THE COMPLETE TECHNICAL PAPER PROCEEDINGS

FROM:

NCTA Technical Papers

250km Transmission of Frequency Multiplexed 64QAM Signals For Digital CATV Backbone Application

Shinji TSUJI　　　　*Yuji HAMASAKI*

64QAM is recognized as a standard for the digital television transmission method on cable plants. This is more robust than the traditional VSB-AM analog method in white noise and nonlinear distortion environment that is typical in the CATV environment. As it is a digital method, it can be regenerated as necessary. Therefore, it will be a better format for long distance transmission using optical fiber. In this paper, we propose an economical CATV backbone solution using a 64QAM transmission system. The system has one digital headend and several analog headends connected by optical fiber cable. The digital headend facility is shared by other headends. We conducted experiments to verify the concept of the digital 64QAM backbone. The experimental system has 250km fiber with 5 cascaded erbium doped optical fiber amplifiers (EDFA) and 11 channels of 64QAM signal or 60CW signals were transmitted as test signals. Such parameters as input/output of EDFAs and modulation depth of laser diode were intentionally changed to find the optimized value. We conclude that it is feasible to transmit 30 channels of 64QAM signals through 250km of optical fiber with EDFAs without regeneration.

QAM DIGITAL TRUNK SYSTEM

Network Architecture

We propose here the QAM digital trunk system. An image of this system is shown in Figure 1. There are some cable operators or cable plants made up of the HFC system with radial optical fibers. If the operator plans to serve digital services individually, he must invest in digital headend such as CS-IRDs, real-time encoders, scramblers, 64QAM modulators and management systems. Especially in the case of the contents provider, satellite operator and cable operator which are all independent. We propose that several operators share one digital headend and transmit digital signals to each operator's analog headend using optical fiber and EDFAs for long distance. This is clearly cost-efficient.

There are two methods of trunk transmission. One is the PCM/ATM method that uses time division multiplexing in the trunk line; at each local headend signals are

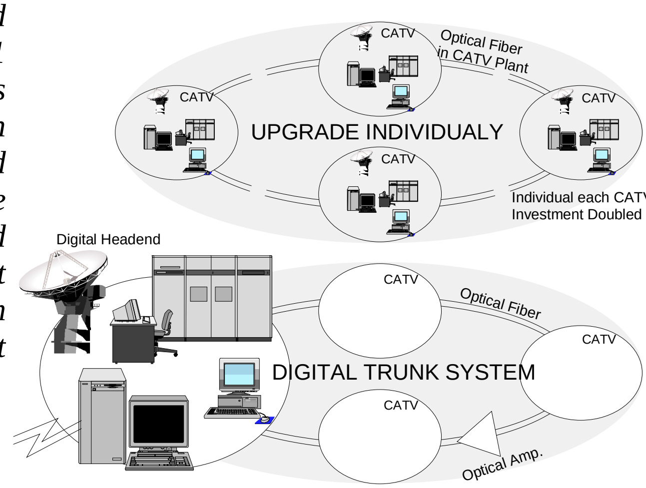


Fig.1　Image of digital services

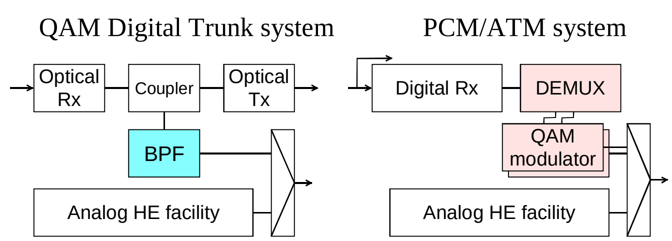


Fig.2a　Example of local headend facility

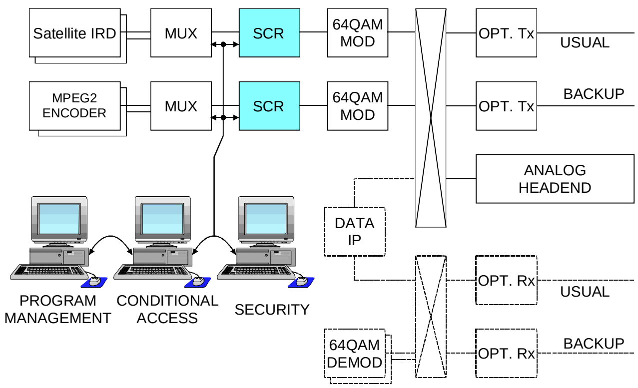


Fig.2b Example of digital headend facility

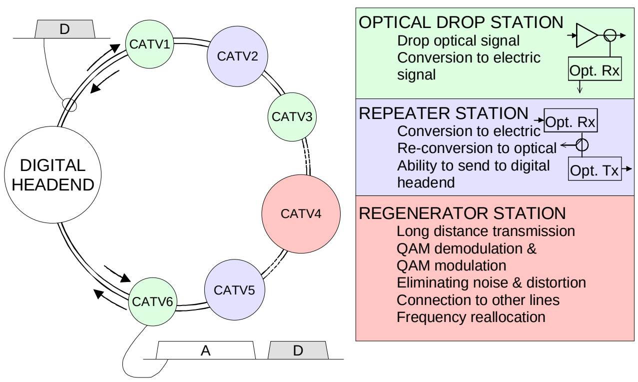


Fig.3 Function of local headends

de-multiplexed and converted to QAM format at every RF channel. An example of the local headend facility is shown in Figure 2a. The other is a method that transmits QAM signals in RF from the digital headend. In this method, each local headend consists of simple equipment including an optical receiver and band pass filter (BPF). The latter is a QAM digital trunk system, which is lower cost at each local headend than the former.

At the digital headend, received and/or served digital contents are transmitted. Figure 2b illustrates the configuration of the digital headend. It consists of satellite and terrestrial receivers, MPEG2 real-time encoders, multiplexers, conditional access equipment, QAM modulators, management system, mixers and optical transmission equipment. For higher reliability, the optical transmission line should be doubled. As the need arises, fast data transmission equipment, ultra capacity cash servers and fast Internet backbone can be connected.

Figure 3 illustrates the system image. The left ring shows that digital signals from the digital headend are transmitted in the trunk line via local headends (CATV1, CATV2,…, CATV6) clockwise and counterclockwise. At local headend, digital signals are combined with analog signals, and both signals are transmitted in the CATV plant. Each local headend is identified by its configuration and has the function shown on the right. The simplest configuration is called “Optical drop station”. It consists of an optical coupler, optical receiver and BPF. Digital signals from the digital headend are filtered to eliminate out-of-band noise and mixed with analog signals here. The next configuration called “Repeater station” consists of an optical receiver and optical transmitter. Optical signal is once converted to an electric signal and again converted to an optical signal. Therefore, if the transmission line is designed in a loop scheme, it is possible to transmit a signal to the digital headend from here. The third configuration is called “Regenerator station”. All QAM signals are demodulated to bit stream and modulated to QAM again. The purpose of regeneration is to remove noise and distortion integrated as long distance transmission. It is possible to add and delete a signal. It is also possible to change frequency allocation.

Here we show application examples using this system. An example in Figure 4 transmits the contents of local UHF broadcasting to other CATV plants. An operator in a local UHF area receives the signal and encodes it to MPEG2 format, then modulates it to QAM signal in R1 channel and transmits it to the digital headend. At the digital headend, the received signal in R1 channel is demodulated and multiplexed with other contents, and QAM modulated to D1 channel. Then subscribers in every CATV plant can enjoy UHF broadcasting by tuning to D1 channel at

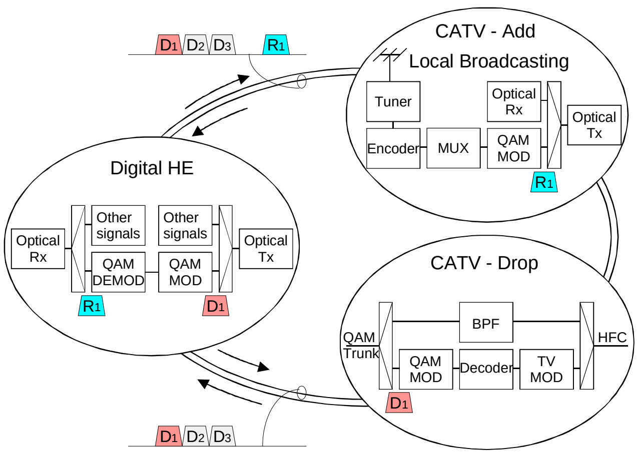


Fig.4 Application – Transmission of local broadcasting to every plants

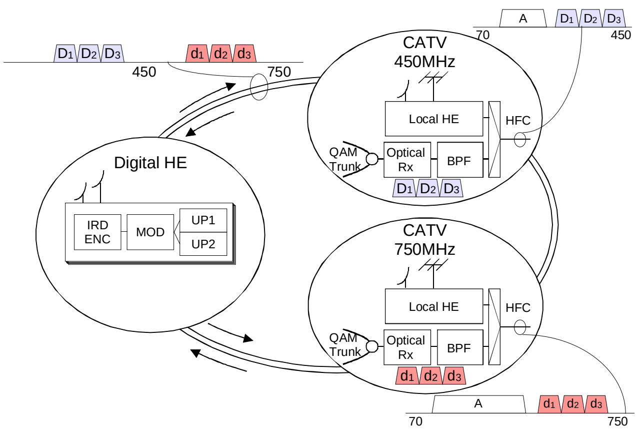


Fig.5 Application – Transmission to combination bandwidth of 450MHz and 750MHz

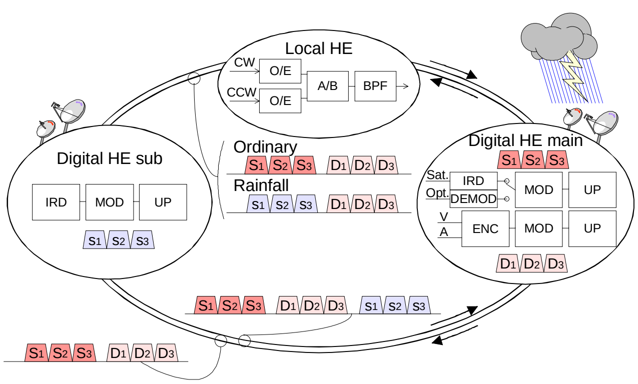


Fig.6 Application – Backup system for eliminating interruption from satellite

as high quality as in the broadcasting area. An example in Figure 5 is a combined system of 450MHz plants and 750MHz plants. In 450MHz plants, digital signals are usually located in higher range such as 300-450MHz.□In 750MHz plants, they are usually 550-750MHz.□In this case, digital signals are transmitted as simul-frequency from the digital headend. At the local headend of the 450MHz plant, signals for 450MHz plants are filtered and transmitted to the plant. Figure 6 is an example of a backup system in case signals via satellites are hindered by rain attenuation. There are two digital headends, separated enough distance to reduce the probability of co-hindrance. The main digital headend selects better signals from the satellite receivers or the signals that are received at the sub digital headend and transmitted through the trunk line. Then disturbed time will be shorter.

System Performance

In this system, QAM signals are transmitted from the digital headend to subscriber terminals via the optical trunk line, local headend, and CATV plant. On the other hand, analog signals are transmitted from the local headend to subscriber terminals. We assume here that C/N of analog signal (4MHz bandwidth) is 43dB. As C/N of 64QAM annex C needs more than 31dB per 4MHz, we distribute the system performance under the condition of 30dB C/N (averaged per Nyquist bandwidth) including margin. When QAM averaged power is NTSC signal –10dB, QAM C/N in CATV plant is 31.8dB. To obtain 30dB as total C/N, C/N in the trunk line should be more than 34.7dB. (Based on 4MHz carrier level, total C/N is 34.9dB, trunk line C/N is 39.6dB. Here expression of C/N means averaged value per Nyquist bandwidth without notice.)

250KM TRANSMISSION EXPERIMENT

Outline of Experiment

We conducted a transmission experiment to investigate the feasibility of a QAM digital trunk system and system design. Another

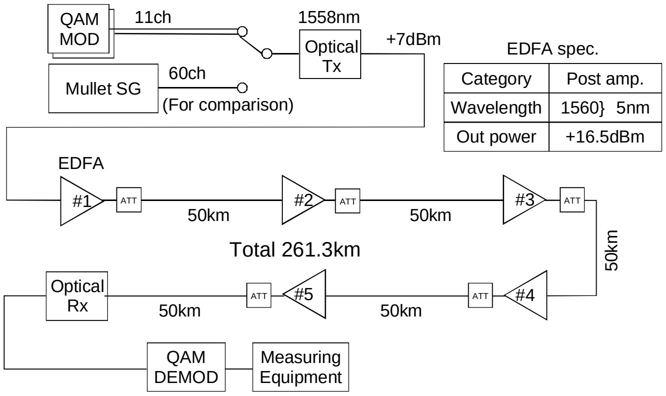


Fig.7 Block diagram of transmission experiment

Table 1 Parameters of the experiment

Symbol Rate	5.274 Msps
Interleave	I=12,J=17
RS FEC	(204,188)
Rolloff	0.13
Fiber zero dispersion	1551 nm
PRBS	2^{23}-1

purpose of the experiment is to obtain information on the number of optical modulation depth (OMD) we can set up, and the distance optical amplifiers can be separated. Figure 7 shows a block diagram of the experiment. Important parameters are listed in Table 1. Test signals are 11 channels of 64QAM or 60 channels of continuous wave (CW). CW is used for comparison. 64QAM signals are generated by our prototype modulators designed for ITU-T J83 annex C (Japan) system. 11 64QAM signals are

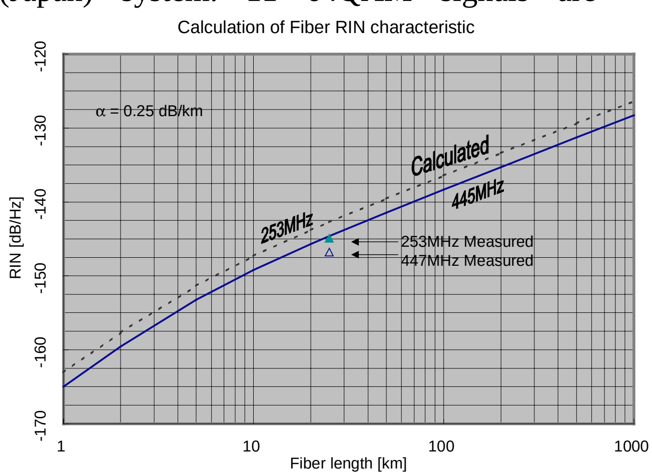


Fig.8 Fiber RIN used in experiment

located 222~288 MHz and 60 CWs are between 91.25MHz and 445.25MHz. We measured 64QAM characteristics at the center channel located at 252~258MHz. This channel modulator is connected to BER measuring equipment and other modulators self-generate PRBS signals individually. Optical transmission equipment uses DFB-LD direct modulation. All of the optical fiber is dispersion shift fiber. And optical bandpass filter is not attached in front of either EDFAs or optical receiver. EDFA is designed for the post amplifier with front and back pumping.

Basic Performance

First we measured the fundamental performance of the optical equipment without optical fiber. The result under 60 CWs and optical modulation depth of each CW is 4% is C/N=54dB, CSO<-72dB, CTB<-74dB. Then we measured fiber RIN. Figure 8 shows fiber length vs. RIN. The dot is measured and the solid line is calculated using parameters of ordinary dispersion shift fiber. Fiber RIN appears to grow as length increases, as shown in the line in Figure 8. This means that fiber RIN influences transmission noise mainly in long distance while the receiver input power influences in short. Figure 9 shows the relationship between optical modulation depth per channel (OMD) and C/N when 11 QAM signals are transmitted. In the range, where OMD is several percent, C/N grows as OMD increases. In the range where OMD~10dB, C/N does not change as OMD changes. In this range, there may be three fundamental elements. The first element is the carrier level and noise level independent of the number of waves. Carrier level and C/N grow as OMD increases. Second is the third order distortion which grows as OMD increases. Since signals are QAM which spectrum is flat over the bandwidth except rolloff, third order distortion is observed as white noise. Third is the noise from other QAM modulators combined with. C/N is constant as OMD

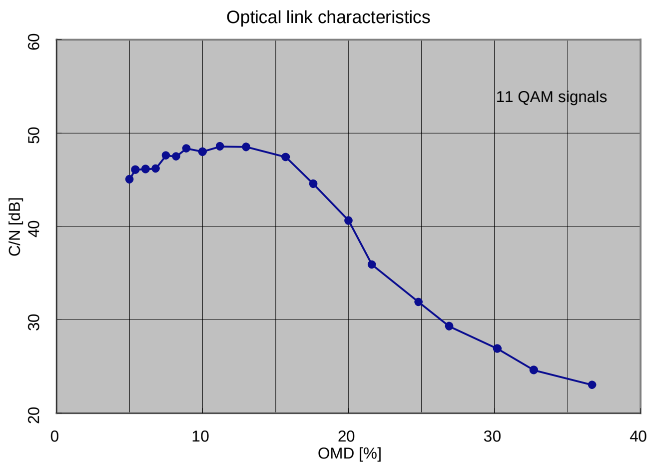


Fig.9 C/N of optical link equipment

changes. In the range where OMD exceeds 15%, C/N reduces as OMD increases. The effect of over-modulation such as distortion and laser clipping seems to cause this. Finally, BER performance of prototype QAM modulator and demodulator in this experiment is 10^{-4} when C/N=26.5dB without FEC.

Performance of EDFA

We outline C/N performance with cascaded EDFAs here. C/N of optical link without EDFA can be obtained from (1). We calculated C/N with cascaded EDFAs from (2) and (3) here. Variables and constants in formulas are as follows.

m: optical modulation depth per channel
Rs: receiver responsibility
P: received optical power
RIN: laser relative intensity noise
B: signal bandwidth
F: noise figure of the receiver preamp
T: temperature
k: Boltzmann constant
R: receiver equivalent resister
e: electron unit
Fi: noise figure of ith EDFA
Pi: input power of ith EDFA

But in actual conditions, because noise not only in the wavelength at the signal but in other wavelength is also amplified by wavelength characteristic of EDFA, noise in

$$C/N = \frac{\frac{1}{2}(mRsP)^2}{(RIN)B(RsP)^2 + 2eBRsP + \frac{4FkTB}{R}} \quad \text{……(1)}$$

$$\frac{F_n}{P_n} = \frac{F_1}{P_1} + \frac{F_2}{P_2} + \cdots + \frac{F_N}{P_N} \quad \text{……(2)}$$

$$C/N = \frac{\frac{1}{2}m^2(RsP)^2}{(RIN)B(RsP)^2 + \frac{2h\nu F_n}{P_n}B(RsP)^2 + \left(\frac{h\nu F_n}{P_n}\right)^2 BoB(RsP)^2 + 2eBRsP + \frac{4FkTB}{R}} \quad \text{……(3)}$$

other wavelength affects C/N of received signal. Figure 10 shows optical spectra. Figure 10a is a transmitted spectrum through one EDFA. Figure 10b is through five EDFAs. Comparing these two figures, it appears that noise of 1560nm – 1565nm is amplified. This noise is called ASE (amplified spontaneous emission). The receiver will receive both the signal and ASE. As EDFA cascades, ASE is amplified and receiver C/N degrades. Figure 11 shows the relationship between input power of EDFA and C/N when an EDFA is in transmission. Here input power changes from –7dBm to +7dBm to keep C/N for QAM transmission. Solid lines are

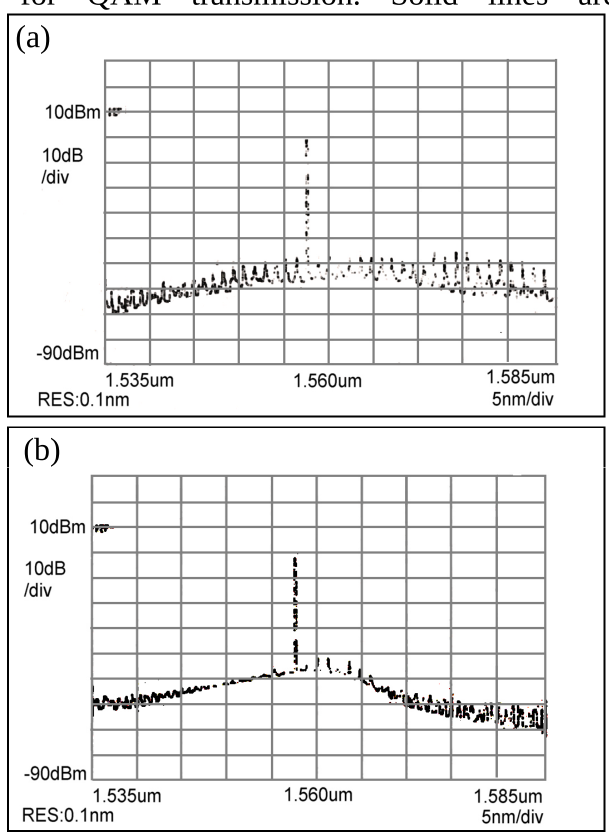
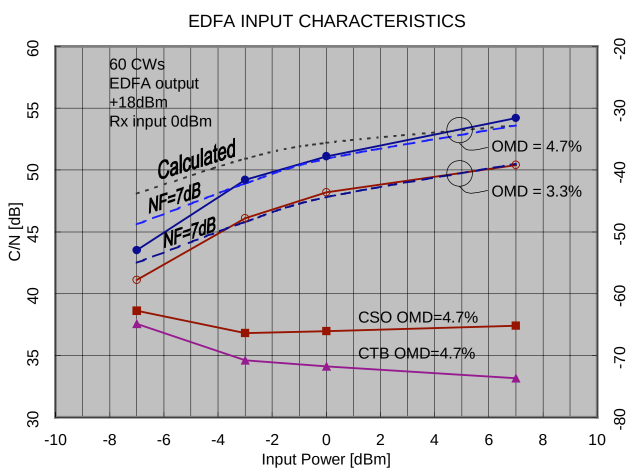
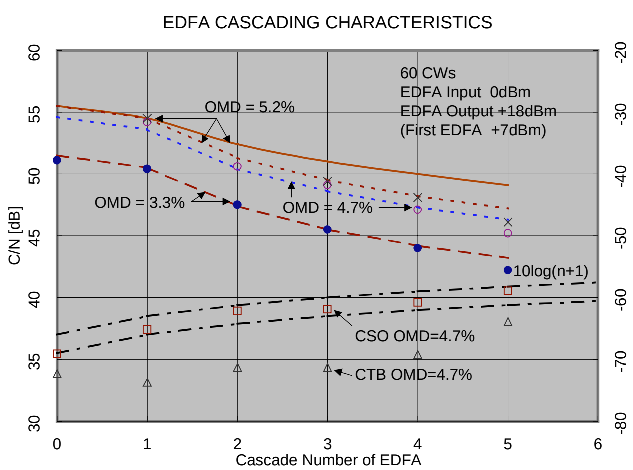


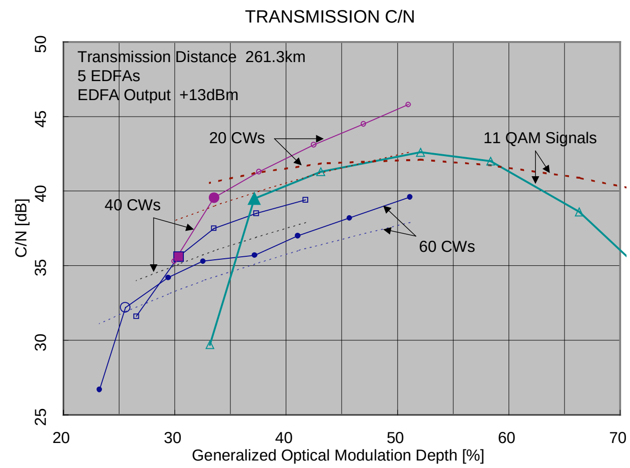
Fig.10 Optical spectrum
(a) 1 EDFA (b) 5 cascaded EDFAs

measured and

Fig.11 Input power characteristics







dotted lines are calculated. At high input range, the calculated line shows good approximation to measured data. However, as input power is lower, there is a difference between measured value and calculated value. This is because relative ASE power increases as lower input power. Noise figure (NF) of EDFA changes as input power; for example, NF is 7dB when input is +7dBm and NF is 4.5dB when input is 0dBm. Dotted line with caption “NF=7dB” in the graph assumes NF is 7dB constant. It seems to be good approximation over –3dBm input power. Figure 12 shows C/N, CSO and CTB as cascading EDFAs. In this graph, we measured using EDFAs and optical attenuators. As to C/N, the graph shows three cases of different OMD. In the case of OMD=5.2%, crosses of measured data degrade more than the solid calculated line. This also seems to be affected by ASE. Dotted lines in the graph are calculated assuming NF=7dB and show good approximation under four or five cascading. For more cascading, according to the condition of the growth of ASE, C/N seems to degrade much more than this approximation. In terms of distortion, CSO and CTB degrades as cascaded. One of the reasons is due to the wavelength characteristic of EDFA, i.e. optical gain is slightly different between the upper and lower edge wavelength of the optical signal. Dotted-solid lines fitted to distortion measured value are curves of 10log(n+1), where n is a cascade number of EDFA.

250km Transmission Characteristic

Now we will discuss the results of the 250km-transmission experiment. Figure 13 shows C/N vs. optical modulation depth where the number of waves is 11 QAM signals, 20 CWs, 40 CWs and 60 CWs. Output power of each EDFA is +13dBm. X axis is a generalized modulation depth calculated $M = m\sqrt{N}$ where m is OMD and N is the number of the wave. Usually, this number is 20% to 30% under normal HFC operation.

Fig.14 C/N and BER performance

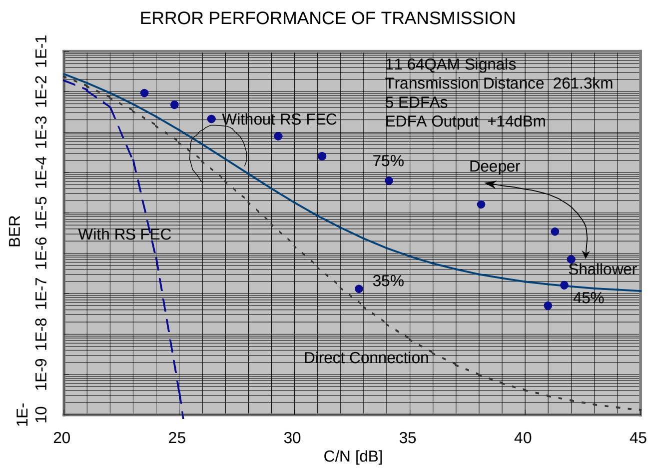


Fig.15 RS (188,204) ability

Fig.12 Cascading characteristics

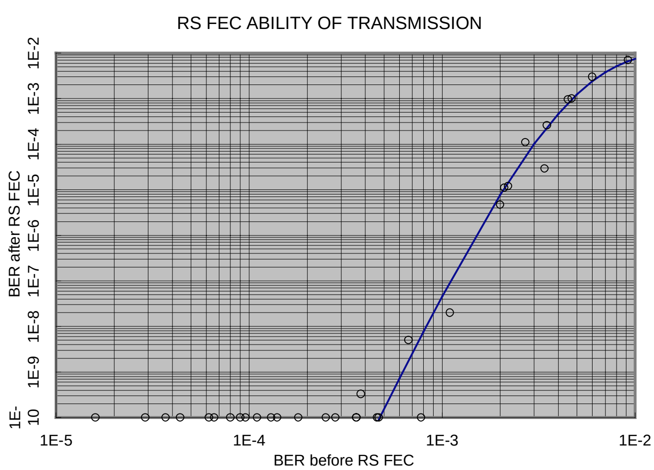


Fig.13 Modulation depth characteristics

At calculation of 11 QAM, the effect of modulator noise and CTB is considered. In the case of 20 CWs, C/N is larger as M is deeper. But this characteristic changes at M=33%. At the range of M<33%, the phenomenon that noise floor rises and distortion increases is observed. C/N degrades substantially. There is a similar phenomenon at M~30% of 40 CWs and at M~26% of 60 CWs. This phenomenon occurs at M~37% of 11 QAM signals as well. This point tends to be shallower as the number of wave increases. We consider that this point is the start of the effect of stimulated Brilloun scattering (SBS). At the shallower M range than this point, operation will be difficult. At the deeper M range, C/N degrades again because of distortion increasing. So we investigated a transmission performance at the range of 37%~M~50%. Figure 14 shows C/N and BER with/without RS FEC performance of long distance transmission. Dotted line in this graph indicates the performance of direct connection of 64QAM modulator and demodulator. Circular dots in this graph indicate C/N and BER as OMD changes. For example, at M=45%, 250km transmission C/N is 41.7dB and BER without FEC is 1.5×10^{-7}. When M is changed, C/N and BER are changed to a deeper and shallower side in the graph. In the graph, two cases of M=35% and M=75% are shown with caption. The dot of M=35% is far away from other dots because this point seems to be observed under SBS condition. Solid line passing the point of M=45% means the calculation with the addition of white noise to observed value, i.e. this line indicates an estimation of BER performance at subscriber's terminal including transmission noise of CATV plants. To demonstrate this circumstance, we measured BER by adding white noise in front of the receiver, and obtained that BER performance is slightly better than this line. From this line, we can estimate that BER is about 2×10^{-5} when C/N is 30dB, and that C/N must be better than 28dB to obtain BER that is 10^{-4} or better. Figure 15 shows BER of transmission with RS vs. without RS. The solid line in this graph is calculated, and data on BER=10^{-10} means no error has occurred during measuring time. This graph, except for rare different data shows general theoretical performance. If BER before RS FEC is less than 10^{-4} then it will reduce to less than 10^{-10}.

Cascading of QAM signal

Fig.16 Calculation of long distance transmission

Here we consider very long distance transmission such as several hundred kilo meters or several times. As QAM is a digital transmission method, it can be regenerated as necessary. The merit of regeneration; i.e. demodulation and modulation is that it eliminates transmission noise and distortion. Transmission noise from modulator, up-converter, optical equipment, optical fiber and coaxial equipment will be eliminated. However, symbol timing is very critical as it is integrated as cascading regeneration.

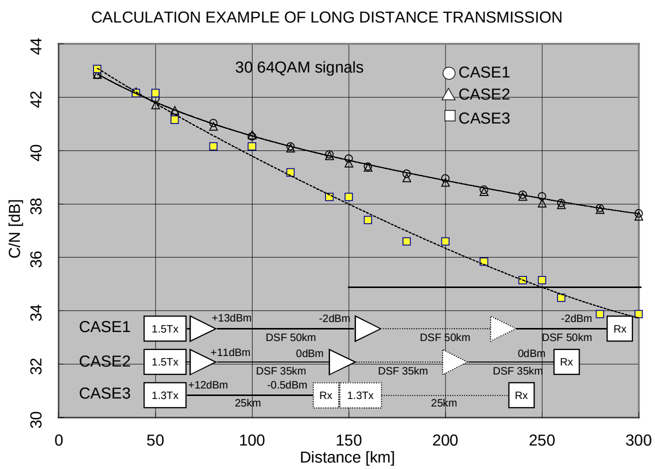


Figure 16 shows a calculation example for long distance transmission. It is assumed 30 QAM signals are transmitted. It is also assumed that 30 signals are allocated in one octave so that no second order distortion falls in signal frequency. We calculate three cases. Case 1 uses 1.5um LD and EDFA with –2dBm input and +13dBm output spanning 50km. Case 2 uses 1.5um LD and EDFA with 0dBm input and +11dBm output spanning 35km. In both cases, input power of PD is the same as that of EDFA. Case 3 uses 1.3um LD with +12dBm output and PD with –0.5dBm input spanning 25km and repeating. From this graph we can read that case1 and case 2 are very similar in characteristics, but that of case 3 deteriorates as distance lengthens. This difference is from the assumption that CTB increases by the summation of power in case 1 and case 2 and by summation of voltage in case 3. The former is from Figure 12 and in the latter it can be safely assumed that all fiber links have similar distortion characteristics. Case 1 and case 2 have different input and output power, but have similar curve. For this reason noise of fiber RIN is principal in long distance transmission and relatively small in that of input power or cascading. Note that the cascade number of case 1 may not be the same as case 2. For example, 4 EDFAs are cascaded to transmit 200km in case 1, but 5 in case 2. As EDFA cascades more than this experiment, according to the condition of ASE, C/N will degrade more. The effect of ASE is not considered in this calculation except for that described above. Under the condition that C/N from the digital headend to subscriber's home is 30dB and C/N between the trunk line is 34.7dB, QAM signals will be transmitted 250km without regeneration.

CONCLUSION

We have proposed a QAM digital trunk system that has several analog headends and a shared digital headend connected by optical fiber. This system will reduce the cost of installing a digital headend facility. We conducted a 250km-transmission experiment with 5 cascaded EDFAs and obtained data. It includes that C/N is more than 40dB and BER without FEC is about 10^{-7} under 11 64QAM signals are transmitted and that calculated

C/N shows good approximation by substituting a constant value for EDFA noise figure. From the results of this experiment, we conclude it will be feasible to transmit 30 QAM signals over 250km with 5 EDFAs and several times via regeneration stations.

REFERENCES

1. M. Shigematsu and M. Nishimura, “The design and fabrication of dispersion compensating fibers for an optical analog transmission and a 1.55um AM-VSB transmission experiment over a 1.3um single-mode fiber by using them”, IEICE OCS92-91, Mar. 1993.
2. H. Dai, S. Ovadia and C. Lin, “Hybrid AM-VSB/M-QAM Multichannel Video Transmission over 120km of Standard Single-Mode Fiber with Cascaded Erbium-Doped Fiber Amplifiers”, IEEE photo. tech. letters, v8 n12, Dec. 1996.

Shinji Tsuji
Senior Staff
Systems & Electronics R&D Center
SUMITOMO ELECTRIC INDUSTRIES, LTD.
1-1-3 Shimaya Konohana
Osaka, JAPAN 554-0024
Tel: +81 6 466 6545
stsuji@rcom.sei.co.jp

Yuji Hamasaki
Manager
Systems & Electronics R&D Center
SUMITOMO ELECTRIC INDUSTRIES, LTD.
1-1-3 Shimaya Konohana
Osaka, JAPAN 554-0024
Tel: +81 6 466 7901
hamasaki@rcom.sei.co.jp

4.5 Mbps Data Compatibly Transmitted in 6 MHz Analog Television

Walter S. Ciciora, Ph.D.
Technology Consultant
Southport Connecticut
WCICIORA@AOL.COM

Abstract

Analog television has a long future. Cable will service these large and important markets for may years to come. The compatible inclusion of digital data in analog television signals will increase the value of the spectrum which must continue to serve analog receivers. This data can be used for inputs to computers and for set top boxes which implement "push technology" information services. The technology described makes possible in excess of 4.5 Mbps of data carriage in an analog 6 MHz signal while not interfering with the normal analog reception of television. Data is hidden in a signal in quadrature with the video carrier, amplitude modulation of the sound carrier, and use of the VBI. This has sufficient capacity to convey one, two, or even three MPEG compressed signals to implement the "Compatible Digital Upgrade".

INTRODUCTION

The National Bureau of Standards, NBS, first proposed data embedded in television signals in 1970 for the purpose of distributing accurate time information nation-wide. While that effort did not succeed, it spawned the Closed Captioning system for the hearing impaired which in turn led to Teletext. More recent efforts to embed data in analog television signals led to the formation of the National Data Broadcasting Committee in 1993 and its consideration of systems by several proponents. On June 28, 1996, the Federal Communications Commission (FCC or the Commission) approved digital data transmission in the video portion of broadcast television transmission in its Report & Order (R & O), "Digital Data Transmission Within the Video Portion of Television Broadcast Station Transmissions", in MM docket No. 95-42. This R & O amends FCC rules to allow ancillary data *within* the video portion of the NTSC signal in four formats.

ANALOG REQUIEM?

It might be assumed that this FCC's action in allowing systems which add data to analog television is too late to be of commercial value. With the emphasis on Digital Television, DTV, some might consider analog television's demise imminent. This is hardly the case!

The recent adoption of an accelerated schedule for DTV deployment in the US has caused some in the popular press to suggest that the sole surviving technology for terrestrial domestic broadcasting will be digital and that analog NTSC will rapidly fade away. There are several reasons this may not happen at all and even more reasons why the notion of an accelerated schedule may not be realistic.

The deployed base of NTSC television receivers in the United States is huge. More than 250 million receivers plus another 175 million VCRs --- all of which are exclusively analog NTSC --- exist in about 100 million American television households. Additionally, Americans purchase about 25 million receivers and about 15 million VCRs each year. The Half Life of these receivers is on the order of 12-15 years! If the average television receiver is a 19" model, it's approximately 15 inch wide screen will be contained in a cabinet about 18" wide. All of the existing U.S. TVs set side by side would stretch 71,100 miles, several times around the earth. And 7,100 miles worth of new sets are sold in the US each year – more than enough to go coast to coast a couple of times! VCRs are in about half that quantity and would stretch about half that distance again.

The American viewing public has yet to be exposed to broadcast DTV. Broadcasters have yet to ascertained a viable business plan for the new capital expenditures. Early predictions by the Consumer Electronics Manufacturers Association, CEMA, of the Electronic Industries Association, EIA, project first DTV receivers to have a cost of around $5,000. New VCRs will be required as well. These will be at least $1,000. The consumer's willingness to spend such unprecedented sums on a large scale for television has not been tested. There is the promise of digital set top adapters which consumers may purchase to obtain the digital programming on their existing receivers. These will be around the cost of cable set top boxes, but with a retailer's profit margin added in. This will make cable's $400 box cost at least $500. What motivation would drive consumers to spend this kind of money to get this programming? If they can afford it, they already have cable or DBS. If they can't afford cable or DBS, they certainly can't afford to purchase an expensive digital set top box adapter or and multi-thousand dollar digital TV receiver. Will those who concern themselves with the plight of the economically disadvantaged allow this ripe cause to pass? How can the government take away the utility of one of the disadvantageds' only sources of entertainment and information? The politics of this situation have not yet been seriously weighed. Long range business plans in such an environment cannot be credible.

It will take at least three years of availability at market for any uptick of consumer acceptance to materialize. Given the optimistic schedule of 40 top market stations on the air with digital signals by mid 1999, the migration won't be a national rush but rather a market by market crawl.

Color TV suffered for thirteen years from its introduction in 1954 until the full color prime time of NBC's 1967 season. Only then did the marketplace favorably respond. It should be noted that monochrome TV was a great improvement over radio. Color TV is a lesser improvement over monochrome TV. The High Definition Television, HDTV, flavor of DTV is a still lesser improvement over color TV. Digital "standard definition" television, SDTV, may be no improvement at all to a good analog NTSC signal!

Surprisingly, there is very little talk of, HDTV. Most of the broadcaster interest seems to be in SDTV. So the offering to consumers is not better TV, but more TV. That may not be an attractive trade off for the high costs of the new television receivers and VCRs.

While the FCC is postulating the return of analog broadcast spectrum, it has no authority to forbid cable from continuing to serve the huge installed base of analog TVs and VCRs. Such a vast marketplace cannot be ignored. Cable will continue to service this important audience.

From the broadcaster's point of view, it will be politically unacceptable to be denied access to analog television receivers if cable operators continue to have that market place.

From the consumers' point of view, it will be politically unacceptable to disenfranchise the consumer's television set. Consider the furor raised over the problems of cable's set top box and the potential interference with features of television receivers. Imagine rendering the television useless and requiring the purchase of an expensive new digital receiver or a digital set top box adapter!

THE MOTIVATIONS

Why at this late date when it appears that everyone is talking about digital television is there a reason to be interested in embedding digital data in analog signals?

Clearly, if analog television is not going to disappear quickly, there is a strong motivation to maximize the utility of the spectrum which must continue to be allocated to serving the existing market of analog television receivers and VCRs.

The Broadcaster's Opportunity

There are a number of opportunities for broadcasters. During the period of transition, when an analog and a digital channel are available, the digital capacity of the analog channel should not lie fallow. It could be put to good economic use.

If a digital signal is hidden in the analog signal, the value of the spectrum increases. Not only does the spectrum continue to serve those who cannot afford new receivers or adapters, but it also serves those who can make such a purchase. The electronics for digital reception is complimentary to that needed to access digital signals hidden in analog. When a digital receiver or set top box is not accessing a digital part of the spectrum, most of those same circuits can be extracting the digital signal out of the analog signal at little additional cost.

The broadcaster may find this double value of the existing spectrum to be a compelling political reason for retaining it when the appointed time for surrendering it arrives.

The Compatible Digital Cable Upgrade

Cable's digital video migration plans usually do not include comprehensive replacement of all analog channels because of the horrific expense of the digital set top boxes. Instead, a hybrid service, part analog, part digital is planned. Two approaches are possible. The first expands bandwidth and uses most or all of the new bandwidth for digital services. This entails considerable construction expense and may not be practical for smaller systems or in tight economic times. In the second approach, there is the intention of converting a few of the channels to digital and leaving the remainder as analog. In this strategy, the channels converted to digital will have previously been occupied by low penetration services. Subscribers wishing to continue with those services will need a digital set top box. If these subscribers take no new services, just the ones they had previously taken, their costs go up considerably while there is no increase in revenue. Subscribers who do not wish the new advanced services will not receive a new digital set top box. However, they will lose programming previously carried on the analog channels which are converted to digital. This can be a serious loss for low capacity cable systems.

Techniques which hide the data in the analog signal are an attractive alternative for the carriage of digital signals. Since television tuners are relatively inexpensive, multiple tuners can be provided so that data can be collected from more than one channel. That data can then be assembled to provide the MPEG streams needed to create new synthetic channels. Statistical multiplexing can be used to advantage. In this approach, all of the analog channels are preserved for those who are satisfied with the existing service. The existing analog subscribers will notice no difference other than the advertising programs promising more programming if they add digital services to their existing analog service. Only those willing to pay for more will incur the extra cost of the new set top box.

In recent years the metaphor of the "Chicken and the Egg" has given way to "Field of Dreams" concept of "Build it and they will come". Each equally points up how progress is frequently suppressed by codependent events. "Build it and they will come" requires an act of faith which is difficult for those who have to sign checks. While there will be some building, the process is helped along considerably if the methods of construction are cost effective.

It will be some time before any meaningful penetration of receivers for the DTV format emerge. It is axiomatic that these receivers will be compatible analog television. The need to support NTSC will be so important that a non NTSC compatible DTV receiver is not expected to be a starter in the US market. While the receiver is tuned to an analog channel, the digital circuits are idle. The Compatible Digital Upgrade allows those circuits to be put to use to provide a data stream for other purposes. This sharing of resources gives extra value to the investment in digital electronics.

The Compatible Digital Upgrade can be the bridge between the massive installed base of NTSC hardware and the yet unproved DTV Service. The building blocks that make up a DTV receiver are intrinsically compatible with the needs of the Compatible Digital Upgrade. A high quality tuner, good signal processing and an MPEG engine provide an ideal environment to support the Compatible Digital Upgrade.

Applications

There are a variety of applications for data in analog signals. Data can be supplied as just data. Alternatively, if sufficient capacity is available, data can be used to deliver digital video or digital audio services. These can result in an expansion of broadcast or cable capacity without interference to ordinary analog reception.

The data can be used with personal computers, special television sets or set top boxes or versions of the "net computer". When computers and computer adapters for television receivers are used, there are two modes: "pull" and "push". The "pull" mode is the traditional Internet approach where sites are accessed and data retrieved. This requires a two way connection. The "push" approach can be implemented in a one way system. Here, the user indicates his fields of interest and the data is retrieved from the data stream and loaded into storage. It then is displayed. In a two way media, "push" can be supplemented with web site access for more details.

A related application involves Datacasting. This is the inclusion of data in the broadcast television signal for use with a personal computer. The most aggressive such implementation is Intercast whose main partners include Intel and NBC. HyperText Markup Language (HTML) formatted Web pages are delivered in the VBI of the television signal. HTML is a method of linking information. Highlighted words or phrases can be "clicked on" and relevant information appears on the screen. In some cases, this is achieved by going to another location in the same document. In still other cases, data from another document is displayed. In other cases, locations on the World Wide Web are automatically accessed and information retrieved. A personal computer with a television tuner receives the signals and displays the video in a small window. The rest of the screen displays the HTML pages. The computer's hard disk can capture and store pages of interest. Since HTML pages are around 50 Kbytes each and most personal computers now come with at least a 1 GB hard drive, capacity is not a problem. In a major application of this technology, the pages downloaded pertain closely to the video programming. When several hundred pages are downloaded, the access speed is governed by the hard disk, not a modem. In affect, the server is built into the personal computer! However, a high speed data link keeps the contents of the hard drive fresh and yields a fast response to a new inquiry for information. The HTML nature of the pages makes accessing different parts of the data downloaded easy and familiar to any Web surfer. The HTML can include embedded links to related Web sites accessed with the computer's regular phone or cable modem. Access to these sites is automatic.

Cable vs Broadcast Data

Since cable's spectrum is much more well behaved than the broadcast spectrum, several significant advantages accrue. A time domain equalizer may not be necessary. If one is included, it may have relaxed specifications leading to lower cost. There is no "airplane flutter", i.e. Doppler effect from approaching or receding aircraft. Because the spectrum is better behaved, less error detection and correction is required for a given level of performance. This was well demonstrated in the Advanced Television Grand Alliance's modulation scheme. While 8-VSB is used for broadcast, 16-VSB was developed for cable allowing two HDTV signals in 6 MHz on cable. 16-VSB does not have twice the data capacity of 8-VSB. The doubling of payload comes because 16-VSB requires significantly less data protection. If this same approach is applied to the techniques proposed for data carriage in analog television signals, more of the raw data capacity can be harvested for payload purposes. This approach has not been well explored and offers a significant opportunity. An additional advantage is cable's availability of multiple channels to carry data. The data carrying capacity of a cable system is just huge!

HIDING DATA IN ANALOG SIGNALS

A few years ago, the question was asked if it was possible to hide enough data in an analog video signal to allow carriage of a separate digital video signal. At first, the prospects seemed dim. There are at least two reasons for this. First, it was thought that 5 or 6 Mb/s would be required for a digital television signal. Secondly, the early analysis was based on binary signals; i.e. signals with just two logic levels.

Subsequently, several things happened. First it became clear that MPEG encoding has improved dramatically so that fewer Mb/s are required for good quality results. Secondly, the use of multi-level coding makes it possible to convey multiple bits simultaneously. Advanced tuner technology makes it possible to consider two or more tuners in a receiver to extract data from more than one channel simultaneously. Lastly, very sophisticated processing is affordable in receivers. This processing allows compensation for problems introduced by extra data signals.

Data Under Visual

The new approach involves a data signal which is double sideband amplitude modulated onto suppressed carrier which is in quadrature phase with the picture carrier. If both the video signal and the new data signal were normal double sideband, they could be separated with synchronous detectors by conventional methods. Since the television signal is not all double sideband, but vestigial sideband, a television receiver includes a Nyquist slope filter to properly weight the upper and lower video sidebands around the carrier so that the correct amplitude is available for detection. A Nyquist filter is one which has an anti-symmetric characteristic about a critical frequency which is called the

Nyquist frequency. The exact shape of the filter is not specified, only that it is anti-symmetric about the critical frequency. The receiver manufacture has some latitude in the implementation of this filter.

The Nyquist filter is a serious problem for a double sideband signal which is modulated onto a quadrature carrier. In the course of its normal functions, this filter would convert a plain double sideband modulated signal (with opposite sidebands equal in amplitude to each other) into a double sideband signal with asymmetrical sidebands. Vector analysis reveals that when the two sidebands are symmetrical, there is only a resultant signal in phase with the carrier. Conversely, when the two sidebands are not symmetrical, there is an additional component in quadrature with the carrier. This newly formed asymmetrical sideband set would have an undesired component in phase with the video carrier which would cause interference. Stated another way, even though the new sideband set was initially placed on a suppressed carrier which is in quadrature to the picture carrier at the origination point of the signal, after being operated on by the receiver's Nyquist filter, a detector extracting the video signal would "see" unwanted components from the new sidebands. Consequently, quadularity would not be preserved between the visual carrier and the added data signal.

This problem can be averted by properly shaping the spectrum of the added data signal so that when it passes through the receiver's Nyquist filter, a double sideband spectrum in quadrature with the visual carrier and possessing equal amplitude sidebands will be obtained. Under these conditions, there will be minimal cross coupling of the quadrature signal's energy to the receiver's video detector. Therefore the receiver's detector will respond essentially only to the video signal. If the receiver utilizes a synchronous or similar behaving detector which inherently is immune to quadrature components, the added data signal will be essentially ignored.

The pre-shaping of the added data signal is done with a compensation network which includes a Nyquist filter representative of those found in the population of receivers exposed to the added data signal. In the event that the population consists of a mixture of differently shaped Nyquist filters, a composite signal optimizing the result can be implemented either with a parallel configuration of Nyquist filters fed with signal strengths in proportion to the numbers of the respective filters in the population or with a Nyquist filter designed to optimize the result using standard filter synthesis techniques.

In order to maintain a relationship that allows synchronous detection to separate the data signal from the video signal, the data signal should be limited to the frequency region over which the Nyquist filter operates. This is $\pm$ 750 kHz around the visual carrier. Two level data will accommodate 1.5 Mb/s of throughput. However, the signal to noise environment required to present acceptable pictures will support better than two level data. Four levels of data will allow two simultaneous bits of data for 3.0 Mb/s of throughput. While higher data rates may be possible, the signal levels required make the problem of avoiding interference with the video very difficult. Since going to three simultaneous bits would involve discriminating eight levels of signal, the point of diminishing returns may have been reached.

