable delay. This approach permits operators to obtain equipment which may be amortized by charges acceptable to subscribers.

---

[1]Pentium is a trademark of Intel Corporation

[2]Power PC is a trademark of Motorola Corporation

References:

(1) “Video On Demand: Architecture, Systems, and Applications”, W.Hodge, S.Mabon, and J.T. Powers, Jr., SMPTE Journal, September 1993.

(2) “A Film Quality Digital Archiving & Editing System” W.Hodge, W.Harvey, and R.S.Block, SMPTE Conference on Advanced Television and Electronic Imaging for Film and Video, New York City, February 5-6 1993.

Copyright © 1994 Hodge Computer Research Corporation, Storage Technology Corporation & National Cable Television Association. All rights reserved

TUNER CHARACTERISTICS OF CABLE-READY RECEIVERS

Joseph Van Loan, Cablevision Industries
David Large, InterMedia Partners

ABSTRACT

Congress has directed the FCC to define the characteristics a television receiver must have if it is to be marketed as "Cable-Ready" This paper describes the joint filing of the receiver and cable industries, along with the author comments, on those characteristics required in order that directly attached receivers A) not interfere with the operation of the cable system and, B) assure viewers of a reasonable degree of performance.

INTRODUCTION

The Cable Act of 1992 included two provisions specifically aimed at solving the technical interface problems between cable systems and consumer electronics equipment. The first directed the Commission to find solutions to such problems as independent tuning access for VCRs and TVs, while the second directed them to define the characteristics which receivers would have to meet in order to be marketed as suitable for direct connection to cable systems.

In its resultant Notice of Proposed Rulemaking, the FCC proposed both short-term and long-term solutions. By the time this is printed, the Commission is scheduled to have considered the extensive comments on those proposals and issued final rules.

The short-term proposals call for cable operators using converters to provide various modifications which will allow customer's receivers to access more than one channel simultaneously. The most common solution is expected to be some sort of bypass switch which will allow the converter to be bypassed for all non-scrambled channels.

The long-term proposals call for the provision, on receivers, of a Decoder Interface connector which will allow descrambling to happen *after* the receiver's tuner. Separate decoder modules on TVs and VCRs would allow completely independent selection of channels for viewing and recording, as well as timed, multi-channel recording of any combination of scrambled and clear programs.

Both short-term and long-term solutions will increase the likelihood that consumers' receivers (both TVs and VCRs will be directly connected to the broadband cable drop. Thus, defining receiver performance has become a critical issue for compatibility.

In actuality, negotiations between the television manufacturers, cable manufacturers and cable operators to define desirable receiver characteristics has been ongoing since the early 1980s, under the auspices of the NCTA/EIA Joint Engineering Committee ("JEC"). In fact an early draft of an interface specification (numbered IS-23) was actually circulated for ballot in 1985.

Under the pressure of the immediate rulemaking and Congressional deadline, the old draft was exhumed and compromise agreements made on nearly all the open issues, so that a joint filing of the industries was made to the Commission suggesting mandatory performance standards. The negotiators attempted to strike a balance between assuring good reception for cable subscribers and increasing the receiver cost so much that few would be manufactured or sold to consumers.

This paper will discuss the various important reception parameters and the logic that lead to the joint filing, as well as the continuing work on the more comprehensive voluntary "son of IS-23" specification which addresses parameters not included in the rulemaking as well as parameters on operation of cable systems. A summary of the parameters, the FCC's proposal, and the JEC recommendation are found in Table I at the end of the paper. Readers are cautioned that this summary does not include all the subtleties and details of the final submission..

For convenience, the parameters are divided between those that can degrade the operation of the cable network (and the reception of neighboring subscribers) and those which primarily affect the reception of the subscriber using the equipment.

CHARACTERISTICS WHICH AFFECT OTHER USERS AND EQUIPMENT

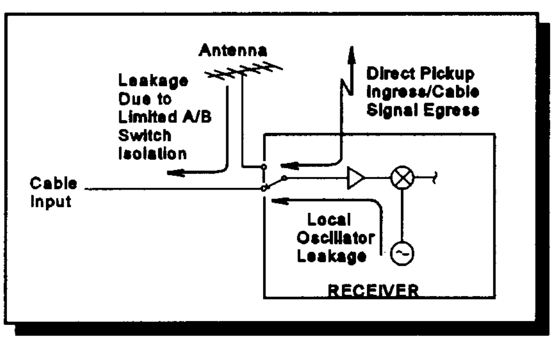


Figure 1: Receiver Antenna Terminal Egress/Ingress Mechanisms

When a receiver is connected to a cable system, it has the potential to affect other equipment connected to the system, as well as users of the electromagnetic spectrum. Those effects can be classified generally as conducted interference, excessive loss and egress.

Conducted Interference

Conducted interference occurs when signals within a TV receiver or VCR are conducted "out" the device's input terminals and onto the cable system. Figure 1 shows the mechanisms by which signals can be transmitted out of a receiver's antenna terminals:

- Signals may be generated within the receiver and inadequately isolated from the input terminals. The most common is the first local oscillator, but other mechanisms are possible.
- Inadequate shielding may cause strong external signals to be picked up and transmitted back out the antenna terminals.
- When receivers include A/B RF input selection switches, inadequate isolation in those switches will cause cable signals to be transmitted out the antenna terminals, or signals from the antenna can be conducted back onto the cable system, thereby disrupting viewing of downstream subscribers.

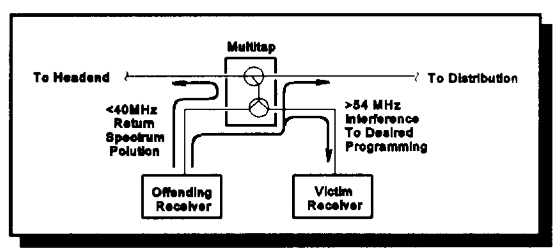


Figure 2: Interference Paths into Network and Other Receivers

These signals can affect the cable network in two ways (Figure 2). Signals which are in the sub-low band (5-30 or 5-42 MHz) will be directly transmitted towards the headend in any fully enabled two-way cable system. The combined effect of many such receivers may reduce the usability of the upstream spectrum or cause operators to have to install high-pass filters at every offending receiver.

Signals which are in the forward spectrum of the system will appear at the input terminals of other receivers, attenuated by the isolation between ports of multi-taps. They have the potential to directly interfere with programs being watched by others.

The first step to setting an appropriate specification for antenna terminal egress in the forward spectrum is to determine how much signal can be tolerated at neighboring receivers. The FCC requires cable operators to deliver at least 0 dBmV at receiver antenna terminals. The industry-accepted standard for "just perceptible" interference from discrete interfering carriers is 55 dB below the desired signal[1]. Thus, interfering signals should be less than -55 dBmV at neighboring terminals.

Cable operators are required to maintain 18 dB of isolation between adjacent tap ports, which would suggest that conducted emissions should be limited to -37 dBmV to guarantee no visible interference. The FCC proposed this level in its Notice of Proposed Rulemaking ("NPRM") for all sources of antenna-conducted egress.

<u>Local Oscillator Leakage</u>. Unfortunately, it is very difficult to provide the degree of isolation in receivers that will reduce antenna-conducted LO leakage to the FCC proposed level. There are, however, several factors that suggest that the worst-case situation will seldom happen in practice:

- There is a relatively low probability (about 2% for an 80 channel system) that receivers connected to ports of the same tap will be on and tuned exactly seven channels apart, which is where the interference occurs within single conversion tuners using a standard 45 MHz IF.
- Taps generally provide greater than the minimum 18 dB isolation and subscribers are further isolated by drop cable losses, which increase with frequency.
- In order to meet the specification on essentially all of their production, receiver manufacturers will have to provide higher isolation *on average*.

Given all these factors, the negotiators settled on an attainable recommendation of -26 dBmV from 54 through 300 MHz, declining to -20 dBmV up to 450 MHz and to -15 dBmV maximum and -20 dBmV average from 450 through 1002 MHz.

Clearly, these values are not sufficient to prevent all interference, but were deemed all that were practical with single conversion tuners.

Not dealt with in the specification to date, but still to be resolved are signals falling in the sub-low band from 5-40 MHz. These will interfere with upstream communications and, combined with those from other subscribers may considerably increase the "crud" in the return path.