The addition of a data signal in quadrature with the video carrier will modulate the envelope of the resulting signal. If the receiver's detector is not a pure synchronous detector, that is, if it exhibits sensitivity to the envelope of the resulting signal, the data signal will contaminate the video. The problem is that there is a huge population of receivers already in existence whose performance must not be seriously impacted.

The solution to this problem is the introduction of abatement signals to counter the impact of the data. Since the phenomena are non-linear, the optimum abatement strategies will involve non-linear techniques. A couple of year's of intensive development work has resulted in successful techniques for achieving this goal. Just as in the MPEG model, the expensive processing is done at the point of signal origination. The receiver circuits are very cost-effective.

Data Under Aural

Additional information may also be added on the aural carrier of the NTSC television format. This is achieved through amplitude modulation of the aural carrier which is already frequency modulated by the TV audio and BTSC signals. The first caveat is that the AM modulation may not be full depth without corrupting the aural program information. The allowable depth of modulation is limited by the worst signal to noise ratio to be encountered in the service area of the broadcasts. To be approximately equivalent in both program material and data performance at the FCC "Grade B" contour a downward modulation depth of about one half voltage (6 dB) which corresponds to 33% modulation is appropriate, but other values can be used. At this depth of modulation multilevel data signals may be employed increasing the data carrying capacity of the channel. Modern television receivers generally perform with full limiting when receiving noisy video signals.

More efficient use of the available spectrum is achieved by multilevel encoding of data. In the aural data system, 2, 4, 8, and even 16 levels are appropriate for different signal quality environments. In an NTSC (type M system) the highest chroma frequencies utilized are 5.43 MHz above the lower band edge. This is arrived at by the visual carrier being 1.25 MHz above the lower band edge, the chroma subcarrier being 3.58 MHz above the visual carrier and the highest frequency chroma sidebands extending up to 600 kHz above the chroma subcarrier (1.25 + 3.58 +0.6 = 5.43). If it is desired to maintain the sanctity of this spectrum, the aural data, 8 level, 1.5 Mb/s signal will be implemented.

With four level data under visual contributing 3.0 Mb/s and eight level under aural contributing another 1.5 Mb/s, a total of 4.5 Mb/s is achieved.

Other Data Resources

Neither the data under visual nor the data under aural interferes with Vertical Blanking Interval (VBI) data transmission. Approximately another half Mb/s can be achieved in that manner. There are other methods for hiding data under consideration by others. Some of these methods can be employed along with these methods to achieve still greater total capacity.

The combined data capacity per analog channel is a significant resource which should not be ignored. The electronics for extracting that data is very cost effective. It is a trivial cost addition to a digital set top box or personal computer mother board. It is well within the customary prices paid for personal computer after-market plug in cards.

When two or more adjacent channels are simultaneously utilized, statistical multiplexing techniques can be employed to pack still more programming into the analog spectrum.

THE AUTHOR

Dr. Ciciora is a technology consultant specializing in Cable Television, Consumer Electronics, and Telecommunications. He serves on several Boards of Directors, Technology Advisory Boards, and does expert witness work.

Most recently he was Vice President of Technology at Time Warner Cable. Walt joined American Television and Communications, the predecessor to Time Warner Cable, in December of 1982 as Vice President of Research and Development. Prior to that he was with Zenith Electronics Corporation, starting in 1965. He was Director of Sales and Marketing, Cable Products, from 1981 to 1982. Earlier at Zenith he was Manager, Electronic System Research and Development specializing in Teletext, Videotext and Video Signal Processing with emphasis on digital television technology and ghost canceling for television systems.

He has ten patents issued, two more allowed but not yet issued and several more pending. He has presented over two hundred papers and published about a hundred, two of which have received awards from the Institute of Electrical and Electronic Engineers (IEEE). His papers have been translated into Japanese, Chinese, German and Spanish. Walt wrote a monthly columns for Communications Engineering and Design (CED) magazine and for Communications Technology (CT) magazine for three years each.

He currently serves on the Executive Committee of the Montreux Television Symposium. He was a member of the board of directors of the Society of Cable Telecommunications Engineers (SCTE) for six years. He was inaugural Chairman of the Technical Advisory Committee of CableLabs for four years and Chairman of the National Cable Television Association (NCTA) Engineering Committee also for four years. He was president of the IEEE Consumer Electronics Society for two years and is a past chairman of the IEEE International Conference on Consumer Electronics. He chaired the Joint Engineering Committee of the NCTA and the Electronic Industry Association (EIA) for eight years. He has served on several industry standard-setting committees. He currently co-chairs the Cable Consumer electronics Compatibility Advisory Group to the Federal Communications Commission and its Decoder Interface subcommittee.

Walt is a Fellow of three engineering societies the IEEE, the SMPTE, and the SCTE. Other memberships include Tau Beta Pi, Etta Kappa Nu, and Beta Gamma Sigma.

Current interests center on competitive technology, the consumer electronics interface with cable, Digital Video Compression, Interactive Television, Multimedia, and High Definition Television.

Walt received the 1987 NCTA Vanguard Award for Science and Technology and was named "1990 Man of the Year" by CED magazine. CED also named him "1993 Man of the Year". He was the Fall 1994 Levenson Memorial Lecturer at the National Cable Television Center at Penn State.

Walt has a Ph.D. in Electrical Engineering from Illinois Institute of Technology (IIT) dated 1969. The BSEE and MSEE are also from IIT. He received an MBA from the University of Chicago in 1979. He has taught Electrical Engineering in the evening division of IIT for seven years.

Hobbies include helping his wife with her horses, reading, wood working, photography, skiing, and a hope to someday become more active in amateur radio (WB9FPW).

Walter S. Ciciora, Ph.D.
45 Hulls Farm Road
Southport CT 06490-1027
Phone: 203-259-5183
Fax 203-259-0556
wciciora@aol.com

An Overview of the DOCSIS Two Way Physical Layer

Bill Kostka

Cable Modems and Cable Modem Termination Systems conforming to the Data Over Cable System Interface Specifications RF Specification provide system engineers with a new set of options that can be used to help ensure reliable data service delivery in cable plants.

A brief history of the decision process used in selecting which options to include is presented, but primarily, this paper describes the options available, in both the RF downstream and RF upstream directions, and helps clarify the criteria to be used when deciding how to apply those options when delivering digital data services.

INTRODUCTION

With the imminent advent of interoperable cable modems that conform to the Data Over Cable System Interface Specification (DOCSIS), cable operators will have many new, powerful options to help manage the delivery of two-way digital data services on their plant. This paper discusses some of the design decisions used to determine which options to include in DOCSIS systems.

Greater detail is provided on those options that can be optimized to suit the needs of a particular plant. Both the plant physical structure and the type of services to be delivered combine to dictate system settings. Options exist, in both the downstream and upstream physical layer, that enable DOCSIS systems to be customized in an optimum manner. The mechanisms employed are explained in enough detail to illuminate basic system operation.

Finally, a common side effect of normal operation of cable modems in general is discussed. This facilitates an understanding of the fact that optimum operation, of any cable modem system including those conforming to the DOCSIS specification, is a result of not just the cable modem system design, but also includes the basic cable plant design.

DESIGN PHILOSOPHY

Priority 1: Reliable Physical Layer Now

The highest priority used to decide what was in and what was left out of the DOCSIS RF specification was that the system must provide a reliable physical layer—NOW! The intention was to minimize the need to develop new techniques. These techniques, some being touted as “advanced” physical layers, are in the labs and in early trials, but would have added significant time to the front-end of any developments for compliant DOCSIS components.

The decision for DOCSIS was to use robust, tested techniques. If there existed a successful tool for delivering reliable data on a cable system, it was evaluated for inclusion into the specification. The best available implementations from the currently deployed field of cable modems were included in the specification

If there were any tried-and-true techniques that could be included, they were also investigated. For example, digital video is being delivered successfully—NOW! The same techniques can be used for delivering reliable digital data to cable modems too. Admittedly, some differences exist in the criteria needed for delivering digital video and those needed for delivering some forms of digital data services, but these are addressed in the specification.

Delivering reliable data was not enough. The specification needed to provide high-capacity in the digital data streams. The specification provides for two different kinds of downstream data channels with raw data rates of over 30 megabits per second (Mbps) or almost 43 Mbps. After removing the overhead for forward error correction (FEC), this still leaves information rates of approximately 27 Mbps or 39 Mbps respectively. Upstream data rates have many more possible formats which include a variety of raw data rates ranging from 320 kilobits per second (kbps) to 10.24 Mbps. These upstream channels are configurable to include a user-specified amount of FEC overhead, including none.

The specification also provides a growth path for cable operators delivering digital data services. Lower data rate, more robust channels can be deployed while plants are upgraded to provide a cleaner environment within which to deploy higher data rate channels when they are needed. Operators can also choose to deploy multiple upstream and multiple downstream channels to deliver a multiplicity of services to a multitude of subscribers.

DOWNSTREAM PHYSICAL LAYER

Downstream Physical Layer Design

The downstream physical layer is based on the International Telecommunication Union, ITU-T Recommendation J.83 (04/97), Digital Transmission of Television Signals, Annex B (ITU-T J.83B). This revision of ITU-T J.83B includes not only the original 64QAM modulation and a fixed depth interleaver used to deliver digital video, but also includes 256QAM for higher downstream channel data rates as well as a variable depth interleaver. DOCSIS compliant downstream channels may occupy any 6 MHz band between 88 MHz and 860 MHz.

64QAM versions of these downstream channels are beyond the test phases. They are deployed and successfully delivering digital video services to cable subscribers. 256QAM versions of these downstream channels have been proven in extensive rigorous tests. This technology is ready for use now.

The reliability of QAM modulated downstream channels is ensured because of the powerful concatenated FEC provided by the ITU-T J.83B specification. Multiple layers of error detection and error correction, coupled with variable depth interleaving to provide variable-length burst error resilience, deliver error rates ensuring customer satisfaction. The high data rates together with the low error rates provide a bandwidth efficient delivery mechanism for digital data delivery.

Downstream Error Protection

The target digital data delivery goal for these downstream channels is to have an error rate that yields less than one error in 15 minutes in a nominal cable system. Even with reduced noise margins, the downstream channels are designed to deliver 64QAM signals with a bit error rate (BER) of less than 10^{-8} at a carrier to noise (C/N) ratio of 23.5 dB. 256QAM channels can be expected to deliver a similar BER at 30 dB C/N. This translates to one error every 3 to 5 seconds. These downstream channels are shared between subscribers, so the error will naturally be distributed randomly to various subscribers. An additional benefit from the strength of this FEC is that it permits operation of the downstream digital data channels 10 dB lower than the nominal level of video carriers on the system. This helps minimize system loading while still delivering robust digital data services.

It is important to note that deploying 256QAM channels requires significantly higher C/N than when deploying 64QAM channels. This is one of the first choices to make when deploying DOCSIS modems: deciding which mode of operation to use for downstream channels. In a clean, new HFC plant that is currently delivering video signal to noise (SNR) of 49 dB or greater to the "big-screen tv" subscribers to successfully compete with the satellite delivery services, then the system may well support 256QAM downstream channels. On the other hand, if the plant consists of longer cascades working nearer to the FCC minimum system specifications while it is being upgraded to a HFC topology, consider deploying 64QAM channels instead.

One of the side effects of the interleaver is it adds latency in the downstream channels. The process of interleaving the outgoing symbols, shuffling the position of the symbols so that normally adjacent related symbols are now separated by unrelated symbols that would otherwise be transmitted later, delays the delivery time of related symbols. The benefit is that a burst of errors that damages adjacent symbols in transmission, damages only unrelated symbols. The FEC can then correct the damaged symbols once they are reshuffled back into their normal order as long as the burst damage did not span too many related symbols. There is an intrinsic relationship between the depth of the interleaving and the latency incurred by the interleaving. The deepest interleaving depth available in the DOCSIS RF specification provides 95 microsecond burst protection at the cost of 4 milliseconds of latency. Four milliseconds of latency is insignificant when watching digital video. If the only digital data services being provided are web browsing, e-mail and Internet access, then subscribers are spending seconds or more watching the message "Host contacted, waiting for reply" so 4 milliseconds is again insignificant. When deploying a near real-time constant bit rate service like IP telephony that requires an end-to-end latency of 20 milliseconds, it may be ill advised to squander 20 per cent of that budget on the downstream interleaver needlessly. The variable depth interleaver enables the system engineer to trade between how much burst error protection is required in the system and how much latency can be tolerated by the services being delivered. This is the next choice to be made when deploying DOCSIS modems: how much interleaving depth is required in this particular cable system.

UPSTREAM PHYSICAL LAYER

Flexible F/TDMA Design

DOCSIS compliant upstream channels provide for both frequency domain multiple access (FDMA) and time domain multiple access (TDMA). FDMA is provided by the ability to have multiple upstream channels simultaneously supporting multiple modems. This is the traditional realm of CATV low split systems wherein upstream channels reside in the spectrum between 5 MHz and 42 MHz. Downstream channels, as noted previously, utilize the spectrum between 88 MHz and 860 MHz.

TDMA is provided by the use of "slotting" on the upstream channels. Each upstream channel is divided into equal-time segments called "mini-slots". The use of each and every mini-slot is controlled by the Cable Modem Termination System (CMTS) at the head end. The CMTS assigns contiguous intervals of mini-slots to individual cable modems, or makes them available for contention by groups of cable modems, or opens them up for contention by all modems. Additionally, the type of communication within the assigned interval is dictated by the CMTS. All DOCSIS compliant cable modems will time-coordinate all their upstream transmissions so that they only transmit within the appropriately allocated interval. This provides the mechanism for multiple access in the time domain.

Flexible Upstream Channel Parameters

In addition to the ability to have different upstream channels on different frequencies and have different mini-slot sizes on different channels, many other upstream channel parameters have options that need to be set to meet individual system needs. Each upstream channel has an assigned bandwidth associated with it. The occupied bandwidth is directly related to the channel's data rate. DOCSIS compliant upstream channels occupy bandwidths of 200, 400, 800, 1600, or 3200 kHz. This corresponds to channel data rates of 160, 320, 640, 1280, or 2560 kilosymbols per second (ksym/sec). The specification provides for both QPSK and 16QAM transmissions upstream which allows us to have either 2 bits per symbol or 4 bits per symbol respectively. Channel data rates are available between 320 kbps (QPSK at 160 ksym/sec) and 10.24 Mbps (16QAM at 2560 ksym/sec). While the bandwidth is fixed for any upstream channel because the symbol rate is defined for that channel, the bit rate on the channel is variable because independent transmissions on that channel can be either QPSK modulated or 16QAM modulated.

Another powerful option available in DOCSIS compliant systems is the ability to flexibly define the amount of FEC protection included with certain types of upstream transmissions. Within an operational system, there are many kinds of packets being exchanged. There are relatively frequent housekeeping messages. The loss of one or more of these messages due to data errors is relatively insignificant because a virtually identical message will occur again in a relatively short time anyway. Other messages may be either time critical, like IP telephony packets, or will need to be retransmitted if lost, like TCP packets. If the packet is large, transmitting it again may consume more time than desirable.

The DOCSIS specified flexible FEC coding enables the system operator to set the size of the error protected data blocks and to set the number of correctable errors within each block. FEC changes can be done while the system is operating normally. In previous proprietary cable modem systems, when impairments in a data channel caused too many errors, the only solution was to abandon that frequency and hope to find a cleaner portion of the spectrum to place the channel. While DOCSIS compliant systems can do that too, the flexible FEC coding option enables the system operator to choose to stay on the same frequency by simply increasing the error protection on that channel. Even though the additional few bytes of error protection reduces the channel information rate a small amount, it makes the overall system capable of a much higher upstream spectral utilization.

The system operator now has the option to tailor each upstream channel to suit plant needs. Multiple modes of operation are available within one plant. Upstream feeds from portions of the plant which have more ingress or more homes passed, can have the upstream parameters set to provide more robust transmissions. The system operator could enable a more powerful FEC setting, lower symbol rates, and choose to use QPSK modulation. For newer nodes in the plant where ingress and noise are less of a problem, the upstream parameters could be tailored for more efficient transmissions. The system operator could reduce the FEC overhead, use higher symbol rates, and choose to use 16QAM modulation.

Table 1 itemizes the upstream physical layer features and the benefits they provide:

Feature	Benefit
Frequency agility	Ingress avoidance
Variable Reed-Solomon forward error correction	Adjustable amount of error correction
Variable Reed-Solomon block size	Individually tailor for large and small packet sizes
Multiple symbol rates	Fit channels into available spectrum
Wide channels with FEC	Ingress mitigation
Narrow channels	Intersymbol interference and reflection mitigation, ingress avoidance

Table 1. Features and benefits available in DOCSIS upstream channels

BURST PROFILES

The mechanism defined in the DOCSIS RF specification for managing the characteristics of upstream transmissions utilizes configurable “burst profile” parameters. Each upstream transmission is independent and is separated by an unoccupied guard time from other transmissions. An upstream transmission is also known as an upstream “burst”. As previously noted, each burst must be exactly timed to occur within an allocated interval of mini-slots.

Burst profiles are defined in the CMTS on a per-upstream-channel basis and are passed to all cable modems using that upstream channel. All cable modems using the same upstream channel will therefore be using the same burst profiles. Cable modems on other upstream channels will use an independent set of burst profiles that may or may not be duplicated by burst profiles in use on any other upstream channel. This is the mechanism that allows a CMTS to set cable modems for most efficient transmissions based on the characteristics of that particular upstream channel.

Burst profiles are stored in the cable modem allowing rapid selection of optimum transmission modes for different types of bursts. Six active burst profiles are stored in each cable modem for the following types of bursts:

- Initial maintenance
- Periodic ranging
- Request
- Request with contention data
- Short data
- Long data

The first two burst types are used for initialization and routine maintenance. The middle two types of bursts are used by the cable modem to request an allocation of mini-slots. The last two are the burst types that carry most of the traffic in the system. A service like IP telephony will generate a steady stream of short packets, while a TCP/IP function like ftp will generate a sequence of short bursts of long packets. It makes sense to FEC encode these two types of packets differently to ensure reliable delivery with minimal overhead. There is no opportunity to retransmit the short packets because all are needed in near real-time. While retransmitting the long packets wastes system resources, it can be acceptable if needed only rarely. Items to consider for optimization include the modulation type, the FEC frame size, and the number of FEC correction bytes.

Table 2 details the burst profile parameters which must be defined for each burst profile. Most of these will be set to match the CMTS receiver needs, but as noted above, some of the parameters are used to optimize the system.

Parameter	Use or possible value
Modulation type	QPSK or 16QAM
Differential Encoding	On or Off
Preamble length	Set for receiver synchronization
Preamble starting point	Set for receiver synchronization
Scrambler seed	Set for receiver synchronization
Scrambler enable	On or Off
FEC correction bytes	Optimize: 0 to 10 bytes
FEC codeword length	Optimize: 16 to 253 bytes
Last codeword length	Normal or Shortened
Guard time	Dead time between bursts
Maximum burst length	Optimize for data services

Table 2. Burst profile parameters to be set in each DOCSIS upstream channel

User Unique Burst Parameters

In addition to the upstream channel parameters, mini-slot size, symbol rate, and frequency, and the burst profile parameters listed above, there is another set of parameters that are adjusted automatically by the CMTS and the cable modems. These parameters are unique for each cable modem in the system. Interesting side effects can occur if deliberate attention is not paid to these effects. Parameter settings unique to each cable modem include the upstream transmit power setting, minor frequency adjustments, and a timing or ranging offset.

Ranging Concepts

Each cable modem in the system is a unique distance from the CMTS. Each cable modem, therefore, will have unique timing and power requirements imposed upon it. The transmitter power required from the cable modem is determined by the home wiring environment, drop and cable lengths, and system components between the cable modem and the first return amplifier. The DOCSIS system design requires the CMTS and the cable modems iteratively communicate details about the power level received at the CMTS, using initial maintenance and periodic ranging intervals. The CMTS will read the incoming power level of a transmission from a cable modem, then determine whether the input level is too low, too high, or within the optimum range. The power level adjustment needed to center future incoming transmissions in the desired power window is transmitted to the cable modem. The cable modem adjusts its transmit power accordingly. This process is iteratively repeated until the optimum received power level of upstream transmissions is achieved by the cable modem. This process is performed during the initial maintenance interval when a cable modem first comes online and is repeated in each regularly scheduled periodic ranging interval. This ensures continued reliable communications between the CMTS and the cable modems.

As with power, there may be subtle differences between various cable modems due to normal component tolerances which cause the upstream transmit frequency to differ slightly from it nominal setting. To accommodate this difference, a frequency offset is also iteratively communicated until the optimum received frequency of upstream transmissions is achieved by the cable modem. The final frequency offset will be 10 Hz or less when ranging is successfully completed.

Finally, during the ranging process, the cable modem is assigned a timing offset that ensures its upstream transmissions arrive exactly within the allocated mini-slot. This is accomplished by the cable modem transmitting earlier than the assigned slot time so that delays caused by interleaving latency in the downstream, propagation in the system, and fixed processing overhead in both the CMTS and cable modem are negated. Ranging requires that the two-way round trip delay be negated because all upstream transmissions must align with the mini-slot timing as viewed by the CMTS. After a CMTS assigns the slot timing, it must then communicate that timing to the cable modem. The communication is delayed in the downstream direction, by latency, downstream propagation delay, and processing overhead. When the cable modem sends a transmission upstream, it is delayed by the upstream propagation delay and processing overhead. The sum of these delays is effectively removed by the ranging process.

Node Combining

The following is not solely a DOCSIS issue because cable modem systems in general have a ranging function that can result in this situation. It is, nevertheless, one of the more interesting side effects of the ranging process that occurs when combining nodes with different optical losses. At the head end, if these nodes are directly connected to a fixed loss combiner/splitter to route upstream signals to the CMTS, and possibly other digital data service devices, then the power level normalizing ranging process can create an undesirable situation.

The CMTS steers the cable modem transmit power to provide a constant RF level at the input of the CMTS. This translates to a constant RF level at the input to the combiner/splitter network that is the same point as the output of the upstream laser receiver. Because the upstream optical paths have different losses, the CMTS steers the cable modems to provide an uneven RF input to the return laser transmitter. While one of the nodes may have an optimum input RF level, all the other nodes will not.

Complications can occur at both ends of the variations. At the link with the most loss, there may not be enough cable modem transmit power available to transmit upstream through the in-home wiring, through a high-value tap, and up to the laser return transmitter. This can result in a lower than expected upstream received RF level even though the cable modem is transmitting at full upstream power. At the link with the least loss, the cable modems will be directed to lower their RF transmit power. This leaves the input to that node's laser transmitter with lower than desirable carrier to ingress ratios and the significant noise levels the input to the combiner/splitter network. The combination of the two ends yields a lower than desired carrier level together with a higher than desired noise level and ingress environment. The result can be an unreliable digital data delivery system.

The situation occurs because the cable modem system carrier level self adjustment has been combined with an unbalanced optical link and the overall system has become unbalanced. The remedy is to balance the optical loss of each laser return link by either adding optical attenuation or, more easily, the equivalent RF attenuation to the output of each return link. Once the return links are matched in loss, the system can be aligned so that not only not only is the input to the CMTS optimized by the ranging process, but with balanced return paths to each node, the system can optimize the input to each return laser link also. The result is a well-balanced system that can deliver reliable two-way digital data services.

GOING FORWARD

Going forward, there will soon be many new powerful physical layer options cable operators can use to help manage the delivery of two-way digital data services on their plant when deploying DOCSIS compliant systems. The options that exist in both the downstream and upstream physical layer enabling DOCSIS systems to be tailored in an optimum manner have, hopefully, been explained in enough detail to illuminate basic system operation. Ultimately, the quality of the delivered service will be perceived through the many facets of sales, installation, and support. With the new options available in these systems, reliability of the physical layer will not be perceived as the weak link in the chain.

Bill Kostka
Project Director
Cable Television Laboratories, Inc.
400 Centennial Parkway
Louisville, CO 80027-1266
(303) 661-3844

Mr. Kostka is a project director at Cable Television Laboratories, Inc. (CableLabs). He joined CableLabs' Digital Network Technologies team in August of 1997. His focus has been on DOCSIS (Data Over Cable Service Interface Specification) cable model physical layer and test issues of the DOCSIS specification.

Prior to joining CableLabs, he was a member of the Hewlett-Packard QuickBurst Cable Modem and the Kayak Digital Set-top Converter projects. His activities with Hewlett-Packard also included managing test and support teams for various Hewlett-Packard product lines since 1987. Prior to Hewlett-Packard, he served in the Cable TV industry with Gill Cable TV in San Jose, CA for over ten years in various capacities, including engineering manager.

CableLabs is a research and development consortium of cable television system operators representing the continents of North America and South America. CableLabs plans and funds research and development projects that will help cable companies take advantage of future opportunities and meet future challenges in the cable television industry.

CABLE HEADEND INTERFACES FOR DIGITAL TELEVISION

Mukta Kar, Ph.D., Majid Chelehmal, Ph.D., and Richard S. Prodan, Ph.D.
Cable Television Laboratories

Abstract

Advanced digital technology, combining compressed digital video with digital transmissions, can deliver multiple programs containing superior quality audio, video, and data, using the spectrum required for a single analog program. Broadcast affiliates will migrate to broadcasting digital signals starting in 1998, per FCC guidelines. By the year 2006, terrestrial broadcasting of analog TV should be retired and replaced by digital television (DTV).

Currently, most television viewers receive terrestrial broadcast signals over cable systems. An interface specification needs to be defined for receiving and processing the broadcasters' DTV signals compliant to the Advanced Television Systems Committee (ATSC) standard at the cable headend and through the subscriber receiver equipment. Such an interface specification must deal with some of the issues related to the integrated delivery of broadcast signals with other analog and digital signals originating from various sources.

This paper will explore the various types of signals, information, and associated interface mechanisms in the headend, and their relation to subscriber receiver equipment.

INTRODUCTION

Television has become the most popular medium for news, entertainment, education, and other information in the United States. Broadcast television signals are transmitted over media such as cable, air (terrestrial, MMDS, LMDS, etc.), or satellite. Broadcasters (cable, terrestrial, or satellite) represent independent enterprises free to make their own business decisions and to choose the broadcast contents desired by their subscribers. But in the national interest, broadcasters provide the general public with news, presidential addresses, and political debates of national importance. In the same light, local news and events must be delivered either by local cable or off-air broadcast services. More importantly, there must be a means to facilitate the pre-emption of regular programming with the Emergency Alert System (EAS) signal, from a central control facility, to all systems connected in a network.

Although there are various types of broadcasting entities in existence today, cable and terrestrial broadcasts reach the overwhelming viewer majority (satellites, and others, reach a small minority). In order to reach most of the U.S. viewers for emergencies, or for news of national and local interest, Congress instituted the "Must Carry" requirements that cable networks carry at least one local channel of each of the major terrestrial broadcast networks (ABC, NBC, CBS, PBS, etc.).

It has been proven in the laboratory, and in cable and satellite field testing, that compressed digital television technology is very spectrum-efficient. At the same time, it delivers superior quality video and audio compared to its analog counterpart. Another benefit of this digital technology is its flexibility. Enhanced television services provide internet access, home shopping, investing and banking from home, games, etc., in addition to passive entertainment. Technologists, businessmen, and policymakers agree that it is time to transition from analog to digital television. To facilitate the transition, the FCC has put forward guidelines requiring terrestrial broadcasters to switch from broadcasting analog television to broadcasting the new digital television standard, beginning in November 1998. Analog broadcasting will be retired completely, and the analog spectrum will be returned by the year 2006.

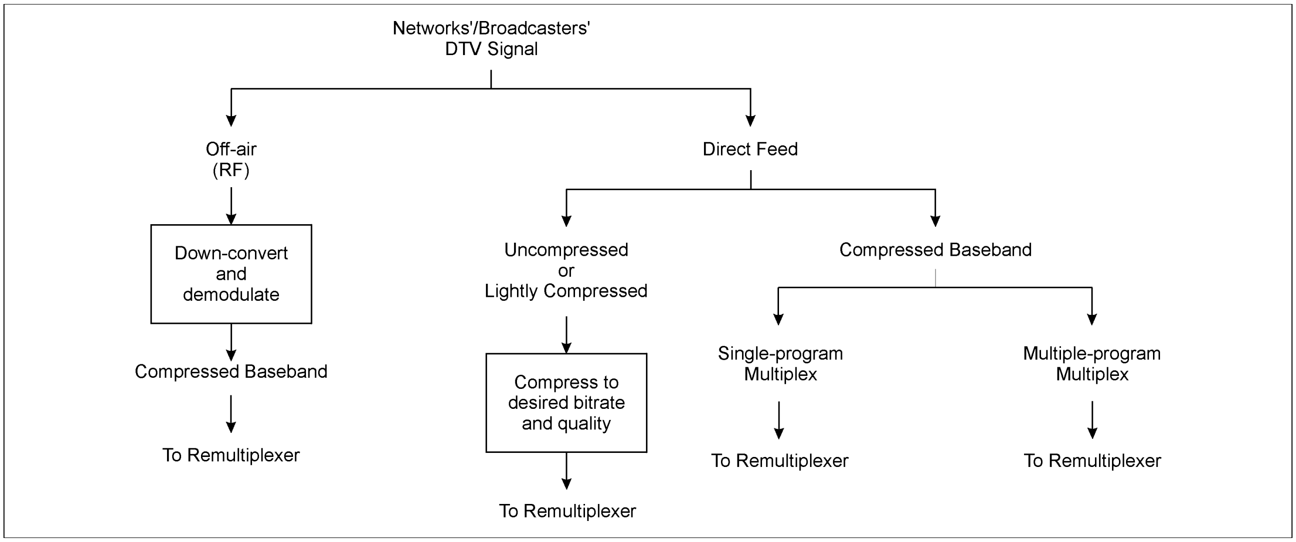


Figure 1. Various forms of digital television (DTV) signals

Terrestrial broadcasters will begin broadcasting a superior quality digital television (DTV) signal, in particular, high definition digital television (HDTV). Many viewers will be impressed with HDTV. Initially, some new HDTV receivers will be available when HDTV broadcasting begins. In addition to delivering superior picture quality, digital technology provides a higher efficiency in spectrum utilization and a higher level of security against theft of services. The gradual switch to digital technology will free up some physical channel capacity, which may be used to add new, or enhance existing, services.

Content providers will switch from distributing programming in analog to digital format. The cable industry needs to prepare to receive digital signals and to interface them seamlessly to headend equipment. Cable telecommunication engineers will have to deal with different signal types, which headends may transport, and which must interface with other headend equipment.

Figure 1 shows a typical combination of DTV/HDTV baseband signals. In the compressed domain, the combined signal is a single- or multiple-program MPRG-2 transport stream. The signals contained in the multiplexed transport stream originate in different formats and must be handled accordingly. Following are descriptions of some popular formats.

COMPRESSION FORMATS

Uncompressed Video Signals

The CCIR 601 digital format is usually the most popular. The luminance and chrominance (two color components) signals are sampled in a 4:2:2 structure with the resultant active picture size of 720 pixels by 486 lines (System M NTSC scanning). Each pixel is digitized with 8 bits (10 bits optionally). The bitrate required to transmit standard definition (SD) video is around 216 Mbits/sec (8-bit video) and greater than 1 Gbits/sec for high definition (HD) 8-bit video. The advantage of receiving uncompressed signals is that they can be compressed at the headend to the desired quality and bitrate. This provides flexibility, but requires an encoder for each uncompressed signal.

Compressed Video Signals

Any video signal, with bitrate reduced from its uncompressed rate by removing redundancy, can be categorized broadly as a compressed signal. Here it means specifically the main profile at main level (MP@ML) compressed signal for SDTV and main profile at high level (MP@HL) for HDTV defined in the MPEG-2 standard. The typical bitrate for a SDTV (720 x 480 pixels) signal is 2 to 6 Mbits/sec, and that of HDTV (1080 x 1920 interlaced) is 12 to 19 Mbits/sec. In either case, a compression ratio of 50 or better has been achieved. The sampling structure for the CCIR-601 standard is 4:2:2, but the MPEG-2 standard uses a 4:2:0 sampling structure. As a result, the MPEG-2 compression method sub-samples the chroma components of the video signal in both horizontal and vertical directions before compression. This is one of the main factors in achieving high compression with a minimal subjective loss in video quality. If this compressed signal has to be processed (decoded and re-encoded a few times), the picture quality of the compressed signal deteriorates significantly. Also, if video signals are compressed, the quality after the first compression cannot be improved by decoding and, subsequently, re-encoding at a higher rate. These shortcomings are mitigated by using lightly compressed signals.

Lightly Compressed Video Signals

A lightly compressed video signal is minimally compressed to around 45 Mbits/sec to reduce signal transmission bandwidth with nearly transparent picture quality. However, this type of signal needs to be decompressed and recompressed to a lower bitrate and quality for final distribution. The handling complexity of these signals is similar to that of uncompressed video signals. To preserve the loss in quality inherent in standard MPEG-2 compression, a profile known as 4:2:2 has been added to the MPEG-2 standard. In the 4:2:2 profile at main level (4:2:2P@ML), picture quality has been preserved in two primary ways. Unlike other profiles, chroma components have not been sub-sampled and the original CCIR 601 signal sampling structure remains. Greater quantizer precision may be used for encoding DC- and AC-coefficients of the 8 x 8 discrete cosine transform (DCT) block. This implies that the bitrate should be fairly high. For these reasons, the typical bitrate for the 4:2:2 profile is 20 to 50 Mbits/sec. It should be mentioned that at a higher bitrate, the 4:2:2 profile can provide a superior quality picture after several decoding and re-encoding iterations.

TRANSMISSION FORMATS

Off-air Signals

Off-air signals may be defined as modulated radio frequency (RF) signals which use free space as the medium of propagation and transmission. In this method, an RF-carrier is modulated by a digital baseband signal using a vestigial sideband (8-VSB) or a quadrature phase shift keying (QPSK) modulation method. Terrestrial broadcasters will use an 8-VSB DTV signal, and satellite networks will use QPSK-modulated carriers to distribute their content. Reception of these signals will be accomplished using an appropriate antenna and an integrated receiver decoder (IRD) to demodulate back to the baseband. The IRD provides the baseband multiplex signal. Some IRDs will have the capability of decrypting the compressed baseband signal. This signal may be used as input to a remultiplexer or as input for some other application. There is another option available for 8-VSB modulated signals—if the headend passes through them, then no demodulation is required. However, the signals need to be frequency-translated to a desired cable channel. This wastes channel capacity, since a quadrature amplitude modulation (QAM) cable signal carries 50 to 100 percent more payload.

Direct-feed Signals

Signals received through a guided medium, such as coaxial or fiber optic cable, to the headend may be considered direct-feed. Direct-feed provides a less noisy signal with a high signal-to-noise ratio (SNR), an advantage for analog signals. Analog signals may be received at the baseband or RF using direct-feed. For digital direct-feed, the RF-modulated signal provides little or no improvement over off-air RF signals above threshold, but can provide a consistent quality signal. Receiving a direct-feed baseband signal does not require an IRD for demodulation.

MULTIPLEXING FORMATS

Single-program and Multiple-program Multiplexes

A single-program transport stream multiplex, consisting of video, audio, data streams, and associated multiplexing information, is known as a single-program multiplex. A transport-stream multiplex consisting of two or more programs of video, audio, data streams, and associated multiplexing information, is known as a multiple-program multiplex. Content providers may distribute their programs to headends and affiliates as a single-program multiplex. Broadcasters may use either a single-program multiplex for HDTV, or a multiple-program multiplex for standard DTV, depending on the time of day. For example, broadcast MP@HL format (1920 x 1080 interlaced) for HDTV requires a bandwidth of 19 Mbits/sec and may be broadcast using a single-program transport stream multiplex. It is also possible to add one SDTV program to a lower bitrate HDTV program and broadcast as a multiple-program multiplex. For broadcasting SDTV programs, broadcasters may bundle several programs in a multiplex.

MULTIPLEXING OVERVIEW

Figure 2 shows a block diagram of an encoding system with an output multiplex bitstream consisting of one or more programs. The encoder compresses audio and video hierarchically into compressed data by removing spatial and temporal redundancies. It then encodes the compressed data using variable length code (VLC, a combination of run-length and Huffman codes). Encoded compressed streams of audio/video are known as elementary streams (ES). The compressed ES is broken into variable length packets containing one or more access units (frames). These variable length packets are known as packetized elementary streams (PES). The packetizer in the multiplexer combines PES packets into fixed-length units of 188 bytes—184 bytes used for payload of PES data and 4 bytes for headers. These fixed-length packets contain a packet identifier (PID) in one of the header fields, allowing easy separation and recombination of the component streams.

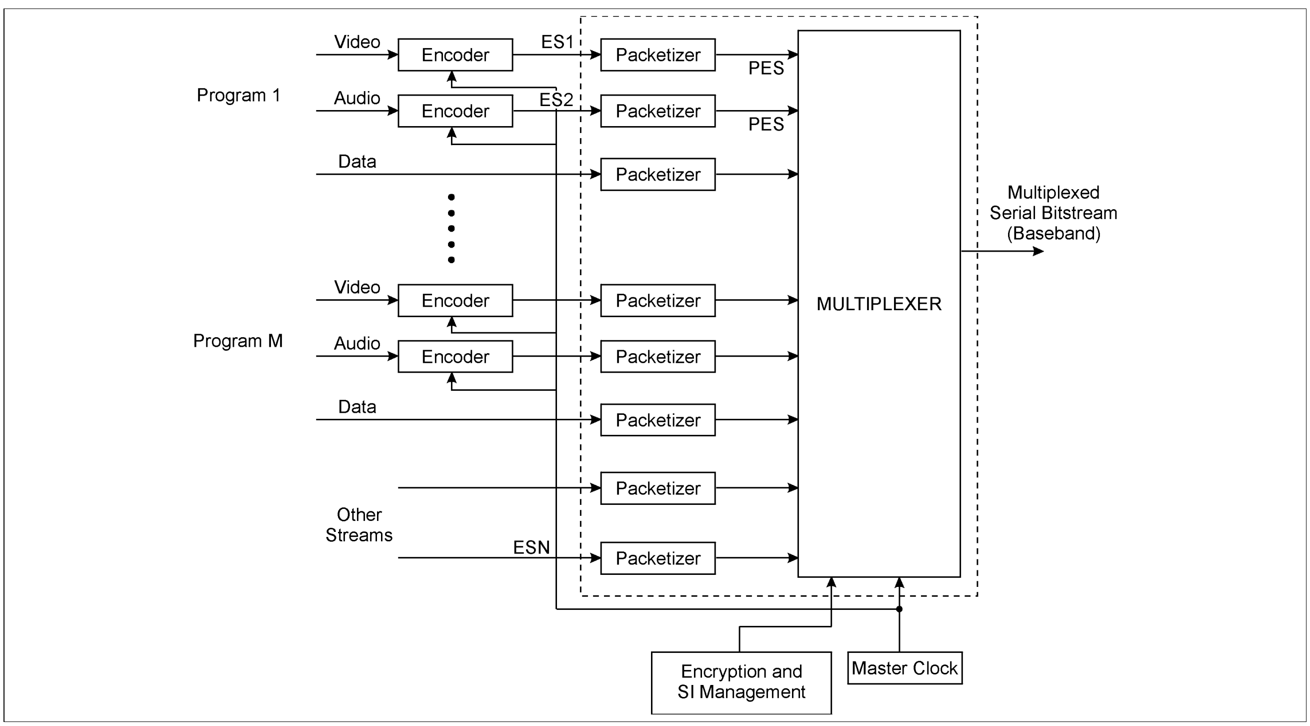


Figure 2. Block diagram of an encoding and multiplexing system

A multiplexer is a device in which various input and output streams may be programmed (*e.g.*, number of elementary streams and their bitrates, grouping elementary streams into programs, PID assignments to each ES, bitrate of the multiplexed output stream, etc.). Input packets (containing audio, video, and data streams) are multiplexed into a serial stream based on the input bitrates of the individual streams. A multiplexer also inserts packets containing system information describing the relationship of the various component streams into programs. This system information is essential at the receiving end for correctly demultiplexing the elementary streams. The multiplexer maintains a constant output bitrate. That is,

$$\sum_{k=1}^{N} bitrate(k) + bitrate(sys) \leq mux\ output\ rate$$

where: N = Number of ES inputs to the multiplexer;

bitrate(k) = bitrate of the individual ES;

bitrate(sys) = bitrate required by the multiplexer to insert system information such as PSI (program specific information), conditional access information (CAI), etc.;

Mux output rate = desired output bitrate of the multiplexer.

The multiplexer is also responsible for synchronizing all audio and video streams with a master clock. This is accomplished by inserting time stamps, such as a decode time stamp (DTS), a presentation time stamp (PTS) for audio/video frames, and a program clock reference (PCR) synchronized to a master clock, into the elementary streams. Another function of the multiplexer is to encrypt video and audio payload and other data (if necessary).

All multiplexed baseband signals are input to a remultiplexer in the headend, as shown in Figure 3. The functions of the remultiplexer are similar to those in the multiplexer, with some differences. A remultiplexer takes one or more transport stream multiplexes as input, and outputs a new multiplex using user-selected programs and data from input. To accomplish this, the remultiplexer has to decrypt and demultiplex the input streams, or simply unbundle the individual programs from the input multiplex. A new package of services is created from the operator-selected programs, with encryption as necessary. During demultiplexing, program specific information (PSI) is extracted from the input multiplexes and new PSI is inserted into the output multiplex in coordination with the headend management system.

Another important function of the remultiplexer is to remove timing jitter in the selected programs from the incoming multiplexes. During transmission due to nonuniform switching delay, jitter in PCR values is introduced. The PTS and DTS values of audio or video access units (AU) may not match with the PCR. For example, the DTS or PTS of the frame expired before the entire frame is available in the decoder buffer. To remove such transmission related jitter, a remultiplexer regenerates a local clock for each selected program in the multiplex. Comparing the local clock with the PCR of the incoming stream and number of bytes received since last PCR values, the remultiplexer can find the variation in PCR values. If the absolute value of the jitter is greater than the pre-selected tolerance, the remultiplexer removes the jitter and replaces the PCR with an adjusted PCR. The remultiplexer readjusts DTS and PTS values according to the adjusted PCR value.

To meet the constraint in the above equation, it may be necessary to reduce the bitrate of some video elementary streams. This may be achieved by using a recoder. A recoder uncompresses a compressed stream and recompresses and re-encodes to meet the desired bitrate. Re-encoding, using previous encoding decisions passed from the decoder, will minimize picture quality degradation by simply re-quantizing. However, one may expect some degradation in picture quality with a reduction in bitrate.

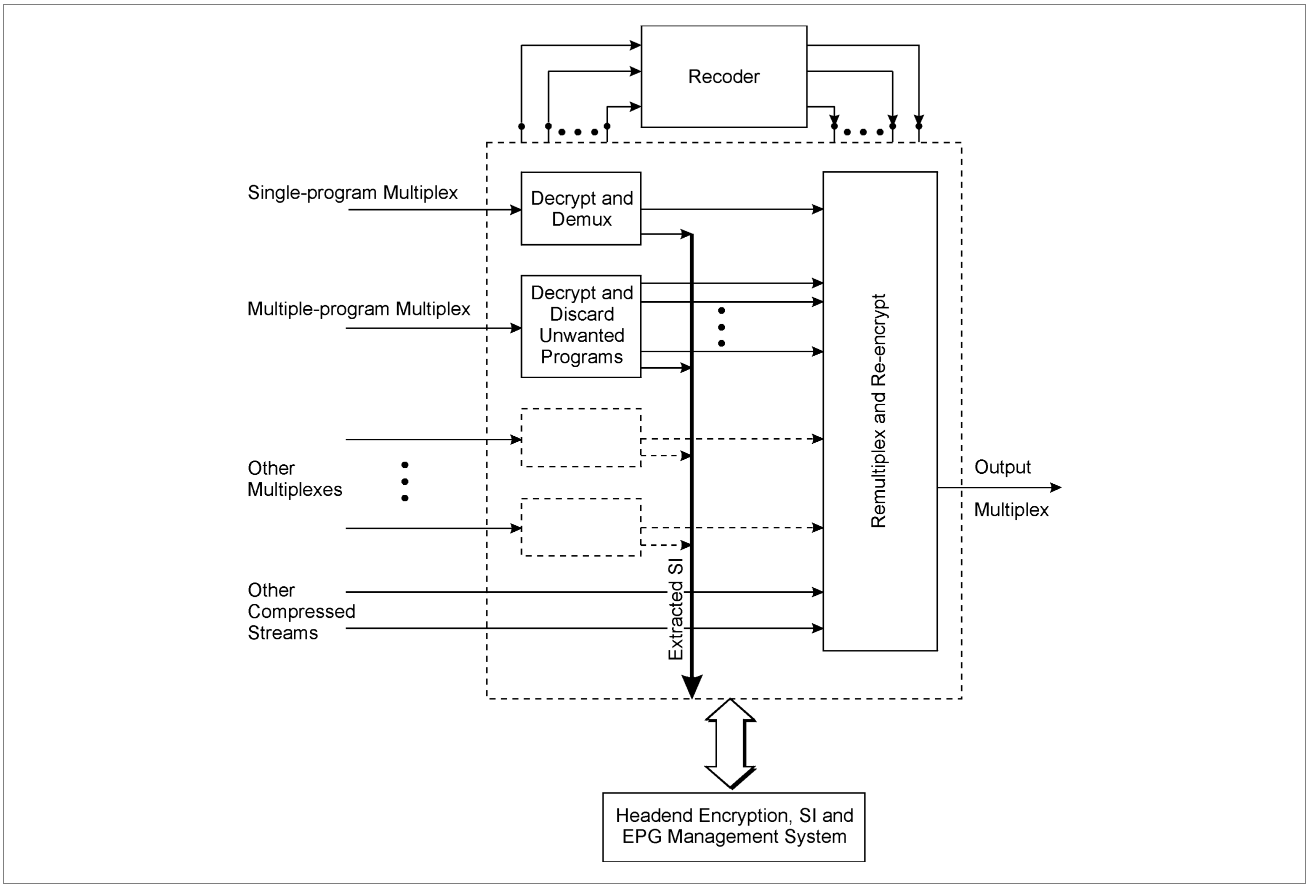


Figure 3. Simplified diagram of a remultiplexer

On occasion, it may be necessary to groom a program from a multiplex. For example, a network program has an overall bitrate of 19 Mbits/sec; 13 Mbits/sec provide actual program content (video and audio) and 6 Mbits/sec provide private data within the video PID or in a separate PID stream. This additional private data may be groomed and the recovered bandwidth may be used for other revenue-generating programs and services. A remultiplexer can simply filter data residing in a separate PID. If the data is not in a separate PID, external equipment may be needed to detect and remove such data.

Typically, a remultiplexer will have different input/output interface cards to accept signals from interfaces such as OC-3, DS-3, parallel and serial DVB, digital headend interface (DHEI), etc. Depending on the interface requirements and bitrates, several different sources may be used for the incoming multiplex streams.

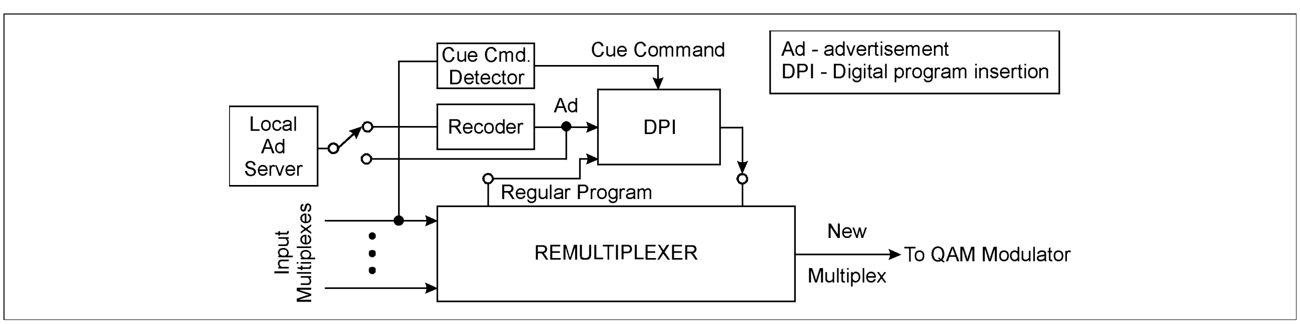


Figure 4. Local program/advertisement insertion

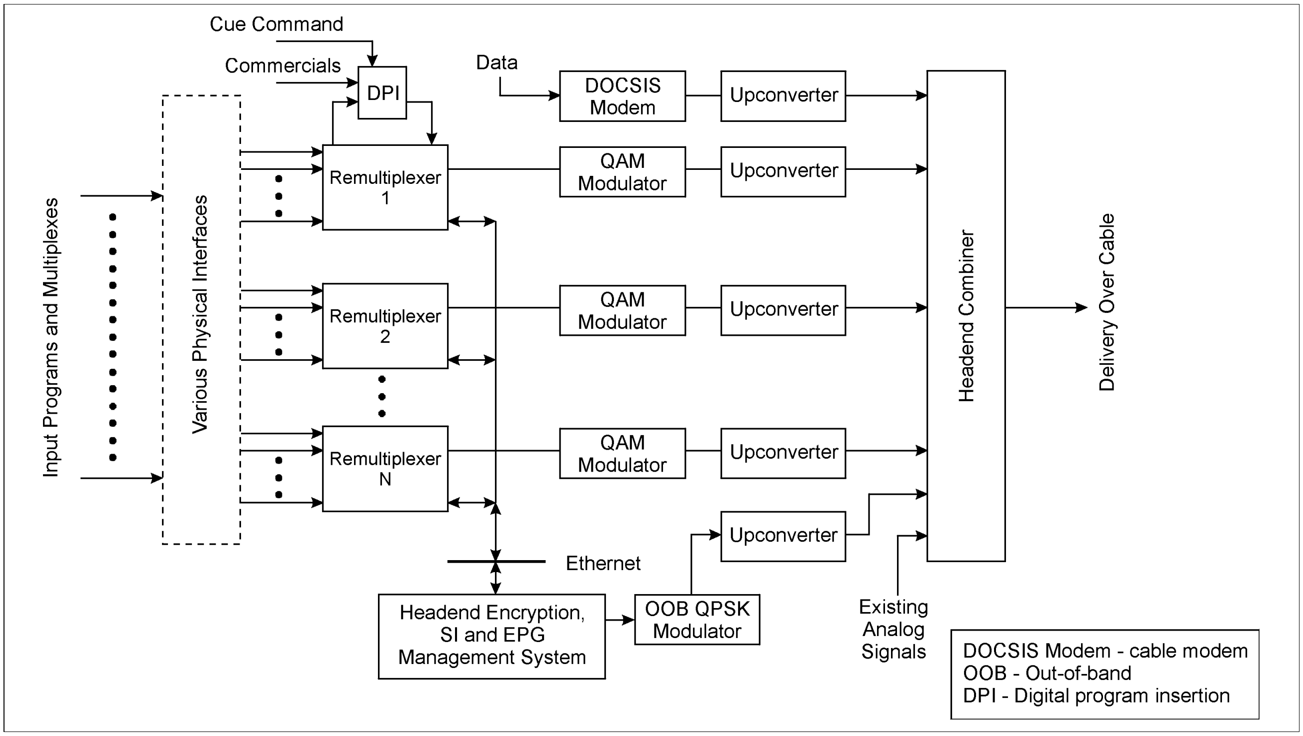


Figure 5. Headend architecture with analog and digital channels

PROGRAM INSERTION

Insertion of locally-generated commercials and short programs has become an integral part of the headend. First, revenue from local ads contributes significantly to overall cable revenue. Second, locally-generated civic and political short programs have become a necessity to local communities.

To insert commercials and short programs, a digital program insertion (DPI) system may be connected externally to a remultiplexer, as shown in Figure 4. A cue-command is detected from the input multiplex for one or more selected channels by a cue-command decoder. The latter output is fed to a DPI for start/stop signaling for an advertising insertion opportunity or avail. Locally-generated materials may be compressed at various bitrates using a non-real-time encoder and stored in a video server. This may be necessary to match the incoming channel bitrate. Alternatively, the material may be compressed at a higher bitrate. A real-time recoder may be used to reduce the bitrate to match the bitrate of the incoming program. The Society of Motion Pictures and Television Engineers (SMPTE) standard on splicing MPEG-2 transport streams includes cue commands (similar to existing analog cue-tones). Design of a cue-command decoder and DPI equipment based on this standard will be facilitated, but the actual insertion mechanism for switching compressed streams is not fully specified.

HEADEND ARCHITECTURE

The basic architecture of integrating digital signals with existing analog signals is shown in Figure 5. Digital multiplexers, carrying one or more programs, are input to a number of remultiplexers. It is possible that, due to some mismatch in interfaces between incoming signals and remultiplexer input interfaces, some hardware may be necessary to make input signals compatible with the remultiplexer input. As mentioned earlier, a remultiplexer will be an important piece of equipment in digital headends. It will be needed to select desired programs from the input multiplexes and to construct new packages (multiplexes) to be delivered over a single

6 MHz physical channel. The remultiplexer will need authorization for de-encrypting the selected channels. The channels, now in the clear, may be used for demultiplexing, recoding (if bitrate reduction is required), or digital program insertion before being remultiplexed in the new package for delivery over cable. The remultiplexer will also encrypt the elementary streams of the new multiplex. The new multiplexes will feed the QAM modulators. Each of these new multiplexes will modulate a 6 MHz carrier. The QAM-modulated carrier will be upconverted to one of the existing cable channels. All upconverted RF signals are combined with the RF analog channels in the headend channel combiner before being transmitted over cable networks.

As stated earlier, the remultiplexer will have authorization to decrypt the demultiplexed elementary streams. The remultiplexer will then extract program specific information (PSI) from the incoming multiplex and will send it to the headend management computer. An integrated program guide for all analog and digital channels can then be built from this information.

An integrated electronic program guide (EPG) is an interactive navigational tool that provides program content and schedule information. It supports selection of any analog or digital channel by direct entry of its channel number or symbolic name (call letters) on an on-screen display. Each of these channels is associated with a physical channel whose frequency range, modulation used, etc., should be known for correct tuning and demodulation. This information is part of the system or service information (SI). The SI is an extended version of PSI. PSI data provides information to enable automatic configuration of the receiver to demultiplex and decode the various program streams within the multiplex. EPG, along with SI, makes navigation of any channel user-transparent. This information is supplied on an out-of-band (OOB) channel. If any 8-VSB signal is transmitted using a cable service physical channel, and by-passes the set-top box (STB) to reach the DTV receiver, implementation of an integrated program guide could be complex. To implement an integrated program guide in a straightforward manner, the STB will demodulate and decode all signals passing through it. The TV will receive all signals and information through an interface, meaning that all digital signals must be QAM-modulated to be demodulated in the set-top box.

STATUS MONITORING

As mentioned earlier, digital technology has remarkable efficiency and flexibility. However, in some situations, it also presents complexity and challenges. One such area is the testing, measuring, and analyzing of DTV signals. First, the signal representation, transmission, and multiplexing characteristics of the DTV signal are quite different from analog TV signals. Second, the testing and measuring equipment currently available on the market, have not been standardized and have not matured. DTV signals may be tested in two parts—as modulated carrier signals and as demodulated baseband signals.

To test the signal quality of the digitally-modulated signal, a modulation analyzer is required. These analyzers provide constellation diagram, eye diagram pattern, error vector magnitude (EVM), and frequency response, from which channel gain, phase error, bit error rate (BER), noise-related problems, etc., can be found.

To test an RF-modulated signal, QAM or VSB analyzers are available from Hewlett-Packard, Tektronix, Applied Signal Technology, Wavetek, and others. At the baseband level, the transport baseband stream can be demultiplexed into component elementary streams. Each elementary stream can be analyzed using a tool like an MPEG-2 conformance verifier. In addition to checking syntax at various MPEG-2 layers, these tools can detect problems related to system information (PSI section), PCR, timing, jitter, and synchronization (PCR, PTS, DTS), and transport stream target decoder (TSTD) buffer overflow. To analyze signals at the baseband or elementary stream level, equipment for real-time monitoring and non-real-time analysis are available from Tektronix, HP, Sencore/Symbionics, Snell & Willcox, and others. Most equipment is PC-based and is somewhat different in terms of features and capabilities. Many software tools, including one from CableLabs, are also available for PC and UNIX platforms for non-real-time MPEG-2 conformance verification.

The noise in a digital video signal manifests itself differently than the noise in an analog signal/video. Any increase in noise will cause the analog signal picture quality to degrade gracefully. If the noise becomes severe and corrupts synchronization signals, the picture will be lost. An increase in noise in DTV signals will cause loss of margin, which in turn causes bit errors. If errored bits are part of MPEG-2 headers, such as sequence headers, picture headers, etc., effects may be visible in the entire picture (broken or blocky picture with irregular colors). On the other hand, if they are part of picture data, the effect may be very localized on-screen. One can expect reasonable video quality up to a certain threshold in signal-to-noise ratio (SNR). If this ratio falls below the threshold level, the noise will be so severe that the video will appear totally unrecognizable and the decoder will fail. Therefore, the signal fidelity in both modulated and baseband domains at the headend will help maintain satisfactory performance to the subscriber.

CONCLUSION

Various forms of digital television signals, headend interfaces, and signal processing requirements have been discussed. Important characteristics of each signal type have been described. A brief overview of each step handling incoming digital signals and downstream processing have been explained. A conceptual architecture has been presented to show how new digital equipment may be integrated, along with the existing analog components, to make the headend a complete program delivery system for both analog and digital channels. The signal and noise characteristics of the incoming DTV signals, as well as methods to test and monitor them, have been discussed. Continuous monitoring of the integrity and quality of the incoming signals will ensure proper signal quality. The equipment and the software from several manufacturers to test, measure, and analyze these signals are becoming available.

All the equipment mentioned above is not required to add digital capability to an existing analog headend. For example, DPI systems and recoders are not required for delivery of digital signals and may be added incrementally as deemed necessary. DPI is of less significance to pay-per-view (PPV) channels. Optional test and measurement equipment supports reliable operation through status monitoring. Also, the equipment needed is proportional to the number of digital channels allocated in the available system bandwidth.

In summary, digital television network interfaces will provide greatly improved capacity, flexibility, and quality to existing analog offerings. Several new interface considerations and requirements presented here will be needed for successfully integrating and distributing these new services.

ACKNOWLEDGMENTS

The authors would like to thank Sam Narashimhan and Bob Eifrig of GI for valuable discussions and information.

REFERENCES

1. CODING OF MOVING PICTURES AND ASSOCIATED AUDIO, ISO/IEC JTC1/ SC29/WG11, 13818-1 (Systems) and 13818-2 (Video).

2. ATSC DIGITAL TELEVISION STANDARD, A/53 (Systems) and A/54 (Video).

3. *Digital Program Insertion for Local Advertising*, Mukta Kar, Majid Chelehmal, and Richard S. Prodan, CableLabs (submitted for NCTA98).

Cable Modem Bandwidth Provisioning and Quality of Service

Shlomo Rakib
Terayon Communication Systems

Abstract

Cable operators today are challenged to provide data services to a broad range of users with differing bandwidth and Quality of Service requirements. The nature of a shared cable network requires systems that allow operators to provision and allocate the bandwidth to multiple tiers of users. Only in this way can operators match bandwidth to the needs of both residential and commercial users, and to build a service model that maximizes revenue. To support such a model, the cable operator should deploy a cable modem system that provides adequate bandwidth capacity in both directions, and the ability to provision the bandwidth to support a broad range of user application needs.

This paper analyzes the performance of Terayon's TeraComm system, operating in its UBR (unspecified bit rate) mode, which along with CBR (constant bit rate), represents the system's two primary MAC layer modes. These test results clearly demonstrate the sophisticated bandwidth management capabilities of the TeraComm system, which support multi-tiered data services. The Terayon system achieves maximum bandwidth utilization and maintains fair allocation of bandwidth among users, maintaining minimum latency for access and data transfer. This capability enables a new generation of data services and broadband applications.

PHYSICAL LAYER

The TeraComm system uses PHY and MAC layers which are both unique in the cable modem industry. With Terayon's Synchronous Code Division Multiple Access PHY implementation, a set of 144 orthogonal codes are used to allow simultaneous transmission from up to 144 individual data streams on each channel. The system defines a separate channel each for upstream and downstream; thus there are 288 simultaneous transmissions streams allowed. The transport rate of each transmission stream is 72kbps. The system reserves 128 data streams for user data. Allowing for the overhead of data cell framing mechanisms, the capacity for data is 64kbps per stream. The remaining 16 data streams are used for system management and access control. This means user data traffic never contends for bandwidth with management or access request traffic (see figure 1).

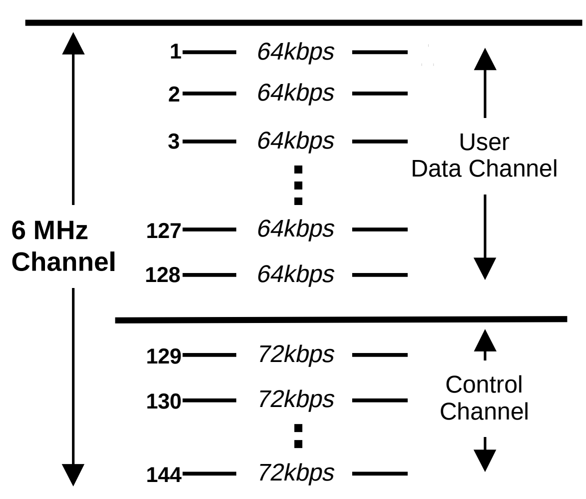


Figure 1

MEDIA ACCESS CONTROL LAYER

The 128 users' data stream codes are managed and assigned to modems by the Bandwidth Manager at the TeraLink headend controller. Figure 2 demonstrates how one or more codes can be assigned on a permanent basis (CBR mode) or on a dynamic basis (UBR mode). Modems requiring access to one of the 128 user data streams will contend for access using 4 slots (see figure 2) in the control channel. Providing the higher transmission rate of 4 slots, rather than a single slot, lowers the probability of corrupting an access request due to impulse noise. These time slots also lessen the probability of a collision during the access process. Each modem will randomly choose one of these time slots to issue its access request. After a successful access has been achieved, additional codes may be allocated based on measured usage within a small time window, rather than via additional access requests.

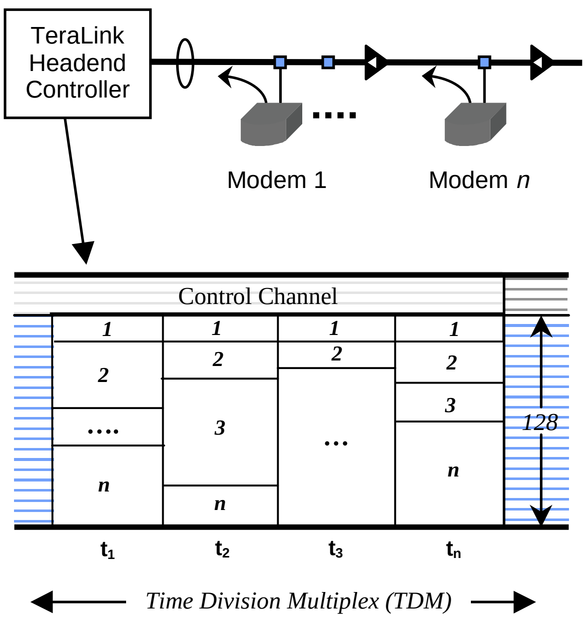


Figure 2

This means that some calculable latency will be incurred only during initial access. In the event that the channel subscription is higher than the available channel rate, data streams will be distributed among multiple modems without requiring further access requests. For this reason, there is zero latency incurred at the MAC layer after initial access in an "undersubscribed" system. MAC layer latency in an "oversubscribed" system is discussed in the next section. While this MAC technique is much like that used for standard Ethernet or other collision-based contention techniques, there are some important differences. By using orthogonal codes for data streams that are separate from those used for access contention ensures that even a high rate of access collisions has no effect on channel efficiency. A standard rule of thumb for Ethernet network scaling is to allow no more than 30% utilization on average, due to the fast break-down of the channel under high contention rates. The actual traffic being sent is used for contention in an Ethernet network, therefore the amount of data (channel capacity) lost and to a collision event can be large. The effect of this data loss is that much of the channel capacity is spent resending data. The TeraComm system can always run at 100% utilization, independent of the amount of access requests being processed.

Finally, within an Ethernet network, the probability of a collision is proportional to the size of the datagram sent. Also, in an Ethernet network all traffic is open to collisions. With the TeraComm MAC, the smallest possible station identifiers are used for access requests, so there is no data lost in a collision event. Due to its small size, the probability of collision is small. Finally, no access requests are required for bandwidth after the initial access is granted in an "undersubscribed system." The Bandwidth Manager task may automatically grant more bandwidth based on usage, system load and operator-defined Quality of Service provisioning.

RESPONSE LATENCY OF AN OVERSUBSCRIBED SYSTEM

A system in which the requested bit rate is higher than the available bit rate is said to be "Oversubscribed.” In a system that supports guaranteed Quality Of Service, over-subscription occurs when the sum of bandwidth reserved for Constant Bit Rate plus the maximum bandwidth allocated to Unspecified Bit Rate applications, (minimum bit rate = 0) is greater than the total bit rate of the system. In a purely contention-based system, over-subscription is merely implied by an access latency which on average is not tolerable by the applications which are running over the system. As collisions reduce the capacity of the data channel in a purely contention-based system, which in turn increases the required access time, access latency will grow non-deterministically.

In an undersubscribed system, the TeraComm MAC uses time division multiplexing of codes to reduce the number of access requests, thus avoiding non-deterministic access latency. Under normal operation, modems that are no longer utilizing granted bandwidth will lose all but their last data code. After the initial code is granted, access latency is zero from then on. In the event that the system is "oversubscribed,” that last code will also be revoked in the event of an access request from another modem. While the system is in an "oversubscribed" state, multiple modems may share a set of unreserved codes which are periodically multiplexed from modem to modem. As the number of UBR modems increases in comparison to the number of unreserved codes, the duty cycle of allowable transmission time decreases.

For example, while 129 modems sharing 128 codes may each transmit 99% of the time, 256 modems sharing 128 unreserved codes may only transmit 50% of the time. The time spent awaiting re-allocation of a code is considered response latency. The average response latency of a time division multiplexed system is then deterministic and is derived from the product of the multiplexing period and the ratio of transmitting UBR modems to unreserved codes.

BANDWIDTH MANAGER RESPONSE IN AN UNDERSUBSCRIBED SYSTEM

In an undersubscribed system, where the number of active modems is less than the number of unassigned time slots, the bandwidth manager will fairly distribute bandwidth among the active modems. Any modem that utilizes more than 60% of its assigned bandwidth is eligable for an increase in allocation. To efficiently use the channel in an environment where data is TCP based, the rate at which capacity is moved between active modems must compliment TCP Slow Start implementations.

MAC TEST RESULTS

The TeraComm MAC implementation performance is superior to other collision-based contention techniques, providing deterministic latency for access and data transfer and for data channel efficiency. This remains true for undersubscribed as well as oversubscribed channel conditions.

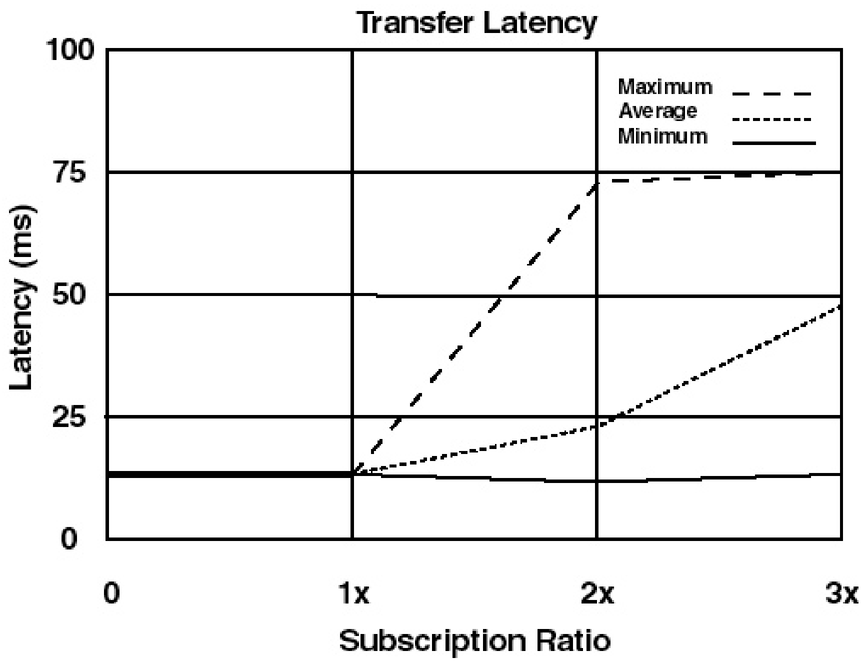


Figure 3: The transfer latency with access in the Terayon system is far superior to alternative solutions due to the architecture of the Terayon MAC and capacity of the channel supported by S-CDMA.

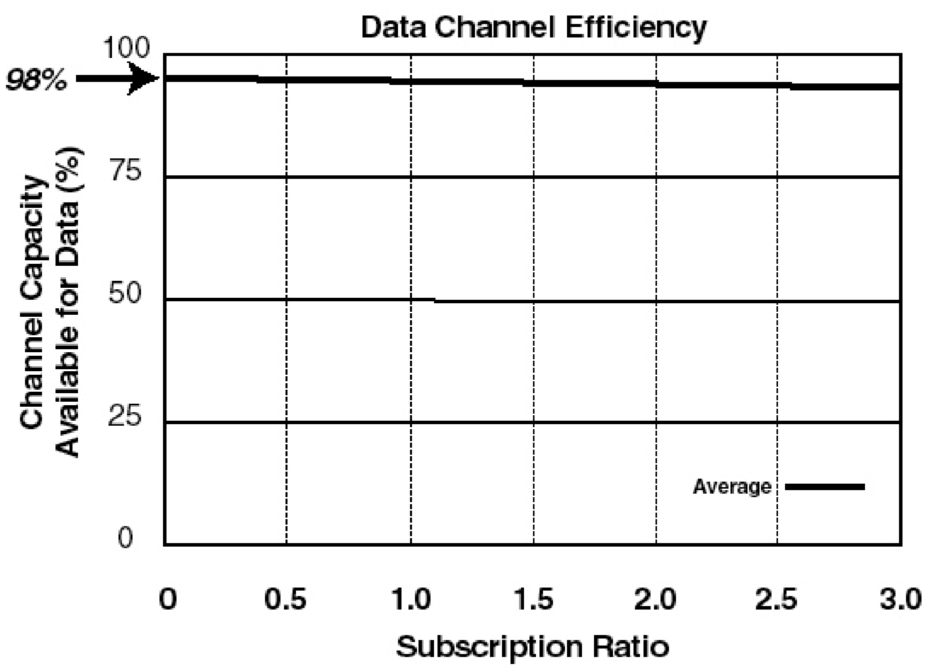


Figure 4: The low transfer latency in the Terayon system makes it well suited for time-sensitive applications, such as IP telephony, video conferencing, and online games. This is even true of best effort class of service systems with high over-subscription ratios.

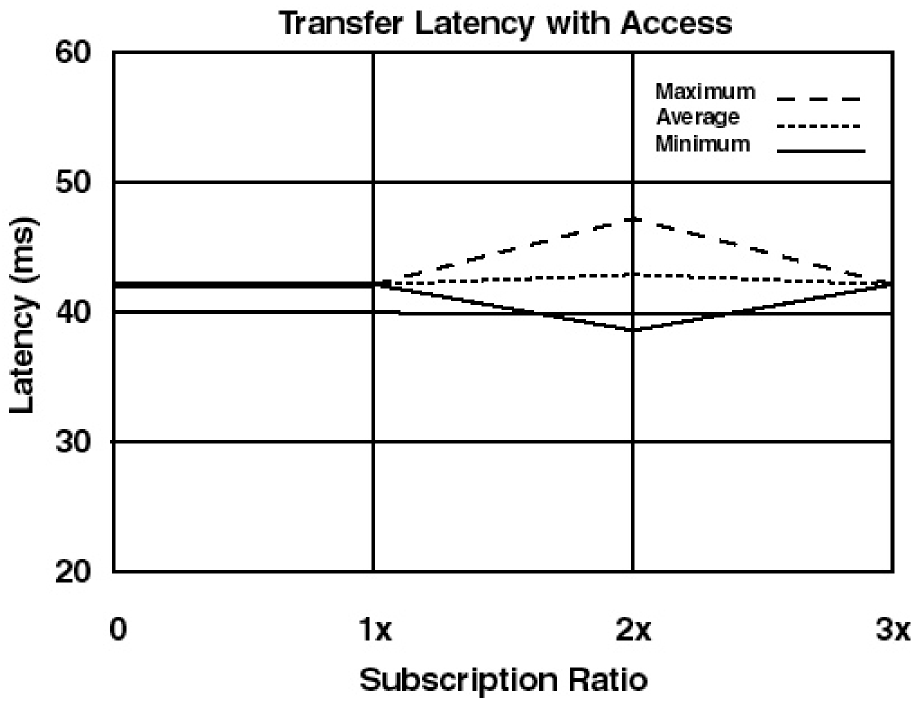


Figure 5Another important performance parameter is channel efficiency, or the impact on data channel capacity as more users are added and access requests to the channel increase. Because of separation of the data and control channels, channel efficiency is only slightly impacted by oversubscription.

AUTHOR'S BIOGRAPHY

Shlomo Rakib
President and Chief Technical Officer
Terayon Communication Systems
2952 Bunker Hill Lane
Santa Clara, California 95054
(408) 727-4400

Shlomo Rakib is President and Chief Technology Officer of Terayon Communication Systems, the cable modem access company he co-founded in 1993. Mr. Rakib developed the technical roadmap for Terayon, based on his vision of the growing need for broadband access to the home and office. Mr. Rakib invented and filed patents on S-CDMA (Synchronous Code Division Multiple Access) technology, which is optimized for robust performance over today's cable systems. Mr. Rakib has over 15 years experience designing products for the cable industry, including 9 years as Chief Engineer at the communications company, PhaseCom. He holds a bachelors degree in engineering from Technion University in Israel.

Cable Modems in the Home Environment

By
Joseph B. Glaab, Rob Howald,
Dean Stoneback, and Matt Waight
General Instrument

Abstract

This paper investigates the "in-home" environment when implementing cable modem systems in two-way HFC networks. Cable modems are hitting the market on a wide scale in 1998, with standardized performance and interfaces per the recently developed MCNS (Multimedia Cable Network System) consortium. Their ability to coexist with the existing downstream infrastructure is crucial to their acceptance as a broadband access tool. However, the modem may be required to operate at high transmit levels, potentially interfering with the downstream TV and/or settop converter unit. Also, the downstream receive equipment may introduce ingress into the return path, which can impair digital communications. Suggestions that address these serious issues are discussed. Measurement results are presented that justify these conclusions.

1.0 Problem Description

1.0.1 Cable Modem Interference on Forward Video

Addressable settop converters and TV receivers, both already deployed and currently being deployed in modern plants, have not necessarily been designed to be compatible with newer cable modems. A typical return path is considered to be 5-40 MHz, with allowances for performance degradation recognized at either band edge due to likely frequency response distortion. Also, the ingress problems at the low end are well documented. Modern cable modems, and particular those being designed to the specification written by the MCNS, will be able to operate up to 40 MHz. Actual return services and bands allocated are, of course, determined by individual MSO's. However, it is all but guaranteed that use of the complete available return spectrum will be expected in the near future to support the new services, which we be counted on as important revenue streams. A traditional converter's return transmitter typically operates in the low end of the return band. Thus, the forward-related equipment in those units was designed to reject high level interference in that band. The CFT-2200 settop, for example, a successfully received GI product which is widely deployed, was designed to be used with any return-path transmitter operating from 5 to 15 MHz at levels up to +57 dBmV with no recognizable video interference. Maximum cable modem levels will be roughly the same, but with wider spectral range. Thus, it becomes important to characterize the effect on video performance against cable modem levels and the wider upstream frequency range anticipated.

Related data has been measured, and is available in the report from Carl T. Jones Corporation, "Non-Video Interference Test Results report" (12/94). That data is important because it characterizes this phenomenon when it is associated with TV receivers and VCR's. A brief comparison of those results will be made with the results obtained in our measurements. Consumer electronics of this type are in the same situation as "old" converters - potentially seeing return band interference that they were not designed to handle.

<u>1.0.2 Return Ingress Issues</u>

In section 1.0.1, the return system's interference on the forward plant was introduced. A second issue is impairments imposed on the return communication link due to forward-related TV equipment. The fact that the home is a major source of ingress, and, in fact, the dominant source, is well known. Many external sources can be blamed for home-generated electrical pollution of the return band. Common examples include sources such as appliances, garage door openers, hair dryers, light dimmer switches, etc. These effects are aggravated by poor electrical practices, such as poor in-home connections and grounding, cheap, flea market quality splitters, and unterminated RF ports. Much effort has gone into the system design efforts for products such as MCNS-based cable modems to assure reliable communications through many of these impairments using advanced signaling schemes and powerful digital receivers in the Headend (HE). These receivers leverage the latest techniques in equalization, forward error correction (FEC), synchronization, and ingress avoidance to meet performance requirements and maintain a suitable percentage of error-free channel availability. However, a perhaps overlooked portion of ingress management is the effect of traditional settop converter units and cable-ready TV's themselves in contributing to the interference. While HE techniques to avoid and/or transmit signals through ingress don't care about where it is coming from, it is important to understand local sources of upstream impairments. In this way, future equipment can be modified and/or augmented properly for maximum return path availability. Also, there is an opposite extreme to very high modem transmit power. The range of output power required per MCNS specifications is 8 dBmV to 58 dBmV. Thus, the low end is quite susceptible to home-generated interference. In addition, for a properly aligned return plant that is heavily loaded, the transmitter should not be made arbitrarily high to achieve detection. For example, a transmitter that just increases its power until heard at the HE, only because high interference is corrupting the path, will not allow for comfortable coexistence with other users sharing the available power load.

It is important to note that both of these important issues have not been completely ignored in the past. In fact, committees exist that are now attempting to come to a consensus on return path emissions. Also, there has been some discussion with the release of the MCNS specification about whether the cable modem interference problem is in the court of the modem developers, or if it is the settop designer who must assure further protection of the video signal from return band frequencies that he does not use. Inspiration for this paper stems from the idea that not enough attention is being paid to the issue at this point, and now is past the time to do so.

2.0 Test Set-ups

<u>2.0.1 Continuous Return Band Interference - Frequency Modulation</u>

The first set of tests used FM to simulate a signal from the home (many traditional boxes implement FSK systems). The modem is simulated by a signal generator that is frequency modulated by an internal +/- 75 KHz tone. To eliminate harmonics from the test signal generator (much care must be taken for these types of measurements) the signal is low-pass filtered (see Figure 1). In fact, for ultimate rejection purposes, two filters are used in cascade, separated by a 3 dB pad. This signal is combined with a would-be cable-drop carrying a 77 channel forward band load. This combined signal is connected to the converter. Other test setups have been used, including one in which the modem simulator and settop converter are connected, as in a real application, to splitter outputs driven by a simulated drop from a tap carrying the downstream signals. However, side issues developed with this test set-up, including differences in splitter isolation, tap return loss, and signal harmonics from test amplifiers needed to boost the simulated modem signal to levels high enough to overcome the splitter port-to-port isolation. Thus, it was determined that the most reliable measure of interference that, in turn, could be used to back out required performance of the passives, was to combine the inputs *into* the converter. Thus, by measuring the level *into* the converter that causes degradation, rather than the levels *out of* the modem that cause degradation, the data is useable with any arrangement of passives associated with the drop and the home. Because it quickly became apparent that second harmonic distortion was the primary impairment, sensitive return band frequencies to test were quickly converged upon.

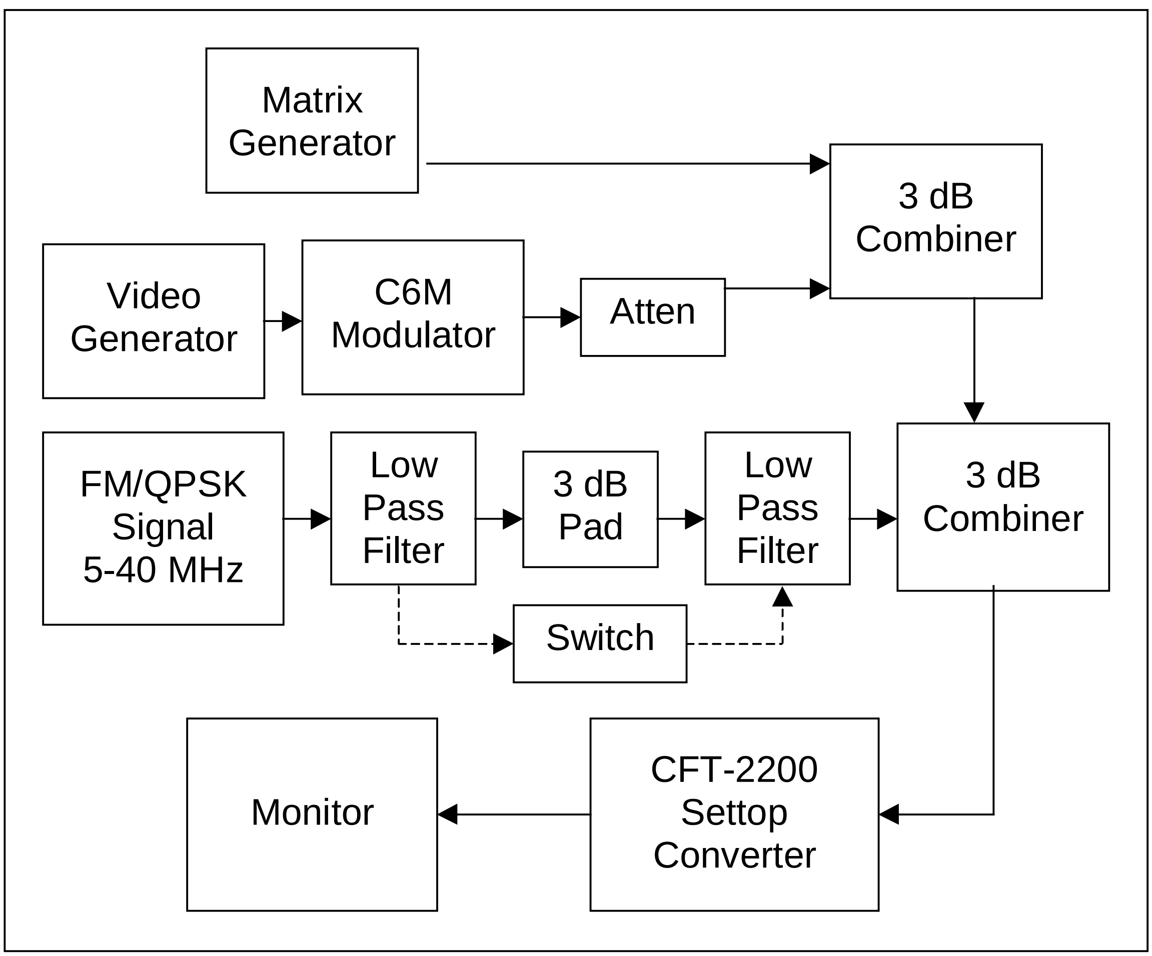


Figure 1 - Setup: Continuous FM or Burst QPSK Interference in Return Band

A 50 IRE test pattern was used, along with a forward matrix-generated load, with the test limit being the threshold of visibility (TOV). The desired signal and the matrix signals were set at 0 dBmV, the minimum specified video carrier level at the converter input. At higher levels, the AGC circuit in the tuner will reduce the return-path signal as well as the desired signal. It is important to note that there will be a correlation between homes that see the minimum video carrier level, and homes who's modems require the high transmit output levels. In this test, the four lowest channels were examined, before subsequently concentrating on the most sensitive for the situation being considered (Channel 5). Test results for this and all cases are given in section 3.0.

2.0.2 Continuous Return Band Interference - QPSK

The above test, using the setup of Figure 1, was repeated with a QPSK interferer for comparison. Two different QPSK data rates were used. One modem was a 2 MBPS unit (1 MSPS), while the second was a 256 KBPS unit (128 KSPS). This contrast of narrowband and wideband digital modulation, it was felt, would be informative. All other information described in the above setup remained the same. Again, care was taken to check the equipment harmonic performance, in this case of the modems, which were wideband and agile. TOV criteria was again the performance criteria (after giving our squinting eyes a break from the previous test).

Other important parameters of this test, and each of the subsequent QPSK tests, continuous or burst, include a forward band loading of 77 channels at 0 dBmV. In the Jones report, it was pointed out that a low end video level was more limiting with regard to interference performance, rather than excess distortion from a fully loaded high end, where all channels are at 15 dBmV. Also, the 50 IRE flat field was used still for TOV observations, and Channel 5 (77.25 MHz) was considered exclusively, for reasons that become more clear after discussing the first set of measurements. However, this is not meant to imply that a complete characterization across all channels should not ultimately be evaluated.

One other item of interest in the test setup is that the limiting distortion is known to be related to second harmonic energy. Thus, modem center frequencies were set close to one-half the Channel 5 frequency. Because of the band of occupancy of QPSK at 2 MBPS, the frequency of the modem in this case was shifted relative to the 256 KBPS case. In order to assure that the interfering energy of the QPSK signal (about 1.5 MHz of RF bandwidth) fell within the video bandwidth, it was offset so that the second harmonic fell at RF video carrier plus 750 kHz (39 MHz). The 256 KBPS unit was centered such that the second harmonic fell at carrier plus 250 kHz (38.75 MHz), a known location of poor TOV. For this narrowband signal, this made the most sense. For the 2 MBPS signal, it made the most sense to make sure the noise energy it represented was both near the sensitive region, and with all of the spectral energy contributing.

2.0.3 Continuous Interference Directly on Downstream - QPSK

In addition to this "overload" testing noted above, whereby tuner nonlinearity was the mechanism to create interference, TOV measurements were also done with direct interference. In other words, QPSK modulation was placed directly under the Channel 5 video signal by summing it in, along with the downstream load (see Figure 2). This has a couple of advantages. First, it allows a rough calibration of "eyeballs". That is, by putting the modems in CW mode, confidence in observations is high, because one can pull out the common chart on interference thresholds of CW interference and compare it to measurements. The second advantage, as previously described, is that it is very straightforward to go from what level of disturbance is actually at the converter that causes poor performance to useful specifications. Unlike the previous description, however, in this case not only is the measurement of tolerable level made into the converter, it is even made in the band of the channel being disturbed (Channel 5 in this case). Thus, it is an accurate gauge of allowable harmonic content in the forward band. For Channel 5, the QPSK modems were set directly at 77.5 MHz (256 KBPS) and 78 MHz (2 MBPS).

2.0.4 Burst Return Band Interference - QPSK

While it is certainly the case that interference all of the time represents a worst case scenario from the standpoint of the behavior of an in-home cable modem, it was considered possible that the burst nature of a transmission may, in fact, be as recognizable or more so. Because the cable modem operates in a TDMA (and FDMA) system, someone "surfing" while another is watching the tube will cause intermittent interference associated with the particular temporal characteristics of the upstream transmissions. This will vary due to overall return plant loading, physical location in the plant, selected modulation and symbol rate for the user, type of web surfer, etc., all which ultimately determine transmit levels, packet lengths, and interpacket periods.

For burst QPSK testing, a logic signal was used to toggle an RF switch. The input to the switch was the QPSK signal from the test modems. Data rates of 256 KBPS and 2 MBPS were again used. The modem burst parameters were 500 usec bursts of QPSK at a 200 Hz rate, roughly consistent with some preliminary studies of the temporal characteristics of upstream traffic. Obviously, these can vary widely. The QPSK signal again was set to interfere with the Channel 5 RF video carrier frequency near it most sensitive TOV point. The QPSK center frequencies were set as in 2.0.2 to land on this known most visible part of the video spectrum. The switch output is lowpass filtered, in an attempt to recognize that a cable modem output itself will be driven through an output lowpass filter, which thus significantly impacts the spectrum in a pulsing situation. It is also worthwhile to point out that, while MCNS does not consider the modem transmit level's ability to overwhelm the input of a converter, the specifications do carefully assure very low (intolerably low, if you ask modem vendors) harmonic distortion and spurious performance at the modem transmit output.

The same test criterion was used in these measurements - TOV. The return band burst test setup is also shown in Figure 1, with the dotted portion representing the addition of the path through the switch for burst testing.

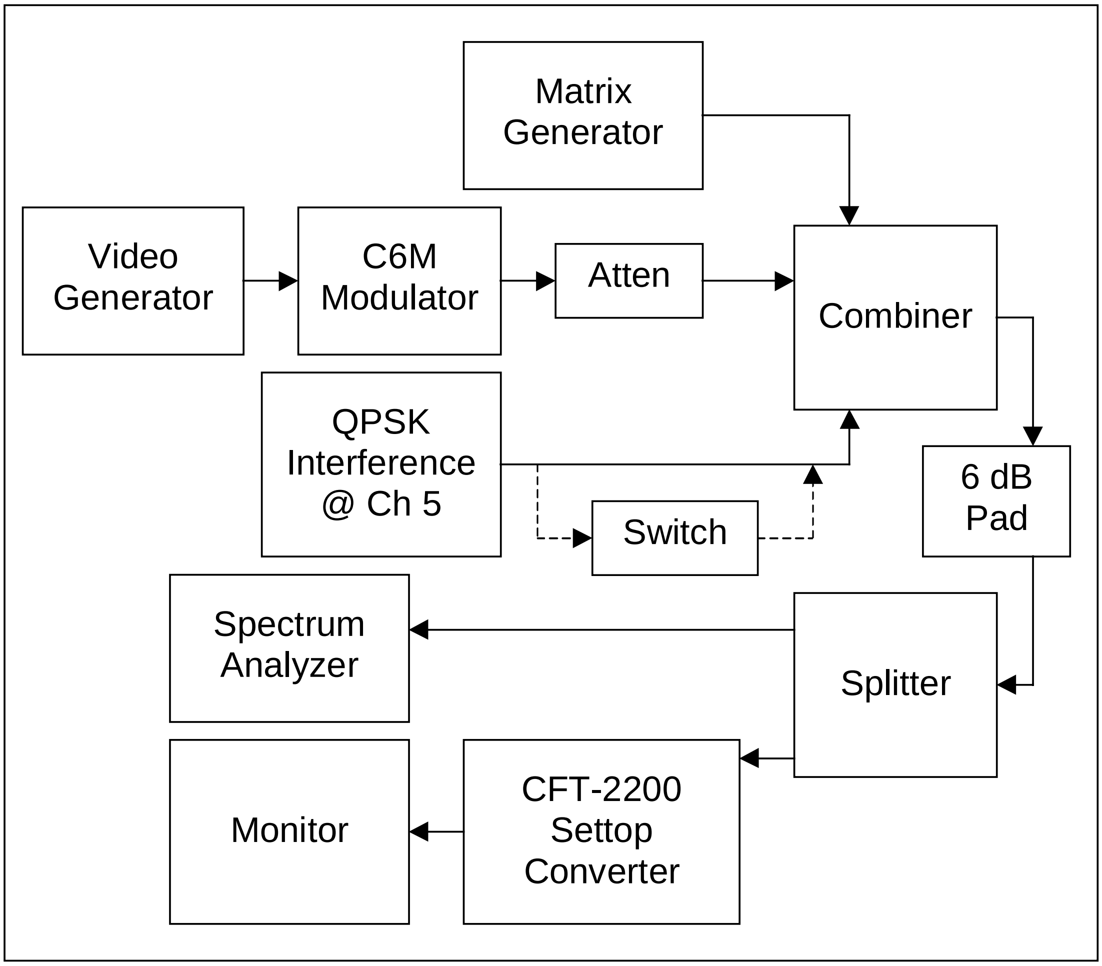


Figure 2 - Setup: Continuous or Burst QPSK Directly on Downstream

2.0.5 Burst Interference Directly on Downstream - QPSK

Finally, the bursty signal was placed underneath the downstream carrier directly, once again at the most sensitive point (Figure 2) Both data rates were implemented, and TOV measured, using the same QPSK center frequencies as 2.0.3.

3.0 Test Results

3.0.1 Continuous Return Band Interference - Frequency Modulation

The results of TOV testing with the continuous FM signal are shown in Figure 3. The primary problem is the high-level signal entering the converter and mixing with itself in the tuner, producing a second harmonic that falls in-band of the tuned channel. Even if the cable modem completely eliminates all harmonics, a fundamental signal that is above 27 MHz may interfere. The levels which cause a problem are consistent with the second order distortion performance of the converter’s tuner.

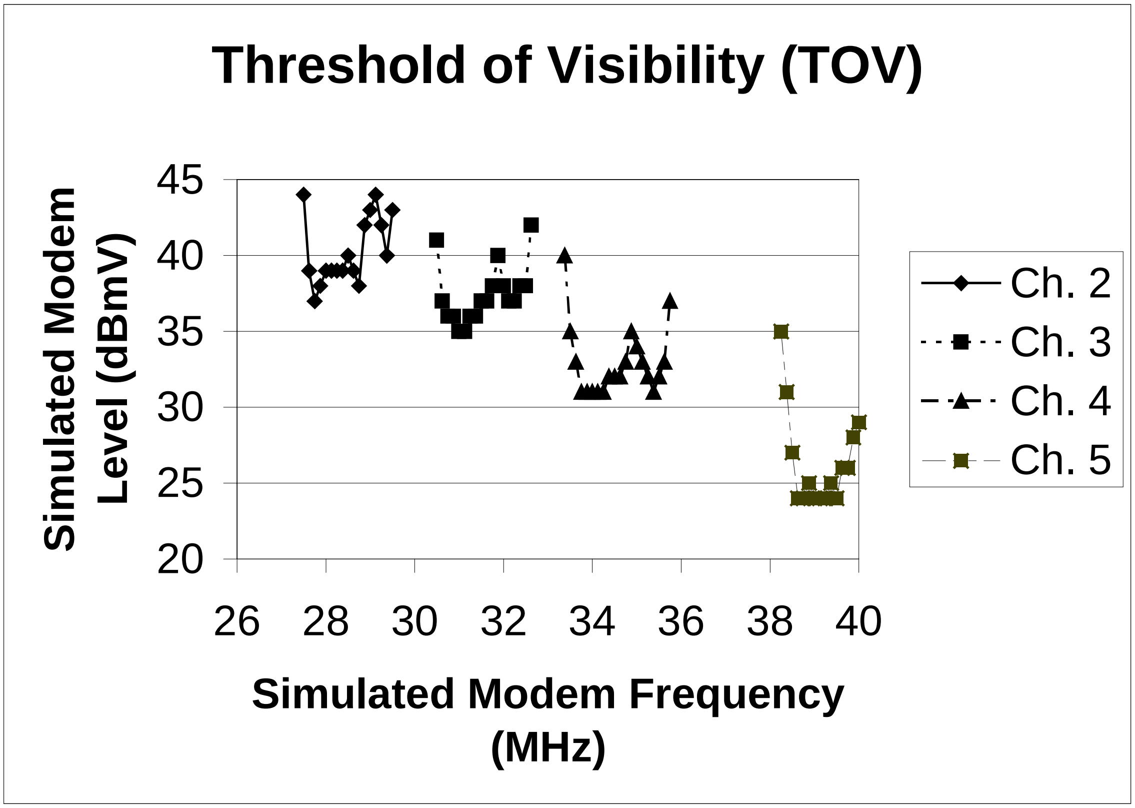


Figure 3 - Continuous FM Interference in Return Band vs. Channel

The highest frequency modem outputs (around 40 MHz) create the most degrading situation when second order distortion falls into Channel 5. Of course, the video TOV is dependent on where within the video channel the interference falls, and in the case at hand, this occurs the worst in the 39 MHz range. It is apparent that, for the measurements taken on Channel 5, the converter has an interference threshold of only 24 dBmV at its input. Thus, a splitter which has only 25 dB of isolation between converter and modem would result in video degradation of Channel 5 if the modem were to transmit above 49 dBmV. MCNS-based cable modems have a maximum requirement to transmit at 55 dBmV (QPSK) and 58 dBmV (16-QAM). And, link analysis will show that these types of levels can be expected in the plant. Both link analysis and the desire to overcome in-home ingress point to the need to transmit from the home as high as possible, while the converter on the other side of the splitter begs for consideration of more tolerable transmit levels to keep the low end video channels clean. For the lowest channel, Channel 2, which is degraded at its worst point by a modem frequency of about 28 MHz, an input level of 37 dBmV establishes the TOV. Here, clearly, a 25 dB splitter isolation would be adequate to avoid video interference (37+25 = 62 dBmV, above modem transmit maximums). Channel 4 becomes an in-between or marginal case, showing a TOV when a 31 dBmV signal is delivered at the converter input. In this case, a 25 dB isolation would just be enough (QPSK), just miss (16-QAM), or simply be very sensitive to the quality of the splitter and the return loss of the tap.

The modem signal level that sees its way to the converter is a combination of splitter isolation and the return loss of the tap. The reflected signal from the tap is reduced by splitter's loss two times, once traveling out to the tap, and once traveling back, for a total loss of somewhere between 6-9 dB on top of the tap return loss. The keys to minimizing problems are to use a good splitter having a port-to-port isolation much better than 30 dB in the high end of the return band, and using taps that have a return loss much better than 20 dB in the same region. Both of these numbers can be demanding relative to what is known about typical values in the field, and relative to some of the poorer commercial quality passives available for use in the home today. The example below illustrates the issues before discussing more about the performance of the RF passives.

Example:

A cable modem transmits at 48 dBmV (MCNS range required is 8 dBmV to 58 dBmV, modulation dependent) into a splitter having 30 dB of isolation and a loss of 3.5 dB. The tap has a return loss of 20 dB. The converter or TV receives the return signal directly from the splitter at (48 – 30) 18 dBmV and from the tap (ignoring return band drop loss) at (48 – 3.5 –20 –3.5) 21 dBmV. The combined signal level is 22.8 dBmV. The tap is the primary problem in this example, but both contribute, and a poor splitter could easily be the dominant source of interference. However, from the measurements taken, this level is below TOV for the worst case situation observed with the converter. Later burst measurements will add yet more insight into what levels to be concerned about.

Clearly, then, it is the case that there is a bit of "competition" between interfering paths, with either being capable of being dominant. It is as well the case that, if they are close to presenting the same level at the converter port, they can add. The exact adding relationship, being the same signal, would be dependent upon the relative phase shifts of the two paths. Certainly, in the worst case, there is the potential for voltage or nearly voltage addition that would create more disturbing peaks than either path alone.