<u>Conducted DPU Ingress</u>. Conducted DPU ingress occurs when signals from nearby TV and FM broadcast, land mobile radio and paging stations are picked up with the TV receivers, VCRs or FM tuners and are conducted out the devices' input terminals and onto the cable system. Using the same logic as above, the negotiators felt that -26 dBmV would be a sufficiently low level to make conducted DPU an unlikely source of viewing complaints at neighbors' receivers. Given that the probability factor is not relevant here, however (the interfering signals will always be present for all strong channels, regardless of where the receiver is tuned and regardless of whether it is turned on or off), that level is not allowed to relax at higher frequencies. That is especially important since

some of the strongest external fields are due to UHF transmitters operating up to 800 MHz.

The most contentious issue has been the external field strength in which to immerse the receiver when measuring the conducted ingress level. A study done by Joe Stern, under CableLabs sponsorship, predicts that 40.8% of television households will experience field strengths of 100 mV/m or greater on at least one television broadcast channel. A similar study done on behalf of the EIA by Jules Cohen predicts that 46.2% of households will experience 100 mV/m. Above that field strength, the studies diverge, with Stern predicting that 6% will experience field strengths of 1 Volt/m, while Cohen predicts 8.4% will experience 300 mV/m, but less than 1% will experience 1V/m. Both studies predict that the probability of UHF interference exceeds that of VHF interference at the highest field strengths which is unfortunate from the standpoint of visual impairment as UHF stations are offset from cable channels in the same frequency ranges resulting in beat patterns which are subjectively more apparent than frequency coherent interference. Neither study included interference from non-television-broadcast sources such as paging transmitters.

There are, however, mitigating factors. For one thing, the test procedure currently under review by the JEC (developed by C.T. Jones under CableLabs sponsorship) measures susceptibility at all receiver orientations relative to the external field. Testing of 35 representative television receivers (plus a number of VCRs and converters) done by Jones suggests that susceptibility is strongly dependent on this orientation. Given that actual receivers may be oriented randomly with respect to external fields, the average susceptibility in homes will usually be less than the tested maximum. Secondly, neither study attempted to predict the attenuation effects of buildings and other structures on the signal strength received by receivers inside dwellings relative to that measured in relatively "free space" outside. While in some cases, the field strengths may actually be higher due to reflected signals constructively combining or due to receivers being located far above ground level (as in a high-rise apartment situation), on average it can be expected that there will be some attenuation affects.

Given all these factors, the JEC adopted a standard of 100 mV/m external field, measured at the orientation of greatest sensitivity using the C.T. Jones approach.

A/B Switch Isolation. This is the amplitude of a signal connected at one input terminal of a switch as measured at the other input terminal. It was felt that few external antennas would produce a field strength at the receivers in excess of +20 dBmV. Given a requirement to keep signals at adjacent receivers below -55 dBmV and a probable subscriber-to-subscriber isolation of about 29 dB, the isolation of antenna selector switches must be maintained at 46 dB or greater. As discussed below, however, this is inadequate to protect reception on the receiver incorporating the switch, so a higher value was chosen.

Radiated Interference

Radiation from Receiver. There has been a disparity between the Part 15 regulations, applying to receivers, and Part 76 regulations, applying to cable systems, with respect to radiated interference. The old Part 15 regulations applied only to signal sources originating within receivers, were higher in absolute levels than Part 76, and did not apply to *re*-radiation of antenna-applied signals. By contrast, the Part 76 regulations apply to cable systems, *including all connected equipment*, and limit leakage to a much lower level.

In its NPRM, the FCC proposed to apply Part 15 limits to re-radiation of antenna-connected cable signals of up to +25 dBmV. The JEC proposed, instead, to apply the Part 76 limits, but at more realistic signal levels of +15 dBmV. They felt that this was a more rational approach.

Radiation from Customer's Antenna. A second source of potential radiation occurs if cable signals are coupled into the customer's antenna due to inadequate isolation in the receiver's antenna selector switch. This issue has been before the Commission previously, and an isolation requirement of 80 dB to 215 MHz and 60 dB to 550 MHz was written into the Part 15 rules. The JEC suggested that the isolation limit between 550 MHz and 800 MHz be 55 dB.

With a likely maximum cable input signal level of +20 dBmV and an isolation of 55 dB, the signal applied to the customer's antenna lead will not exceed -35 dBmV. With typical feedline losses and consumer antenna gains, this should not result in leakage in excess of Part 76 limits of 15 μV/m at 30 meters.