Also, the above example is not representative of every situation. For example, very good splitters may have 40 dB of isolation over a portion of their band. This would be what is necessary to stay below TOV for a 58 dBmV signal in the example above. Fortunately, these better values typically exist nearer to 40 MHz, where the video interference problem is the worst, and they are poorer, for example 25 dB, at 5 MHz. A bigger problem is consumer TV home-improvement efforts. The well-infiltrated consumer quality brand can have isolation as poor as 10 dB in the worst part of its band, and perhaps averaging 15 dB across the band. Other splitters exist of quality between these two extremes. Clearly, this type of home environment will significantly aggravate the problem. A comparison of two-way splitter data of consumer quality and professional quality models is given in Tables 1 and 2, respectively.

	5 MHz - 40 MHz			50 MHz - 750 MHz		
	Worst Case			Worst Case		
	Level	Freq	Avg	Level	Freq	Avg
S11	-15	5 MHZ	-17	-13	600 MHz	-16
S22	-10	5 MHz	-15	-8.95	750 MHz	-14
Insertion Loss	-4	5 MHz	-4	-4.55	750 MHz	-4
Isolation 1-2	-12	5 MHZ	-18	-25.47	52 MHz	-27
Isolation 2-1	-10	5 MHZ	-15	-25.72	52 MHz	-27

Table 1 - Consumer Grade 3 dB Splitter Measurements

	5 MHz - 40 MHz			50 MHz - 750 MHz		
	Worst Case			Worst Case		
	Level	Freq	Avg	Level	Freq	Avg
S11	-24	5 MHZ	-32	-24	750 MHz	-27
S22	-21	5 MHz	-30	-28	750 MHz	-33
Insertion Loss	-3	5 MHz	-3	-4	750 MHz	-4
Isolation 1-2	-26	5 MHZ	-39	-33	631 MHz	-37
Isolation 2-1	-26	5 MHZ	-39	-33	631 MHZ	-37

Table 2 - Professional Grade 3 dB Splitter Measurements

In addition to splitter values, the tap return loss described in the example also can vary significantly. The one advantage here is that this is an item out of reach of the consumer. However, it is still the case that, in the return band, tap return loss values of about 18 dB may be more typical. Again, however, to our advantage, better return loss numbers in the upstream band occur closer to 40 MHz. These results indicate that typical the tap products made by the industry are not well suited to coexist with cable modems. Because of this, additional converter input filtering will probably be needed in the future.

Another tap-related concern, aside from the trouble that a modem in the home will do to a TV in the home, is what the same modem may do to his neighbor’s TV. The level issues are the same with regard to what can be handled at the input to the neighbor's settop, but the specification of interest with regard to the passives becomes the port-to-port isolation of the taps. Here, a typical value is on the order of 25 dB. At lower frequencies of the return band, it may be as poor as 20 dB. It is important to recognize that, as with any RF passive, both return loss and isolation performance can be sensitive to the VSWR's seen at the other ports of the tap.

Ignoring any drop losses, the 25 dB isolation number results in interference still a few dB above TOV for the tested converter when a 58 dBmV modem output is assumed (58-25-7 = 26 dBmV). A solution that solved the return loss problem at the settop in a subscriber's home would correspondingly work, technically, for the isolation problem to his neighbor. However, the headache of possibly moving into multiple consumers' homes for adding one new subscriber is not a pleasant scenario. Note that use of return path equalizers would reduce the isolation problem and likely aid with VSWR as well, if implemented in the tap arms. This, however, is not the preferred implementation. But, this complexity must be weighed against the significantly complex effort of possibly needing major performance improvement in existing RF passives to support cable modem introduction. In addition, a settop from another manufacturer showed a TOV criterion of only 18 dBmV, another 6 dB more sensitive than the unit described above. Clearly, this settop is sensitive to cable modem levels that are output closer to the middle of their transmit power range.

<u>3.0.2 Continuous Return Band Interference - QPSK</u>

The results for the Channel 5 tests in each of the QPSK cases are summarized in Table 3. There, it is apparent that the QPSK interference tested was less visible than the FM in the previous tests. There are two likely reasons for this: the wider bandwidth (albeit slightly in the 256 KBPS case) and the noise-like modulation (as opposed to the FM, "more" periodic tone-modulated waveform). However, the differences are relatively small (3-4 dB), enough such that, along with the subjectivity of the test, they could be considered not very significant. There does seem to be a pattern established that the broader band modulation that spreads the energy away from the sensitive TOV, and across the video band, is less visible.

<u>3.0.3 Continuous Interference Directly on Downstream - QPSK</u>

The CW video interference charts have been around for a long time. These graphs show that the location of the interference is important to what level it can be to upset performance, and is the reason that specific interfering frequencies are setup as previously described. Most charts show the lowest visible level is about -57 dBc near the carrier offsets already mentioned. In other locations, it can be as high as -40 dBc. In the baseline done here, the CW threshold observed was -63 dBc, possibly due to higher quality monitors of today, and the close range, intensive observations. The modulated data effects on video observed for the tests here are best compared to the relative measured TOV of -63 dBc.

As in the previous section, the QPSK modulated signal is able to be higher than the CW carrier, as shown in Table 3. Again, the reason for this is likely related to bandwidth and periodicity, as the "snow" effect of digital noise fades more easily into the background than discrete lines. Both the 256 KBPS and 2 MBPS cases have this "snowy" interference characteristic. Also, as in the previous case, the differences are rather small between the two, but show an impressive 7 dB to 9 dB more forgiveness to interference when modulated, relative to the -63 dBc baseline. Based on the discussion so far, it would have been expected that modem at 2 MBPS would be less discernible, but, in fact, the opposite is true from the measurements taken. However, it is felt that, with the small differences and different days of measurement, this could easily be attributed to measurement error.

3.0.4 Burst Return Band Interference - QPSK

The concern that transient bursts of interference would be more troublesome was put to rest in this round of tests. As can be seen in Table 3, another 5 dB or 6 dB of transmit level, as measured during a burst, was tolerable. The likely reasons are perhaps, again, spreading of spectral energy due to the pulsing, or simply a lower average interference power at the detector. In fact, it is suspected that truly random interference, such as a cable modem user in a home would induce, would be less visible yet. The interference in these tests had line-like qualities, with snow "within" them, and these periodic characteristics would be removed in a real application. It is likely that the ability to recognize what disturbance was being looked for effected the ability to see it.

It is very important to note the absolute numbers measured for this part of the test. Consider this 5 or 6 dB improvement, and assume a 5 dB or 6 dB improvement in tap return loss over currently accepted typical industry performance. Assume a high, but achievable, splitter isolation. All of this could allow the prior example to remain below TOV, even with a modem output at 58 dBmV. In addition, the true randomness of transmissions will likely provide margin above that measured here. However, improved tap isolation is unlikely given that wider tap bandwidths are being accommodated, causing VSWR compromises. Return path pads may be the only solution. Also, it is unlikely that great strides will be made in consumer grade splitters.

These important results emphasize the need to examine more cases of picture and modulation under more traffic (burst) conditions. This will provide a better handle on the realistic magnitude of the problem.

3.0.5 Burst Interference Directly on Downstream - QPSK

The results of section 3.0.4 were supported by the measurements taken using a direct QPSK burst interferer. In this case, 6-7 dB of additional transmit level was tolerable over the continuous and direct interferer.

	TOV on Channel 5	
	Interference Characteristic	
	Continuous	*Burst*
Interference Type		
FM @39 MHz	24 dBmV	
QPSK (Return Band)		
256 KBPS @ 38.75 MHz	27 dBmv	32 dBmV
2 MBPS @ 39 MHz	28 dBmv	34 dBmV
QPSK (Direct C/I)		
256 KBPS @ 77.5 MHz	-54 dBc	-48 dBc
2 MBPS @ 78 MHz	-56 dBc	-49 dBc

Table 3 - Summary of TOV Data for Channel 5

3.0.6 Comparison of TOV Measurements with Jones Report

The test report "Non-Video Interference Test Results Report", prepared for the EIA by the Carl T. Jones Corporation in 1994, describes measurements taken on ten different TV models and ten different VCR models which, as of 1994, were of "recent manufacture". EIA compliance requires meeting specifications only up to 30 MHz, the traditional limit of return band signaling. In the Jones report, "non-video" means CW interference. The inspiration for the Jones report work was basically the same as this paper. That is, an increase in upstream services means more of the band will be used, and at high signal levels. Thus, the report closely augments nicely the work that has been described above for settops by concentrating on TV's and VCR's. Also, Jones defined the TOV numerically, as a 55 dB C/I, a number that some may question based on subjective perceptibility data. However, it is in the ballpark of most reasonable attempts to characterize this subjective phenomenon. Jones also used Channels 2-6, because of their low-end location, and their relative relationship to return band signal harmonics. The test uses CW interference up to 48 MHz, to characterize IF interference effects, which turn out to be the worst kind. That is not the topic of our experiment, however, and we will concentrate on Jones' results for second harmonic distortion, the most deleterious also in the measurements of consumer electronics.

Summarizing the relevant results, the report concludes that 90% of the "recent" TV's have TOV problems for around 34 dBmV, with 32 dBmV being about the bottom of all tested. For VCR's there is a bigger cause for concern, as 90% bottom out at around 22 dBmV, with 18 dBmV being the lowest tested. This later was also the lowest interfering level to cause a TOV problem among the settops measured. Thus, the conclusions that can be drawn do not significantly differ. In order to provide reliable return services that can comfortably co-exist with existing equipment, it appears some changes will have to be considered to the existing forward-related hardware, whether it is by CATV equipment providers, or consumer electronics equipment providers.

4.0 Ingress on the Return

Traditional return equipment was very limited in capability and in spectral occupancy. Egress from such sources, and from television sets, had no reason to be a major concern to MSO's, because there was such limited use of the spectrum, and signaling techniques, while not very bandwidth efficient, were very robust (i.e. FSK). However, the situation going forward is once again decidedly different than the one that existed during mass deployments of traditional set-top converters and televisions. It is generally expected that more and more plants will become two-way, and that more services covering more return bandwidth will be offered. Eventually, spectral real estate will be at a premium. While ingress will always be a problem due to uncontrollable sources in the home and surroundings, interferers that are imposed by the CATV equipment and consumer electronics gear that make return channels unavailable will not be tolerated from the manufacturers. It is likely that this problem will come to the forefront as older plants move into upgrade phases, in search of providing much more traffic on the returns. Some systems use blocking filters for all inactive returns. This certainly helps ingress management, but is inherently self-defeating, as adding revenue from the return requires adding as many new subscribers as possible. Thus, the filters ultimately will be mostly removed if the return band is used successfully.

4.0.1 TV's and VCR's

The egress associated with existing consumer equipment is designed only to meet FCC requirements. Unfortunately, consumer equipment will do just that - meet these "minimal" targets and do little else. Our region of interest for the return is, of course, the 5-40 MHz band. FCC cable ready consumer electronics specifications are detailed about what can emanate from the port above 54 MHz, but there is little information to go by for return band interference. About the only recognizable "rule-of-thumb" idea is that interference from the port does not interfere with services in the band (CB radio, for example, is in the 27 MHz range). However, this is normally associated, for example, with situations such as TV's receiving broadcast services, but with an unused cable port left open that may output RF interference

When actually using the cable network, the requirement becomes a more complex assortment of possible leakage regions, including cables, connector, and tap emissions. In essence, however, the specifications are about assuring no existing services, such as CB, are disturbed, and not related at all to disturbing any services that do or may eventually return on the cable. Recognizing this, one could peruse the FCC regulations for other "unintentional radiators", or other radiated emission violations, which are measured by field strength at various distances. A possible calculation could then back into what that emission specification would say about the level of interference on a CATV port of a settop, TV, or VCR. The main point, however, is that nothing in the FCC spec will relate at all to considering how interference on that port in the return band effects what services are in that band and on that cable. This is not the fault of the FCC, it is just that, until recently, there has never been a need for such consideration. But, now there is.

4.0.2 Settop Equipment

A second issue of concern is consumer-grade CATV equipment in the field, primarily the traditional set-top converter, and its potential for spewing signals onto the cable drop. For set-top boxes, such egress can be aggravated more so than with VCR's and TV's. This is because homes within a plant may employ various TV and VCR models from various vendors. This variable means that noise contributors are likely to vary, depending on where across the band the different equipment spews interference. However, for settop units, there is a strong likelihood that everyone in the plant will be using the same or a similar model unit. As such, the contribution to upstream interference will be very close to the same frequency for every box, with the difference being related to the reference crystal's drifting, and possible the channel being tuned to. This is simultaneously both good news and bad news. The good news is that the interfering frequencies are more predictable, and thus more able to be filtered out, at the expense of not using those bands, of course. The bad news is that, because of this and the noise funneling effect of the return system, there is more of a likelihood of interference increasing in a more destructive fashion up towards the node.

4.0.3 Addition of Spurious Signals in the Feeder System

The first step in calculating how the spurs add in the feeder system is determining the loss from each tap port to the nearest amplifier. The loss from each tap port can be used in conjunction with the number of homes connected to each tap to determine the total signal power at the amplifier. It is sufficient to calculate the loss to the nearest amplifier since, in a properly aligned plant, the net gain between amplifiers is zero. The loss between each tap port and the nearest amplifier was calculated for a sample of actual plants with low, medium, and high densities. The results are shown for densities of 24, 97, and 230 homes per mile (HPM) at 5 and 40 MHz in Figures 4 and 5. As expected, there are less high value taps in the rural area (24 HPM), since there is almost always some cable loss between the amplifier and the nearest tap. There are only minor differences between the distribution at 5 MHz and at 40 MHz.

The distribution of tap losses can be combined with the number of homes connected to each tap to determine the average loss from an average home to the amplifier port. This calculation was performed for all three densities at both 5 and 40 MHz and the results are shown in Tables 4 and 5. All losses are from the tap port to the nearest upstream (toward the node) amplifier station port. Table 4 shows what the total spurious level would be if the spurs add on a power basis. Table 5 shows the results if the spurs add in voltage. Measurements taken show a tendency to add higher than as power. However, in a statistical sense, over many plants, the power addition probably represents "most" plants, but the capability to add more destructively, the case that must be designed for, certainly exists. Since like crystals tend to behave like one another with regard to drift and aging, considering them as independent and random may be too lenient. Further, because they move slowly, any existing summation condition is likely to last for prolonged periods of time.

Example:

Consider a medium density (approx. 100 HPM) node with 1000 devices each emitting a signal at -35 dBmV at 5 MHz. If the signals add on a power basis, the total power at the node will be -35 + 11.63 = -23.37 dBmV. Note that Gain per Home (-18.37 dB) + 1000 homes (30 dB) = 11.63 dB. Thus, the number of homes is weighted by averaging it over the loss distribution. It is easy to show that, for large number of HPM, significant loading of the return due to interference alone can occur.

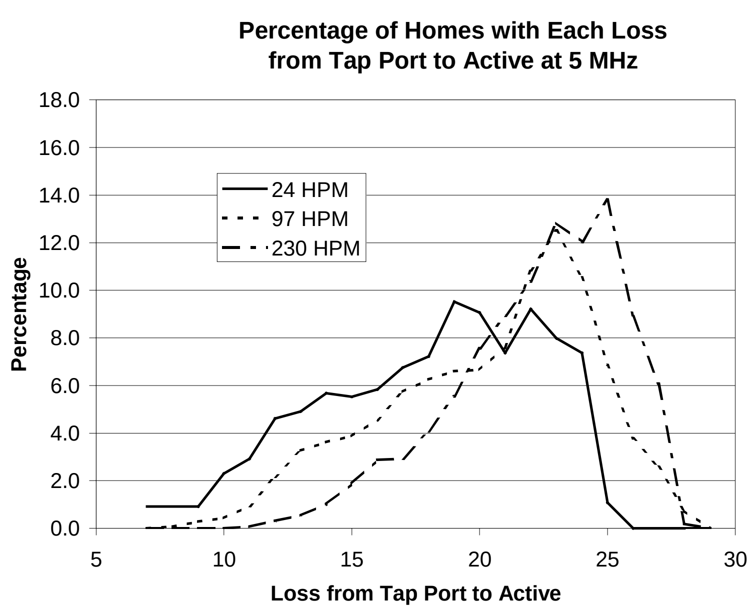


Figure 4 - Statistical Distribution of Return Path Loss @ 5 MHz

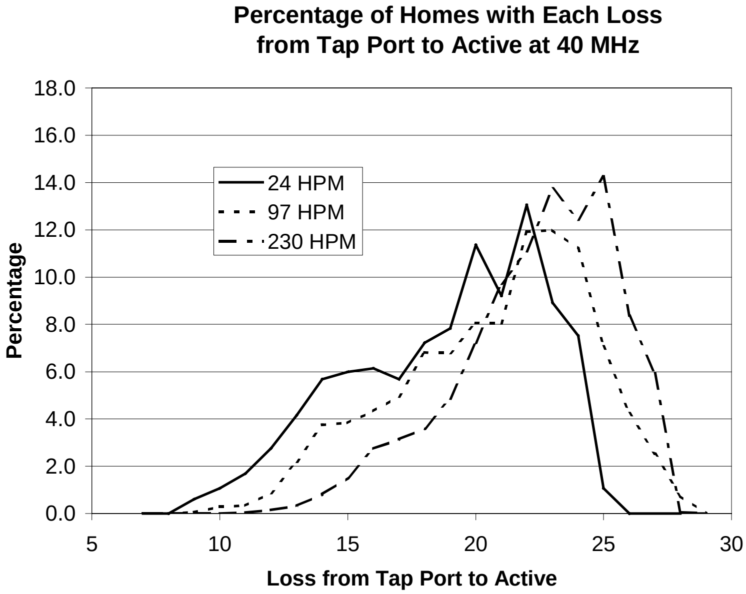


Figure 5 - Statistical Distribution of Return Path Loss @ 40 MHz

Frequency	5 MHz			40 MHz		
Homes per Mile (HPM)	24	97	230	24	97	230
Homes Passed in Sample Area	217	1306	1772	217	1306	1772
Total Gain (dB)	7.48	12.79	11.55	6.18	12.09	11.30
Average Gain per Home (dB)	-15.88	-18.37	-20.93	-17.19	-19.07	-21.19
Gain for 500 Homes (dB)	11.11	8.62	6.06	9.80	7.92	5.80
Gain for 1000 Homes (dB)	14.12	11.63	9.07	12.81	10.93	8.81
Gain for 1500 Homes (dB)	15.88	13.40	10.83	14.57	12.69	10.57
Gain for 2000 Homes (dB)	17.13	14.65	12.08	15.82	13.94	11.82

Table 4 - Power Addition of Spurs

Frequency	5 MHz			40 MHz		
Homes per Mile (HPM)	24	97	230	24	97	230
Homes Passed in Sample Area	217	1306	1772	217	1306	1772
Total Gain (dB)	29.75	42.92	43.33	28.69	42.39	43.15
Average Gain per Home (dB)	-16.98	-19.40	-21.64	-18.04	-19.93	-21.82
Gain for 500 Homes (dB)	37.00	34.58	32.34	35.94	34.05	32.16
Gain for 1000 Homes (dB)	43.02	40.60	38.36	41.96	40.07	38.18
Gain for 1500 Homes (dB)	46.54	44.12	41.88	45.48	43.59	41.70
Gain for 2000 Homes (dB)	49.04	46.62	44.38	47.98	46.09	44.20

Table 5 - Voltage Addition of Spurs

5.0 Conclusion

It has been demonstrated that the expansion of CATV networks to include broadband communications may have troublesome consequences if the pitfalls are not recognized. Specifically, return path transmitters have the potential to disturb forward services, a definite no-no. This can occur both for concurrent users in the same household, and possibly in a neighboring household. Aggravating the situation is the lack of quality off-the-shelf gear that finds its way into consumers' homes. A second important issue is the potential for ingress in the return band that is caused by the CATV equipment and consumer electronics. While upstream garbage is a well-known fact of life, any further obstacles in the upstream that are generated by this equipment should no longer be tolerated in new designs, in anticipation of the return path bottleneck on the way. Also, a strategy to deal with already deployed offending hardware must be thought through to make the most of the limited available return bandwidth.

Acknowledgments

The authors would like to thank Mr. Detlev Wundershock, Mr. Mike Short, and Mr. Dick Gresko for their invaluable contributions to the successful completion of this paper.

Mr. Glaab and Mr. Waight work in the Advanced Network Systems unit of General Instrument. Mr. Stoneback and Dr. Howald are with the Transmission Network Systems unit. GI's new address is 101 Tournament Drive, Horsham, Pa. 19044 (800-523-6678).

Considerations for Optical Dense Wave Division Multiplexing in Digital and Analog CATV Network Applications

John Holobinko, Dr. Jim Farina and Robert Harris

ADC Telecommunications, Inc.

The use of optical dense wave division multiplexing (DWDM) is well established in high speed, digital networks in both telephony and CATV networks. However, DWDM has not been deployed in broadband linear systems for CATV distribution to date. This paper explores the potential applications and benefits of DWDM in large scale CATV networks. A comparison is made between the economics, technical merits and technical obstacles of employing DWDM in digital backbones and deployment of DWDM in broadband linear distribution systems. Current technical limitations in broadband linear systems transmission are defined and explained (for example Raman and SBS effects), and empirical data is presented. The paper concludes with predictions on future functionality and economics of DWDM in both broadband linear and digital CATV network applications.

INTRODUCTION

The year 1998 finds the demand for transmission information capacity continuing to accelerate. Some carriers report a doubling of capacity requirements every five months. This puts a tremendous demand for new fibers on the physical fiber optic plant. In wide area fiber networks, the cost of the physical fiber plant represents a major portion of network costs. Some networks have exhausted their installed fiber capacity. Therefore, a tremendous need exists to reduce the cost of adding transmission without the need for investment in additional fiber cable installation, in order to provide new capacity and additional revenue generating services cost effectively.

Dense wave division multiplexing (DWDM) refers to the process by which multiple independent optical signals differing only by their respective optical wavelength (i.e. "color") can be transmitted simultaneously through a single optical fiber. DWDM technology effectively multiplies the transmission capacity of a fiber by the number of simultaneous wavelengths that can be effectively transmitted through that fiber. (Holobinko, NCTA Technical Papers 1997)

The deployment of DWDM technology for high speed digital optical transmission is now widely established in both long distance telecommunications and cable television regional interconnect applications. The total market for these systems is widely estimated to be $500 Million to $1 Billion in 1998. In the cable television industry, deployment of four and eight simultaneous wavelengths has been made by ADC Telecommunications, Inc. with the DV6000 2.4 Gb/s system, representing a total capacity of up to 20 Gb/s on a single fiber for transmission of video, data and voice services. Other systems designed primarily for long distance telecommunications carriers have demonstrated up to eighty (80) simultaneous optical wavelengths over a single fiber (Lucent Technologies, January 26, 1998).

Figure 1 illustrates a typical cable television regional interconnect network architecture. A digital fiber optic transmission backbone connects the primary hubs and the master headend. A combination of digital fiber transmission and broadband linear fiber transmission connects the primary hub and secondary hubs, while broadband linear transmission is used exclusively to connect the secondary hub with the nodes.

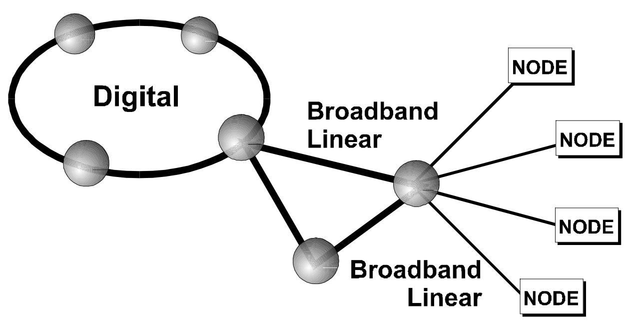


Figure 1

- DWDM is an enabling technology that has the potential of greatly affecting CATV networks. For example:
- What is the potential impact of DWDM technology on this network architecture?
- Can broadband linear transmission capacity be accomplished using DWDM?

- What will be the role, if any, for pure digital transmission, if DWDM using broadband linear techniques can be accomplished over long distances?
- What if any, are the network implications as fiber reaches further into the network and node sizes become significantly smaller?

One possible application for broadband linear DWDM transmission is in the forward path from the primary hub to the node. If one fiber is used for all broadcast services, a second fiber can be used to carry a number of narrowcast channels to the secondary hub, and even to individual nodes served from the secondary hub. This is represented in figures 2 and 3.

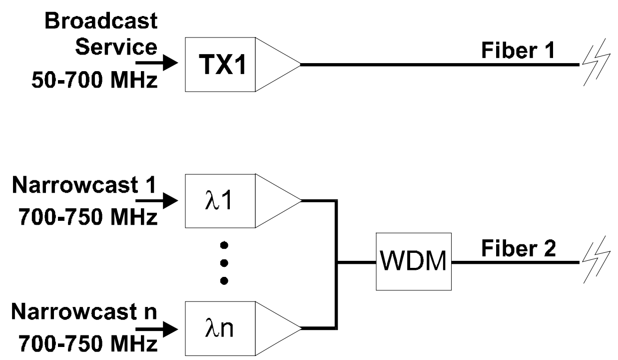


Figure 2.

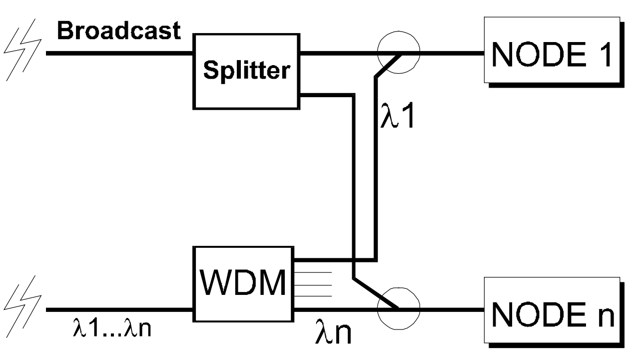


Figure 3.

In CATV systems comprised of a broadcast architecture with a narrow cast overlay, WDM can employed in the implementation of the narrow cast portion of the system. In these systems, ADC has assumed that the narrow cast optical signal consists of a linear transport scheme consisting of QAM modulated carriers. Multiple narrowcast signals can then be combined to comprise a WDM stream. This might lead one to assume that the system performance and limitations can be modeled as a purely digital system. However, because the QAM signals must be combined and coexist with the broadcast AM-VSB signals in the optical domain, the interaction of the QAM noise and distortions must be treated much the same as an analog problem in so far as these artifacts spill over into the AM-VSB bandwidth. Therefore the usual suspects in both linear and nonlinear signal degradation must be considered. These include: Stimulated Brillouin Scattering (SBS); Self Phase Modulation (SPM); and dispersion related CSO and CTB. In addition there are a few optical domain mixing processes such as Four Photon Mixing (FPM) and Stimulated Raman Scattering (SRS) which must also be considered.

Stimulated Brillouin Scattering (SBS)

In a system such as the one being considered here there need not be any special considerations for SBS. In particular, the AM-VSB and QAM optical signals can be considered separately since the wavelengths are to be intentionally different. In the QAM WDM link the energies are spread over discrete but separate wavelengths thus insuring safe SBS operation since no one optical carrier will rise above any SBS threshold particularly, in the usual first defense, phase or frequency dithering, is employed.

Self Phase Modulation (SPM)

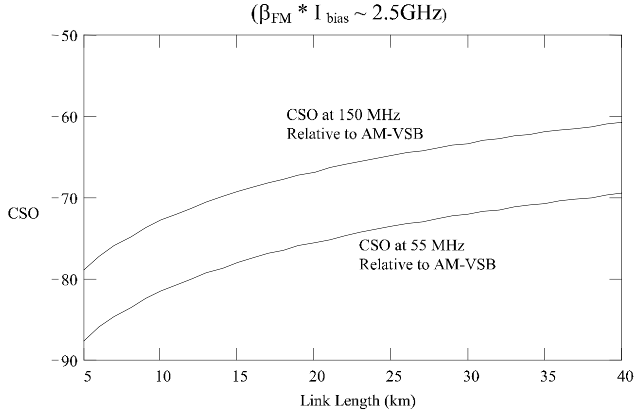
It is the opinion of the authors that there need not be any special consideration given to controlling SPM in this implementation other than that given to any other analog link.

Dispersion

In optical links, dispersion can result in significant CSO performance degradation in propagation over fiber. This effect will dictate the modulation mode for both AM-VSB and QAM carrying optical signals. In the case of the AM-VSB broadband optical link, dispersion will dictate external modulation because of the required link CSO performance. In most instances, AM-VSB requires external modulation because it is difficult to achieve the <-65 dB CSO with commonly used fiber (Corning SMF-28) with D~17ps-nm/km at 1550nm.

For the QAM signals, the situation is not quite as clear. In this case, the CSO specifications are not so stringent so as to prohibit the use of direct modulation on dispersive fiber. Figure 4 shows an estimate of the upper and lower CSO for a set of 10 carriers situated above 550 MHz. Note that the worst case is at 150 MHz. Since the dispersive effect is greater at high frequencies, the spill over into the lower frequency, AM-VSB signals is somewhat less than the worst case CSO for the QAM, which occurs at the highest frequency. Also in this

Figure 4 - CSO Due to Dispersion with Direct Modulation
(β_{FM} * I_{bias} ~ 2.5GHz)



figure, we estimate the CSO contribution from the QAM at 550 MHz relative to the AM-VSB carrier situated there. This assumes that the QAM carrier levels are -6dB relative to the AM carriers and the OMI for the QAM carriers is about 10%. Since the CSO in a QAM modulated system takes the form of a broadband signal spread over the whole channel bandwidth, one could argue it behaves much like additive noise to the AM-VSB signals. Therefore, if the added CSO "noise" can be tolerated, direct modulation of low chirp lasers on the ITU grid can be used for the QAM WDM narrow cast overlay and meet the overall network performance objectives.

Four Photon Mixing (FPM)

In WDM systems, FPM can result in significant signal degradation through an optical mixing process. This effect allows mixing in the optical domain to occur thus giving rise to third order optical products. This mixing process mimics the RF analogy in that the worst case for nonlinearities is where the optical carriers are evenly spaced. One strategy for the minimization of this effect is strategic placement of the wavelengths in the WDM system on the ITU grid. An additional remedy may be implementation of polarization diversity, which has also been reported to achieve a high degree of suppression of FPM.

Perhaps the most important factor governing the efficiency of this process is dispersion. The FPM process relies upon a phase matching condition and thus is spoiled by any chromatic dispersion present in the fiber. Therefore, we do not believe that FPM will be a major contributing factor to system performance. Measurements need to be performed to confirm this and assess the impact of this FPM on the WDM QAM link.

Stimulated Raman Scattering (SRS)

SRS has been a consideration of WDM system design for two reasons. The first is the inherent gain tilt caused by the Raman gain process. This is reasonably easily mitigated by tilting the signals at the primary hub to compensate for the SRS induced effects. The second and more troubling is a process which is somewhat like the optical analog of cross modulation. Because the Raman gain process relies upon at least two wavelengths being present and one providing the energy for the amplification of its companion, variations in the amplitude of the pump wavelength will result in gain variations and thus cross mod. Based on the magnitude of the carriers and their composition, ADC does not anticipate at this point that SRS will impose any severe limitation on the WDM link being considered here.

Effects of Other Components

Direct modulation of the QAM carriers promises to be the most economic solution to forward path narrowcast WDM transmission. If direct modulation is employed in the WDM QAM link, any component which has an optical frequency dependent transfer function will impart odd or even order distortions depending on its symmetry. In the case of this link, there are two potential contributors: the EDFA and the WDM muxes and demuxes. The EDFA gain profile can be minimized, to some degree, by gain flattening. However, the WDM mux/demux problem is not as simple. The optical filters in a DWDM will always possess some tilt or curvature particularly when the number of WDM channels becomes dense enough to require very sharp skirts for adjacent channel rejection. This can result in both CSO and CTB depending on the exact filter shape and the placement of the center wavelength of the laser within the filter response. This effect is potentially the most difficult to deal with and may prohibit immediate deployment of a full complement of sixteen ITU wavelengths because of the narrowness of each channel filter.

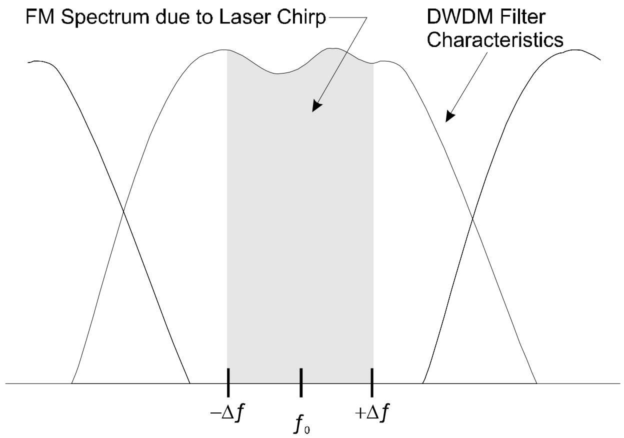


Figure 5.

The in-band figure to the optical filters in the DWDM demux will result in non-linear distortion.

Broadband Linear DWDM Component Economics

Figure 6 illustrates the three major elements which dominate the costs in a broadband linear DWDM network: Narrowcast optical transmitters are used to transmit different sets of channels down the same fiber at different optical wavelengths; DWDM passive optical multiplexers and demultiplexers for combining and separating optical wavelengths on a given fiber with a minimum loss of optical power; and gain flattened optical amplifiers which are used to amplify the entire optical spectrum on the fiber to compensate for losses of signal strength due to fiber attenuation, splitting losses, etc.

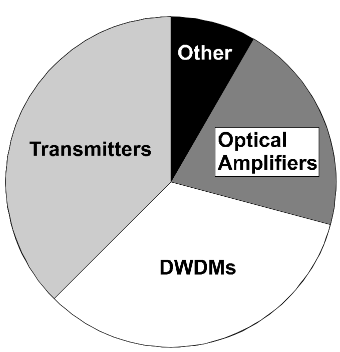


Figure 6

DWDM System Component Costs

Currently, the price of a narrowcast transmitter is roughly twice the cost of an 860 MHz broadband transmitter with ten times the output power. DWDM multiplexers and demultiplexers tend to be priced similarly to narrowband transmitters. Optical amplifiers are approximately three times the cost of the other components, but based on rapidly increasing volumes, the price of optical amplifiers is rapidly decreasing. If optical amplifier prices can be reduced by at least 50% from current levels, it will become practical to use standard optical couplers replacing DWDM multiplexers in many applications. At this point, the cost of lost power due to optical coupling will be less than the power savings minus the cost of the dense wave division multiplexor. This will reduce the overall cost of the WDM network, and also simplify physical configuration of the network. Fewer DWDMs also means less considerations of optical filter roll-off and impact to CSO and CTB.

Other Forward Path Limitations

As the number of narrowcast channels grow, expansion of the narrowcast spectrum beyond ten channels and to twenty, thirty or even forty channels will place significant demands on the narrowcast network. This paper focused on relatively low narrowcast channel counts. Further research needs to be conducted on means of controlling SRS and other optical phenomena as the number of narrowcast channels grow. The potential expansion of some domestic CATV networks to 860 MHz represents the anticipation that more narrowcast channels may be anticipated in the future.

Broadband Linear DWDM versus Digital Transmission in the Primary Ring

As optical amplifiers become more powerful and offer improved noise performance, and 1550 based broadband optical transmitters become more linear, the question is raised as to the viability of eliminating digital in the backbone altogether.

Broadband linear DWDM relies on the fact that the broadcast channels going to the nodes will be identical. From the secondary hub this can be the case. But from the primary hub a number of factors may dictate that the channel content at each secondary hub is different. These factors include but are not limited to the following:

- Revenue generation from narrowcast advertising
- Channel line up variations due to local must carry rules
- Imbedded VBI information in revenue generating services
- Local origination programming
- Data oriented services for SOHO and residential delivery

As the number of custom channels expands and channel configurations vary, combining optical narrowcast signal streams to create specific, differentiable broadband channel line-ups becomes technically difficult and expensive. The addition of more than two optical signals together results in noise floor penalties, and thus lower delivered signal quality. Simultaneously, new data applications call for expanding the data capacity to the primary hubs. (See Figure 7).

Based on these same factors, in regional networks the migration of broadband linear transmission back toward the primary headend is highly dubious based on both economics and technical considerations.

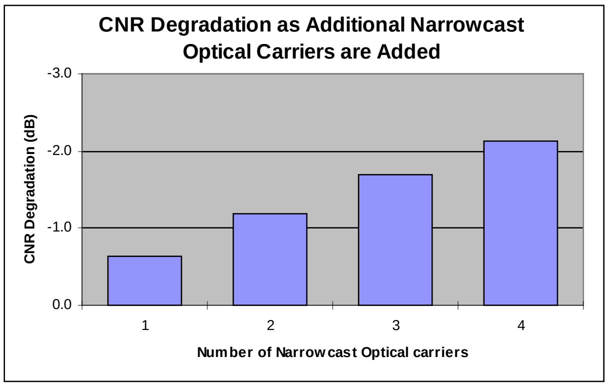


Figure 7.

Digital DWDM in the Primary Ring

The greatest potential for upside revenue in CATV networks today arguably is in non-traditional, i.e. non-entertainment based services. Some examples of these include

- E-Mail
- Internet Access
- High Speed Data
- Web Site Service
- Corporate VLANs
- Cable Telephony

The ability to provide new services to existing customers and to expand the customer base to include small office/home office (SOHO) and even larger corporate customers represents both new revenue opportunities and new network challenges. Given this scenario, CATV networks will soon find themselves in the same situation as other network providers: namely, digital traffic will grow exponentially and begin to consume enormous amounts of network capacity.

In this environment, the ability to expand network capacity incrementally, and gracefully is paramount. For example, if doubling of network capacity every six months is required, replacement of routers and transmission path equipment is unacceptable both in terms of cost and network service disruption. Many classic data networks are built on a distributed routing architecture as shown in Figure 8. This architecture is highly susceptible to this problem.

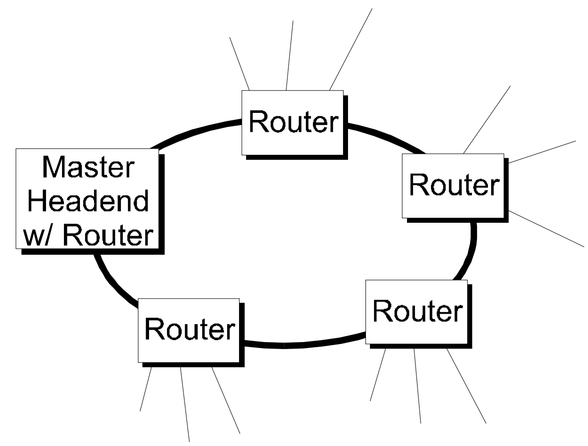


Figure 8.

Ring Network with Distributed Routers at Every Hub

An alternative architecture is to provide centralized routing with distributed switches that have some intelligence to provide data "sniffing", e.g. the ability to recognize multicast addresses, as shown in Figure 9. The advantages of this configuration are:

- Ability to segment the network into multiple parallel rings via WDM without replacing any individual switches;
- Through centralized routing, more balanced loading and greater ease in upgrade than with distributed routers.

In this scenario, data rates can be upgraded with minimum disruption to the network. Any equipment changeout is virtually limited to the master site where the network router is located. The use of DWDM and distributed switching with centralized routing provides a network solution for handling video, data and voice in an environment in which traffic is growing at high rates.

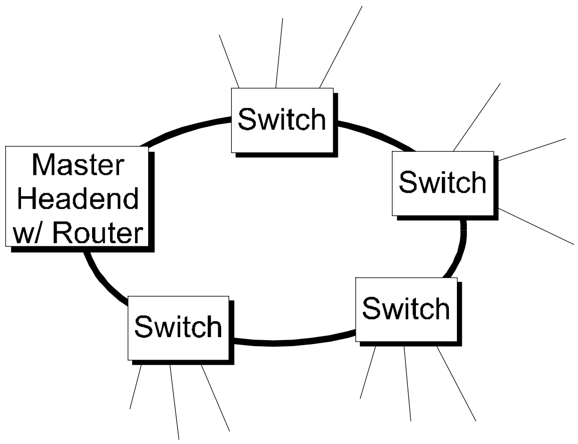


Figure 9.

Ring Network with Centralized Router and Distributed Switches at each Hub

Return Path Considerations

Given elimination of main service equipment at the secondary hubs, the return path information must be transported not only from the nodes back to the secondary hub, but on to the primary hub as well. In CATV networks, return path information takes the form of RF carriers of different bandwidths and modulation schemes, simultaneously occupying a total bandwidth of approximately 40 MHz. Dedicating one fiber per node to transport return path information to the primary hub would be cost prohibitive based on the transmission distances and number of nodes involved. Correspondingly, although DWDM addresses the fiber count issue, it is not practical based on the current cost of the DWDM devices. (The return path in essence would cost as much as the forward path fully loaded with narrowcast services.) Today, RF spectrum block conversion is an alternative for addressing this issue, albeit not an ideal one, since each step of frequency upconversion and downconversion introduces undesirable artifacts such as noise, group delay, etc.

A Look Towards the Subscriber

As optical amplifiers and other broadband linear DWDM components become more cost effective, the cost of bringing fiber closer to the customer will become a reality. Given that delivered VSB/AM quality to the residential tap is typically 49 - 50 dB CNR with -54 dBc distortions, an optical node which serves eight to thirty two homes passively can perform with very low signal level and relatively high distortions, since there are no other active devices such as amplifiers between the node and subscriber. At the point where no active devices exist between the optic node and subscriber, a look at return path methods will be required. Firstly, the number of return paths is inversely proportional to the number of homes served per node. For example, for a system with fifty homes per node, there are ten times the number of return paths as a system with five hundred homes per node.

CATV networks use an RF return path based on traditional architectures in which there are multiple active broadband devices required between the hub and the subscriber. As a result, for every return path signal on the network, format conversion is required at both the customer origination and the receive point, in the form of RF modulation and RF demodulation equipment. Yet, in contrast to the forward path where most channels are still analog video, virtually all information in the return path is digital.

As the node moves closer to the subscriber, there will be a time where a reevaluation of the use of RF carriers for transport of return path is necessary. If signals in the return can be sent to the node digitally, then multiplexed at the node, very low cost optics can be used to transport large amounts of data back to the hub, with no additional loss of signal quality, and a reduction in problems associated with RF interference. Given appropriate digital encoding techniques, the current 5-42 MHz return path represents an opportunity for a very high speed digital return path or shared digital and analog return path from the subscriber back to the headend.

CONCLUSIONS

DWDM is an enabling technology that allows digital backbones to expand capacity gracefully in response to exponential growth in new traffic types. Within the distribution system, DWDM represents a potential means to reduce the cost of delivering narrowcast services to secondary hubs and nodes, where the number of narrowcast services is limited to a few channels. As optics migrate closer to the subscriber, additional complexities in the return path may challenge current thinking on the accepted transmission technology of the return path.

REFERENCES

Holobinko, John and Hartmann, Bill, ADC Telecommunications, “Optical Network Technology: Future Impact on CATV Networks, 1997 NCTA Technical Papers, pp. 284-289.

Lucent Technologies, January 26, 1998 News Release, “Lucent Technologies delivers record-breaking optical networking capacity; five times greater than current systems.”

Author: John Holobinko
Director, Business Development
ADC Telecommunications, Inc.
999 Research Parkway
Meriden, CT 06450
PHONE: (203) 630-5771
FAX: (203) 317-4160
email: john_holobinko@adc.com

Co-authors: Dr. Jim Farina
Director of Technology,
BCD Division
ADC Telecommunications, Inc.
999 Research Parkway
Meriden, CT 06450
PHONE: (203) 630-5700

Robert Harris
Program Marketing Manager,
Digital Transmission
ADC Telecommunications, Inc.
999 Research Parkway
Meriden, CT 06450
PHONE: (203) 630-5732

COST VERSUS FLEXIBILITY IN DIGITAL CABLE HEADENDS

Paul Harr
Scientific-Atlanta, Inc.
Atlanta, GA

ABSTRACT

Cable Programmers are faced with a multitude of challenges as they move to digital. Maintaining the flexibility of their existing analog systems today must be weighed against the cost of the equipment that it will take to provide that same flexibility in a digital architecture.

An architecture that allows the cable programmer the same flexibility as analog is one that incorporates high definition (HD) and standard definition (SD) digital encoders. In this architecture the program element streams are demultiplexed and switched to allow for commercial insertion. This option however, will be expensive as the cost of decoding and encoding is high today will likely remain so for the foreseeable future.

On the other hand, lower cost options will exist that include processing the incoming programmer and broadcast signals at the transport stream level. This option, however, provides less flexibility than program stream switching. Program streams are not demultiplexed and therefore do not allow seamless insertion of commercial ads. Technologies are becoming available which provide switching between multiple signals at the transport layer.

This paper will compare high cost / high flexibility technology with lower cost / lower flexibility technology for cable headends.

INTRODUCTION

Digital TV promises to change the complexion of television. It will surely change the appearance of the cable headend infrastructure. From simply transmitting a single digital signal to managing a dynamic digital multiplex, cable headends will have many different looks.

The deployment of Digital TV offers the cable operator many options to convert his station to digital. Some cable operators have already begun the transition to digital and are receiving pre-packaged digital programming from services such as Headend In The Sky (HITS). This type of service, while providing many channels tightly packed into a satellite transponder for easy conversion to quadrature amplitude modulated (QAM), is very inflexible for adding advertising or simply subtracting a single service.

Soon, other programmers will also be transmitting their own digital signal, offering an alternative to HITS. Broadcasters likewise, will be transmitting their signals in digital starting this year. Both will transmit some programming in high definition television (HDTV) and other programming in the form of standard definition television (SDTV).

What this all means to the cable operator is that he will ultimately have to find bandwidth to carry these new services while trying to achieve the same flexibility he has today. Competition is driving cable operators to add the new digital services. DBS, Telco and broadcasters transmitting HDTV will force the cable operator to keep up and convert his system to digital. While must-carry for digital services is likely to be hotly debated, if passed, it will further drive bandwidth requirements for the cable operator.

Digital was the promise of virtually unlimited bandwidth for cable. It was believed at least 500 channels could be carried on a single cable. However, with HDTV, the need to add local advertising and carriage of other digital services, the operator must weigh optimizing his bandwidth against the costs to manage that bandwidth.

TRADITIONAL CABLE HEADEND

Compared to the complexity of new digital cable headend architectures, today's analog architecture appears refreshingly simple. Today's analog cable network is built around transmitting just one video channel in one RF channel at a time. Programming is received from typically two sources; cable programmers over satellite and broadcasters over-the-air or via a direct fiber link. Figure 1 illustrates this.

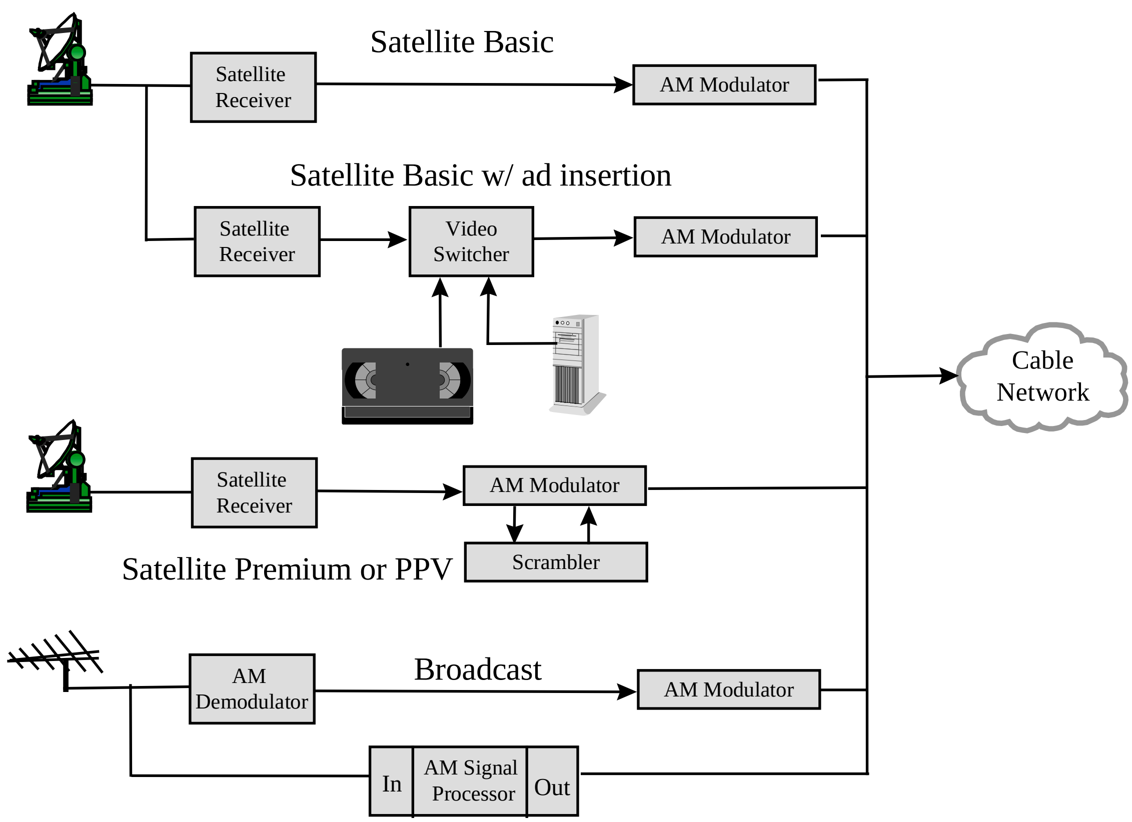


Figure 1. Traditional Headend

Cable programming takes one of several forms. It is either a basic, premium, or pay-per-view (PPV) service. If it is a basic service, then it will either allow local ad insertion or be passed straight through over cable. Premium services are scrambled or secured passively (traps) while providing no ad insertion opportunities. PPV is always scrambled and authorized addressably through a set top box. No advertisements are allowed on the PPV channel except on video barker channels.