Excessive Signal Loss in VCRs and Converters with Bypass Functions

The majority of homes have a VCR connected in series with at least one television receiver. Since cable systems may deliver minimum signal strengths of 0 dBmV on at least one channel, the reception on a TV receiver connected to a VCR output is materially affected by the loss of the VCR. Similarly, under the new regulations, many converters will have bypass filters or switches of some sort designed to deliver clear signals directly to following receivers. As with VCRs, the loss of the bypass mechanism is important to assuring noise-free reception.

In the simplest configuration, VCRs incorporate a two-way splitter at the RF input and a switch at the output for selecting between its internal modulator and the bypassed input signal. VCRs have up to 10 dB or more of insertion loss, but in the ideal case, VCR through loss should not exceed 4-5 dB. However, it was felt that the specification should allow slightly more loss than the expected ideal to take into account manufacturing variances. Finally, both splitters and switches have increasing loss with frequency.

While recognizing the concerns of manufacturers, cable operators wished to make the specification sufficiently tight that a manufacturer could not use an unequal ratio splitter (to favor its own tuner) or a resistive splitter. They compromised on 6 dB to 550 MHz and 8 dB to 1 GHz. It will be incumbent upon CE manufacturers with high loss VCRs to insure the tuner has sufficiently a low noise figure to provide noise-free reception with low input levels!

CHARACTERISTICS WHICH DEGRADE THE USER'S OWN RECEPTION

Although some receiver characteristics can degrade the reception of other subscribers, others affect primarily its own performance in a cable environment. That environment is more benign than over-air reception in some respects (such as relatively uniform signal levels) and more challenging in others (such as tuner overload and use of adjacent channels). The specifications which follow are designed to assure that a cable-ready receiver will offer reasonable reception when connected to a cable system.

Adjacent Channel Rejection

In assigning off-air channels, the FCC has carefully avoided having broadcasters in one area operating on channels which are on adjacent frequency bands[2]. Thus, receivers intended for purely broadcast reception do not typically have to deal with lower or upper adjacent signals. Cable operators use the full spectrum of available channels, but control levels of adjacent channels within 3 dB and lower aural levels to 10-17 dB below the associated visual carrier.

In the worst case, therefore, adjacent visual carriers could be 3 dB higher than the desired visual carrier and adjacent aural carriers as high as 7 dB below the desired visual carrier. In order to prevent visible interference, neither of those conditions should result in products falling within the tuned channel and having amplitudes greater than -55 dBc.

In considering this, the JEC determined that the dominant problem was the lower adjacent aural carrier and so simplified the specification to that single parameter. It further deemed it unlikely that operators would run aural carriers as high as -10 dBc and that, if they do, that it would be uncommon for the 3 dB channel difference and -10 dBc aural levels to occur simultaneously. For that reason, the final recommendation to the Commission was that a lower adjacent aural carrier 10 dB below the tuned channel's visual carrier would not cause in-channel products of greater than -55 dBc.

Tuner Distortion Products

One of the most fundamental differences in the signal environment of over-air reception and cable reception is in the overall RF spectrum presented to the tuner. When connected to an antenna, a receiver typically has to deal with far fewer signals, but of widely varying amplitude. Thus, off-air tuners are characterized by the use of low-noise preamplifiers and wide AGC ranges.

When a receiver is connected to a cable system, the minimum signal levels are much higher and the total dynamic range is generally much less, but the total RF power is significant because all or nearly all channels are simultaneously used. Furthermore, the total power is increasing as cable systems are built to wider and wider bandwidths. The result is often visible second and third order beat generation in the preamplifier and mixer stages.

In order to reach a consensus, the JEC had first to deal with the fact that maximum signal strengths are not specified under the Part 76 rules except for a blanket "don't overload" statement. Additionally, the rules do not take into account non-video signals that may be present in the delivered cable spectrum and therefore impinge on the tuner. Examples are positive trap scrambling carriers, data carriers, FM carriers and digital audio signals.