Broadcast channels are received using a signal processor. A signal processor receives the broadcast signal on frequency, cleans up the out-of-band noise and amplifies it to approximately 60 dBmV. The processor either converts it to a different frequency for cable distribution or keeps it on-channel. Ad insertion is not allowed on over-the-air broadcast channels.

Ad insertion, which accounts for an increasing portion of cable systems' revenue today, is inserted on basic channels with available local spots. Ad insertion has already gone the way of digital in many headends. Ads are queued on a server and played out at the appropriate time and decoded back to analog before being inserted into the analog channel.

In analog, cable headends are simple and flexible. Digital cable, on the other hand, now limits the operator's operational flexibility and the insertion of advertisements into local spots.

DIGITAL CABLE CHALLENGES

Digital Cable presents several challenges to the operator. The first is where does the operator find the bandwidth to carry all these new digital HDTV and SDTV programs? The second challenge is how can the operator optimize his bandwidth to extract every bit possible? The final challenge is how does the operator achieve the same flexibility he has today in an analog system?

Pipe Analysis

If we consider the multiple transmission formats or "digital pipes" that affect cable we see that the inbound and outbound data rates are not the same. Because digital modulation formats used by satellite, broadcast and cable are all different, the information rates do not equal. Satellite transmission uses QPSK modulation over either 36MHz or 27 MHz transponders. Different transponder bandwidths are used in an attempt to optimize cable's information rates, however, this is an inefficient use of the transponder if the transponders' bandwidth is not optimized. Digital broadcast transmission will use 8VSB modulation which provides an information rate of 19.39Mb/s. Cable's outbound digital pipes will use either 64 or 256 QAM.

Different spectrum bandwidths and modulation efficiencies result in digital cable bandwidths that will not be easily optimized. Consider Table 1. Cable programmers using satellite distribution must decide how they transmit their service to the cable operator. They must consider the number of services they are trying to deliver versus the transponder spectrum bandwidth available. For example, if the programmer is planning to deliver 10 channels each at 4.0 Mb/s, then he will overrun the cable operator's capability to transmit a 40Mb/s multiplex transport stream. As a result, operators cannot carry a 40Mb/s transport stream because it does not have a large enough out-bound pipe. Therefore, the operator will need to "break up" the multiplex.

Pipe	Inbound 8VSB	Inbound QPSK	Outbound 64QAM	Outbound 256QAM
Broadcast	19.39Mb	-	-	-
Satellite (36MHz @-40dB)	-	42.74Mb	-	-
Satellite (23MHz @-40dB)	-	27.27Mb	-	-
Cable	-	-	27Mb	38.81Mb

Table 1 Digital Pipes' Information Rates

Broadcasters, on the other hand, will be using 8VSB which provides a maximum of 19.39 Mb/s information rate at the full transport level. Broadcasters have said they will use this spectrum to transmit either HD or SD programs. Only one HD program at the highest scan rate of 1920 X 1080I can fit into the broadcast transport stream. Whereas, four or five SD programs are possible in the same bandwidth. In either case 19.39Mb/s does not fill-up the out-bound pipe creating wasted bandwidth.

Digital Pass-Through

Operators must decide how they will convert their digital signals to digital cable signals. The lowest cost method is called pass-through. Cable operators will want to pass-through the signals received from satellite and broadcasters because only a minimum amount of equipment is required. Programming from satellite can be received by an integrated receiver decoder capable of decoding all programs, simultaneously. The output of this multi-decoder will be an over-the-air transport stream. The transport stream will have local conditional access added and then be modulated on the cable by a QAM modulator.

Programming from broadcasters will be received either off-air or from a dedicated link from the broadcaster. If the signal is received off-air, then it will be received by an 8VSB demodulator capable of providing the transport stream from the broadcast signal. This output will then be remodulated by QAM modulators. Conditional access will not be needed for the broadcast signal.

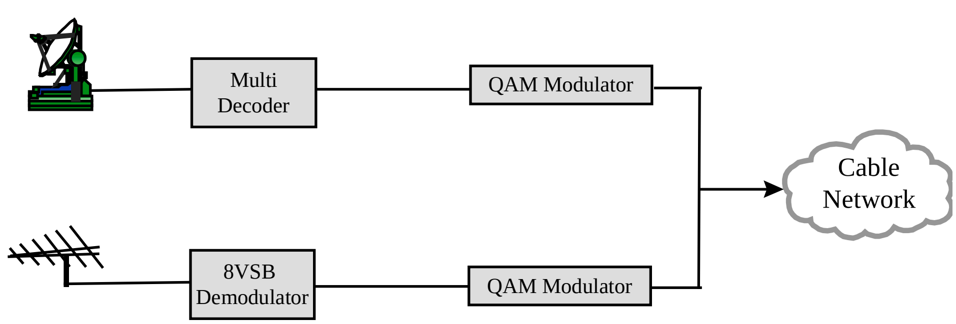


Figure 1. Pass-through

Bandwidth Analysis

So, why is simply passing-through the signals a big problem? Less equipment is needed and therefore it must cost less. The cable operator can use the digital equivalent of satellite receivers, modulators, and signal processors to convert satellite and broadcast signals to the cable format. Unfortunately, the problem is that digital compression has changed all the rules. An analog video signal is the same whether it is satellite, broadcast, or cable. In digital, the same video signal can be 19Mb/s of HD, 5Mb/s of high quality SD, or 2Mb/s of low quality SD, depending solely on the content originator.

What this means then is that depending on the content data rate and whether it is received from satellite or broadcast, the cable operator may be sacrificing more bits than is necessary. Table 2 illustrates a hypothetical cable system's bandwidth environment. The incoming broadcast and programmer services have different numbers of program streams. The cable operator's outbound pipe can be either 64 or 256 QAM and is compared to the satellite and broadcast inbound pipes. What becomes noticeable is that in a pass-through implementation the input and output pipes are not optimized. Satellite delivered services, unless tweaked down in data rates, overruns the output pipes. Broadcast signals, on the other hand, underrun the output pipes. In the case of broadcast programming the wasted bits begin to add up to real bandwidth.

			Unused Bandwidth	
Service	**PES**	**Inbound Bandwidth**	**64QAM**	**256QAM**
CBS	1	19.39Mb	7.61Mb	19.42Mb
ABC	1	19.39Mb	7.61Mb	19.42Mb
NBC	1	19.39Mb	7.61Mb	19.42Mb
FOX	4	19.39Mb	7.61Mb	19.42Mb
HBO	2	27.27Mb	-.3Mb	11.54Mb
Discovery	3	27.27Mb	-.3Mb	11.54Mb
Turner	2	42.51Mb	-15.54Mb	-3.7Mb
MSG	2	42.51Mb	-15.54Mb	-3.7Mb

Table 2

For example, the information rate difference between a single 8VSB broadcast signal and one 64QAM cable signal is 7.61Mb. This is enough bandwidth to carry at least two program streams of reasonable quality video and even more with lesser quality. At 256QAM the difference is more dramatic. As cable systems begin to carry digital signals operators cannot afford to waste bandwidth and must examine ways to recover this unused bandwidth.

Grooming

One technique that allows the operator to overcome the bandwidth inefficiencies is "grooming." Grooming allows an operator to receive an array of compressed program streams from different sources and multiplex them in his own unique way.

Of course, grooming requires more equipment to implement and therefore is more expensive. However, operators are likely to reach a point when delivery of digital programming dominates its bandwidth and that bandwidth becomes scarce. When this happens, the operator will need a way to squeeze out every possible bit for more services whether it be video, audio or data. Figure 2 illustrates how to accomplish grooming.

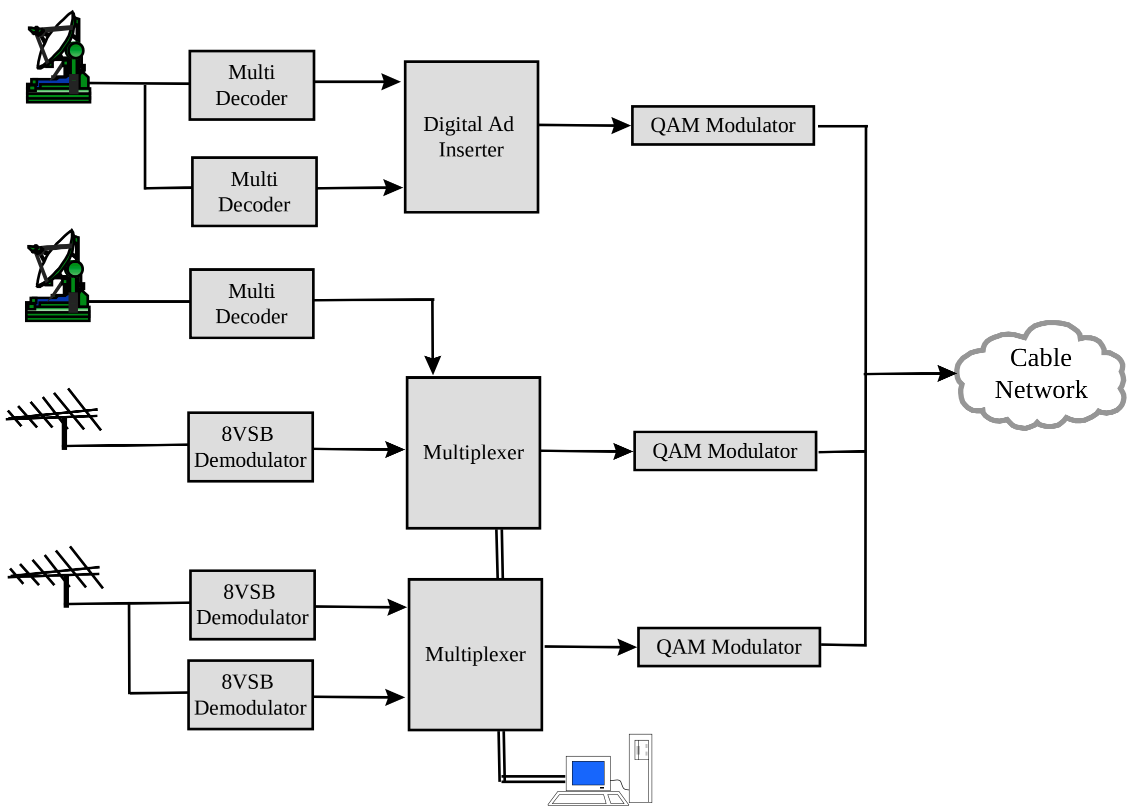


Figure 2. Grooming w/ Digital Ad Insertion

With the use of multiplexers, operators can manage the bandwidth of their digital network. Multiplexers allow "cherry picking" program streams in a way to maximize the data in the transport stream. Program streams that are not desired can be easily deleted. Transport streams that do not fill up the "cable pipe" can be combined with other transport streams. Where there is bandwidth left over, data can be inserted into the transport stream to provide for digital set top boxes or TV applications.

Ad insertion will be accomplished using a digital ad inserter. Compressed program streams from the IRD will be input into the digital ad inserter. Ads will have been already compressed and queued up on a server. At the appropriate time ads will be switched into the transport stream while the program stream is temporarily deleted. At the end of the ad, it will be switched from the transport stream and the program stream re-inserted. Switching at the program element stream (PES) layer is called splicing.

Splicing can be non-seamless or seamless. Non-seamless splicing means there is a momentary visible artifact or glitch in the picture as the IRD needs an I-frame from the new program stream to lock-up. Seamless switching means the inserter knows where the I-frames are in both streams, buffers up any difference, and switches when they are equal or close to it. Splicing is the optimum way to switch in the compressed domain, however, it requires more sophisticated multiplexing equipment and pointers in the MPEG streams to identify splice points. Today, splicing is in early development but will eventually be commonplace as the demand for seamless switching increases.

Conclusion

Digital, while it offers cable operators increased channel capacity, also threatens to consume much more bandwidth than it needs. Because the size of the pipes are not naturally equal, the operator must take measures to minimize bandwidth waste and protect his assets.

The operator's challenge is recovering wasted bandwidth and then adding or combining other services. Pass-through offers the lowest cost alternative to trans-modulate digital signals for broadcasters and programmers, but does not allow the operator to recover unused bandwidth. To overcome the limitation of pass-through, operators will employ multiplexers with grooming capabilities to manage systems that will add, drop, and combine program element streams and optimize the digital pipe.

Initially, switching between digital programs will be crude and cause minor impairments in the video. As we employ tricks like fade-outs and fade-ins we can minimize glitches. Eventually, techniques such as splicing will make program switching and ad insertion seamless in the digital domain.

In the past, a cable system's infrastructure has been the cable operator's most important asset. In the future, bandwidth will be the most important asset in a digital cable system. Tools such as multiplexing systems will help cable operators manage this asset more effectively and overcome bandwidth inefficiencies created by new demands placed on that bandwidth.

Countdown to DTV

Bob Rast

General Instrument

Abstract

The new broadcast digital television service (DTV) will launch this year. The presentation focuses on where things stand, what will likely happen, and how that will impact cable digital service.

Tick, Tick ...

So much to do, so little time. That is true for broadcasters as they start this year to replace, end-to-end, their 50 year old analog television service with DTV, the new, digital service based on the ATSC standard. At long last, following 10 years of government-fostered standards setting, 1998 is D-Year.

The focus of this presentation is to update and address many of the burning questions regarding broadcast DTV as it relates to cable. Insights should fly off the screen as we summarize what amounts to the reinvention of broadcasting, and how cable plans to handle those new signals.

It all starts this year. Twenty-six TV stations in major markets across the country have pledged to launch DTV service this fall. Starting in 1999, FCC-mandated deadlines dictate a forced march to DTV, with commercial broadcasters having to be on air no later than May 1, 2002, and non-commercial (public) stations on air by May 1, 2003. The FCC has declared that analog broadcast service will cease in 2006, after which the recovered spectrum will be auctioned. Few believe the transition can occur that quickly.

Questions, We've Got Questions ...

What will the service be? Most predict a mixture of HDTV and SDTV, depending on day parts. Given a lower data rate than is possible on cable, what is the capacity of the broadcast channel? Will it be interlace, or progressive? How about 480P?

Will the new digital TV receivers be ready? Will the networks be ready? How about the local stations? Will they just pass through the network feed, or will there be digital local content? Where will they get the programming? Will the service be free, or pay/subscription?

Cable Impacts

How are digital broadcast signals going to get through cable systems, and into the new digital TV's, as well as the legacy analog TV's? How compatible are the ATSC digital broadcast signals with cable digital systems now being deployed? What about VSB to QAM, PSIP, and on screen displays? Is there ever going to be a digital interface, to allow compressed digital HD signals to be passed from the set top into the new DTV's?

Will cable HDTV programming be available? When, and by whom?

Competitive

How is satellite dealing with HDTV? DirecTV announced an HDTV service to launch late this year. What is that service and what about the other satellite providers?

If HDTV succeeds, cable is strategically advantaged. Why? Why is satellite disadvantaged?

Regulatory

Must carry is a major FCC issue this year, and can dramatically affect both cable and broadcasting. What is likely to happen?

Consumer

DTV's were introduced at the Consumer Electronics Show in January. Will they be ready and shipping in time to support service introduction this year? What is the outlook for prices? Will consumers buy? Though the volume of TV's sold may be small, those consumers are likely to be inordinately influential. Can cable afford to be on the sidelines?

Is This For You?

If you have questions about DTV on cable, this talk should help. If you already know the answers, come share some of them with the rest of us.

DELIVERY OF NEW PACKET-BASED SERVICES OVER CABLE: THE PACKETCABLE™ CONCEPT

Steven C. Craddock, Comcast Corporation
David Reed, Cable Television Laboratories, Inc.

Extended Abstract

CableLabs, and its member companies, are developing a packet network to overlay on cable, which will deliver a new family of services. Generically called "PacketCable™," these services could be any useful features capable of being transported by a packet network. The technical underpinning of this overlay network is the Internet protocol (IP) running over standard cable modems. The following extended abstract discusses the conceptual framework and focus of the PacketCable development effort.[1]

PACKETCABLE SERVICES

Any data or information service that lends itself to packet-based transport is a potential PacketCable service candidate. We anticipate that the initial PacketCable service set will be Packet Voice and Packet Video for business and residential markets. Our approach to providing these services is described below. Over time, it is likely that other IP-based business and residential services will be added to the product line. Key factors differentiating the PacketCable concept from the market are: the North American footprint represented by CableLabs' member companies; cable networks' ample bandwidth, enabling broadband packet-based services such as video; and the capability of packaging a unique set of packet-based services customized to users' needs.

As envisioned, Packet Voice will provide a PSTN-like (Public Switched Telephone Network) experience to users by providing all the enhanced features of circuit-switched telecommunications systems. Consumers will originate and receive voice calls using either their existing telephones or a multimedia personal computer (PC). Using the service, consumers will encounter normal PSTN-like audible tones, such as a dial-tone when going off-hook, a busy signal if the receiving station is off-line or engaged in another call, or ringing. In addition, consumers will receive the equivalent of today's CLASS Services (*e.g.*, call waiting, call forwarding, and 3-way calling).

Packet Video is an audio and video conferencing service that will enable consumers to place and receive video calls. At the customer premises, we envision delivering a symmetrical service through a cable modem, and terminating in one of three premises device options:

1. a special adapter we call a Multimedia Terminal Adapter (MTA), interconnecting a television set, video camera, and a standard telephone;
2. a cable television OpenCable™ digital set-top, equipped with a PacketCable client, interconnecting a television set, video camera, and a standard telephone; or
3. multimedia PCs equipped for video conferencing (plus any required peripherals).

We also envision some asymmetrical applications using one-way Packet Video with two-way audio. For example, live video conferencing with a call center representative where the customer sees and hears the Customer Service Representative (CSR), and the CSR only hears the customer.

---

1. The authors note the contributions of many other individuals, collectively representing the membership of the PacketCable product definition group, for their contributions to the production definition concept.

PACKETCABLE NETWORK SUPPORT

Conceptually, the PacketCable development effort focuses upon the higher application layer, with the presumption that network-layer services, and those below, will be provided by the DOCSIS modem or the OpenCable set-top box delivered through a hybrid fiber/coax (HFC) cable network. The exception is the requirement for quality of service created by the isochronous[2] PacketCable services.

Within this context, our development effort assumes PacketCable services will utilize a "cable intranet," or private IP network, that uses edge devices, known as clients, to perform protocol conversions that permit communications between subscribers directly connected to those devices. It also allows PacketCable subscribers to originate and terminate calls to stations connected to the PSTN through other edge devices, known as PSTN gateways.

A "cable intranet" consists of interconnected IP networks operated locally by the individual cable companies and interconnected, managed-IP backbones (such as @Home and RoadRunner/MediaOne Express). Such interconnect capabilities will give the PacketCable service a North American footprint (*i.e.*, allowing PacketCable subscribers to make calls throughout North America to other subscribers without having to interface with the PSTN). In instances where connectivity with the PSTN is desired, such connections may be accomplished through conventional interconnect means. This may include leveraging the existing presence of some MSOs and their affiliates in the CLEC (Competitive Local Exchange Carrier) and CAP (Competitive Access Provider) markets to provide these interfaces.

2. A descriptive terminology to describe a datastream that has a guaranteed, constant time relationship, or a signal in which the time intervals between consecutive bits have the same duration.

PACKETCABLE INFRASTRUCTURE

The infrastructure required to deploy PacketCable services in a local geographic area, consists of three categories of equipment:

- a logical gatekeeper, consisting of a set of servers, to manage most aspects of the service connections;
- clients homed on the Packetcable server group; and
- local gateways used to enter or exit the IP network to the PSTN.

This architecture can be expanded to cover a larger "On-Net" footprint by interconnecting local PacketCable networks using a hierarchy of directory servers and IP gateways. The actual media connections between local IP networks, over which the voice and video packets are transmitted, are made through a series of IP gateways, which function as private network access points called C-NAPs (Cable - Network Access Points). The PacketCable directory-server hierarchy and C-NAP gateways permit calls to be originated by a PacketCable client in one gatekeeper zone. The calls are terminated to a PacketCable client in any other gatekeeper zone served by PacketCable, regardless of who operates the system.

SPECIFICATION EFFORT

Whenever possible, our preference in creating a PacketCable specification is to utilize existing technical standards or known service-quality parameters where they exist. For example, Packet Voice is an isochronous service, the quality of which is very sensitive to packet loss, "jitter,"[3] and overall network latencies (delay). The PacketCable development effort will specify an acceptable range of values for these parameters. Moreover, we require Packet Voice be compatible with the basic functions of ITU recommendation H.323. Similarly, Packet Video is also an isochronous service, the quality of which is very sensitive to packet loss, "jitter," video/audio "skew,"[4] and overall network latencies. This two-way symmetrical service will be based primarily on ITU H.263, supporting different grades of video quality.

Our investigation of these relevant technical standards, however, reveals that they are incomplete to specify and to support fully the deployment of interoperable IP voice and video services over cable networks. Consequently, the PacketCable committee is currently in the process of developing a functional specification to deliver the services noted above that "fills in the gaps" of these technical standards. The PacketCable committee hopes to complete this specification with the goal of delivering commercial products for consumer and business applications within the next 9–12 months.

---

3. The time elapsing between the emission of the first bit of a data block or packet by the transmission client, and its reception by the receiving client, is known as the network transit delay. The variation of this delay over time, or the extent to which it varies from packet to packet, is known as "jitter."

4. Video/audio "skew" refers to the stringent type of synchronization that occurs when speech is played out at the same time the image of the speaker is displayed. In the case of video conferencing, this is also known as "lip-synchronization."

Designing “Outside The Box”: An Alternative Solution for High-Density Architectures

Jim Mabry
Network Specialist
Philips Broadband Networks

Abstract

Today’s HFC designs are limited by three main factors: End-of-line (EOL) performance, cascade limitations, and homes passed per node. In high-density architectures (urban areas or MDUs with greater than 150 subscribers per mile), the number of homes passed is the primary issue. This factor, coupled with the practice of dividing a node into sections for future fiber migration plans, leads to shorter amplifier cascades instead of maximized cascades. Why are the forward path system designs of today being limited by unclear, future return path usage? Why is the optical receiver/node a bottle neck for return signals? Can we eliminate these limitations by using readily available equipment in an asymmetric cascade?

This paper suggests an alternative, cost-effective broadband network design for high-density architectures that allows the operator to use fewer forward transmitters to serve more customers today, while building a future-proof system for tomorrow.

TODAY’S ARCHITECTURE

Over the past few years, the future of digital services and the return plant has been the subject of much debate, which is certain to continue. In the meantime, we continue designing CATV networks based on these primary limiting factors: end-of-line (EOL) performance, cascade limitations, and total homes passed per node.

EOL Performance and Cascade Limitations

Currently, the Federal Communications Commission (FCC) requires a minimum EOL performance of: 43 dBc* Carrier to Noise ratio (CNR); -51 dBc Composite Triple Beat (CTB), Cross Modulation (XMOD), and Composite Second Order (CSO) distortions. (*Note: CNR is to be raised to 45 dBc.)

Most CATV operators, however, aim to provide a better-quality picture signal to their customers than merely the minimum performance. Typical EOL performance for a 750 MHz design (77 analog channels and 550-750 MHz reserved for digital) is: 48 dBc CNR; and -53 dBc CTB, XMOD, and CSO.

With readily available nodes and radio frequency (RF) amps, one fiber optic link (typical performance) from the headend to the node easily can provide a cascade of the node plus four actives (see Figure 1). Many cable operators, however, have reasons – such as total system reliability – for keeping the RF amplifier cascade as low as possible; also, these operators can achieve desired EOL performance with reduced fiber optic link performance and/or higher RF output levels. This paper discusses the cascade options available to operators who desire to keep their options for future services flexible.

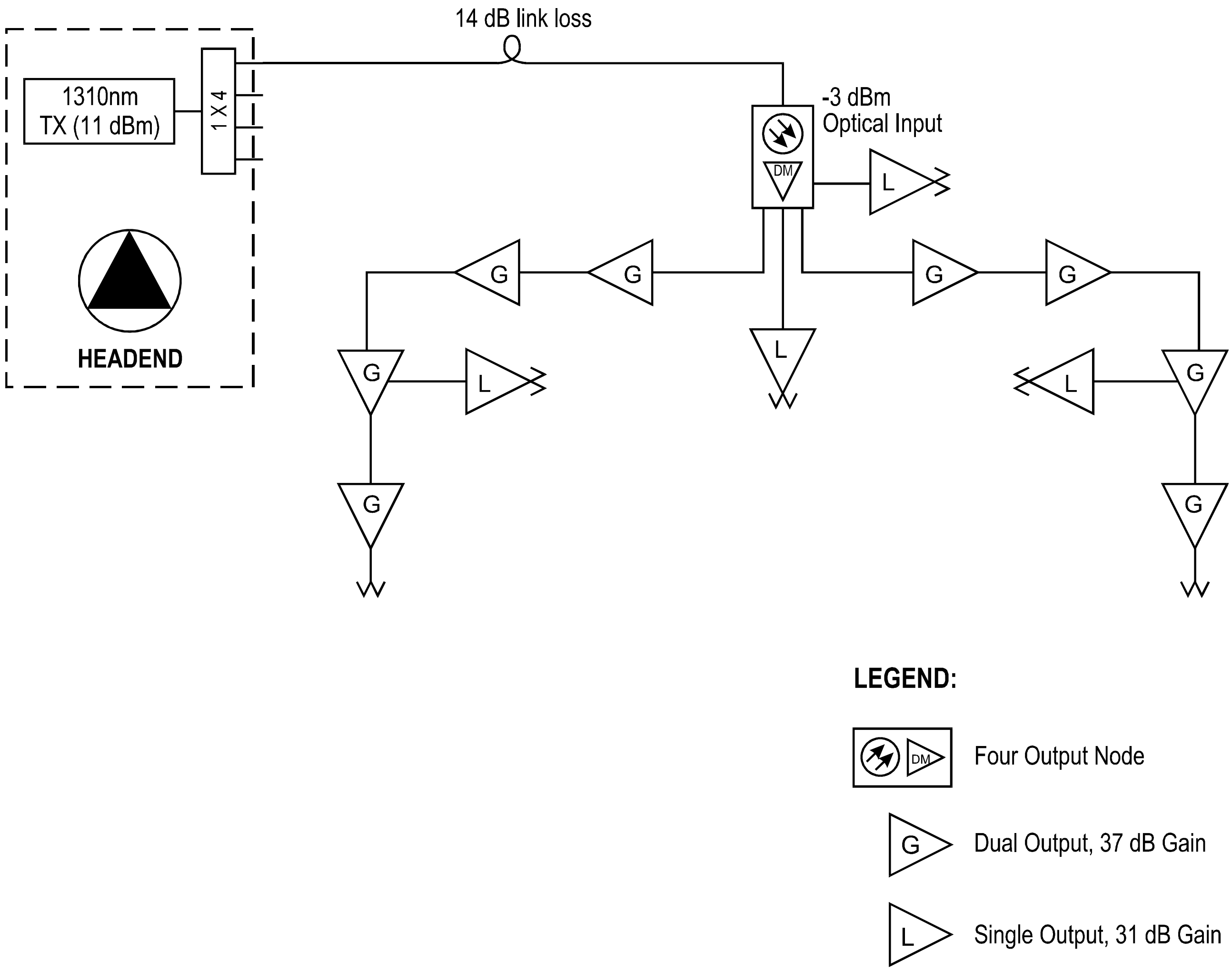


Figure 1. HFC Cascade: Node Plus Four Amplifiers (48 dBmV output level for all devices)

In this cascade, the RF output of all actives is +48 dBmV. The operating gains of the dual- and single-output amps are 37 and 31 dB, respectively. Typical performance of the fiber optic link provides for: 51 dBc CNR; -68 dBc CTB and XMOD; and -64 dBc CSO. The calculated EOL performance for the cascade in Figure 1 is: 49 dBc CNR; -54 dBc CTB and XMOD; and -60 CSO.

<u>Total Homes Passed Per Node</u>

Again, apart from system reliability, the issue of the maximum number of homes passed in any given node is related primarily to anticipated future usage patterns of the return path for telephony or other digital services. Figure 2 uses a Plain Old Telephony Service (POTS) chart to illustrate the relationships between available return bandwidth and concentration level.

Available Return Bandwidth	Concentration Level: none	2:1	3:1	4:1	5:1	6:1	8:1	10:1
4.5 MHz	72	144	216	288	360	432	576	720
9.0 MHz	144	288	432	576	720	864	1,152	1,440
15.0 MHz	240	480	720	960	1,200	1,440	1,920	2,400
19.5 MHz	312	624	936	1,248	1,560	1,872	2,496	3,120
24.0 MHz	384	768	1,152	1,536	1,920	2,304	3,072	3,840
30.0 MHz	480	960	1,440	1,920	2,400	2,880	3,840	4,800
34.5 MHz	552	1,104	1,656	2,208	2,760	3,312	4,416	5,520
39.0 MHz	624	1,248	1,872	2,496	3,120	3,744	4,992	6,240
45.0 MHz	720	1,440	2,160	2,880	3,600	4,320	5,760	7,200
49.5 MHz	792	1,584	2,376	3,168	3,960	4,752	6,336	7,920

Figure 2. System POTS Line Capacity per Available Bandwidth

The bandwidth is based on the fact that a T1 line provides 24 subscriber lines and requires approximately 1.5 MHz of bandwidth. The concentration level is the ratio of the number of subscribers that the system is designed to handle versus the actual number of subscribers on line simultaneously. For example, if only 15 MHz was available in the return spectrum, and the system was designed to handle 100% (1:1) of the subscribers (for example, on Mother’s Day), then a node with only one return transmitter and no frequency up converter should have no more than 240 homes passed in it. As shown, many different configurations can be derived from this type of predicted usage pattern. Contemporary node sizes generally range from 500 to 2000 homes passed. Also, most designers try to create some sort of node segmentation to enable future fiber migration plans without the need for much redesign or additional cable.

Example of a Traditional System Design

A cable operator plans to upgrade a system to 750 MHz (77 analog channels and 550-750 MHz reserved for digital), with no more than 1800 homes per node or 450 homes per quadrant of a node, using a four-port optical receiver/node.

This example assumes the following: 300,000 potential subscribers (total homes passed); 1,500 plant miles; and an average density of 200 homes per mile.

In a perfect plant, which has a consistent density throughout the entire system, 167 nodes would be needed. This assumes that the cascade length selected (node plus four amplifiers, as in Figure 1) will be able to reach the extents of every node. Since the total number of homes passed per node is the most likely limiting factor in high-density areas (such as in this example), the longest cascade in a node often will be only the node plus three amps.

In the headend, assuming that one transmitter will feed no more than four nodes, this system will need at least 42 high-powered, 1310 nm transmitters. If long fiber optic links or other factors reduce the ratio of nodes fed from one transmitter to 3:1, then the number of transmitters needed will increase to 56. Table 1 displays these numbers and compares them with the alternative system design.

As the first step in the system upgrade, the network is “cut” into sections to create the nodes. In the traditional fiber migration plan, sections are divided up by the maximum number of subscribers per node, or 1800 homes passed. Then, the designer creates the subsections while laying out the design.

As system demand (mostly return path bandwidth) grows, and/or return noise and ingress increase, the cable operator has a few options: 1) dividing the return path into smaller sections by adding one or more optical return transmitters into the node housing itself; 2) employing a form of return frequency block up-conversion at the node and then down-converting back at the head end; or 3) upgrading the sub-sections, or quadrants, to complete, stand-alone node stations. Most migration plans recommend installing reserved, or “dark,” fiber in the system to easily facilitate this transition. (Refer to the 1997-8 CED Cable TV Fiber Topologies Comparison for examples.) Also, many manufacturers have nodes capable of adding extra forward optical receivers and return transmitters to increase bandwidth both downstream and upstream.

ALTERNATIVE INCREMENTAL SYSTEM DESIGN

Why are the forward path system designs of today being limited by unclear, future return path usage? Why is the optical receiver/node a bottleneck for return signals? Can we use an asymmetric cascade? CATV operators without concrete plans for future return services may find building a future-proof network through an incremental system design process more cost effective than traditional methods.

With the very first step, the architecture’s optimization should be based on all three main limiting factors – not one individual factor. These are: 1) EOL performance; 2) amplifier cascade; and 3) number of subscribers per node segment.

EOL Performance and Amplifier Cascade

This design process will not lower the desired EOL performance. The signal’s target performance from the headend transmitter through the single fiber optic link to the node and through the RF cascade will still be 48 dBc CNR, and -53 dBc CTB, XMOD, and CSO.

Figure 3 shows a cascade of the node plus six RF amplifiers. The fiber optic link performance remains the same in both scenarios. As compared to Figure 1, this cascade extends each node's reach by two actives; this is accomplished by lowering the RF output levels of all amplifiers by 1 dBmV, from +48 dBmV to +47 dBmV. The EOL performance for this system is: 48 dBc CNR; -53 dBc CTB and XMOD; and -59 dBc CSO.

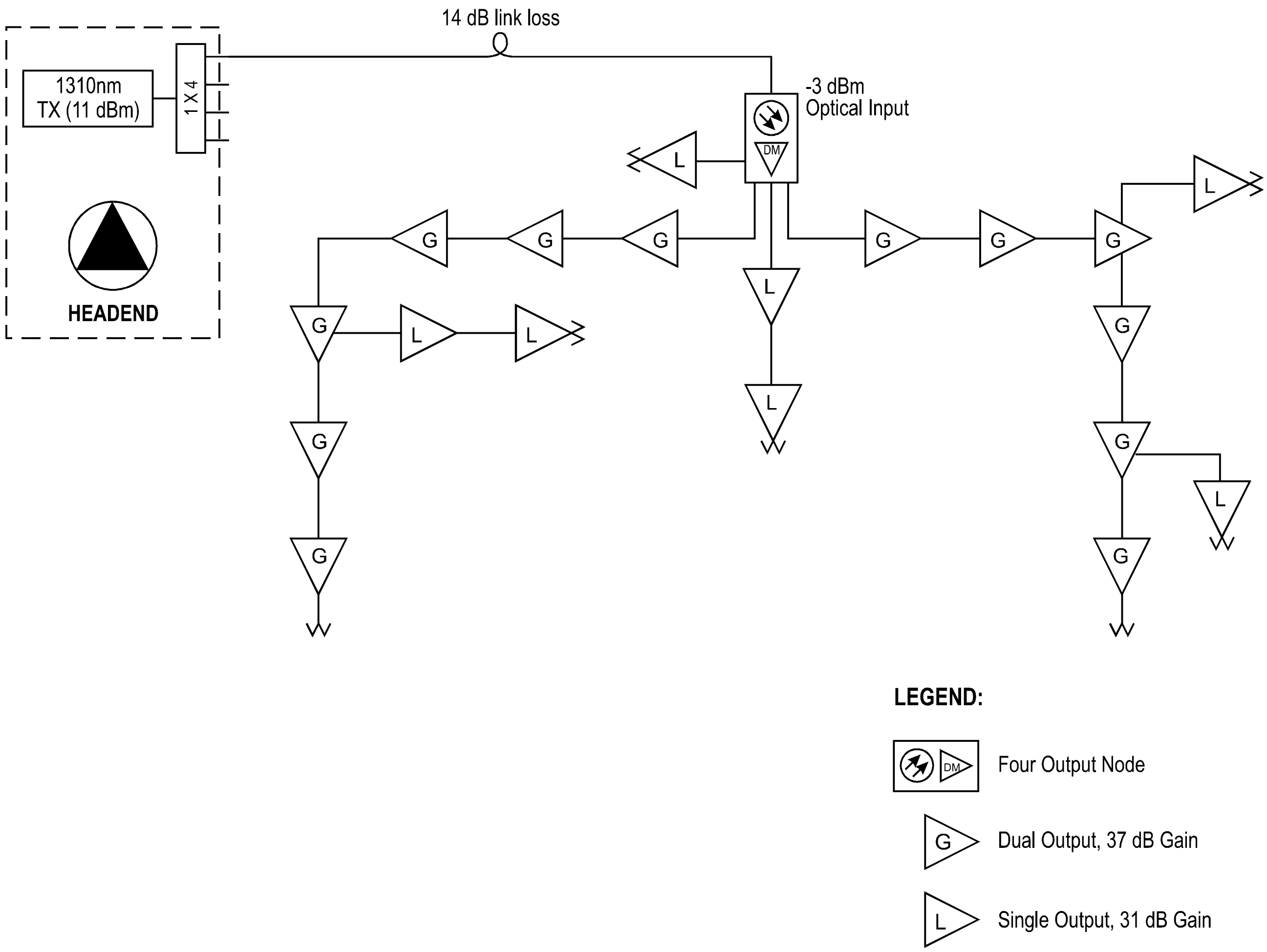


Figure 3. HFC Cascade: Node Plus Six Amplifiers (47 dBmV output level for all devices)

<u>Number of Subscribers Per Node Segment</u>

The next step of this system design “cuts” the network into sections to create nodes. Since this is a time-consuming process and the ultimate goal of the incremental architecture, the entire system should be divided into sections based on the smallest desired service area. This could be 300-500 home pockets. Then, the nodes can be placed at optimum locations to use the full reach of the amplifier cascade, with minimal regard to the total number of homes passed. Important to note, at this step, is that each pocket does not have to originate from the node. The node can be at a location where the last four actives of a six-amp cascade form a 300- to 500-home pocket. For example, in a highly populated area, this sort of node may now feed 2000-3000 homes, with a node plus six-amp cascade. Figure 4 gives an example of this system layout.

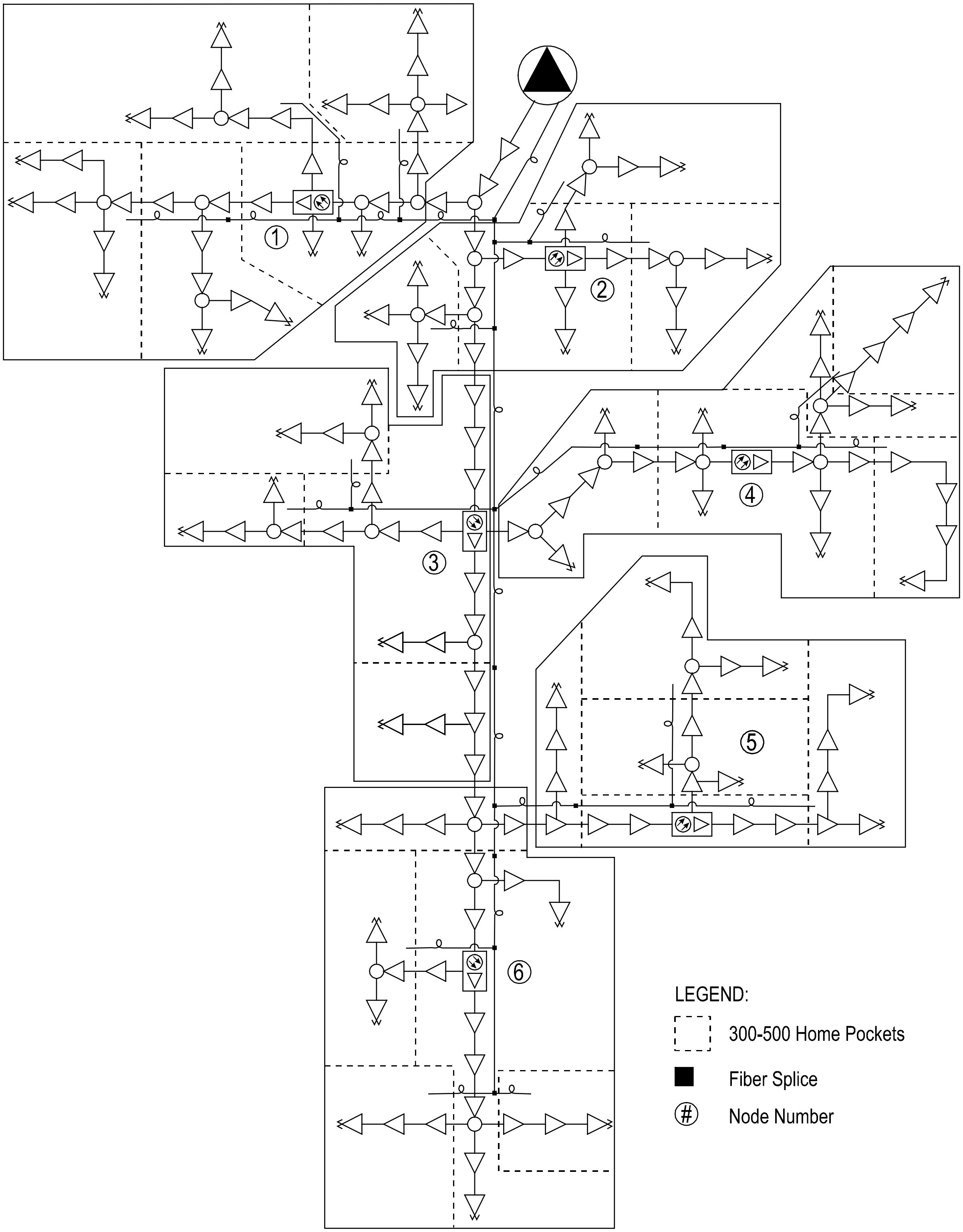


Figure 4. Segmented HFC Plant with Fiber Optic Overlay

As in the traditional fiber migration plan, several dark fibers should be placed into the system for each node as the main fiber cables are installed. This includes running the extra fiber out to the smaller home pockets. The costs of fiber optic cable and electronics may increase or decrease with time; however, labor costs certainly will increase.

The next steps involve the incremental increase in system bandwidth per home based on subscriber usage patterns, return noise funneling, and ingress. While all the above-mentioned upgrade options are still available, this plan offers an *extra* option.

The three premises of this incremental design are that: 1) most readily available nodes will not hold more than three forward receivers and two return receivers, in addition to any status monitoring or extra features; 2) return frequency block conversion currently is neither readily available nor cost effective; and 3) the forward path is not a source of congestion.

<u>Asymmetric Cascade</u>

That being said, as data usage increases demand for return bandwidth, and noise/ingress becomes more of a problem, a return transmitter can be placed into an amplifier station currently downstream of the node!

Thus, the forward and return signal paths will be of different cascade lengths in each node service area. The forward video and downstream data still travel through the node and the complete RF cascade; however, the return path is reduced. The return transmitter sends the upstream signals directly to the headend without bottlenecking all of the signals in the optical node itself! Figure 5 details how this works.

As more downstream digital bandwidth is needed, this system offers two options:
1) adding extra receivers in the node for narrowcasting; or 2) fully converting the subsection of the original node, which was upgraded with an optical lid and a return transmitter, to an independent node.

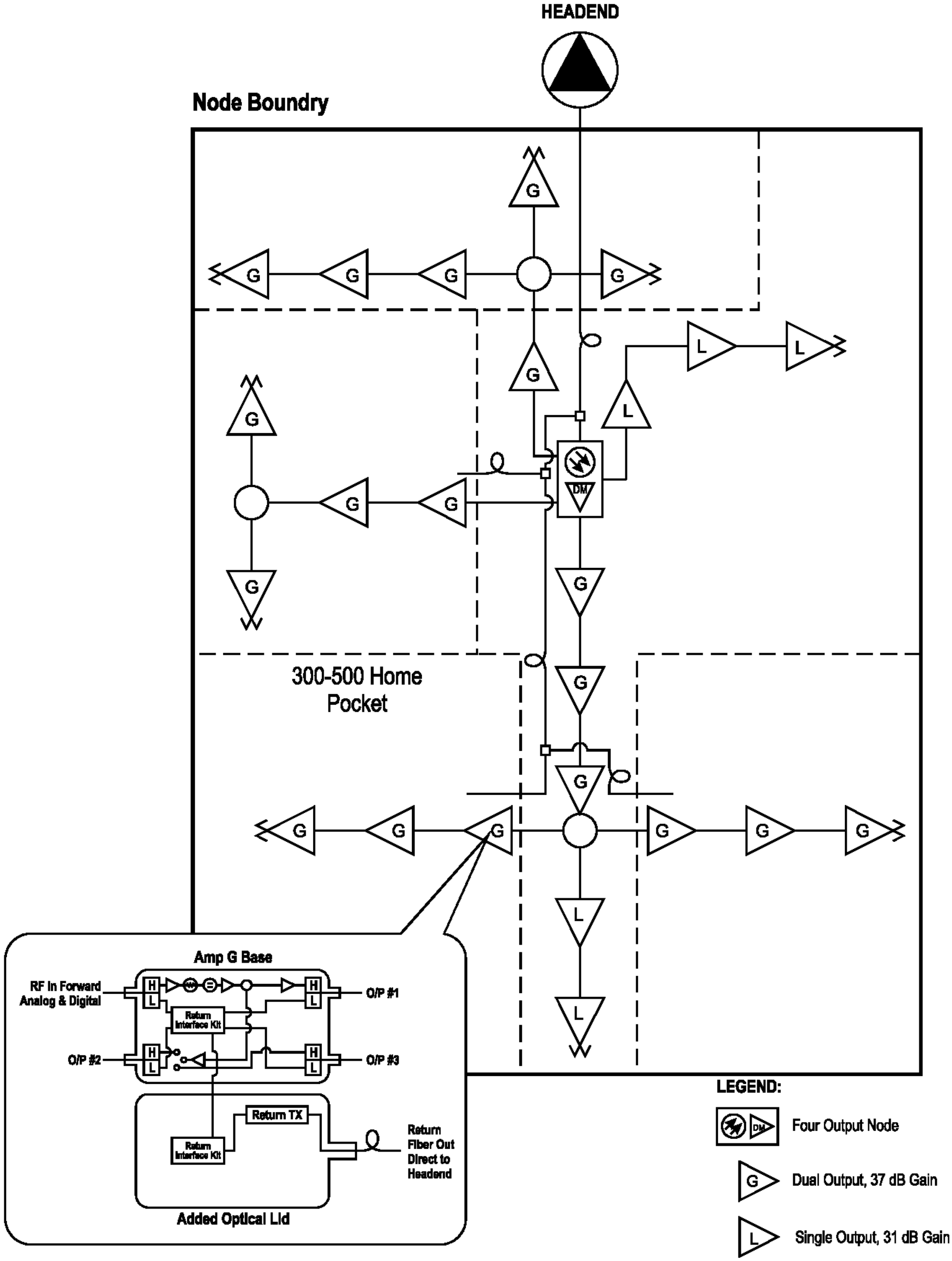


Figure 5. Proposed Alternative Design Solution

Homes per Node		Number of Nodes	Number of TXs 4:1 Ratio	Number of RXs 3:1 Ratio
	1800	167	42	56
	2500	120	30	40
% Change (2500 home node relative to 1800 home node)	+39%	-28%	**-29%!**	**-29%!**

Table 1: Sample Comparison Between Homes Per Node and Number of Forward 1310 nm Transmitters Required for Initial System Upgrade

What does this save the cable operator?

The bulk of the cost savings would be seen in the initial forward fiber optic laser deployment. In the above example, assuming that the nodes now are expanded to serve an average-sized node of 2500 homes and that the system is equally dense, at least 120 nodes are necessary. Applying the same generalizations to the node-per-transmitter ratio, between 30 and 40 transmitters are necessary, as compared to 42-56 transmitters if the nodes were smaller. This represents a *29% decrease* in forward transmitter requirements! (Refer to Table 1.)

Extra initial savings result from the reduced need for return receivers and associated hardware in the headend, along with *28% fewer* optical node stations in the field.

The extra material costs of adding the return transmitter downstream of the node will be limited to the return transmitter, a return receiver (if spare port is not available in the head end), a return interface kit, an optical lid, and possible status monitoring or ingress protection accessories. This option, however, is still less expensive than completely upgrading an RF station to a stand alone-node.

CONCLUSION

This unique alternative incremental architecture is not consistent with many telecommunications operators' plans. Companies that already have a solid idea of what they want from their system in terms of bandwidth capacity and reliability are already past this phase. Likewise, many low- to medium-density systems have all the return bandwidth necessary. These system operators ask manufacturers for equipment that will work in longer cascades at 750 MHz. The asymmetric cascade described in this paper, however, may offer some of the high-density system operators a cost-effective architectural option when upgrading existing networks or planning a completely new system.

ACKNOWLEDGMENTS

I would like to thank my colleagues at Philips Broadband Networks, Inc., especially John Caezza, Manny Serrano, Tony Rossi, and Greg Davis for their advice and support of this concept.
Phone: 800-448-5171
e-mail: jmabry@pbn-us.be.philips.com

DIGITAL PROGRAM INSERTION FOR LOCAL ADVERTISING

Mukta Kar, Ph.D., Majid Chelehmal, Ph.D., Richard S. Prodan, Ph.D.
Cable Television Laboratories

Abstract

Current advertising insertion systems enable cable headends and broadcast affiliates to insert locally-generated commercials and short programs in a remotely distributed regional program before delivery to the home viewer. The revenue generated by the local ads and short programs is very significant. Current ad insertion systems are analog in nature, switch between uncompressed video sources, and use digital video compression for local storage only.

Digital compression provides digital audio, video, and data with superior quality and efficiency to existing analog means. The application considered in this paper involves insertion of locally-generated compressed digital commercials and short programs into a digital channel containing previously multiplexed and digitally compressed programs. However, combining compressed video streams presents additional challenges for seamless insertion of ancillary programs. It is difficult to splice a compressed digital bitstream compliant to the MPEG-2 standard without adversely affecting the resultant display due to decoding failures at compressed bitstream discontinuities.

This paper presents a brief overview of existing analog and digital program insertion systems.Various solutions, including limitations not present in current analog systems, are discussed. Flexibility, implementation complexity, and network operational constraints are also discussed for various potential digital program insertion methods.

INTRODUCTION

Cable advertising generates significant revenue for the cable industry. National ad-sales (advertising sales) and local ad-sales in early 1980 were $50 million and $8 million, respectively. National ad-sales for 1997 are expected to top $5 billion. Revenues from local ad-sales are expected to top $2 billion. These figures underscore the growing importance of advertising for Multiple Systems Operators (MSOs).

Current advertisement (ad) or commercial insertion methods are essentially analog, using strictly uncompressed video. A hybrid system stores local spots in compressed form, but decodes upon playback into uncompressed video prior to insertion. Analog insertion stores the commercial on tape in analog format. Hybrid insertion stores commercials in digital and compressed format on a server and converts them back to analog format before insertion into analog channels.

The advent of television signal compression technology and its standardization (MPEG-2), coupled with digital modulation, have ushered in a new era in the television broadcasting industry. One of the most important benefits is the bandwidth efficiency in spectrum utilization compared to its analog counterpart. An existing 6 MHz analog channel can be used to send multiple digital channels with equal or better picture and sound quality. This indicates that the existing limited number of analog channels can be transformed into a larger number of viewing or logical channels. These logical channels can be used for audio, video, and data services. To take advantage of this benefit, MSOs are gradually moving to a digital tier and adding more programs.

Cable operators also expect that their ad sales revenue will go up by inserting additional commercials targeted to a more specific audience or market segment in the added digital channels. Unfortunately, the devices used for insertion of commercials in analog channels will not work in compressed digital channels without significant

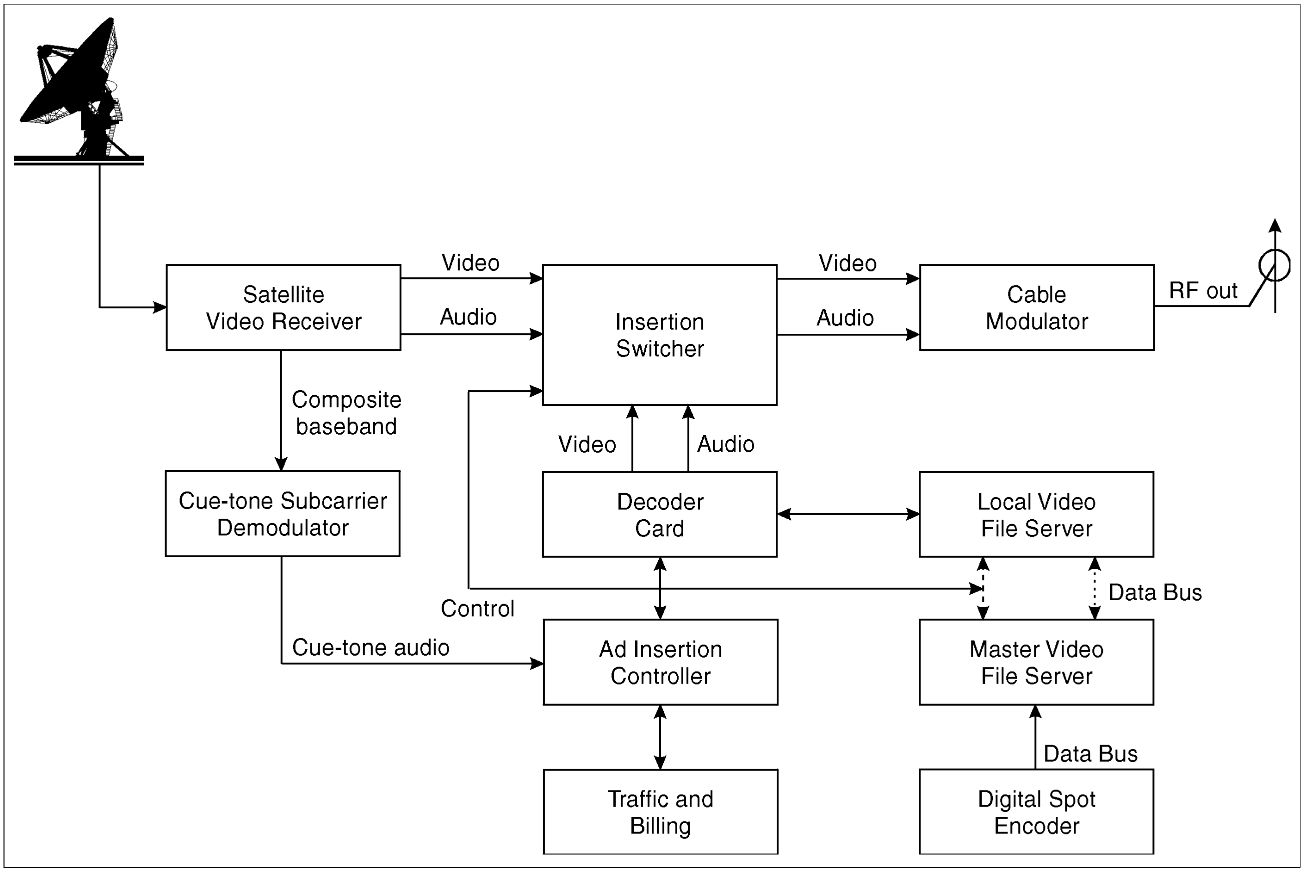


Figure 1. Hybrid Insertion System

modifications. Although operators are adding digital channels to their headends, a true digital ad insertion system is not commercially available. Currently, the added digital channels are used mostly for pay-per-view content. Very few commercials are inserted locally into those digital channels. Commercial insertion devices and systems will become more important as the transition from analog to digital program delivery progresses.

Advanced digital technology brings efficiency, flexibility, and other benefits over its analog counterpart. However, in some applications (*e.g.*, splicing MPEG-2 bitstreams) it can present more problems. Splicing is the fundamental technique used to insert commercials or short programs in channels, in editing audio/video content in post-production houses, and channel switching in headends and other broadcast stations.

ANALOG PROGRAM INSERTION

In the analog domain, splicing between two NTSC video sources or TV programs is relatively easy. A switch between two video sources at any frame boundary can be accomplished, since the frames are equal in size and time duration and are independent of one another. Analog insertion maintains synchronization among video sources prior to splicing. The resultant video, with the switched video source at the insertion or splice point, will appear seamless.

Figure 1 shows the block diagram of a typical commercial insertion system in cable headends and broadcast stations. In existing analog systems, networks distribute their content to cable headends and broadcast affiliates via satellite using a baseband bandwidth close to 10 MHz. As shown in Figure 2, the lower portion of the transmitted signal spectrum is used for NTSC video. The upper portion of the spectrum above 4.2 MHz contains a few FM subcarriers which are used to transmit mono, stereo, multi-language audio, and data. One of the subcarriers is used to transmit cue-tone signals. The subcarrier is modulated with a cue-tone signal using frequency modulation (FM) or frequency shift keying (FSK). Some networks send cue-tones using a pair of FSK tones (*e.g.*, 25 Hz and 35 Hz) within one of the audio channels. The integrated receiver-decoder (IRD) receives the RF-modulated signal, demodulates the cue-tone subcarrier signal, and provides an output signal. Either a dual-tone multiple frequency (DTMF) audio cue-tone sequence and/or a relay contact closure signal is provided to the insertion equipment. The networks send their schedules and spot availability (avail) in advance for local advertising insertion in a program. The precise location of the spot is signaled by a cue-tone during program broadcasts. The cue-tone consists of a sequence of numbers indicating the start and the end of an insertion opportunity. For example, the Discovery Channel uses a DTMF local avail cue-tone with 826* indicating the start, and 826# indicating the end of the spot.

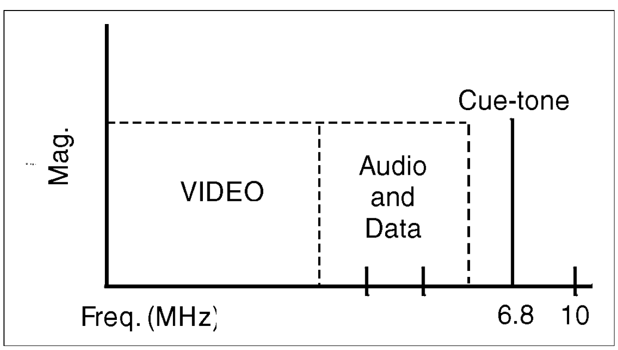


Figure 2. Analog Baseband Used by Networks

DIGITAL PROGRAM INSERTION

Compressed digital video frames are unequal in size (number of bits) due to the variable length coding compression techniques employed. This complicates locating the video frame boundaries. Not all frames in MPEG are independent of one another since substantial compression is achieved using temporal prediction between successive video frames. Also, compression algorithms may reorder the sequence of frames. The MPEG-2 standard has additional complexities, such as encoder-decoder clock synchronization, and decoder buffer management.

In an MPEG-compressed video elementary stream, the size of the compressed frames (access units), in terms of the number of bits, varies considerably. The buffer in the decoder smooths out the changes caused by this compression characteristic, as shown in Figure 3. For example, the size of an intra-frame (I-frame) may be 100 kbits, while that of a predictive-frame (P-frame) and a bidirectionally predictive-frame (B-frame) could be 35 kbits and 25 kbits, respectively. Hence, the compressed frames take unequal amounts of time to travel from the encoder to decoder at a fixed bitrate. However, the output of the decoder produces uncompressed video frames (presentation units) of equal size and duration. The buffer removes the effects of the variation in the arrival of the compressed frames on the non-varying periodic video decoding and display.

The minimum buffer size for a decoder must be specified for the buffer model of an MPEG-2 profile (*e.g.*, Main Profile at Main Level requires 1.8 Mbits). The decoder has no control of the decoder buffer fullness. When MPEG bitstreams are encoded, there is an inherent encoder buffer occupancy at every point in time. The same buffer fullness is reflected exactly in the decoder buffer. The constant transmission rate of the compressed video bitstream fills the buffer and the video decoder removes variable size compressed frames every display period, as shown in Figure 3.

Any discontinuity caused in the original bitstream, by instantaneously switching to a new bitstream, can cause the decoder buffer to overflow at some time in the future. The buffer fullness determines the delay or, equivalently, the time data spends in the buffer.

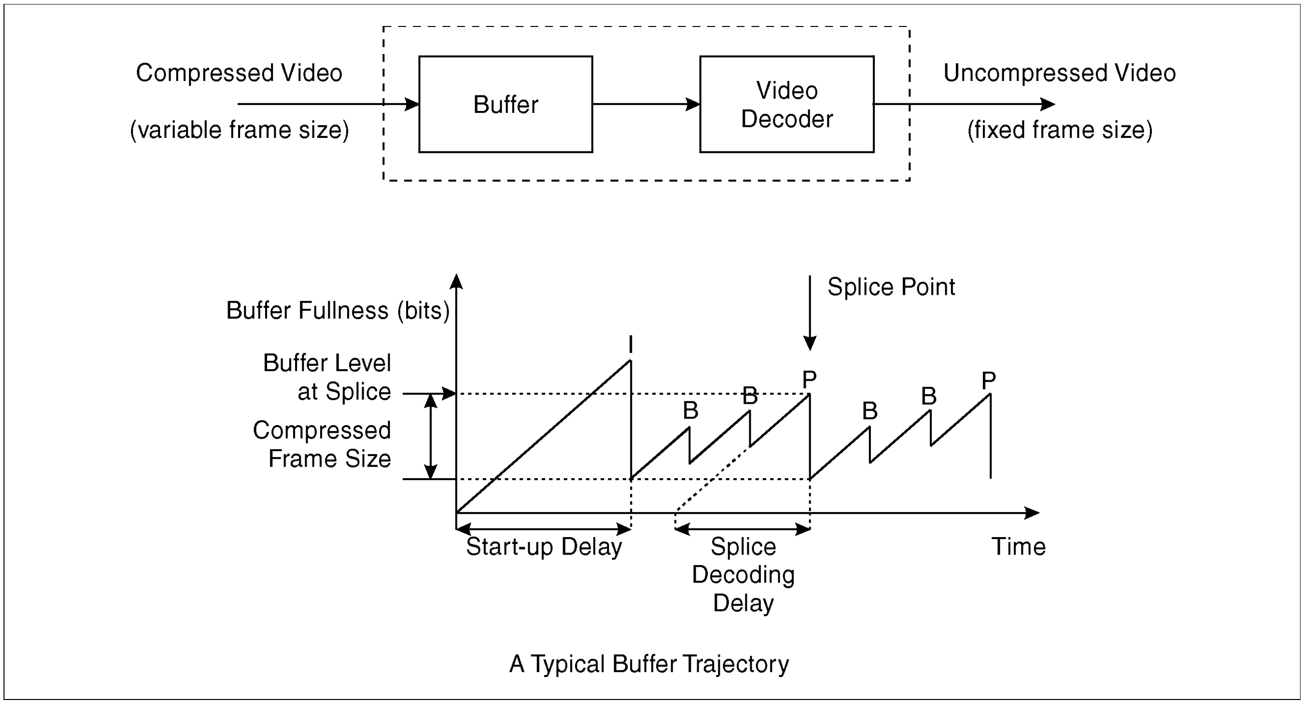


Figure 3. Simplified Representation of a Video Decoder

Seamless splicing requires that the encoder match the delay at the splicing point. Non-seamless splicing shifts the constraint from the encoder to the splicing device. The splicing device is responsible for matching the delay of the old and new streams as closely as possible without causing the buffer to overflow. Alternatively, controlled underflow of the decoder buffer, while displaying a repeated frame or several black frames before inserting the new stream, appears nearly seamless and prevents buffer overflow. Either splicing method may cause the audio buffer to overflow, resulting in an audio artifact. In most cases, this may be avoided by muting for a few audio frames.

As a result, splicing in the compressed domain is somewhat complex and needs additional care. Splicing points have to be identified in the bitstream at the time of encoding prior to the actual splicing operation. Without easily identifiable splice points, splicing would be more complex to implement. Some of the requirements include simplicity, low cost, and, above all, preserving current functionalities (such as cue-tone signaling) as much as possible. Identification of splice points is supported in the MPEG-2 standard, but operation of the splicing mechanism is not specified. Splicing can be implemented in a number of ways, deterring interoperable splicing devices. One of the important goals of standardization is to ensure that equipment is interoperable. This will allow cable operators to buy equipment from a competitive marketplace and not be tied to a single vendor. A program and commercial insertion system consists of three basic subsystems: insertion, billing, and trafficking. To standardize all these subsystems in the cable headend, the SCTE Digital Video Subcommittee (DVS) has created the Digital Program Insertion (DPI) *ad hoc* group.

The PT20.02 group of the Society of Motion Picture and Television Engineers (SMPTE) created an *ad hoc* group to devise a splicing standard. Similar efforts have also been undertaken by the Society of Cable Telecommunication Engineers (SCTE), the European Broadcast Union (EBU), and the Association of Radio Industries and Businesses of Japan (ARIB).

CableLabs has sponsored the Advertising in TV (Ad TV) group among its member cable companies to provide requirements and operational information to equipment manufacturers. CableLabs has contributed to the SCTE DVS group regarding some functionalities, features, and parameters desired for the next generation digital program insertion system. This contribution summarizes some of the comments received by CableLabs in response to the *Request for Information on Digital Program Insertion Systems* issued in April 1997. The comments address the following specific areas:

Bitrate

Many responding companies have voiced the need for mandatory support of constant bitrate insertion. Of the responses that also supported variable bitrate, most stated that the amount of bitrate variation required upper and lower bounds. Some suggested that an industry bitrate guideline be reflected in an insertion standard. Other responses suggested that the insertion system should have the capability to reduce the inserted program bitrate through recoding. In this case, some loss of video quality may result.

Audio

Mandatory support for Dolby AC-3 was suggested by many companies. There was also a request that the sampling rate be constant between old content and new. If audio and video splice points do not coincide, the audio splice point should be selected after the video splice point. Momentary muting of audio during the splice may be an attractive method to conceal audible artifacts occurring near the splice point. The audio of the new stream should be the first byte of an audio frame.

Addressability

Although many vendors stated their intent to support geographic and demographic addressability, they were concerned about substantial complexity and increased cost associated with adding this functionality. Low revenue-generating channels could be turned into more profit-generating channels by offering targeted advertising.

Emergency Messages

Mandatory support for emergency messages was confirmed by many companies. Consideration of FCC requirements about emergency messages should be taken into account where applicable.

Timing Reference

Some responses proposed the use of a global timing reference for scheduling the splicing operation. This constraint would require the cooperation of all program providers.

Interoperability

Most of the companies acknowledged the need for standardization to facilitate interoperability, one of the important goals of the cable industry. The responses generally included detailed physical interfaces that utilized formal and *de facto* standards. Consideration of interfaces to proprietary analog insertion systems was suggested. Such interfaces could allow replacement of “core” splicing technology and associated control within existing analog insertion systems. There were requests to integrate existing standards for various interfaces to take advantage of economies of scale.

To facilitate interoperability at the subsystem level, CableLabs proposed a Logical Topology for the next generation DPI system in their Request For Information (RFI) of April 1997. This architecture, shown in Figure 4, has been modified and submitted to both the SCTE DVS and SMPTE groups.

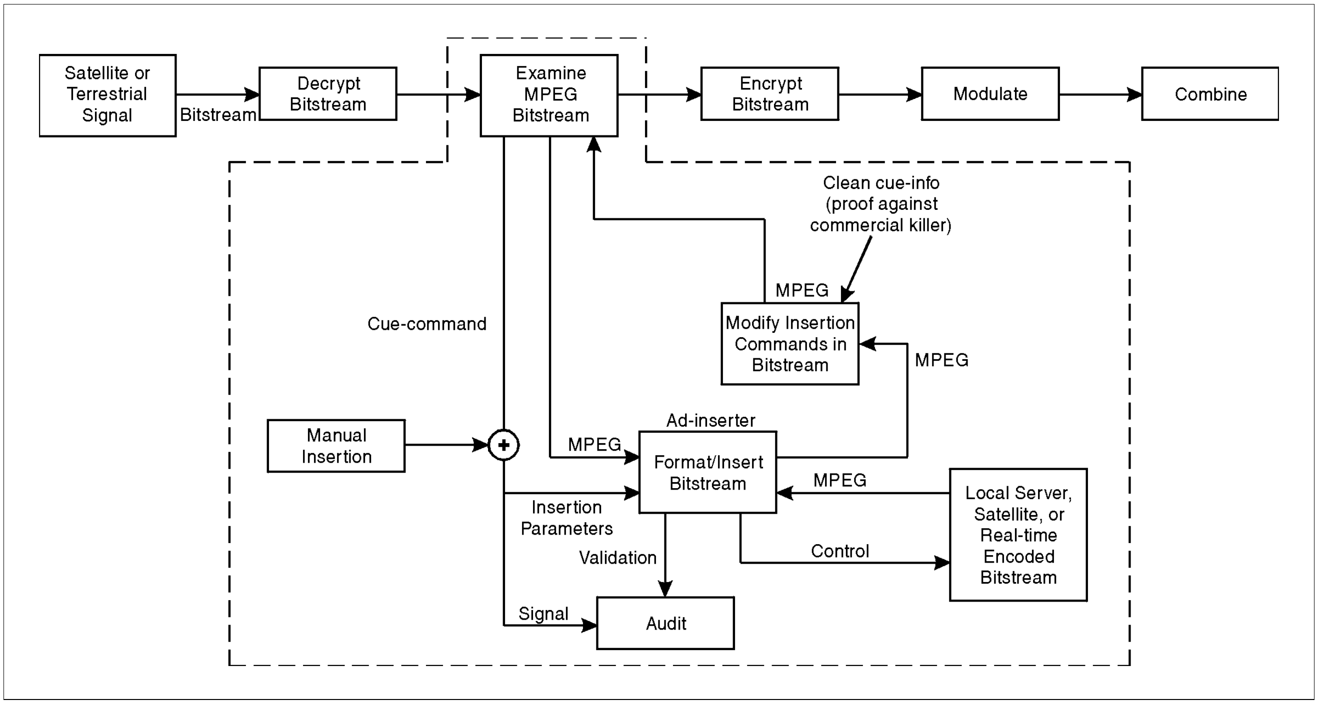


Figure 4. Typical Architecture of a Digital Program Insertion System

INSERTION SOLUTIONS

SMPTE also collected requirements from other application sectors, primarily in the television production environment. Two of these sectors, studios and post-production houses, have been influential in designing the new SMPTE standard. Production facilities prefer seamless splicing, whereas cable headend and broadcast affiliates prefer a less complex (non-seamless or near seamless) inexpensive design.

The SMPTE splicing method switches between two video streams in the compressed domain. To minimize complexity while reducing stream overhead and cost, SMPTE made major assumptions which constrain the compressed bitstreams for splicing as follows:

Splicing will be performed between unencrypted streams (implying splicing will be performed in the clear).

1. The streams must have equal bitrates.

2. The streams must have identical raster formats (*i.e.*, an equal number of pixels per line).

3. All legacy decoders (decoders already deployed) must be transparent to splicing for commercial insertion and other purposes.

4. Splicing will leave no signature in the post-spliced stream to alert potential commercial killers.

5. A cue-tone signaling mechanism (analogous to existing analog methods) has been included to indicate a specific splicing point, start and duration of a single avail, schedule of avails over a period of time, and other information within the MPEG-2 Transport Stream for one or more programs.

The SMPTE splicing method switches between two video streams in the compressed domain. As a result, some of the disadvantages of the SMPTE splicing standard are:

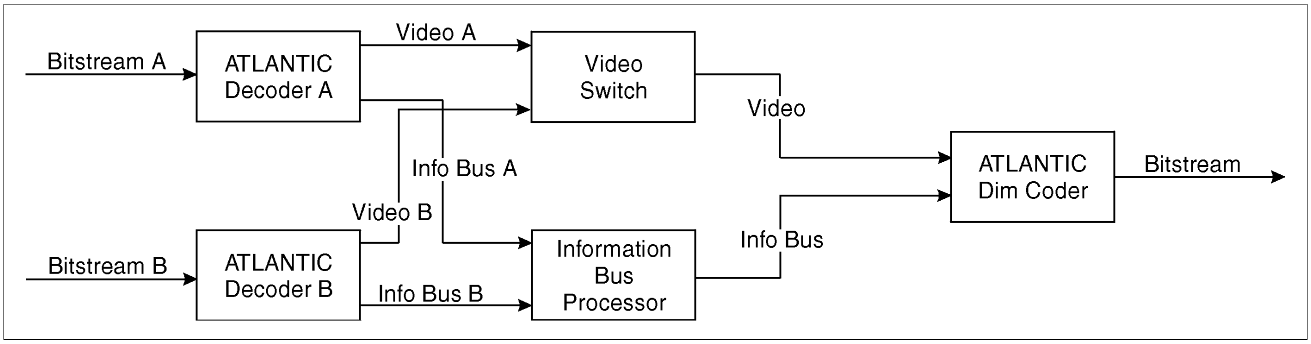


Figure 5. Simplified Diagram of an ATLANTIC Video Switch

1. Identification of splice points has to be inserted at the encoder, including additional buffer management, with a possible concurrent loss in compression efficiency.

2. Splicing cannot be achieved at every video frame boundary—a feature inherent in the analog method—because of the dependency of the compressed frames on each other.

3. Encoding of a splice point requires an increased number of bits. Increasing the number of splice points in a bitstream increases overhead and effectively increases the bitrate.

Similarly, the European Broadcast Union (EBU) has been working on projects involving the interface and manipulation of MPEG-2 compressed bitstreams. One of their efforts addresses compressed video bitstream switching and editing under the ATLANTIC project. As previously mentioned, switching MPEG-2 bitstreams is complex, and splicing of compressed bitstreams does not offer the flexibility required for all switching applications. The ATLANTIC project proposes a solution based on the decoding of the MPEG bitstreams with all mixing, switching, or other production operations carried out on uncompressed video streams prior to re-encoding into a new compressed bitstream.

Naive decoding of the compressed bitstream to uncompressed video, and then re-encoding (recoding) back, degrades video quality due to cascaded lossy compression. Using a lower compression factor on the material to be switched and re-encoded can reduce these artifacts. A disadvantage of this approach is that the initial compression requires a high bitrate and, therefore, would not be appropriate for efficient broadcast delivery.

The ATLANTIC approach uses a partial decoding technique. To minimize degradation in the output bitstream, the coding decisions (temporal predictions encoded as motion vectors, quantization, and information in the picture header), used in compressing the sequence of frames, are directly used or passed through, when possible, in the recoding process. As shown in Figure 5, the uncompressed decoded video streams are routed to video switches, while saved coding decisions are sent to the information bus processor. A few frames around splice points may change temporal prediction modes, and motion vectors may need to be recalculated in these instances only for a few frames. However, a vast majority of the frames may remain unchanged and the recoder can use the motion vectors available from the info_bus processor. Therefore, recoding is much less complex than full encoding, and no noticeable degradations in quality are observed even after several decodings and recodings at the same bitrate.

The ATLANTIC project proposes a method of mixing the information in the info_bus (the so-called MOLE™) with the decoded uncompressed digital D1 signal to enable unconstrained post-production effects and editing. Therefore, once the bitstream is decoded, the initial encoding decisions are embedded in the uncompressed stream. This information will be reinserted during re-encoding. The same stream could be decoded and re-encoded a number of times using a simpler and less complex encoder without noticeable degradation in picture quality. The EBU has submitted a proposal to SMPTE for standardizing MOLE™ technology so that interfaces for standard production equipment, enabling transparent production and editing of compressed video bitstreams, may be standardized and manufactured by various vendors. The real advantage in this method is its flexibility and lack of constraints. No splice points have to be created prior to splicing, and splicing can be performed at any frame boundary. In fact, any type of production effect can be achieved since editing is accomplished in the uncompressed domain.

The authors believe that this technology is attractive and has many potential applications. However, recoding is substantially more complex than operating solely on compressed bitstreams. Therefore, implementing this technology with the components available today may be more expensive. An implementation using customized VLSI processors (currently not available), such as media processors, may be more economical and, therefore, more widely available.

CONCLUSION

As mentioned earlier, splicing and switching standardization efforts are in progress. SMPTE and EBU have standardized the basic splicing of MPEG-2 bitstreams in two different approaches with different constraints, flexibility, and complexity trade-offs. Efforts are also underway to standardize the use of basic splicing in digital program/commercial insertion and its other integral aspects, such as trafficking and billing. The SCTE DVS DPI *ad hoc* group and the AdTV (Advertising in TV) group are working on these two areas. Still, a few issues remain to be addressed by the cable industry, such as changing the original compressed bitrate of the ad spot, ad insertion into a statistically multiplexed variable bitrate stream, and use of constrained bitstreams per the SMPTE method or unconstrained streams per the ATLANTIC method.

Today's average channel bitrate is 3 to 4 Mbits/sec (sports up to 6 Mbits/sec). For quality reasons, the bitrate of the compressed commercials may be higher than the average channel bitrate of the program targeted for ad insertion. The question remains—how to insert a higher bitrate ad in a lower bitrate channel? For efficient use of the spectrum, it is better to keep the channel multiplex unaffected. In this case, a trade-off may be required to code (or recode if already compressed at higher bitrate) the ad at the same channel bitrate with a corresponding reduction in picture quality.

In statistical multiplexing, the channel bitrates specified are the average rate and the maximum and minimum limits of the bitrate variation. If an ad is encoded at a constant bitrate, the variable rate of the statistically multiplexed channel targeted for insertion has to be maintained at a constant matching rate during insertion. This may affect the multiplex efficiency and, consequently, the video quality of the other channels.

For basic splicing, the SMPTE method is simpler than the ATLANTIC method. But the SMPTE method requires several bitstream constraints, and splicing cannot be performed at all frame boundaries. The ATLANTIC method is more flexible, does not have any bitstream restrictions on frame boundaries, and bitrate reduction can be accommodated. However, it is more complex than the SMPTE method. Advanced programmable processors may make this implementation more cost-effective by sharing economies of scale with other applications.

The cue-command mechanism provided in the SMPTE splicing standard preserves and extends the cue-tone features available today. Designing a device to detect cue-commands in a transport multiplex should be straightforward.

In summary, several operational issues need to be resolved that impose constraints on the delivery of compressed programming targeted for the insertion of local advertising. Relaxation of some of these constraints requires a more complex solution. Standardization is needed for interoperable digital program insertion equipment from multiple vendors. If the cable industry can achieve consensus on these issues, such standardization will facilitate the timely introduction of such new digital commercial insertion systems.

ACKNOWLEDGMENTS

The authors would like to thank Don Morgan and Carol Derr of TCI, Joe Davis of Digital Video, Reem Safadi of GI, and Steve Riedl of Sky Connect for many useful discussions and information, as well as their ongoing participation in the various digital program insertion technology development and standardization activities.

REFERENCES

1. *Cable Advertising*, white paper, March 1997, Cable Television Laboratories, Louisville, CO.

2. *Cable Ad Sales, Business Goals, Objectives, Strategy and Technical Vision*, November 1992, Cable Television Laboratories, Louisville, CO.

3. *Digital Program Insertion, Request for Information*, April 28, 1997, Cable Television Laboratories, Louisville, CO.

4. *Splice Points for MPEG-2 Transport Streams*, SMPTE Standard, December 1997.

5. *Seamless Concatenation—A 21st Century Dream*, M.J. Knee, Snell Wilcox Limited, U.K., N.D. Wells, BBC Research and Development, U.K., September 1997.

HDTV Deployment: A funny thing happened on the way to the decoder interface....

Robert M. Zitter
Senior Vice President, Technology Operations
Home Box Office
New York, NY
robert.zitter@hbo.com

Michael Adams
Senior Project Engineer, Advanced Engineering
Co-Chair, Digital Standards Working Group, JEC
Time Warner Cable
Denver, CO
michael.adams@twcable.com

Abstract

The decoder interface has, for quite some time, been viewed as an important technology for cable -- first for consumer-friendly deployment; then by force of law; now a critical requirement for the deployment of High Definition Television (HDTV).

This paper describes how HDTV (and SDTV) digital transmission needs the decoder interface -- and what obstacles lie ahead. In the focus of cable technology over the years, the "last mile" has always been a critical hurdle to overcome. In the cable deployment of HDTV to digital TV's the critical technology hurdle has become the "last six-inches."

The specifics of digital interfaces such as 1394 "FireWire" are discussed as are business requirements such as digital copy protection.

THE BEGINNING

Program streams, in any form, start from an origination facility. The complexity of a contemporary digital origination facility is beyond the scope of this paper -- but it forms a critical foundation for the transition to a digital end-to-end delivery system for cable programming. Purely creative reasons have led to a fully-digital origination facility -- the complex graphics and promotional elements in most program networks could not be done without sophisticated digital environments.

Similarly, economics have drawn program networks toward a purely-digital origination model. State-of-the art, reliable, automated, and high-quality tape and server formats are all digital. In fact, improvements in picture quality (e.g. component digital) have also yielded benefits in the "Mb/s" -- consumed bandwidth for a given picture quality.

When HBO embarked on a transition to full end to end-digital transmission in 1994, all of these benefits were capitalized on in a state-of-the-art digital facility completed in 1997. At that time HDTV was not part of the mid-term plan.

However, as plans were made for cable-industry digital infrastructure, it was evident that cable had the opportunity to leapfrog terrestrial broadcasters -- and attain parity with upstart DBS providers who had barely begun their business plans. Certainly, HDTV -- then a gleam on the horizon -- would fall into place behind a solid end-to-end digital infrastructure. Operators, such as Time Warner Cable were well into their upgrades to 750 MHz HFC – Time Warner Cable’s upgrade is currently 50% complete.

THE PLAN

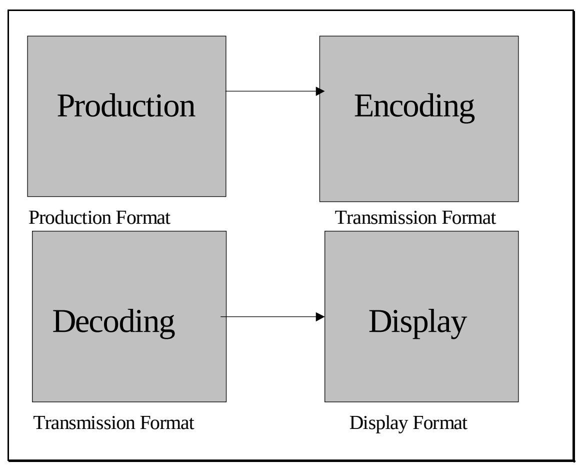
As cable and terrestrial HDTV plans became more of a business needing study and analysis, it was clear that the complex infrastructure for digital cable was falling nicely into place. The Time Warner Orlando Full Service Network had laid groundwork for service offerings and proved that computer-based cable equipment could leverage the ever-plummeting computer cost curves.

But the end-to-end infrastructure was originally viewed as a closed-system architecture. Then-deploying MPEG-2 equipment such as DVD was enjoying a ground swell of adoption -- and much of the cable infrastructure was not compatible. The major vendors and users deployed an MPEG-2 system layer -- which turned out to be a wise decision.

In short, an end-to-end infrastructure is now in place so that cable can leverage the already-deployed digital architecture to deploy HDTV in a timely manner.

HARD-FOUGHT SUCCESSES

The design and deployment of digital infrastructure should not be hidden from view as is the foundation of a major skyscraper. The cable digital infrastructure deployed today took tens-of-millions of dollars and years of research and product development. But that hard work will make the deployment much easier than that of terrestrial broadcasters -- who have an intrinsically analog plant, a transmitter and tower on the "wrong frequency" and origination facilities based on analog designs of a scanning format that was not to be used in the future. Digital HD and SDTV will be deployed by cable using the groundwork already deployed.



1) Origination. Digital infrastructure based on MPEG-2 is the core foundation for DTV. Program networks that are compressed for delivery to set tops can easily add HDTV-compatible compression cards to their existing encoding systems. Put simply, in an existing transport multiplex the compression system merely has a

channel with a greater hunger for bandwidth.

2) Format: The ATSC digital standard for terrestrial broadcasting has been adopted by the FCC and supports no fewer than 18 different image formats. While each of these formats have requirements for source material and different difficulties for processing in the display device -- they all are transported by the same digital highway. While there has been much public debate about the "best and only" digital image format for DTV, the end users are mostly indifferent to the actual format (since they are all supported in the ATSC chip set), the only substantive difference is a format's particular "hunger for bits" which varies by program content and picture quality objectives.

Bandwidth

Scanning Format	**Frame Rate**	**Bandwidth (Estimated)**
1080I	30	13+ Mb/s
720P	60	12 Mb/s
480P	60	6+ Mb/s
480I	30	3-4 Mb/s

3) Transmission and Encryption: Broadcasters continue to ponder how conditional access can be added to their proposed digital transport systems -- and, must also deal with the political issues associated with their offerings of non-free services. The conditional access and digital transport systems currently being deployed by cable are compatible with DTV transmission to set tops. Although some modifications may be necessary to support the recently-adopted broadcast system information and navigation system, it is generally expected that the headend Integrated Receiver Transcoders (IRT's) of today will support satellite-delivered DTV transport without replacement. Additional headend equipment will be required to process the terrestrial VSB signal and convert it to QAM.

4) Plant: 64 QAM is being deployed today. Some systems have deployed 256 QAM and most HFC rebuilds can easily support this very-efficient modulation scheme. Cable's ability to support two terrestrial 8 VSB signals into one cable QAM-modulated 6 MHz channel is a significant benefit during the "transition years" from analog to digital end-users. While some operators may practically have to transport 8 VSB, the Program System Information Protocol (PSIP) used by broadcasters is very complex and may raise many transmission issues for cable operators who heterodyne VSB and / or convert to 64- or 256-QAM

Modulation

Modulation	Capacity
8-VSB	19.3 Mb/s
64-QAM	26.97 Mb/s
256-QAM	38.8 Mb/s

5) Consumer Interface: In the analog world, the set top provided the "common denominator" of VHF-3 to the television set. All conditional access, navigation, user interface (e.g. character display or overlay) was provided upstream. As mainstream digital

MPEG-2 decoders dropped in price ("it's only silicon"), the affordable set top transitioned from advanced analog to sophisticated digital set top. However, this affordability does not scale to HDTV. Initial low-quantity chip sets are, to say the least, higher cost. HDTV processing -- RAM and resolution -- is many multiples of SDTV NTSC. It was economically impossible to layer HDTV processing on top of the costs of advanced digital interactive set tops. A more necessary (and desirable) approach was to separate the conditional access -- leaving it in the set top -- and passing a decrypted stream to the already-purchased DTV receiver for processing and display. Such is the way of NRSS and OpenCable.

Consumer Interface (Issues)

However, the rosy story of compatible, deployed infrastructure temporarily came to a screeching halt at this point. Unfortunately, several technical and political issues clouded the consumer interface.

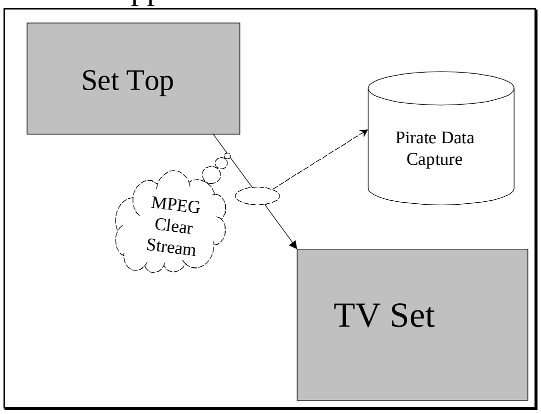
First, the Consumer Electronics manufacturers -- sensing zeal on the part of broadcasters (real or not) designed the first-generation receivers without an interface to cable. VSB was left as the common denominator interface for digital cable -- and cable had clearly stated that QAM was the key enabling technology for the digital cable system of the future. No digital baseband interface was planned. Consumer Electronics manufacturers wish to build only to adopted standards. The ATSC was the organization of record for DTV – and they had adopted VSB as a transmission standard. Cable's selection of QAM was not viewed as mandatory by the CE industry. This is was a major "disconnect" by the CE industry.

Second, the cable industry did not send a clear message that a baseband interface would be acceptable. Previous decoder interfaces and even NRSS have not been deployed on a widespread basis. Practical cable business issues such as user interface (ranging from I/R pass-thru to NRSS-B extensions) have not been understood by the CE industry. Cable use of a baseband interface would likely represent a "short-term" solution with the implementation of NRSS representing "mid-term" and the "Open Cable" initiative representing the "long-term" solution. The cable industry was focusing on configurations to be implemented several years after the HDTV introduction.

Lastly, copyright owners (e.g. movie studios) have seen improvements in digital consumer technology such as powerful computers, digital recording, video compression and Internet bandwidth. Such advances are enabling technologies for very high quality copyright piracy -- if not on the wholesale scale of Chinese disc factories certainly one clip at a time. Copyright owners, computer system developers and DVD manufacturers worked out a series of copy protection guidelines for DVD equipment -- and a cable baseband digital interface was fertile ground for PPV and other premium television content to be potentially captured and distributed. "Bypass" distribution of pirated content is not healthy for the content providers and certainly not for cable.

Copy protection proposed includes control over the ability to not record based on the following business practices:

- PPV = No Copy Permitted
- Premium (Pay) = One Copy (Time Shift)
- Ad Supported = Unrestricted



Consumer Interface (Solutions Underway)

Each of the issues raised has a solution in process. As of this writing, the solutions are close enough at hand to enable (with some assurance) that cable can deliver DTV on a competitive par with terrestrial broadcasters -- and, we hope, with recently-announced satellite broadcasts of HDTV.

First, the CE industry took a look at the demographics of American television purchasers -- and the early adopters who would likely purchase DTV receivers were already cable subscribers. With virtually no installed base of unproven "digital rabbit-ear antennas" it is highly improbable that a "non-cable-compatible" DTV set would enjoy much retail activity.

Second, the cable industry looked at the economics and politics of OpenCable -- and now, more than ever before, it does appear that both retail-sale of "set top devices" could practically occur and it also appears that NRSS-B could provide the necessary user interface (UI) functionality for a consumer to continue their transactional business relationship with their cable provider whilst fully-utilizing their high-priced, high-powered DTV receiver.

Lastly, the synergy of these solutions has led to some rapid decisions for the form and function of the interface itself and the copy protection implementation that is a practical requirement for a contemporary digital consumer product:

a) The 1394 "FireWire" architecture will be used for consumer / cable interconnection. Extensions to 1394 for high-speed operation as well as interfaces to other CE devices are underway;

b) A copy protection system will be provided for the CE and cable "nodes" in the home electronics infrastructure. In the copy protection system, the "source" of video streams in the home (e.g. DVD, set top, etc.) serves as the host for a LAN-based encryption and authentication system. The system proposed is that proposed jointly by “5-companies”—Hitachi, Intel, Matsushita, Sony and Toshiba -- which provides for a public- and private-key authentication and encryption system.

c) Additional aspects of the interface such as the control protocol and extensions to support cable's necessary transactional user interface are under discussion. While also on the critical path, the implementation of a) and b) had system design impacts much more critical than user interface adaptation. As a result some first-generation HD receivers will be incompatible with cable at introduction, others will only be so in a subsequent release. This was done at the last minute of interface design of some manufacturers receivers some designs had already been frozen.

CONCLUSION

In conclusion, the foundation laid by the deployment of digital infrastructure by cable has resulted in cable's ability to assume a leadership role in the deployment of valuable high-profile digital programming to consumers. HBO and our cable affiliates look forward to providing "cable's best" programming in the best -- and highest quality HDTV format.

High-Speed Data Multiplexing over HFC Networks

Mark Laubach
Com21, Inc.

Abstract

The deployment of high-speed data services over all-coax and Hybrid Fiber-Coax (HFC) networks is forcing cable operators to quickly develop an understanding of high-speed data multiplexing. Traditional Internet engineering focuses on local area network design, backend network technology and topology, and connection to wide-area or regional networks. In addition, the emergence of different varieties of Cable Modem Termination Systems at the head-end has met with some interesting challenges related to incremental growth, scalability, and matching data channels to RF trunks in a scalable and cost effective manner. This paper overview engineering design issues associated with aspects of high-speed data multiplexing and the incremental re-ordering of the networks to effectively meet growth starting from sparse subscriber deployment and continuing to large take rates.

Getting Started and Growing Larger

A cable operator or Internet Service Provider (ISP) working with a cable operator has a variety of ways to get started with deploying Internet services. The basic starter kit has simply been a router with one Ethernet port and a T1 connection (1.544 Mbps) to a larger Internet Access Provider (IAP) or ISP, a cable modem head-end system, an Ethernet LAN. This back-end LAN connects the cable modem equipment with the router, and a truck full of cable modems. This starter set has elements common to all Internet access scenarios since the early 80's: a connection to the bigger world, a back-end network, and a baseband or broadband local area network connecting the users to the backend network.

What is not the same as in the 80's (even with broadband modems) is that there has been a technology revolution (several times) and an explosion in the number of users who can potentially take advantage of new access to incredible amounts of bandwidth. Since these users are located in homes, the incremental growth from starter residential access network system to largely deployed systems is both staggering and highly variable on a market or fiber node basis. Fortunately, cable modem equipment has followed an anticipated advancement path however, their flexibility with respect to data multiplexing has in some cases outstripped the CATV plant.

Mentioned briefly, but not the subject of this paper are the other new services, abilities, and challenges that a cable operator and ISP have jointly discovered:

- Management and deployment of Dynamic Host Configuration Protocol (DHCP) servers
- Management of large blocks of IP addresses and changing these over time, web proxy and caching servers and where to place them
- When and how provide news, mail, and FTP servers
- How to manage the entire data network effectively
- What type and size routers needed
- Where and when to use ATM
- What to do about simultaneous data and voice transport?
- When to where to use multicast services
- What standards do I need to be concerned with?.

Generally, cable operators and their ISPs, seem to fall into one of three business model mentalities:

- "flat rate pricing is just fine",
- "I want multi-tier pricing", or
- "I want multi-tier pricing and the ability to sell bandwidth pipes over the RF to other ISPs or corporate concerns"

The requirements for high-speed data multiplexing are different for each model.

FLAT RATE PRICING IS JUST FINE

This business model begins with the Internet "starter kit" model. I also call this model Home Box Internet. The topology of the backend and regional networks form a tree. A cable head-end is connected to other head-ends via a large backend networks or each head-end is connected to a regional network center, if available. There may or may not be a nationwide backbone.. Large networks such as @Home do have their own backbones.

A data multiplexing point (or concentration point) exists where the "child" connects to the "parent" in the tree; i.e. link from head-end to regional hub, link from regional to nationwide access, etc. When the link costs money to provide, then the goal is: spend as little money as possible on the cost of the link but maximize the number of users accessing the link through the child. The link is sized to meet the needs of the statistical peak load of the users and not on the sum of bandwidth available to each individual users. This is called statistical multiplexing. Practically, this means putting caching servers and other servers on the child side of the link. This reduces the overall number of individual "connections" that need to flow through the link. So long as the price of the caching servers is less than the cost of the next size link, this model holds. Sizing the link is not rocket science. But we do have rocket science modeling tools to predict the size of links. If the link is too small, users will complain. If the link is too large, then the Chief Financial Operator (CFO) will complain. The trick is sizing within the subscriber/CFO tolerance envelope.

Within a head-end, the choice of local area network equipment is going to be chiefly decided by the Internet model being deployed. If the downstream world looks like one big Ethernet, than just a single high-speed Ethernet port is needed between the Cable Modem Terminal System (CMTS) and the router. However, this model varies greatly based upon the desires and design of each cable operator and ISP. For the basic system, one Ethernet is needed and one IP subnet is needed. As the system grows larger, than the number of Ethernet ports and IP subnets will vary. This effects the type and configuration of the cable modem equipment and will require changes while the size of the subscriber base grows. For North America, plan on a CMTS with a single downstream RF channel moving about 25 Mbps full-duplex when fully loaded. If using Ethernet, than getting started with 10BaseT works, but needs to go to 100 BaseT very soon. If using ATM as the backend connection from the CMTS, then the size is either OC3 (155 Mbps) or DS3 (45 Mbps) meaning the connection doesn't need to change over time between the CMTS and the router - there is just unused bandwidth. Since that link is free from monthly cost, this is ok. However, if the ATM link (from a router or from the CMTS) is being back hauled over a SONET network, than there many be monthly costs and the data multiplexing ability of the router/CMTS must be capable of filling the OC3 link. This means supporting multiple downstream and upstream data channels over the RF via the single link.

I WANT MULTI-TIER PRICING

In the context of this paper, multi-tier means that the cable operator has the ability to give different subscribers different allocations of bandwidth based upon how much monthly service fees they are going to pay per month. This capability provides the operator with a much better revenue generation model than single flat tier pricing, as there are always customers willing to pay more for more bandwidth or reduced delay.

From a data multiplexing point of view, the model for supporting this type of billing and tiered ability is very similar to the flat rate model. It is just that the size of the link from the child to the parent has to be increased to accommodate that bandwidth that has been sold to the multiple tiers. The assumption is that the users at the same pricing level (billing class) can be statistically multiplexed together, so it is not a sum of all bandwidth per user game. The issue at hand is that users are paying more money for better perceived performance (lower delay) and there needs to be capacity allocated over the links between the head-end and the Internet in excess above a flat price scenario. Also, the same rule applies to stay within the subscriber/CFO tolerance envelope. The added dimension is that there are now several classes of subscribers (related to service tiers) and expectations must be set per tier.

I Want The Ability To Sell Bandwidth Pipes Over the RF

This model is a bit different than the previous two types. While the single tier or multi-tier scenarios are part of the data multiplexing puzzle, there are added requirements:

- cable modem equipment must be able to “carve out” portions of data and/or RF bandwidth that can be allocated to a specific user, group of users, or virtual LAN
- there must be added routing and data segregation in the back-end networks
- companies or other ISPs may be connecting at different places in the network hierarchy

This type of business model is very close if not the same used by Competitive Access Providers (CAP). The cable operator may actually not be in the direct business of offering ISP services for subscribers, rather they are providing access to one or more ISPs, which in turn provide Internet, access to their subscribers. There may be several ISPs in operation over the same RF channels in this model. In addition, companies may buy bulk bandwidth for their own backend network needs or for supporting a closed group of employees for telecommuting. Every one is sharing the same RF channels, the cable modem equipment is providing segregated multiplexing to keep the traffic separate from one another: in a routing sense, a security/privacy sense, and from a bandwidth management sense.

The motivation for this type of business model is driven by the recognition that cable operators can compete with local telephone companies for providing bulk bandwidth connections. It is expected that the cost of service over cable is and will remain a better value deal than a corresponding telephone service.

For this business model cable modems must be able to be placed into closed user groups. Closed user groups need to be given different amounts of bandwidth over the RF, the data traffic from closed user groups must be able to be kept separately partitioned from the traffic from other closed user groups for privacy, management, and billing/accountability reasons. Cable modem equipment that supports this type of operation can be configured to support the single tier and multi-tier types. Cable modem equipment designed to support either of the other types cannot support this CAP model.

In addition to the above, connections to ISP or companies can occur as various places with the hierarchy of the cable operator’s network. There are many varieties of scenarios and too numerous to detail however, each one impacts the amount of segregated bandwidth that needs to traverse a link.

Also, the same rule applies to stay within the subscriber/CFO tolerance envelope. The added dimension is that there are now several classes of subscribers (related to service tiers) including ISPs or companies who have purchased bulk access.