As a result, although the current rulemaking does not involve reconsideration of the Part 76 rules, the voluntary IS-23 specification does contain limitations on maximum and average amplitudes of both video and non-video carriers delivered to receiver inputs. In general, individual carriers are limited to +20 dBmV and average carrier levels across the spectrum are limited to +15 dBmV.[3]

Given these restrictions on delivered cable signals, the test condition that was recommended for evaluating the distortion performance of receiver tuners was a continuous comb of CW carriers extending from 54 to 750 MHz at a uniform carrier level of +15 dBmV. Under these conditions, the JEC recommended that the tuner should generate no in-band products in excess of -51 dBc. Since this is a CW test, the expected performance under modulated signal conditions should be better by 6 dB for second order products and 12 dB for third order products.

DPU Ingress

Arguably the most contentious issue for the many years of these negotiations has been shielding performance. From operator's standpoint, this has certainly been the most frequent cause of complaints for systems located within the "Grade B" contour of TV broadcast stations or other radio transmitters, especially paging transmitters.

The results of the studies done by both industries to support their arguments are discussed above. The study by C.T. Jones of existing equipment not only showed susceptibility to be dependent on receiver orientation, but also established that few current production televisions or VCRs come close to acceptable performance when measured at the most sensitive channel and orientation.

The compromise agreed upon by the JEC negotiators was that, when measured at the most vulnerable of several channels spread across its tuning range and at its most vulnerable orientation with respect to the external field, a receiver will not experience DPU interference greater than 50 dB below a 0 dBmV input signal when immersed in a 100 mV/m external field (with some relaxation at the highest frequencies).[4]

Although this might appear to represent no improvement over the existing Canadian ingress standard (which also uses a 100 mV/m external field), there are important differences:

- The Canadian standard is measured only over 300 MHz
- The test condition for the Canadian evaluation uses a time and frequency coherent interfering signal so that the threshold of visibility (it is a subjective viewing test) is only about 40-45 dB desired to undesired ratio.
- There is no requirement in the Canadian test for testing at the most vulnerable orientation with respect to the external field.

Given these differences, the proposed new standard arguably represents at least a 15 dB improvement over the Canadian standard. Whether it is sufficient to eliminate the effects of DPU interference in urban cable systems remains to be seen.

Image Rejection

Single conversion television receivers using the common 41-47 MHz IF frequency also have a potential response to signals higher than the desired channel by twice the IF frequency, or approximately 90 MHz. In VHF off-air reception, this is not a problem because of spectrum assignments for the low and high band channels. In UHF, it can occur, but is statistically unlikely. By contrast, cable systems use the entire spectrum, so that virtually every potentially interfering image channel is present at the tuner input.

FCC regulations allow cable signal levels to vary by 10-17 dB across the spectrum, depending on the system total bandwidth. In the worst case, therefore, the image signal could be 17 dB higher than the desired carrier and 72 dB of image rejection would be required to insure a 55 dB (barely perceptible) image response. More realistically, though, the variation across 90 MHz is unlikely to be that great and, with newer fiber-to-the-neighborhood architectures, the variation is unlikely to exceed 5 dB.

Based on this, the JEC settled on a compromise of 60 dB image rejection up to 650 MHz, reducing to 50 dB through 900 MHz. The reduction at the high end reflects the feeling of the negotiators that frequencies above 750 MHz will most probably be used for digital signals whose amplitude will be reduced at least 5-10 dB relative to analog visual carrier levels. Certainly, this is a marginal performance specification, but reflects the difficulty in achieving higher image rejection in single conversion tuners without significantly greater cost and without affecting in-band frequency response. Should cable industry practice ultimately lead to full amplitude analog or digital carriers through 1,000 MHz, this specification will turn out to be inadequate.

Antenna Selector Switch Isolation

Just as inadequate A/B switch isolation can result one customer's antenna input signals being coupled into the network and affecting a neighbor's reception, it can also affect his own reception.

Given that cable signals may be as low as 0 dBmV and antenna input signals may be easily as high as +20 dBmV, an isolation of at least 75 dB would be required to suppress the antenna signals to 55 dB below the cable input.

The FCC has previously considered this issue and, as a result, existing Part 15 regulations require that antenna selector switches maintain 80 dB of isolation through the VHF broadcast band, and at least 60 dB through 550 MHz. They asked, in this NPRM, what the isolation should be between 550 MHz and 1,000 MHz.