Oop's, I've Added Another Different Type of Digital Service

This small section just mentions that if the cable operator wants to install an additional type of service (e.g. a digital toll quality voice system or digital video transport) down the road, then they might want to give consideration when they are building their networks for high-speed data for Internet. For example, if the cable operator used a backend ATM network, they would find that it costs a little more, but multiplexes a wide variety of services and scales to meet a variety of configurations and capacities for high-speed deployments. If they deploy an additional service type that could not be run over Internet service (e.g. true toll quality voice or MPEG2 video streams), the existing switching equipment should be able to handle the capacity, but they may need to buy additional port cards or re-size a link. The same network can be used for multiple services. If the cable operator has deployed Ethernet based services and IP links between routers for Internet use only, they are likely faced with having to replace the IP routing equipment with high capacity routing equipment. They likely must also beef up the interconnection links or they are going to have to deploy a second network just for the new service. Both options are costly when compared to the former ATM model.

The Impending Data Ports Versus Cable Plant Ports Mismatch Problem

Initial deployment of high speed data services on all coax plants can typically be accomplished using one CMTS for the entire plant. Existing CMTS equipment today come in one of two scalability architectures: "fixed" scale configuration with one downstream port with only one upstream port, or "flexible" scale configuration with one or more downstream ports with one or more upstream ports. Incremental growth to meet new subscriber demand or capacity is different for fixed versus flexible architecture.

For incremental growth with a fixed scale CMTS, when more high-speed data capacity is needed in either the downstream or upstream, a new CMTS is required; i.e. the cable operator needs to purchase another CMTS box.

For incremental growth with a flexible scale CMTS, when more upstream capacity is needed, the cable operator can add an additional upstream channel demodulator (demod) card to the CMTS in addition to current channels. When more downstream capacity is needed, the cable operator can add an additional downstream channel or purchase a new CMTS box.

There is a large difference between fixed and flexible scale CMTS systems. With a fixed configuration, the operator must recombine upstream trunks into as few as many ports as possible to avoid having to purchase and abundance of fixed scale CMTS boxes. With a flexible configuration CMTS, each upstream channel may be connected to a different upstream trunk, eliminating any need for the recombination of trunks. This has two benefits: firstly, the cost of an additional upstream channel is generally less than the cost of a fixed configuration CMTS box, and secondly, the noise floor is reduced at the upstream port. The operator is free to distribute to trunks and combine upstream trunks with a flexible scale system.

At some point in the growth of service deployment, more downstream capacity will be required to meet subscriber demand. In a fixed scale CMTS, a new downstream channel is required for every upstream channel added and vice versa regardless of whether the downstream or upstream channel capacity has been filled by demand. In a flexible scale CMTS, the relationship of the downstream channels to the upstream channels within a single CMTS box are separately scalable, allowing the addition of downstream or upstream channels to follow subscriber demand. In addition, this flexible scale-ability allows for capital expenditures to more closely match revenue growth, and also allows for noise impairment to be better controlled by use of more upstream ports per downstream channel. This latter point is very important in that the cable operator has much more flexibility in managing the recombination of upstream trunks and subsequent noise funneling issues.

When the downstream channel capacity has been exceeded and not enough RF spectrum is available in the cable plant, the operator has the option of upgrading the plant to HFC. The upgrade to HFC will produce more downstream trunks and more upstream trunks. If previous CMTS installments matched capacity and revenue growth, there is likelihood that the existing installed CMTS equipment will match the newly available trunks and subsequent ports. Note that in this incremental HFC upgrade scenario, the cable operator has the option to do upgrades only where high-speed data capacity is needed, i.e., where the active subscribers and revenue is coming from. Upgrading the entire plant to HFC is not required.

Recombining return trunks at greater than four to one (4:1) causes noise funneling contribution and reduces the Carrier-to-Noise Ratio (CNR) below a 25 dB margin at the upstream return port. Several cable operators use this ratio. Converting the upstream lasers from FP to Direct Feedback (DFB) lasers allows the upstream return trunks to be recombined at a ratio of up to ten to one (10:1) which is attractive. If the plant currently has FP lasers, the cost differential to go to DFB is substantial and in most cases prohibitive.

High noise floor interrupts all upstream modulation schemes in an HFC plant. The ability to recombine upstream return trunks is limited by the lowest capable interactive service; for example, impulse pay per view, interactive two-way node management protocols, etc. The recombination problem affects more than just high-speed data services for Internet.

Solutions for the initial sparse deployment scenario are few. Either buy sufficient CMTS equipment to cover the upstream return ports or look into solutions that recombine data but do not recombine noise. Look towards CMTS solutions that support a large number of upstream return ports per downstream port.

The use of a Reverse Path Multiplexer has been shown to be effective in allowing the data from a larger number of return plant ports to be recombined without recombining the noise as compared to existing passive combiner techniques. As an active device, the RPM is placed downstream from the upstream channel port card in a CMTS. The RPM receives communications from the port card in response to when packets are expected to be received on a particular trunk. This high-speed switching system is effective at allowing a 200 microsecond packet burst with a small lead time and a rapid turn off.

Summary

The multiplexing of high-speed data over CATV networks and the business models of the cable operator are closely tied together. Different business models have different requirements for growth and capacity planning. The single flat rate per month is very straightforward to implement and grow. Multi-tier models and bulk bandwidth models each require more backend multiplexing than the simple mode and have benefit of increased revenues. The bulk bandwidth model is the most comprehensive of the business models, but allows a great deal of flexibility in how data is handled in networkp. The future may bring the need for additional digital data services to be multiplexed through a head-end. It is desirable to share the same backend switching network. This is advantageous as capital is only needed for incremental capacity increase and not for building a second parallel network for the new service. Data multiplexing over the RF has challenges for initial sparse deployments that can be overcome with CMTS equipment that supports many upstream channels and reverse path multiplexing.

References

[1] S. Beauchamp, "Throughput Analysis of Com21 Service Levels and ISDN," Application Note, Number 280-0024-D1, August 4, 1997.
[2] Kathleen M. Nichols and Mark Laubach, "On Quality of Service in an ATM-based HFC Architecture," IEEE ATM Workshop 96, August 27, 1996
[3] Kathleen M. Nichols, "Improving Network Simulation with Feedback," Internal Paper, Bay Networks.
[4] Mark Laubach, "Deploying ATM Residential Broadband Networks", NCTA Cable 96 Conference, Los Angeles, CA, 30 April, 1996.
[5] Kathleen M. Nichols and Mark Laubach, "Tiers of Service for Data Access in a HFC Architecture", SCTE Emerging Technologies Conference, January 1997.
[6] Mark Laubach, "ATM over Hybrid Fiber-Coax Networks", 2nd International Workshop on Community Networking – Integrated Media to the Home, IEEE Communications Society, 20-22 June, 1995, pp 27-33.
[7] Mark Laubach, "To Foster Residential Area Broadband Internet Technology: IP Datagrams Keep Going, and Going, and Going….", ConneXions, Softbank Expos, 2 February, 1996.
[8] Mark Laubach, "High Speed Data Systems Architecture and Tools for Managing the Upstream Noise Floor in HFC Systems", IEEE/SCTE, High Integrity Hybrid Fiber-Coax Networks Workshop, 7-9 September 1997.
[9] Mark Laubach, "Serving Up Quality of Service", CED Magazine, December, 1996.
[10] Mark Laubach, "Avoiding Gridlock on the Data Infobahn: Port Mismatches Pose Challenges", CED Magazine, April 1997.

Author Information

Mark Laubach
VP of Technology, Chief Technical Officer
Com21, Inc.
750 Tasman Drive
Milpitas, CA 95035
Laubach@com21.com
(408) 953-9175 (direct)
(408) 953-9299 (FAX)

IMPROVED OPTICAL FIBER AMPLIFIER FOR 1.3 µm AM-VSB SYSTEMS.

Vincent MORIN and Edouard TAUFFLIEB

PHOTONETICS S.A., Marly-le-Roi, France

We present an innovative, compact, powerful fiber amplifier operating at 1.3 µm and dedicated to CATV transmission. Its output power is +20 dBm and its noise figure is 6 dB. Supertrunk transmission of 30 channels over 94 km is reported.

INTRODUCTION

Most CATV systems throughout the world use the hybrid fiber-coax architecture, with SubCarrier Multiplexing (SCM) in the Amplitude Modulation-Vestigial Side Band (AM-VSB) format. The fiber backbone of these networks operate at the wavelength of 1.3 µm with laser diodes using direct modulation.

An optical amplifier compatible with the demands of AM-VSB, i.e. high output power, low noise and high linearity, offers an increase of loss budget which opens new possibilities of network architectures, in terms both of geographical extent and connectivity.

We present an innovative Praseodymium Doped Fiber Amplifier, which combines optimized ZBLAN fiber with advanced splicing technology, pumped with a compact fiber laser. +20 dBm output power is obtained, and full CNR, CSO and CTB degradations are measured, for a 94 km straight-line supertrunk transmission.

DESCRIPTION OF THE PDFA

PDFA's have been studied for sometime now[1], [2] , and they use a fluoride ZBLAN fiber doped with Praseodymium. The ZBLAN host, (compared to silica), reduces non-radiative transition of the Praseodymium ion, which is a key feature for the amplification.

The amplifier uses a contra-propagating configuration, with a simple pass of the signal. The ZBLAN doped fiber has a 1000 ppm weight Praseodymium concentration, a 12 meter length, a 1.8 µm core diameter, a 0.35 N.A. and a 0.05 dB/meter scattering loss. The angle-cleaved ends of the ZBLAN fiber are butt-coupled to a high-N.A. silica fiber, which is then fusion-spliced to a low N.A. silica fiber. All fiber splices are of low reflection. The pumping source is an Ytterbium-doped cladding pumped fiber laser emitting 600 mW at 1030 nm. The very strong confinement of the ZBLAN fiber is required because of the unfavourable spectroscopy of Praseodymium, (at least compared to that of erbium), namely short metastable level lifetime and low cross-sections. The choice of the 1030 nm pumping wavelength is a compromise between the best pumping wavelength of Praseodymium (1020 nm) and the peak emission wavelength of Ytterbium (1060 nm). The detailed theoretical and experimental investigation of the structure of the amplifier has been discussed in [3]. The scheme of the PDFA is given on Figure 1.

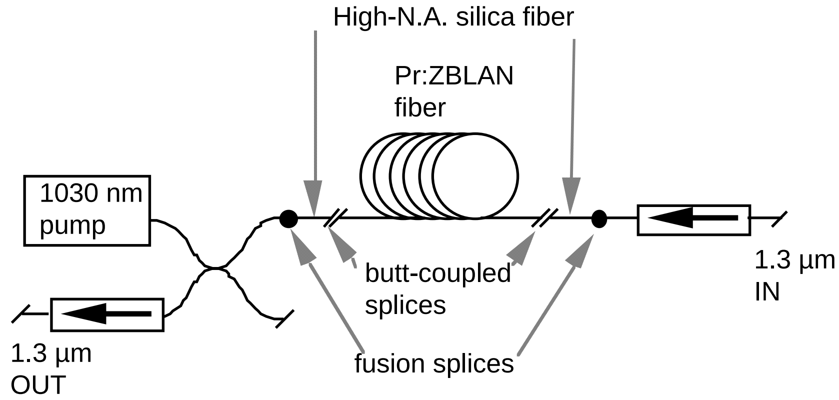


Figure 1
Scheme of the amplifier

GAIN CHARACTERISTICS

The output versus input power is shown on Figure 2, for the peak gain wavelength of 1300 nm, and input powers ranging from -10 to +10 dBm.

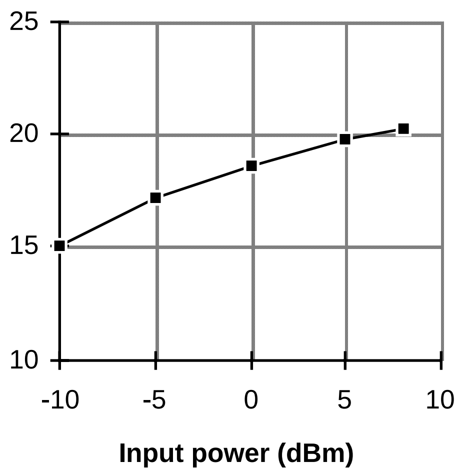


Figure 2
Output versus input power
Wavelength: 1300 nm
Pump power: 600 mW

The input power range around 0 dBm corresponds to the values compatible with the high CNR required for CATV systems. Although the output power increases with input, there is a significant saturation of the gain of the amplifier, since the 18 dB change in input signal results in only 5 dB increase in outptut. Taking into account the losses to and from the doped fiber, the internal quantum efficiency at maximum input power is higher than 30 %. The noise figure, measured with the optical spectrum analyzer method, is 6.2 dB for 0 dBm input signal at 1300 nm.

TRANSMISSION OVER 94 km

We have recently reported the application of this type PDFA's to distribution networks, and have shown that less than 2 dB CSO and CTB degradation were obtained with a high-quality link[4] . This asseses the use of PDFA's with AM-VSB signals, contrary to the results reported in [5]. However, application to long fiber spans remains to be confirmed, because of possible effects of fiber scatterings or non-linearities.

We report transmission over 94 km, with results of CNR, CSO and CTB, for a 30-channel multiplex spanning from 136 to 823 MHz, according to the France-Télécom frequency map. The reference and amplified links are represented on Figure 3

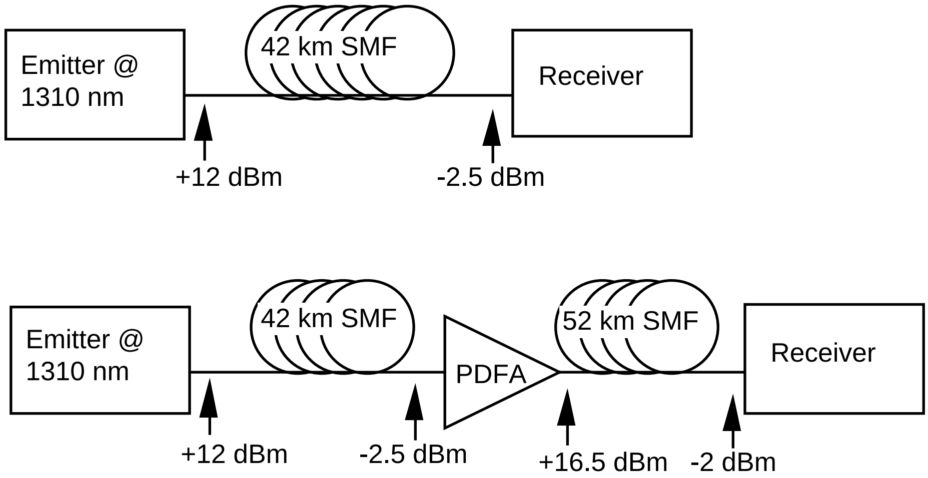


Figure 3
Reference and amplified links.

The average attenuation of the G.652 single-mode fiber (including splices and connectors) is 0.35 dB/km. There are 6 in-line FC-APC type connectors for the amplified link.

The CNR, CSO and CTB versus carrier frequency are given respectively on Figures 4, 5 and 6.

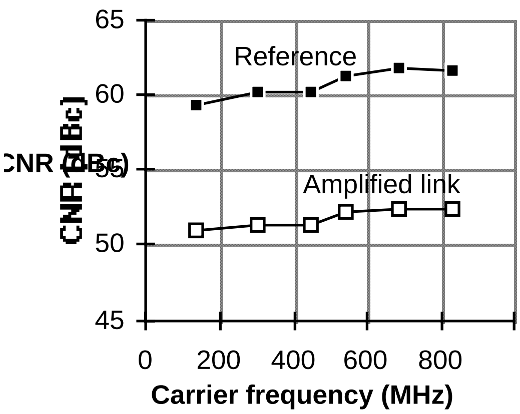


Figure 4
CNR versus carrier frequency

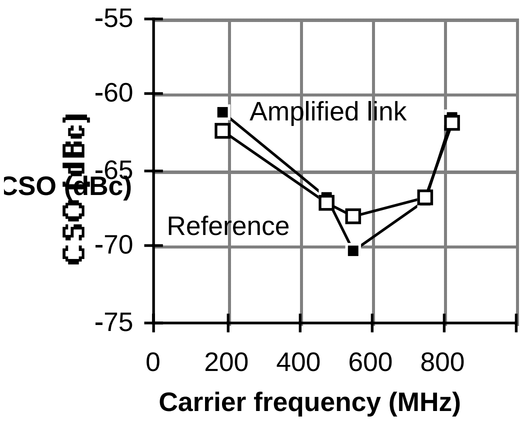


Figure 5
CSO versus carrier frequency

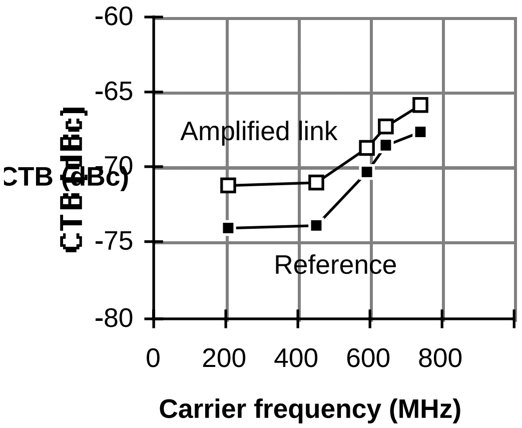


Figure 6
CTB versus carrier frequency

The CNR remains higher than 50 dBc over the whole band and its degradation is 9 dB. There is no measurable CSO degradation; CSO remains lower than -60 dBc and CTB lower than -65 dBc with about 3 dB degradation. Primarly, transmission is limited by the CNR, which in-turn is given by the gain of the amplifier for a given span. It should be noted that these results are obtained with no electronic regeneration, and that the Rayleigh and the possible non-linear scatterings in the fiber are taken in account. Concerning Brillouin scattering, it is estimated that direct modulation of the emitter results in an increased linewidth of the source, (due to the chirp of the laser diode), which lowers the Brillouin scattering efficiency. This is a significant difference with 1.55 µm based systems, where the high dispersion of standard fiber makes external modulation mandatory with a smaller linewidth and thus an increased sensitivity to Brillouin scattering.

CONCLUSION

We have demonstrated supertrunk transmission over 94 km of single-mode fiber, using an innovative Praseodymium Doped Fiber Amplifier. 30 subcarrier multiplexed channels are succesfully transmitted, with CNR's higher than 50 dBc, and CSO and CTB lower than -60 and -65 dBc respectively. This demonstrates that PDFA's are suitable for CATV networks with a large geographical span, or for the extension of already installed systems.

ACKNOWLEDGEMENTS

We would like to thank Michel BOITEL at FRANCE-TELECOM/BD/DTD for the distorsion mesurements, Stéphanie NOBILI at Photonetics for technical support when building the amplifier, and Hervé LFEVRE at Photonetics for encouragement throughout the project.

CONTACT

Vincent Morin and Edouard TAUFFLIEB are with PHOTONETICS S.A., 52 avenue de l'Europe,
BP 39, 78160 Marly-le-Roi, FRANCE.
Telephone: +33 (0) 1 39 16 33 77; Fax: +33 (0) 1 39 16 56 06.

REFERENCES

[1] Y. DURTESTE, M. MONERIE, J.Y. ALLAIN and H. POIGNANT, *"Amplification and lasing at 1.3 µm in praseodymium-doped fluorozirconate fibers",* Electronics Letters, **27**, pp. 626-628 (1991)

[2] T. WHITLEY, *"A review of recent system demonstrations incorporating 1.3 µm praseodymium-fluoride fiber amplifiers",* J. of Lightwave Technology, **13**, pp. 744-760, (1995)

[3] V. MORIN and E. TAUFFLIEB, *"High Output-Power Praseodymium Doped Fiber Amplifier Single-Pumped at 1030 nm: Analysis and Results",* J. of Selected Topics in Quantum Electronics, **3**, pp. 1112-1118, (1997)

[4] V. MORIN, E. TAUFFLIEB and H.C. LEFEVRE, *"30-channel AM-VSB CSO-free transmission through Praseodymium Doped Fiber Amplifier",* paper TuO1, OFC'98

[5] K. KIKUSHIMA et al. *"Distorsion and noise properties of a praseodymium-doped fluoride fiber amplifier in a 1.3 µm AM-SCM video transmission systems"*, Photonics Technology Letters, **30**, pp. 1431-1432, (1994).

Increasing HFC Capacity: Design and Field Test of Return Path Frequency Stacking

Robert Howald, Michael Aviles, and Frank McMahon
General Instrument Corporation

Abstract

Thinking ahead during HFC deployment and upgrading is the key to assuring the network's long-term success. In the race to build the information superhighway, it is important to recognize that now may be the time to add the few extra lanes that could make the difference between smooth sailing and gridlock. For the case of HFC, this means making the most of the performance-constraining return path. One cost-effective way to expand return capacity is to allow each node port to be translated to its own 35 MHz of RF bandwidth using frequency conversion. This paper describes system analysis, key hardware, and summarizes the key results of the first known field trial of a complete node-based frequency stacking system.

Fattening the Pipe

For Hybrid Fiber Coax (HFC) infrastructures, there are a couple of ways to assure that the network is future-proof as the subscriber base for two-way communications and multimedia services continues to increase. HFC utilizes analog fiber optic transport between the Headend and the neighborhood node. Fiber optic nodes (optical to RF transducers) located throughout the community output the broadband downstream onto coaxial cables, through which it is then transported to subscribers as in an all-coaxial network. In the return path, signals travel this same coax, and return on an upstream fiber. A single neighborhood node branches into multiple coaxial outputs from the downstream fiber, serving from hundreds to as many as 2000 homes.

In the return band, all 5-40 MHz node upstreams share spectrum on a single fiber. As more services are deployed and subscribers lined up, the constrained bandwidth will become a sure bottleneck, unless the network is designed ahead of time to be ready for the onslaught. One solution is to add multiple fiber and lasers, such that there is a return laser and receiver for each RF port on the node. However, this is not a particularly low cost or low power solution, and is wasteful of the generous bandwidth available to implement a laser to transport only about 35 MHz of bandwidth. A more cost effective and efficient method of expanding capacity allows each subset of the subscriber community sharing a fiber optic node port to have their own 5-40 MHz return. Frequency division multiplexing the returns to the node using frequency stacking, also called block frequency conversion, can do this. Then, each port on the node accommodates a unique 35 MHz of bandwidth, providing N times the capacity for an N-port node, and also isolating port-to-port ingress. This composite signal can then modulate the single laser, and the upconversion function can be inverted at the Headend by downconversion. The frequency stacking system (FSS) concept is shown in Figure 1. This paper will outline the relevant issues involved in the design and implementation of a robust, high performance, FSS.

Communications Issues

The design goal of any hardware added in the middle of the pipe is to make sure that nobody on either end knows it is there. This means allocating specifications such that any degradation introduced goes unnoticed by any application. Ideally, any application's link budget will be negligibly effected. Obviously, this is an imposing goal. To fulfill it would require knowledge of every potential application, the modulation technique used, and more information on the HFC plants themselves. The latter two items can at least be quantified to generate numbers based on some assumptions, while the first item would require a visionary in marketing (always a challenge!). Today's fast-paced markets in telecommunications and wireless often place technology companies and their product developers back on their heels, and predicting services and take rates for two-way HFC is one such instance.

There are several primary enabling technologies that will be big players as HFC digital transport mechanisms. Among these transmission and modulation schemes are QAM, in conjunction with TDMA and FDMA, and possibly CDMA and OFDM. Designing FSS components to support complex modulations puts significant constraints on phase noise, amplitude ripple and flatness, group delay variation, among other important parameters. In order to maintain a quality uncorrected BER in the return path (our measure of performance), the converters are each allocated a portion of the tolerable amounts of the various degradations. Care must be taken in recognizing which impairments particular transmission schemes may be especially sensitive to.

Link Design Philosophy

It must be determined what effect FSS will have on the existing link budgets, which are the key to any successful communication system design. Consider the cable modem application. Degradation of 16-QAM performance can become quite substantial with relatively small amplitude and group delay distortion, since 16-QAM has information in its amplitude. The ability to correct minor disturbances with a simple equalizer is well known, and it is not cost effective in the design of the transmission network to impose difficult specifications that would be otherwise mitigated by proper modem design. Similarly, it is well known that the impulse noise problem is best handled by proper code and interleaver design. Optimum implementation strategies would utilize concurrent development of the system and the plants. Unfortunately, existing infrastructures are already in place.

Three primary link parameters of interest for FSS are thermal noise and signal-to-noise ratio (SNR), intermodulation distortion (IMD) and spurious, and phase noise.

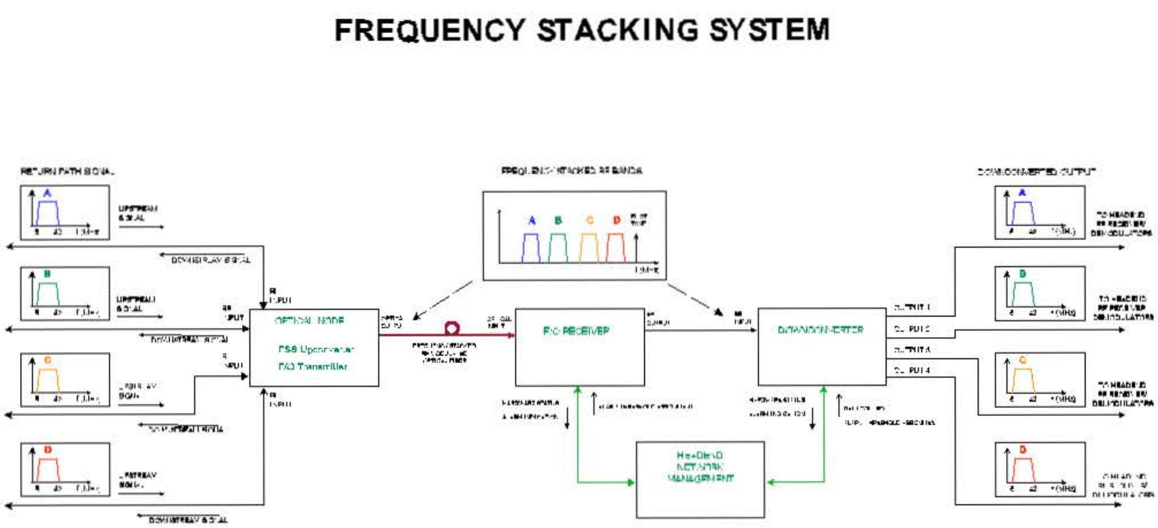


Figure 1 - A Four-Band Frequency Stacking System

Thermal Noise

Perhaps the most significant channel quality on the beneficial side for HFC networks is its inherently high SNR, where SNR implies the ratio of signal to thermal noise (AWGN). In the upstream, and even more so in the downstream, HFC networks are capable of quite high SNR's, which translates into high theoretical channel capacity. Exactly how much of this capacity can be taken advantage of is a function of how well the other impairments can be mitigated through proper modem design. For HFC networks, the primary contributor to the thermal floor is the fiber part of the system. Other noise contributors include the coaxial part of the plant, particularly in the noise funneling upstream, the noise figure associated with the upconverter at the head of the RF cascade, and, to a lesser extent because of location, the downconverter and demodulator.

The Fiber Optic Link

The fiber optic portion of the network, consisting of the laser transmitter, fiber optic cable and optical receiver, typically dominate link SNR capability. The RF/cable portion of the network, and the post-optical receiver electronic hardware, usually contribute in only a small way to overall SNR. Fiber optic limitations generally reveal themselves by two means: unacceptable minimum SNR and distortion effects, including those associated with excess loading, causing clipping. The initial setup and operation of a return link requires a different philosophy than that for the forward path, where a fixed number of signals are located within known video bandwidths at constant levels. The simplest assignment of return signal levels, although not ideal from a communication link perspective, is on a per bandwidth basis. The approach has important implementation advantages, such as its setup and testability. Also, it yields a constant SNR for all channels regardless of bandwidth, and allows operators to be prepared for eventual full channel loading without having to adjust signal levels. The total power allocated for return services is determined by the recommended composite signal level at the laser transmitter needed to maintain acceptable SNR yet avoid clipping effects.

Optical Parameters Effecting SNR

The main contributor in HFC links to SNR degradation is the laser diode used in the transmitter. Unless operating through the longer fiber networks, the transmitter diode's internal noise limits SNR. The noise is quantified as relative intensity noise (RIN). The RIN of a diode is expressed in dBc/Hz and depends on the type of laser used. The two types in use for HFC are Distributed Feedback, or DFB, and Fabry-Perot (FP). The RIN is typically anywhere from -110 to -160 dBc/Hz. DFB's normally have lower noise characteristics than FP's. Thus, for high quality digital communications, DFB's appear best suited to providing good SNR at the low signal levels desired to avoid laser overdrive and clipping in heavily loaded returns. Using RIN, together with the optical modulation index (OMI), the SNR associated with the laser diode section of the optical link can be found. Figure 2, CNR vs. Fiber Length (non-FSS), shows this value, identified as CNRtx, to be a straight line at 41 dB (carrier-to-noise, or CNR, is often utilized to be more consistent with analog video CNR, already familiar to the industry). CNRtx performance tracks on a dB to dB basis with RIN variation.

At the optical receiver, shot noise in the photodiode (typically PINs) contributes to SNR degradation. Shot noise limits are determined by the optical power and diode responsivity. Post detection RF amplifier circuitry also contributes to SNR degradation, its effect usually defined by an equivalent input noise current (EINC). Degradation due to shot noise tracks on a dB to dB basis with the optical input power level, and the variation due to EINC tracks on a 2 dB to dB basis with the optical power. In Figure 2, the two contributors are combined and the value is identified as CNRrx. The plot shows CNRrx to vary about 20 dB versus length. Overall link performance can be found by combining transmit and receive performance. As can be seen from the Total CNR, for fiber lengths out to 20 km, the laser transmitter is the dominant contributor to link degradation in this network. The receiver is not a major factor until distances of 25 km and longer are reached. From a BER perspective, the lowest SNR at the longest link is well above the minimum required to support the return services anticipated.

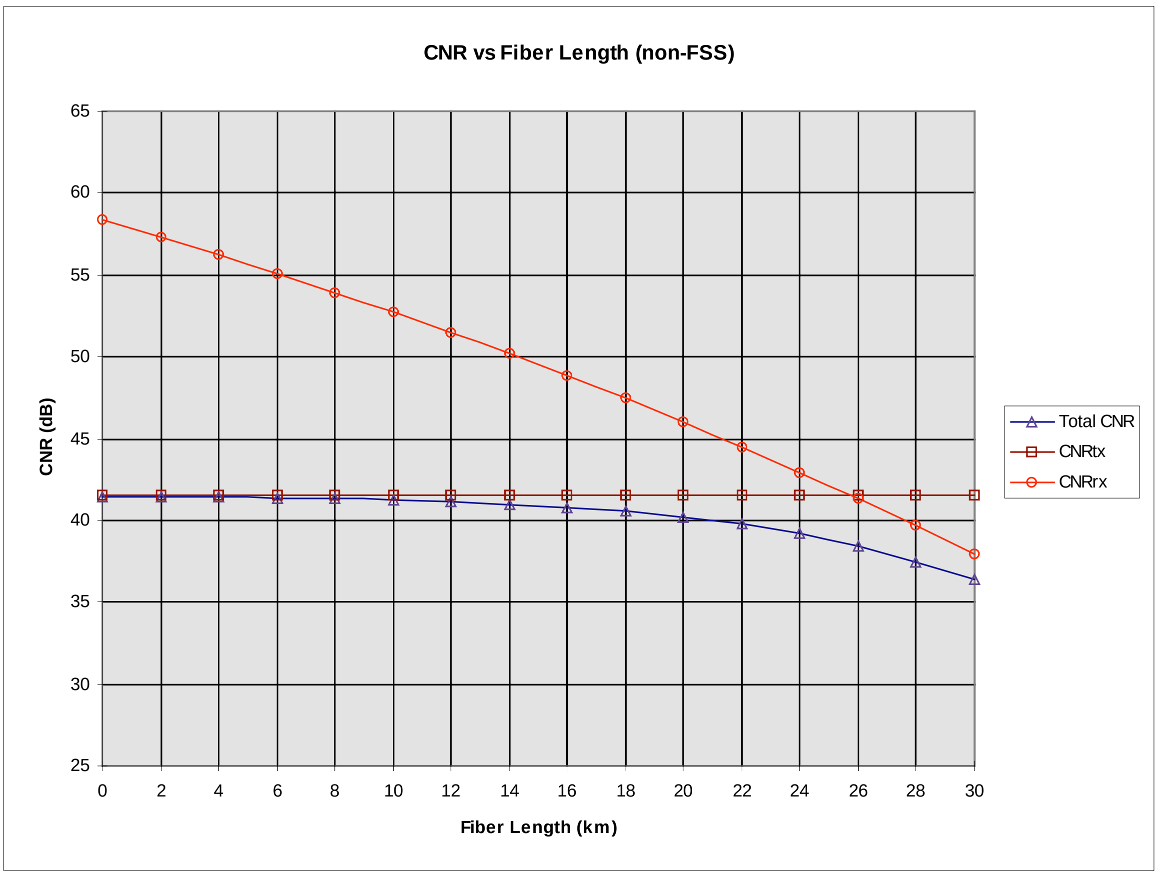


Figure 2 - Performance w/o FSS

Link Performance with Frequency Stacking

In Figure 2, without FSS, the link was shown capable of providing greater than 35 dB SNR always, and typically > 40 dB. Now consider the additional noise generated by the up and downconverter hardware. Figure 3, CNR vs. Fiber Length (FSS) shows what the addition of frequency conversion hardware does to the overall link performance.

At the shorter fiber lengths, for constant downconverter input power, optical receiver gain control attenuation settings are highest, degrading the subsequent CNRrx from the non-FSS system as shown in Figure 3. This has little effect on the link since the laser diode noise is still dominant. For these shorter links, the contribution of the FSS upconverter causes minimal degradation, by design, to the CNRtx under all conditions. At the longer links, the added noise associated with the receiver/downconverter hardware is masked even more by the equivalent noise degradation due to fiber losses. The result is even less difference in total SNR between FSS and non-FSS. Thus, incorporation of a properly designed FSS has minimal impact on an operator's ability to provide quality return path services while significantly increasing subscriber density per node.

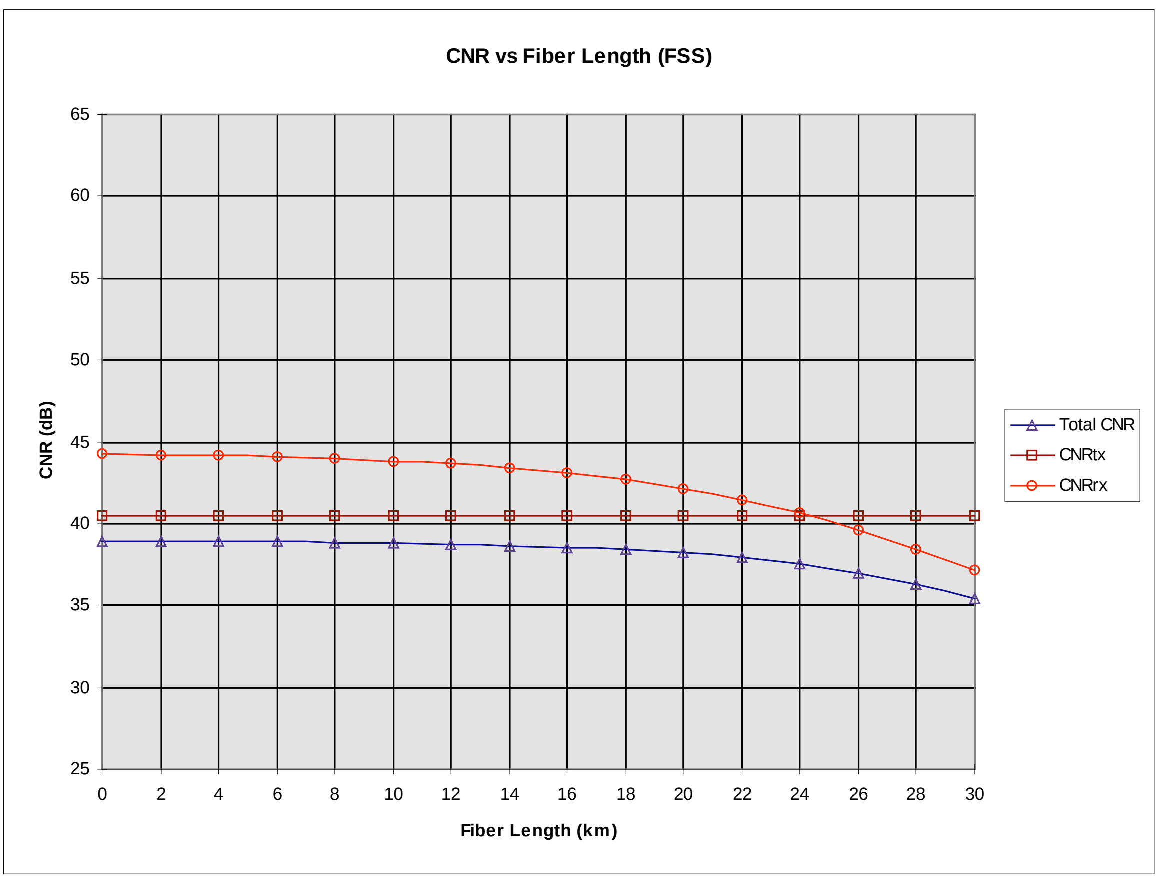


Figure 3 - Performance with FSS

Channel bandwidths above are not discussed, because the power-per-unit Hz allocation equalizes SNR in any channel. However, to discuss performance in terms of BER and data rate, SNR and bandwidth are needed. For video, this bandwidth is about 4 MHz. This bandwidth, at the SNR's calculated above, is adequate for high performance M-QAM with significant margin of between about 10-20 dB for QPSK through 64-QAM. For example, 16-QAM at 1e-8 symbol error rate requires about 22 dB of SNR. Naturally, QPSK needs less SNR, 64 QAM more, etc. The 4 MHz of RF bandwidth would represent at least 8 MBPS for 16-QAM.

Intermodulation Distortion (IMD) and Spurious

Use of frequency synthesis and conversion hardware results in the need to analyze and quantify intermodulation and spurious performance. Because of the broadband nature of the system, multiple intermodulation beats exist, and the number grows drastically as the number of signals increases. Of most interest are products that contribute to the degradation of digital communications performance by causing a significant signal-to-interference ratio (S/I). The products that dominate broadband performance can be either second order or third order beats, in contrast to a narrowband system which can often ignore second order products. Any part of the RF chain called on to process a broadband input and produce a broadband output must be concerned with second order products. Between these items, any filtering that reduces second order products will benefit second order performance. For HFC, the second order performance is typically laser dominated. Another advantage of frequency conversion is the freedom to design a frequency plan that helps mitigate second order interference.

The third order intercept (TOI) is typically used to characterize third order intermodulation characteristics of RF components. Third order products require consideration of the effects of multiple carriers, as this degrades the overall third order intercept of the cascade relative to the common two-tone reference. Unlike noise figure, cascaded intercept point is typically dominated by elements at the end of the chain.

In broadband systems, care must be taken in understanding the many possible sources of spurious frequencies. Unwanted signals can point to many areas: undesired higher order mixing products, frequency synthesizer related and local oscillator (LO) leakage spurs, intermodulation distortion in active components, DC power distribution, RF leakage, etc. Spurious contribute to S/I degradation, decreasing link margin. Locally generated spurious in the upconverter in the node can have the capability to be large relative to incoming signal levels. Proper design for adequate S/I is a combination of proper RF chain gain and intercept allocations at full load, and quality RF circuit board design.

Phase Noise

Traditional CATV frequency synthesis techniques result in typically phase noisy carriers, because analog video requirements are non-demanding in this regard. Thus, low cost, direct divide

PLL synthesis is dominant, and, combined with fine resolution of desired channel frequency outputs, results in high division ratios and the resulting noisy output. One of the key items to recognize in any digital communications system being implemented over HFC is that noisy synthesizers are generally not well suited to reliable digital communications, particularly of the high M-QAM variety. And since phase noise is a burst-type error mechanism, to mitigate via FEC requires interleaving, burst correcting codes, or both. Much can be gained by making relatively simple modifications to frequency generation. Once again, with the freedom to choose a frequency plan, designs that minimize phase noise are possible. Because of the myriad of applications to be served, the most useful specification of phase noise is to quantify its rms jitter performance over each decade of offset. This allows ease of identification of the portion of the phase jitter spectrum of significance to each application modem.

In addition to generating low noise local oscillator signals, implementation of frequency tracking benefits every application. The amount of frequency error that is acceptable varies by application, since there are so many different types of modems both existing and being developed. Lack of standardization for return path transmissions has resulted in the proliferation of various techniques, including modulations such as FSK, QPSK, QAM, and signaling strategies like CDMA, or OFDM and its wavelet-based cousins. A zero-frequency error approach requires a tracking PLL in the downconverter, and eliminates on both sides the need for very high stability references. In addition, the pilot recovery PLL serves to track out some of the upconverter's phase noise contribution. Zero frequency error means the FSS can be ignored in the complex analysis and allocation of requirements of synchronization, and, in particular, for sensitive burst modems, CDMA, and OFDM applications. Designs based on free-running crystal references contribute to frequency offset that must be handled by an application demodulator, which means that the FSS has become intrusive.

Upconversion in the Node

In terms of RF hardware, frequency conversion is a considerably mature technology. It is, however, unique in end-to-end design to nearly every application. For this case, upconverter design constraints include size, environment, power dissipation, induced phase jitter, spurious generation, and nearby image frequencies. What the design can take advantage of is that the signal levels are not high in this part of the plant, and that there is a zero gain requirement.

A dual conversion (see Figure 4) approach provides a good compromise of straightforward filtering of images and LO leakage, mixer product spurious management, and commonality of parts in getting from the multi-octave 5-40 MHz input to UHF outputs in four isolated bands. Converting all four bands eliminates multi-octave RF design in the node, and allows ease of laser implementation using ordinary forward-band units.

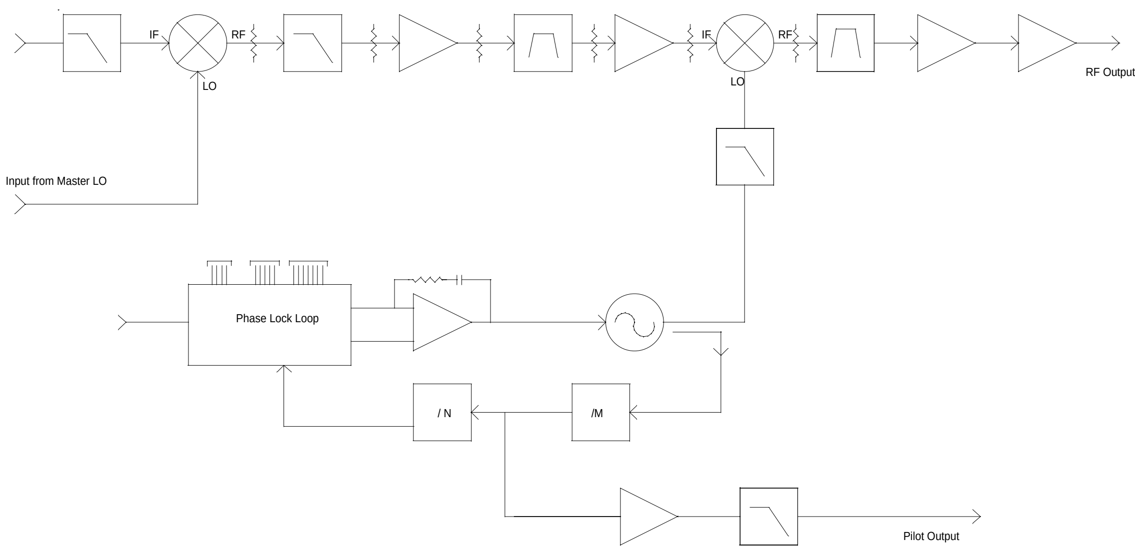


Figure 4 - Dual Conversion RF Design

Because of its location in the chain, upconverter noise figure can contribute to overall RF-related thermal noise. As such, it is important for cascaded gain blocks to be mixed into the front end and evenly dispersed. However, for IMD, filtering of wide out-of-band mixer products prior to the amplification is important, while simultaneously providing sufficient termination of the mixer ports so its performance does not degrade. Because of the multi-octave input, the first mixer and associated circuitry are critical to spurious performance, and therefore it is important that a quality mixer be used, and that the amplifier after the mixer have good dynamic range. These elements and isolation of circuitry drive the spectral purity of the design, provided proper matching around the high selectivity IF filter and output RF bandpass filters is maintained.

Other critical performance parameters in the RF path include the amplitude and group delay responses. The use of high selectivity filters can have consequences in both aspects, and can be troublesome should equalization of sensitive modulations be ignored. Another key RF parameter is the forward path isolation. Because the upconverter shares a compact, highly integrated RF environment in the node, isolation of return and forward signals, mostly analog video, is very important.

The frequency synthesis part of the design uses a common PLL synthesizer IC for all LO's, such that only varying of the divide ratios in the PLL are required. To be compatible with M-QAM, the design of this subsystem was based on maximizing phase comparison frequency for minimum divide ratio, and optimizing loop filter design values. In addition, because frequencies are not required to be programmable, low cost, low noise, narrowband discrete VCO's can be implemented. Integrated rms phase jitter on the order of less than one degree rms is typical over a 100 Hz to 1 MHz range.

Finally, zero frequency offset is achieved in the link by using a pilot tone, as previously described. With the flexibility of choosing a frequency plan, the signal can be placed well out of band of payload traffic.

Downconversion at the Headend

At the Headend, the purpose of the downconverter is to take the return path RF signal, consisting of four individual bands stacked in frequency, extract them, separate, and downconvert each to the original frequency bands. The downconverter is also typically required to interface with Headend network management equipment. As described, one important characteristic of block conversion is the frequency error introduced. For zero frequency offset at the unit's output, the downconverter implements the pilot tracking PLL, which is used to exactly re-derive the LO frequencies generated at the upconverter.

A critical design requirement for the downconverter is its spurious performance. Since the output signal band of the downconverter may exceed three octaves, special care must be taken in assuring highly linear amplifiers and mixers. Other important parameters, again, include amplitude flatness, phase linearity, induced phase jitter, noise figure, gain control, electrical isolation and power consumption.

Spurious signals can be generated in the downconverter through the mixing, nonlinearity in PIN diode and FET attenuators, and in switches. These distortion products need to be low enough not to interfere with desired transmissions. An important step in minimizing the levels of in-band spurious is in the analyzing of mixer products for the chosen frequency plan. Having specified the approach, functional block performance allocations are defined using the system level requirements for gain, noise figure, output signal level, second and third order intercept performance, gain control range, etc.

Determining the downconverter's output signal level is dependent on the number and types of return path services to be supported. The larger the number of different services accommodated, the larger the number of RF power splits, and correspondingly the higher the RF losses between the downconverter and application demodulators. The output signal level required is defined by the range of level requirements for the various demodulators, adding the splitting losses for present and future services, and then providing some level of margin. In order to provide sufficient output signal level in each output arm, the downconverter may have to deliver output signals on the order of 40 to 50 dBmV per converter channel (i.e. 35 MHz bandwidth). In order to accommodate various Headend configurations, it is desirable to provide some level of gain control within the module.

Performance in the Field

In the fall of 1997, GI's FSS was deployed in a field trial of the new SG 2000 node. The upconverter was installed in an existing four-port node configuration (the upconverter in the node was designed to be field replaceable as an identical form fit to an existing passive combiner RF board). Two RF ports on the node were connected to a functional plant, and the second two were wired to a motel room, where a QPSK modulator, taking in a pseudorandom bit stream, was located. At the Headend, located at the end of a relatively short fiber optic link, a QPSK demodulator followed the downconverter. The returns in this case implemented 5-42 MHz bands. Performance testing consisted of measuring the error rate statistics and average BER performance on one band for a period of time, and subsequently rotating through each band repeatedly. Because the network in question had no operating return prior to the testing, only ingress characterizations prior to running the error rate testing were available to gauge the nature of the return being used. Note that each band (i.e. each node port) had the QPSK upstream signal summed in, but only one band at a time was measured.

Unmaintained Plant

Before performing any plant upgrades in anticipation of employing return services, data was taken with the QPSK operating, providing a measure of the raw networks' readiness for digital communication. It is well known from return HFC characterizations, ongoing for about five years now, that major impairments that exist include noise funneling, narrowband ingress, frequency response distortion, impulsive noise, and 60 Hz interference coming in both hum and impulsive varieties. Because of these known problems, equipment being designed today is employing sophisticated equalization and error correction techniques to provide mitigation. Some equipment manufacturers are implementing advanced modulation and signaling approaches, such as the CDMA and OFDM, built specifically to mitigate these known impairments. For this test, we selected the simplest practical scheme anticipated for modern advanced services, QPSK, and did not implement any error correction. Thus, raw data availability parameters could be obtained, as well as important BER data, to help characterize the plant's capacity for digital communications. For the length of the test, the QPSK data rate was 2 MBPS, using about 1.5 MHz of RF bandwidth.