The receiver manufacturer's problem is that these switches are typically either made with PIN diodes or reed switches and it is very difficult to achieve consistently high isolation at the high end of the band.

Fortunately, few customers currently use antenna selector switches for alternating between roof-top antennas and cable system inputs so a lack of isolation in current receivers has not been a frequent cause of customer complaints. When an interference situation does occur due to high external levels, a cable operator can put an attenuator in the antenna input and eliminate the problem by reducing the external level.

Based on the above, and that average receivers will have to considerably exceed the standard in order for all receivers to pass, the JEC agreed to drop the required isolation to 55 dB for frequencies in excess of 550 MHz. In the authors' opinion, this is a very marginal specification.

REFERENCES:

1. This was recently confirmed by a Bronwen Jones study dated October 12, 1993 and published in the CableLabs document, Customer Premises Equipment Performance and Compatibility Testing, Chapter 5, "Direct Pick-up Interference Subjective Tests".

2. Channels 4 and 5 are not actually adjacent, nor are 6 and 7.

3. In the special case of upstream signals, those limitations are unrealistic. It is not uncommon for set-tops to generate signals as high as +60 dBmV. Given the finite isolation of multi-taps, this could easily result in levels at neighboring receivers of nearly +40 dBmV. The draft IS-23 allows for this level up to 30 MHz. Still under dispute is the expansion of the return band from 30 MHz to 42 MHz.

4. Shielding is typically characterized by Relative Equivalent Length (REL) which is calculated as:

REL=ExternalField(dBmV)-InputSignalLevel((dBmV) +Carrier/InterferenceRatio (dB)

Using this definition, the proposed REL is 90.

TABLE I
SUMMARY OF RECEIVER REQUIREMENTS

Characteristic	FCC NPRM Proposal	JEC Proposal	Final FCC Rules
Local Oscillator & other Conducted Interference	54-1002MHz: -37 dBmV	54-300MHz: -26 dBmV 300-450MHz: -20 dBmV >450 MHz: -20 dBmV Average, -15 dBmV peak.	
Egress Conducted Interference	-37 dBmV	-26 dBmV @ 100mV/m field, 54-1002 MHz	
Antenna Selector Switch Isolation	54-216MHz: 80 dB 216-550MHz: 60 dB >550 MHz: no proposal	54-216MHz: 80 dB 216-550MHz: 60 dB 550-800MHz: 55 dB	
Re-radiation of Cable Signals	Part 15 limits @ +25 dBmV input	Part 76 limits @ +15 dBmV input	
Tuner Overload	Distortion products below -55 dBc, unspecified conditions	51 dB CTB & CSO with 110 CW input carriers +15 dBmV	
Adjacent Channel Susceptibility	"Just perceptible" with + 3 dB adjacent signals	-55 dBc with lower sound @ -10 dBc	
Image Rejection	No proposal	54-650MHz: 60 dB 650-1002 MHz: 50 dB	
DPU Interference	"Just perceptible" @ 100 mV/m	54-800MHz: 90 dB REL .8-1GHz: 80 dB REL	
VCR Loss	6 dB	54-550 MHz: 6 dB 550-1002 MHz: 8 dB	

ISBN 0-940272-01-6; 0-940272-08-3; 0-940272-10-5; 0-940272-11-3; 0-940272-12-1; 0-940272-14-8; 0-940272-15-6; 0-940272-16-4; 0-940272-18-0; 0-940272-19-9; 0-940272-20-2; 0-940272-21-0; 0-940272-22-9; 0-940272-23-7; 0-940272-24-5; 0-940272-25-3; 0-940272-26-1; 0-940272-27-X; 0-940272-28-8; 0-940272-29-6; 0-940272-32-6; 0-940272-33-4; 0-940272-34-2; 0-940272-35-0; 0-940272-36-9; 0-940272-37-7; 0-940272-38-5; 0-940272-39-3; 0-940272-40-7; 0-940272-41-5; 0-940272-42-3; 0-940272-43-1; 0-940272-44-X; 0-940272-45-8; 0-940272-46-6; 0-940272-47-4; 0-940272-48-2; 0-940272-49-0; 0-940272-50-4; 0-940272-51-2; 0-940272-52-0; 0-940272-53-9; 0-940272-54-7

© 2015 National Cable and Telecommunications Association. All Rights Reserved.