Single Signal Testing

On the raw plant, two weeks of QPSK data showed obvious impairments to uncorrected transmission. There was a very high degree of channel availability, as measured by the percentage of error-free seconds (EFS) and severely errored seconds (SES). In other words, there are very long periods with no errors (typically measuring 99.75% of the time), followed by an impulsive burst of errors (about .2% of time). The percentage of time without SES was therefore greater than 99.95%. This is important, because even rudimentary error correction can fix errors not related to severely errored seconds, because they tend to be more randomly distributed. The severely errored seconds, which occur in bursts, are more difficult to correct. This is the reason for the strong recognition of sophisticated error correction in the Multimedia Cable Network Systems (MCNS) specification (a standard for cable modems). The forward error correction (FEC) to be employed consists of a concatenated trellis and Reed-Solomon implementation, as well as interleaving. Interleaving is the technique by which the symbol sequence is transmitted such that some designed number of symbols, ideally associated with the expected burst statistics separates adjacent symbols. Thus, it is implemented specifically to aid in burst correction.

MCNS FEC specifications were built around the anticipated statistics of return burst-type interference, of which very recent studies have indicated a strong presence of the power-line related type. These findings indicate an important need to provide quality AC distribution and grounding in the plant, as well as the need to consider the nature of home-generated disturbances associated with major appliances. RF ingress levels associated with HFC returns, both steady-state and impulsive, have been accumulated by many sources for statistical analysis. The sources of ingress are quite well understood, and most will be present on every two-way HFC plant, no matter how clean. However, it is important to point out that ill-maintained plants, be it by poor grounding, poor in-home wiring, and/or poor connections, will significantly aggravate ingress levels at the Headend, potentially effecting the EFS performance and the ability of errors to be corrected.

Finally, looking at average BER during this test is also somewhat informative. For the length of the single signal testing, the QPSK power was set such that it represented the level as if the modem had to share a fully-loaded channel using a power-per-Hz allocation methodology. In other words, the input signal used was about 14 dB below (10log(1.5/37)), the maximum port input suggested. Looked at yet another way, the QPSK level was scaled to match its percentage of band occupancy in the 37 MHz return. Raw data indicates that, not counting SES periods, average BER's on the order of E-10 occurred during quiet times, on the order of E-8 and E-9 during nominal periods, and on the order of E-7 in the worst periods. Including the SES periods, there was variations on a per-day basis of days as good as E-9, nominal in E-5 and E-6 range, and E-4's at the poor extreme. Measurements were auto-recorded from the BER tester, and it is important to note that the lowest extreme (the zero-error limit) of the BERT is, in fact, 1E-10. Thus, periods of zero errors would correspond to this average BER as measured by the BERT.

Noise Loaded Testing

After about two weeks of gathering data on the unmaintained plant, the MSO implemented a well-coordinated effort to go through the plant methodically, tightening down all connections, and assuring good plant grounding and powering. Not all details were immediately available about every maintenance item addressed, as an outside contractor performed much of the work. Following the plant upgrade, the single modulated signal test was repeated, and it was immediately apparent that there were zero errors 100% of the time. This behavior continued on upstream band 1, until, after continuing to count zero errors, verification on the other three bands was done to make sure things were connected and ready to go to the next test step - noise loading.

To further demonstrate the capability of the FSS-2000 platform, the unit was subjected to a full loading of every band. The input of each port has the maximum suggested input level, uniformly spread across the full 37 MHz of bandwidth. The white noise occupied the entire 37 MHz band, except for a small portion. Using a notch filter, part of the spectrum is cleared out, and inserted in this open real estate is the QPSK signal. A plot of this "noise notch" transmit signal is shown in Figure 5. This loading configuration harshly tests the FSS and fiber link dynamic range capability, as Gaussian noise samples have a higher likelihood of producing large peaks capable of clipping the laser as well as driving RF amplifiers into saturation briefly. The fully loaded spectrum also fully stresses the ability to provide highly linear 5-40 MHz RF outputs from the downconverter at a high signal level.

Results of this test were also extremely encouraging. The same basic procedure was implemented, where the error rate measurements were accumulated on one band at a time. This test lasted for about one week, with data again recorded around the clock. On bands one and two, there were nearly identically zero errors during all measurement periods (it is reasonable to assume that each band had one fourth of a week of measurement time). Band one still actually had zero errors, and band two counted two bit errors in 40-some hours. On bands three and four, there were more errors, and logically so, since these two bands were connected to the operating plant. As such, they were exposed to the sources of upstream ingress. Even given that, over 40 hours of monitoring band three produced only about 350 bit errors, while band four showed only 651 bit errors. This was quite astonishing, but, given that the plant was relatively small (about 150 homes, 75 on each port), there were fewer sources of home-generated ingress. With ingress correspondingly reduced, excellent BER resulted, even without error correction.

While QPSK is a very robust modulation, this performance, without any error correction, was better than had been anticipated. The dynamic range of the FSS had been thoroughly and harshly tested, and performed admirably at the high end. Noise power ratio (NPR) tests show virtually identical performance (about 41 dB) between a system with FSS and without FSS using the suggested power loading.

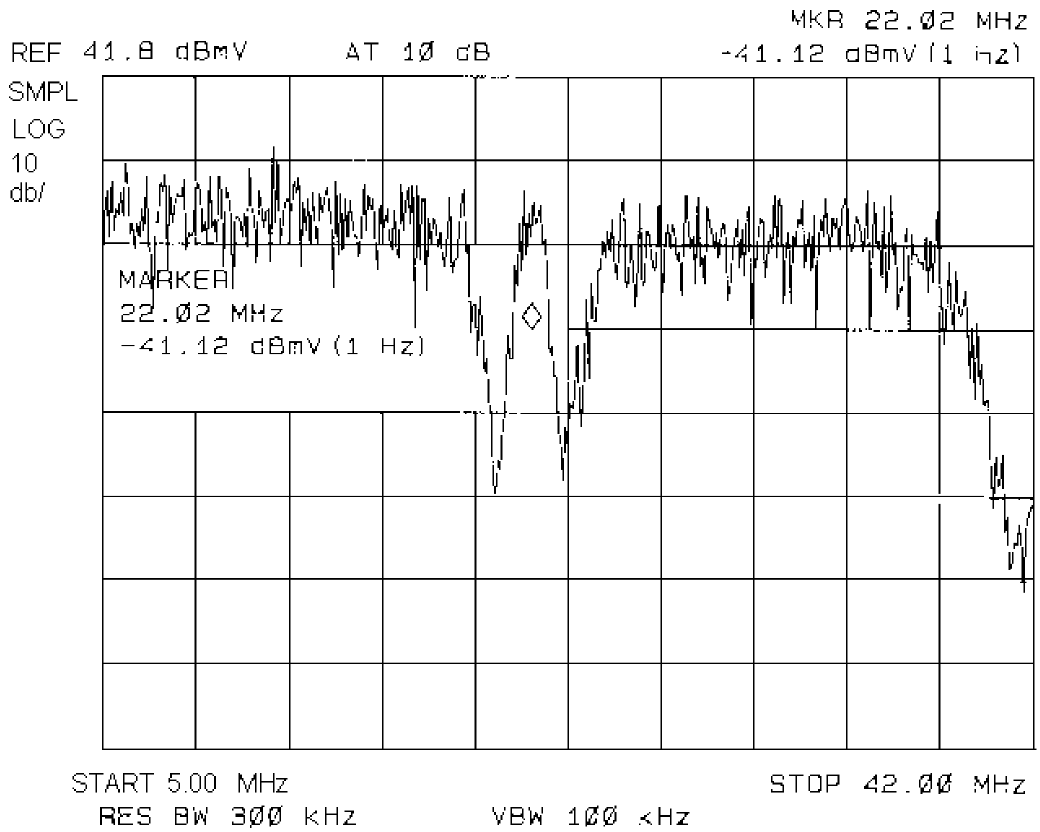


Figure 5 - Noise Load Testing

Conclusion

This piece has discussed a straightforward and cost-effective way to add capacity in the return path for HFC networks. More importantly, FSS has been proven through link testing, both in the lab and in the field. While telecommunications roll-out onto cable networks has been slow, if MSO's are committed to the growth of digital data transport in the return path as an important revenue stream, now is the time to provision that network for this growth. Technologies such as FSS effectively multiply the shared bandwidth, allowing the revenue stream to grow without sacrificing network quality.

Acknowledgements

The authors would like to recognize Mike Short and Gary Hitchner, whose efforts were critical to the completion of this paper.

Dr. Howald (215-323-2276) is the group leader of Systems Engineering in GI's Transmission Networks Systems unit. Mr. Aviles (215-323-2269) is the lead systems engineer for the FSS product line. GI is located at 101 Tournament Drive, Horsham, Pa., 19044.

Invisible Hub or End-to-End Transparency

Oleh J. Sniezko & Tony E. Werner

During the last decade, the tree-and-branch cable TV architectures have evolved into HFC architectures. These changes were inevitable due to advanced service requirements for increased quality, increased reliability, and interactivity.

Most of the advanced HFC networks introduce secondary hubs with a star configuration from secondary hubs to the nodes and a ring between the primary hub and the secondary hubs.

This paper analyzes the transport layer choices in this ring. Four basic alternatives are presented and compared. The three of them have been analyzed before, the fourth, dense wavelength division multiplexing (DWDM), became a feasible alternative in 1997. The paper compares advantages and disadvantages for all four of them and their capital and operating costs. Finally, it presents a possible implementation path for DWDM transport system.

NETWORKING IMPERATIVE

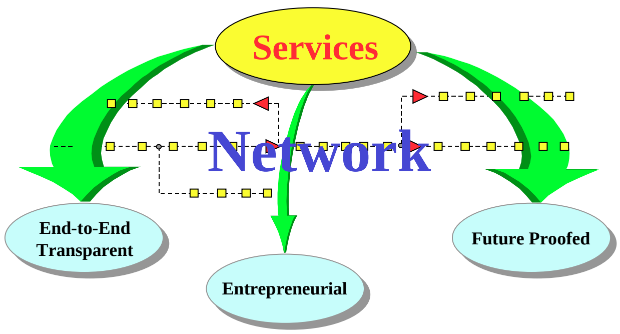
Networking Paradigm

One of the major objectives of telecommunications engineers is to design and build a network that is transparent, scaleable, and future-proofed. Such a network would allow us:

- to introduce any services that we can anticipate
- as soon as the demand for them is high enough to justify the investment
- in a timely fashion to outdistance the competition
- by adding required terminal equipment at the customer premises and signal processing centers only with no or almost no changes to the network.

The authors have presented the thoughts listed above before. However, they are repeated here to emphasize the leading imperative of network design. This imperative never became so obvious as during the implementation of the digital TV services over the HFC network. The entrepreneurial character of our network and its transparency allowed for fast implementation at significant savings.

Figure 1: Network Design Paradigm



Smart or Not-So-Smart Network

One of the most heated debates relates to the question of how smart the network should be. To be more accurate, the question is whether the smartness should extend to the final user or should stop at some higher network level. The HFC operators tend to design the network that does not require a complex OAM&P (Operations, Administration, Maintenance and Provisioning) systems beyond the primary hub or headend. Except for limited monitoring (secondary hubs, optical nodes, and stand-by power supply) and redundancy switching in the primary and secondary (optional) hub rings designed for improved reliability, the network operations, maintenance and provisioning relies on smartness in terminal equipment and in transmission protocols.

On the other hand, telecommunications network operators invest significant effort and capital in designing and building intelligent networks with the OAM&P elements deployed to the level extending to the very last interface. Even the customer interfaces in some proposed solutions include switching and multiplexing and provision for a multitude of physical layer protocols. This approach involves high risk of network obsolescence caused by technology progress, and requires high level of capital invested in the network (fixed cost). Moreover, this approach is not scalable. The fixed costs that have to be born in the initial stages of network provisioning are prohibitive to any single telecommunications company. The broadband access intelligent network plans are being abandoned as soon as the players have to put their money where their mouth is. The same players build the broadband access network based on copper plant with smart terminals (xDSL). On the other hand, asking the public to finance this network through a tax system involves very high risk of spending public funds on the network that can be obsolete before it is ready.

The situation with these two different approaches to the network intelligence reminds many other similar dilemmas. The closest parallel can be drawn between this debate and the debate about centralized computing (very high capacity mainframe computers with "dumb" terminals connected to it) and distributed computing (smart terminals interconnected to create a network). The pace of progress in processing power and storage capacity so far favors the latter approach, especially in residential and business environment.

HISTORICAL OVERVIEW

The tree-and-branch cable TV architectures served the public with great success that was rooted in the perfect match with the services demanded. Over time, they have evolved into HFC architectures to satisfy the advanced service requirements for increased quality, reliability, and interactivity. This evolution was enabled through the deployment of fiber optic technology deep into the plant. This node based deployment has also satisfied the requirement for spatial division multiplexing (SDM) that allowed for effective reverse path problem management, and for effective traffic engineering.

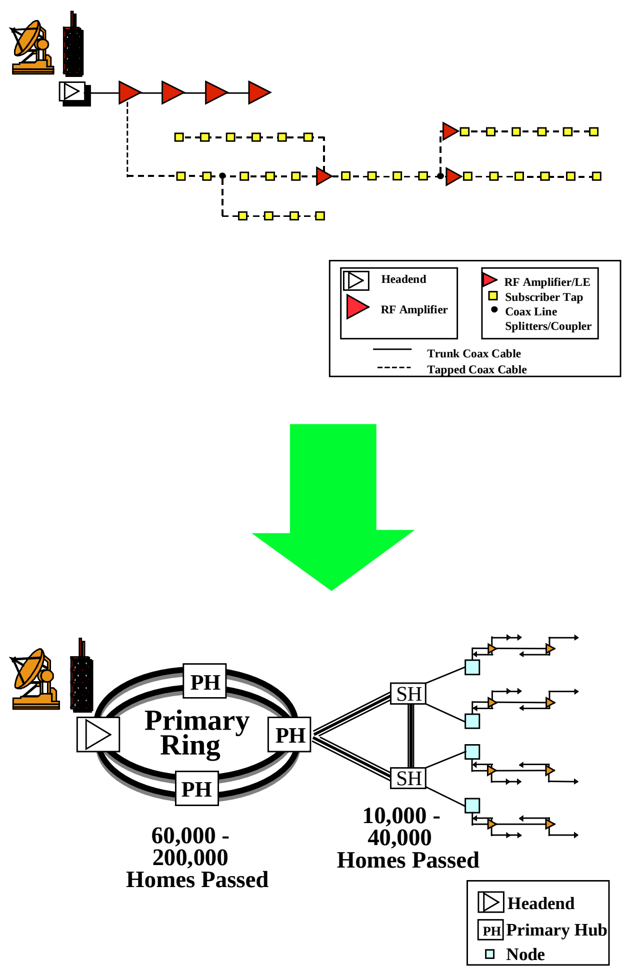
Table 1: Perfect Historical Match

Services	*Architectures*
Broadcast TV Broadcast Radio Addressable Services: • addressable tiering • PPV • games • digital radio Home Shopping	Coaxial Tree & Branch Fiber Supertrunk Fiber Backbone Fiber-to-the-Feeder

Table 2: New Services -- New Solutions

Services	*Architectures*
Services: • targeted advertising • targeted entertainment • telephony • high speed data • full multimedia **Competition**: • DBS • telephone companies • multimedia mergers	**Requirements**: • superior reliability • competitive quality • competitive price **Architectural changes**: • fiber supertrunking and fiber backbone • regional hub ring • redundancy (secondary hub rings) • deep fiber deployment & segmentation

Figure 2: Network Evolution: from Tree-&-Branch to HFC



Most of the HFC architectures closely resemble CableLabs' Active Coaxial Network Architecture. In the largest metropolitan areas, numerous headends are most likely to be connected in a ring to provide a fully redundant and survivable platform. In many cases, primary hubs are established to maintain signal quality delivered to the distribution network. These hubs serve from 60K to 100K homes passed. In most implementations, these rings deploy one of the following transmission technologies:

1. Proprietary Digital Systems,

2. Synchronous Optical Network (SONET), or

3. 1550 nm optical links with Erbium Doped Fiber Amplifiers (EDFA).

A choice of the particular technology is based on the market size, distances, and network complexity. In most markets, the transport system is based on a digital baseband (TDM) transport and, in many cases, deploys both SONET and proprietary digital transports.

SECONDARY HUB RING

Topology

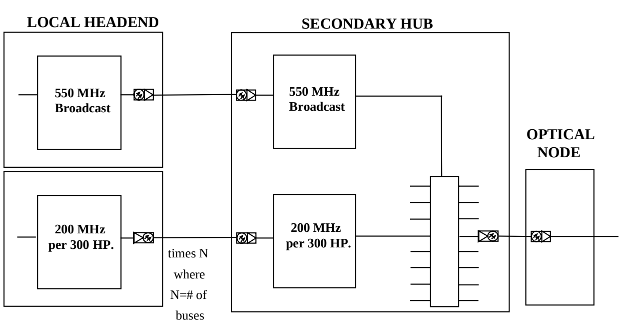
The choice of the transmission technology for the secondary hub ring is not that apparent. It is strongly dependent on the topology selected for the optical section of the HFC plant. Two basic topology choices are:

1. star architecture (home run forward and reverse fiber) from headend/primary hub ring (ring-star-bus architecture) to the nodes, or

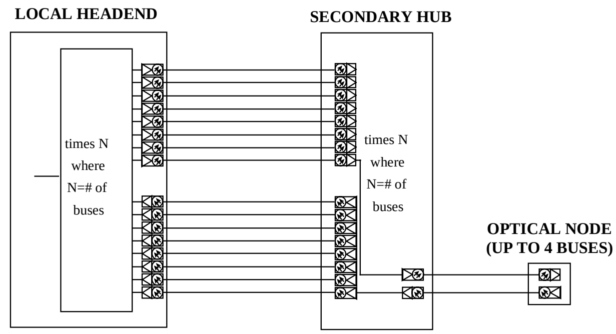
2. ring architecture where secondary hubs are interconnected with the primary hub in a ring, with star architecture from secondary hubs to the nodes (ring-ring-star-bus architecture).

Figure 3: Home-Run Optical Links

a) Forward Split Bandwidth, TSD Overlay



a) Forward and Reverse, Combined Bandwidth



The first alternative results in a simple architecture that is largely both passive and transparent between the headend or primary hub and the optical node. Unfortunately, it employs high fiber count cables that increase capital cost and allow for single point of failure with long mean time to repair. This architecture is also impractical to employ in a ring configuration for path redundancy. Other technical challenges are related to the reverse path implementation.

Secondary Hub Ring Technologies

The ring topology provides an opportunity for a cost-effective and highly reliable network with a limited number of fibers between the primary and secondary hubs. and is deployed in some form by all major MSOs. However, there are several technology choices that can be used to implement the ring configuration:

1. Analog FDM for broadcast and FDM with frequency conversion (frequency stacking) overlay for targeted signals;

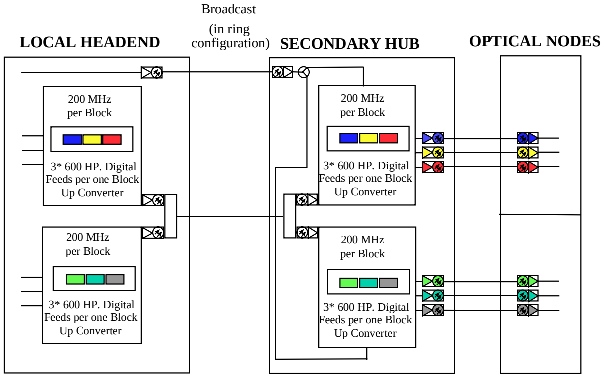
2. Hybrid analog FDM for broadcast and SONET or Ethernet (TDM) for targeted signals;

3. Analog FDM for broadcast and all optical DWDM (dense wavelength division multiplexing) for targeted services.

Furthermore, the choice of the technology for downstream transport can be different than the choice for upstream transport. The final decision will depend on many factors.

Figure 4: FDM & Frequency Stacking

a) Forward Split Bandwidth, TSD Overlay



b) Reverse

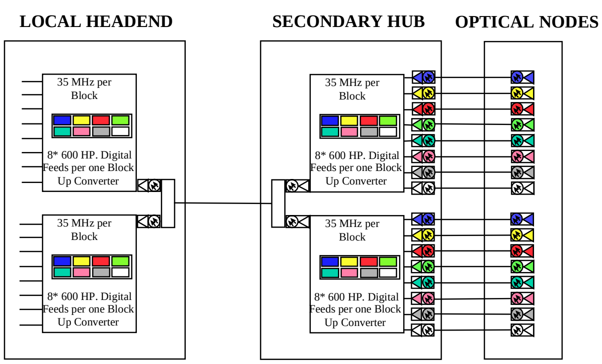
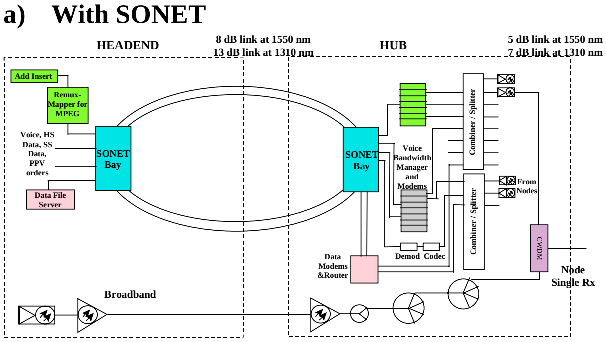


Figure 5: Hybrid Analog and Digital

a) With SONET



b) With 100BaseT

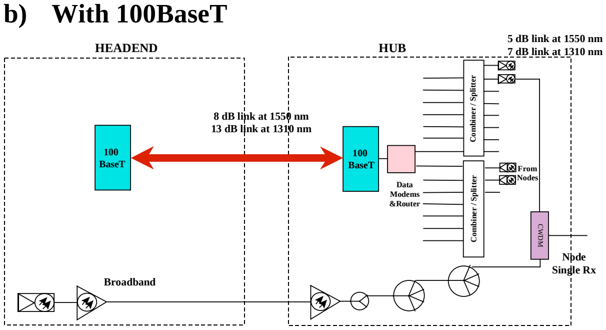
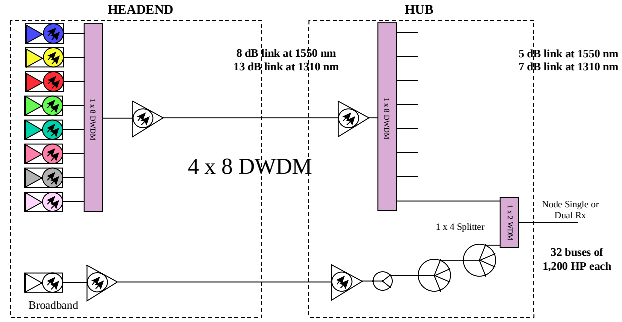
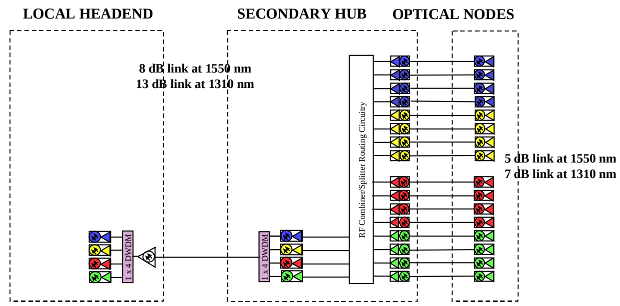


Figure 6: DWDM

a) Forward



b) Reverse



Comparison

When choosing a technology, we have to take into account the network design paradigm but at the same time we have to consider the technology status. The DWDM technology has been considered by the authors as a very desirable solution for some time [1] yet not mature for field deployment with FDM signals in cable TV environment. During the last couple of years, the technology matured and became practical. The DWDM has been applied first to the baseband digital transport and systems with 40 and 64 wavelengths multiplexed in telecommunications backbone links are commercially available. Our industry has been using this technology in regional and primary hub interconnects for the last two years. The scalability and flexibility of this technology makes it almost a perfect match for the HFC network. The advantages of this technology are compared against characteristics of the remaining alternatives in Table 3.

Table 3: Comparison between Technologies

Feature	Space Division Multiplexing	FDM	SONET/100BaseT	DWDM
Transparent	Fully transparent, only O/E repeaters at SH	Partially transparent, FDM equipment at SH	Partially transparent, RF/SONET interfaces at SH	Fully transparent, all optical network PH-to-Node
Entrepreneurial	Services added at any time (O/E repeaters to install)	Services added with little upgrade at SH (for further segmentation)	Only new service equipment added at SH (if interfaces are available & standard)	Services added with little upgrade at SH (for further segmentation)
Future Proofed	Basic network is service independent, problem with reverse (long distances)	Problems with frequency conversion, locked into frequency bandwidth in forward and reverse	In predictable future	Basic network is service independent, fully flexible frequency allocation in forward and reverse, flexible reverse/forward split
Comments	Very high initial cost and full cost, redundancy impractical	Moderate cost, partially scalable (high fixed cost)	The most cost-effective, only partially scalable (very high fixed cost)	Moderate cost, highly scalable, steep cost curve; allows for further segmentation with frequency conversion in the nodes

The quick review of Table 3 clearly indicates that DWDM should be the preferred choice for secondary hub rings in HFC networks. It provides the required level of segmentation for targeted services with limited requirements for fiber. It matches the advantages of the home-run architecture while avoiding its pitfalls of high fiber counts and related to it problems with providing redundancy. The next two tables add to the comparison of the three secondary hub technologies.

Table 4: Comparison of Positives

Desirable Feature	FDM or Block Conversion	SONET	DWDM
Low cost interface to RF	+		+
Many vendors		+	+
Standard system		+	+
Standard network management		+	
Limited number of fibers required	+	++	++
Interfaces with digital systems	+	+	+
Good reliability record		+	+
Survivability	+	++	+
Same system for forward and reverse		+	
Drop/add capability		+	under development

Table 5: Comparison of Negatives

Undesirable Feature	FDM or Block Conversion	SONET	DWDM
Possible problems of instability	✔		
Potential of becoming obsolete	✔		
Potential of becoming single-vendor product	✔		
Fixed system frequency bandwidth split between broadcast and targeted services	✔		
High cost of RF interfaces		✔	

Cost/Scalability

Besides comparing qualitative characteristics of the technologies, the authors prepared a case study to compare the cost of these alternatives. The following system was analyzed:

1) One primary hub feeding 160K homes passed;
2) Four secondary hubs with 120K homes passed (2 with 40K homes and 2 with 20K homes) configured in a ring (optional);
3) Distances:
   - ⇒ 14 miles from primary to 40K secondary hubs,
   - ⇒ 26 miles from primary to 20K secondary hubs (12 miles from 40K secondary hubs to 20K secondary hubs),
   - ⇒ 6 miles between 20K secondary hubs (to close the ring)
4) Optical nodes off each secondary hub with 1500HP per node and three buses of 500HP per bus;
5) The following scenarios were analyzed:
   - ⇒ low segmentation case (A) without redundancy in the secondary hub ring (today's prices),
   - ⇒ low segmentation case (B) with redundancy in the secondary hub ring (today's prices),
   - ⇒ high segmentation case (C) with equipment prices at initial level and without redundancy in the secondary hub ring,
   - ⇒ high segmentation case (D) with equipment prices at initial level and with redundancy in the secondary hub ring,

⇒ high segmentation case (E) with price projection for 3 years and with redundancy in secondary hub ring.

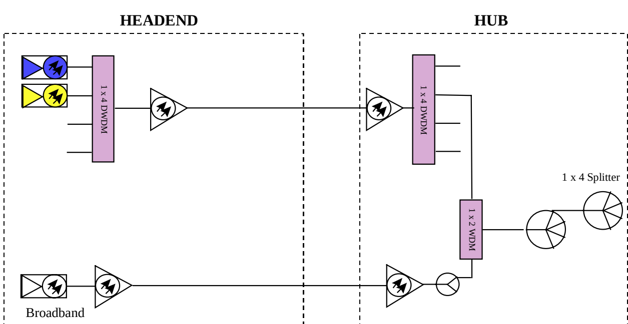
The equipment prices were collected from two vendors that have the DWDM systems ready for deployment. Table 6 summarizes the results of the analysis. The shaded areas indicate recommended deployment strategy.

Table 6: Cost per Home Passed

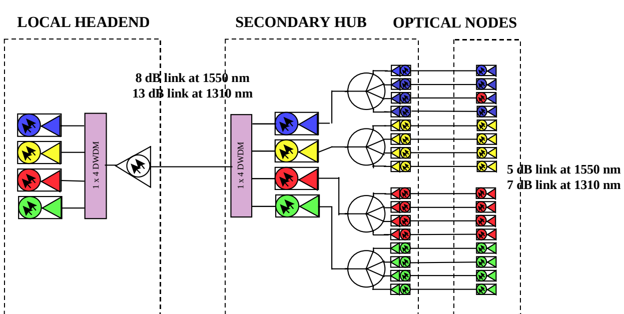
Technology	A	B	C	D	E
FDM with frequency conversion	$11.48	$15.21	$22.71	$29.07	$24.38
Hybrid with SONET	$15.59	$16.73	$24.35	$25.32	$21.91
Hybrid with 10BaseT	$14.07	$20.45	$20.48	$29.39	$27.47
DWDM with 4 wavelengths/fiber	$8.93	$14.28	$34.93	$45.54	$29.83
DWDM with 8 wavelengths/fiber	$8.95	$12.56	$35.37	$42.52	$26.67

Figure 7: Scaled Down DWDM Configuration

a) Forward



b) Reverse



The results indicate the high scalability of the DWDM technology (see Figure 7). Even at today's prices for DWDM elements, the cost per home passed for this technology is significantly lower than the cost of any other technology. The DWDM element prices are on a very steep part of the price curve, which resembles the situation with 1310 analog technology between 1991 and 1995. Within these four years, prices for 1310 analog systems dropped by more than 50%.

The comparison between technologies was performed for a similar capacity/home passed provided by each technology. Table 7 compares the capacity for the technologies for low and high segmentation cases. As presented in [4], the capacity provided at the low segmentation scenario should be sufficient at very high HFC resource usage level. Although the segmentation analyzed in scenarios C through E is provisioned by the network design, fiber count, and node location and configuration, the resources provided by the high level of segmentation will not be required for several years.

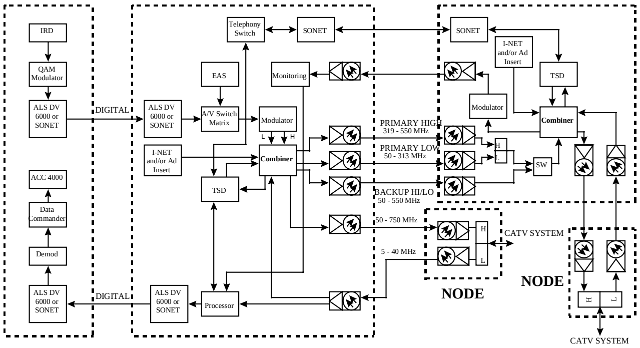
Table 7: Network Capacity per Home Passed

	Capacity (kbps)							
	64QAM/QPSK				256QAM/16QAM			
	Fwd	R	Fwd	R	Fwd	R	Fwd	R
Technology	Low segmentation		High segmentation		Low segmentation		High segmentation	
FDM	16	5	162	54	20	9	204	108
Hybrid with SONET	21		167		21		167	
Hybrid with 10BaseT	20		200		20		200	
DWDM with 4 wavelengths	22	5	216	54	27	9	272	108
DWDM with 8 wavelengths	22	5	216	54	27	9	272	108

To achieve the capacity required while using SONET or 100BaseT transport, caching was assumed at secondary hubs to stop 75% of the traffic generated in the secondary hub area. This will further increase the cost of deployment for these two technologies since caching will have to be deployed in very early service implementation stages. This deployment will lower the efficiency of caching (the same information will be cached at many locations).

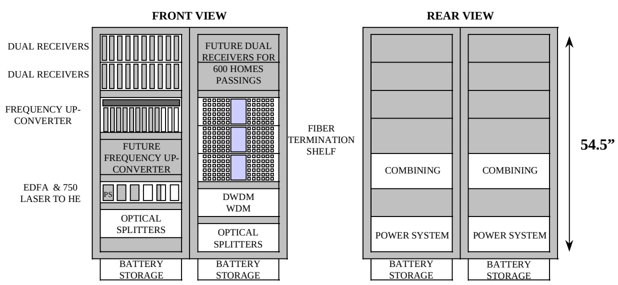
The data in Table 6 indicates that the SONET based architecture has some cost advantages at the high segmentation level and with full redundancy (low incremental cost of redundancy due to inherently redundant SONET transport systems). However, high scalability of the DWDM technology and the extremely high operating cost of secondary hubs, with signal processing equipment and SONET transport installed there, make the DWDM technology extremely attractive. Figure 8 depicts the complexity of SONET based network at the secondary hub level.

Figure 8: SONET Based Network



The DWDM equipment, on the other hand, requires little maintenance and little space. The secondary hub DWDM equipment, scalable to feed 40K homes, fits comfortably into 3x4x5 feet air-conditioned enclosure (see Figure 9).

Figure 9: DWDM Secondary Hub Equipment



Technical Challenges

Before a transmission technology can be deployed, one must analyze all possible impairment sources, design the testing scenarios to test each source independently (separation of impairments), and collect data to prove the technology or to indicate its shortcomings and means to correct them.

The review of technical publications and the discussions with engineering R&D teams from vendors resulted in a list of the following possible impairment sources in DWDM transmission systems:

1. Nonlinear Effects:
    - 1.1. stimulated Raman scattering (SRS),
    - 1.2. stimulated Brillouin scattering (SBS),
    - 1.3. self-phase modulation (SPM),
    - 1.4. cross-phase modulation (XPM),
    - 1.5. four-photon mixing (FPM),
    - 1.6. linewidth dependent frequency roll-off;
2. Linear Effects:
    - 2.1. CSO caused by laser chirp and fiber dispersion,
    - 2.2. PM-IM conversion on amplifier tilt and filter slope.

Nonlinear effects in fiber are often interdependent. Moreover, an effective correction of one problem may lead to another problem becoming a dominating contributor. Examples of such interdependency are:

1. SBS mitigation may lead to noise floor rise at higher frequencies due to PM-IM conversion,
2. Four-photon mixing is lower when the laser chirp and the fiber dispersion are higher, on the other hand higher chirp and dispersion result in CSO.

The theoretical analysis indicates that FPM and SBS can be disregarded for the fiber type (dispersion) and power levels present in the DWDM system. Moreover, SPM considerations for DWDM system with digital signals are no different than the same considerations for analog single wavelength system.

SRS and XPM will result in a crosstalk between the same channels carried on different wavelengths. Both effects increase with the total optical power in the fiber. SRS effect increases with the difference between wavelengths. Fortunately, both effects affect only the RF channels shared between the wavelengths.

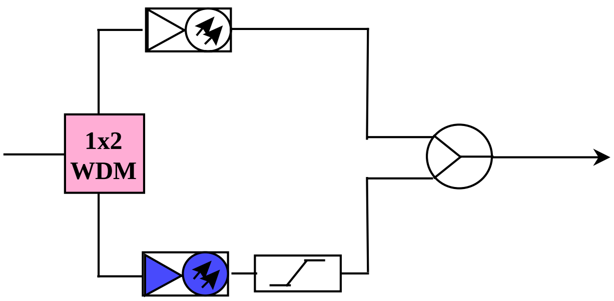
The last nonlinear effect, frequency roll-off, should not be a contributing factor but should be monitored during testing.

Linear effects may cause higher level of impairments and may be more difficult or costly to control. The most visible should be CSO caused by laser chirp and fiber chromatic dispersion. The control of the CSO level is possible by controlling either the sources (for example by using narrow linewidth external modulated lasers or non-zero dispersion shifted fibers) or the effects (by using dual receiver and clearly defined frequency assignment with adequate filtering, see Figure 8 for details). However, the means to control it may be too costly, impractical, or limiting the flexibility of the system operator.

The PM-IM conversion effects (such as higher noise floor at higher frequencies and CSO) should be also monitored during the testing process.

Besides controlling the impairment level, the system designer must define the alignment process (optical power levels at different points of the system, optical modulation index, etc.) that would allow meeting the performance requirements.

Figure 10: CSO Elimination in Dual Receiver



- **Dual RX**
  - ✓ **Broadcast F1-F2**
  - ✓ **Narrowcast F3-F4**
  - ✓ **F4-F3<F3**
  - ✓ **F3>0.5*max(F2,F4)**

Performance Requirements

Any new technology introduced to the HFC network should provide at least the same or equivalent level of performance as the technology being replaced. This approach allows for preserving the remaining sections of the HFC plant without a major redesign or upgrade. The following performance requirements were established for DWDM technology as equivalent to the performance of the complete optical link(s) between primary hub and the nodes (usually two analog 1310 nm links cascaded):

Forward:

1) Broadcast (@ node receiver output) for 50-870 MHz/82 analog channels:

   ⇒ CNR≥51.5 dB,

   ⇒ C/CTB≥65 dB,

   ⇒ C/CSO≥63 dB;

2) Narrowcast (@ node receiver output):

   ⇒ OOB C/N+I≥50 dB,

   ⇒ IB C/N+I≥46,

   ⇒ Flexible frequency allocation;

3) Combined (@ node receiver output):

   ⇒ Broadcast for digital signals at -10 dBc in respect to analog channel equivalent levels:

      i) CNR≥51 dB,

      ii) C/CTB≥65 dB,

      iii) C/CSO≥63 dB;

   ⇒ Broadcast for digital signals at -6 dBc in respect to analog channel equivalent levels:

      i) CNR≥50.5 dB,

      ii) C/CTB≥65 dB,

      iii) C/CSO≥63 dB;

   ⇒ Narrowcast:

      i) C/N+I≥40 dB.

Reverse:

1) RF:

   ⇒ CNR≥40 dB over temp. range,

   ⇒ ≥15 dB DR, 7 dB optical loss at 1310 nm plus 8-10 dB optical loss at 1550 nm,

   ⇒ level stability within ±1 dB;

2) BER performance:

   ⇒ ≤10^(-9) over operating range,

3) Reliability:

   ⇒ 15 years of MTBF,

   ⇒ redundancy for the key shared elements.

Test Results

The technology was tested in several stages:

1) R&D tests were performed by two vendors,

2) Technology feasibility test was performed in the lab environment with all system elements,

3) System tests were conducted on complete forward and reverse systems, and included thermal cycling.

The next stage (in April) will be performed in the field during pilot implementation of the technology.

All the results collected so far support the theoretical analysis. The main concerns were related to optimizing the alignment to achieve adequate CNR after combining and to minimizing CSO caused by laser chirp and fiber chromatic dispersion.

Second order distortions will introduce some limitations to frequency allocation. These limitations will disappear with an advance of low-cost directly modulated lasers (expected in the third quarter of 1998). In the interim, directly modulated lasers with low chirp (≤100 MHz/mA) will provide adequate performance as long as:

- the number of channels is limited (to 10 channels),

- the channels are placed above the middle frequency of the forward operating bandwidth, and
- the difference between the lowest and highest frequency of these channels is kept to a minimum.

The last limitation is not a critical one but will allow for second order intermodulation noise to fall below the forward bandwidth or at very low frequencies (CSO caused by chirp and dispersion at these frequencies is lower). Alternatively, the targeted signal channels can be randomly distributed to avoid multiple beats (accumulated intermodulation noise) at any particular analog channel. The maximum number of beats at any particular channel with 10-channel load will not exceed five beats.

Figure 11: Second Order IM Noise[1]

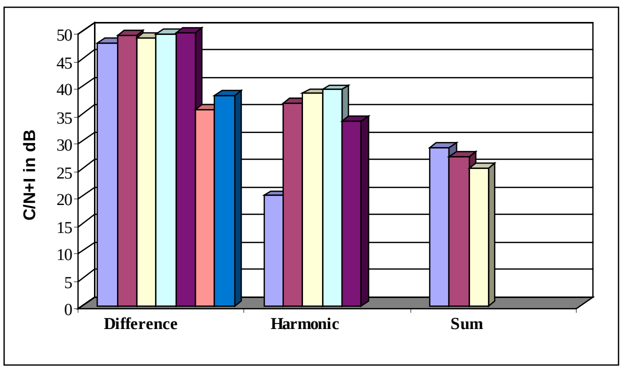


Figure 12: Narrowcast Signal Impact[2]

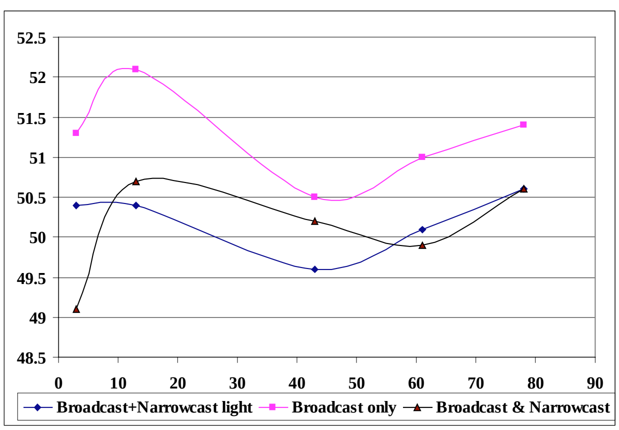


Figure 11 indicates that the second order difference products can be controlled as long as their frequency is below 200 MHz, even with lasers of 300 MHz/mA chirp. The second harmonics and second order sum products are relatively higher even at as low frequencies as channel 2. Placing the targeted service channels above the middle frequency would place these distortions above the highest operating frequency.

Figure 12 illustrates the impact of the second order distortions on the C/N+I. In the test case presented, the targeted signals were placed to produce four difference beats in channel 2. Even in this case (with lasers of 300 MHz/mA chirp and the alignment far from optimal), the performance was quite acceptable.

Alignment and CNR optimization process. Figure 13 shows the sensitivity of the target signal performance to the alignment parameters. The optimal choice of optical modulation index and receiver input power is crucial. For a constant RF output level (constant product of optical input power and OMI), the input power can be optimized to lower the impact of the target service signal laser RIN and fiber RIN on the total CNR.

Figure 13: Alignment Optimization[3]

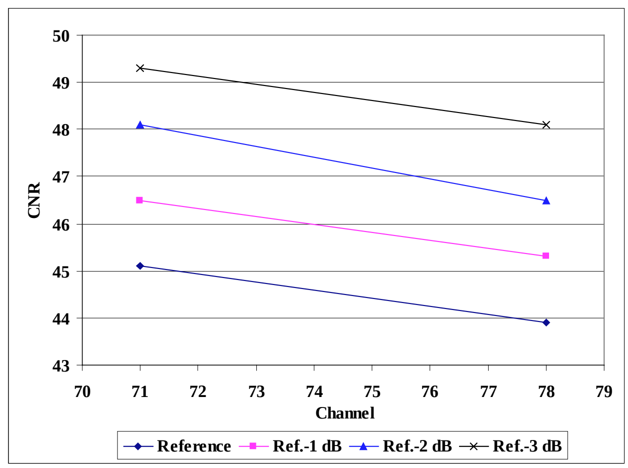
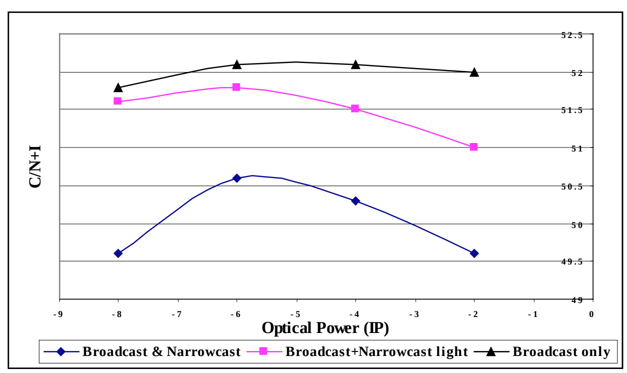


Figure 14: C/N+I Performance @ Channel 2[4]

a) Targeted Signals 10 dB Lower than Analog



[1] Courtesy of Antec Network Technologies.
[2] Courtesy of Antec Network Technologies.
[3] Courtesy of Antec Network Technologies.
[4] Courtesy of Harmonic Lightwaves

b) Targeted Signals 6 dB Lower than Analog

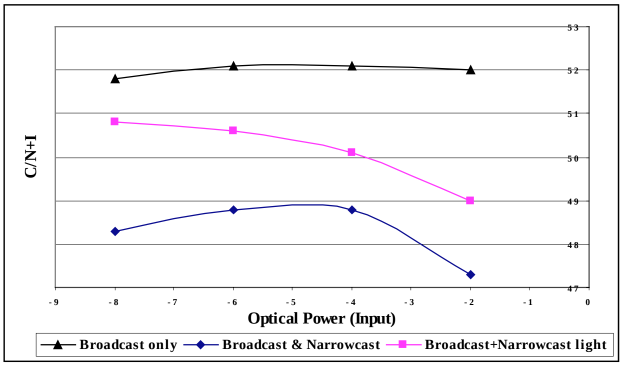
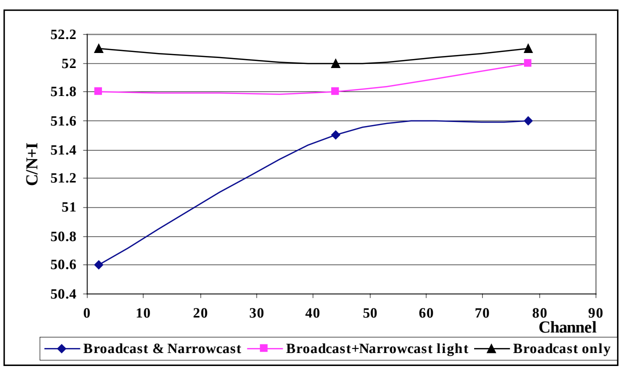


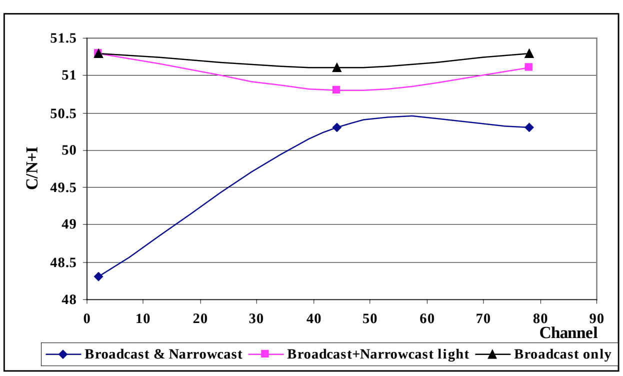
Figure 14 presents the alignment optimization process for targeted signals lower by 10 and 6 dB from the analog broadcast signals. The optimization was performed on the worst channel (channel 2 with second order intermodulation noise). The optimization for the 10 dB lower signals yielded very good results for wide range of optical input levels (CNR≥50 between –7.3 and –3 dBm). The alignment process for 6 dB lower signals achieved 49 dB CNR (worst case channel) for optical input level of –4.7 dBm.

Figure 15: CNR at Optimal Setup[5]

a) Targeted Signals 10 dB Lower than Analog



b) Targeted Signals 6 dB Lower than Analog



[5] Courtesy of Harmonic Lightwaves

Figure 15 illustartes the C/N+I performance of the WDM system afetr the optimization. The graph is misleading since only a few first channels were affected by second order intermodulation noise. All channels above channel 4 met the required 51 dB CNR for targetd signals lower by 10 dB.

Similarly for 6 dB lower targeted signals, all channels above channel 4 approached the required 50.5 dB CNR within the measurement error.

Reverse path testing yielded all the performance required without a significant optimization effort. The only potential risk was related to second order intermodulation noise caused by laser chirp and chromatic dispersion. The test results proved that the system had a performance safety margin.

CONCLUSIONS

The theoretical analysis of the DWDM systems and the testing performed during the last six months proved that the technology is mature for field deployment. The advantages of this technology over any other technology deployable in the secondary hub rings and the affordability and high scalability of this technology, makes it the most desirable alternative for the transport system in this section of the HFC network.

The most difficult to control impairments in the transport system based on this technology is related to second order intermodulation noise. The means to control this type of impairments are available today. However, the most effective ones are expensive or impractical. As long as the frequency allocation is reasonably managed, the second order intermodulation noise can be maintained at low levels with directly modulated lasers of low chirp.

The other major challenge is related to the CNR optimization. The understanding of the technology gained during the testing sessions allows for achieving the optimization during the designing stages in the same way that applies to designing 1310 and 1550 nm optical links.

The outcome of this activity is the authors' conviction that the technology offers major advantages in the secondary hub ring in HFC networks and is field deployable today.

ACKNOWLEDGEMENT

The authors wish to thank John J. Kenny, Dogan A. Atlas, and the lab personnel from Antec as well as David Piehler, John Trail, Xingyu Zou and the lab personnel from Harmonic Lightwaves for sharing the results and spending long hours with the authors on testing the DWDM systems. Further, the authors want to express their appreciation to Mani Ramachandran of Optical

Transmission Labs for sharing his experience on technical challenges of the DWDM technology and for discussing the testing framework.

REFERENCES

(1) Thomas G. Elliot and Oleh J. Sniezko, Transmission Technologies in Secondary Hub Rings — SONET versus FDM Once More, 1996 NCTA Technical Papers.

(2) Oleh J. Sniezko, Signal Transport Technologies in Modern HFC Network, 1996 SCTE Engineering Conference.

(3) Oleh J. Sniezko, Video and Data Transmission in Evolving HFC Network, 1998 OFC Conference.

(4) Dan Pike, Tony E. Werner, On Plant Renewal Strategies, 1998 NCTA Technical Papers.

(5) Tony E. Werner, Regional and Metropolitan Hub Architecture Considerations, SCTE 1995 Conference on Emerging Technologies.

(6) E. E. Bergman, C. Y. Kuo, and S. Y. Huang, Dispersion-Induced Composite Second-Order Distortion at 1.5 μm, IEEE Photonics Technology Letters, Vol. 3, No. 1, January 1991.

(7) Rolf Heideman, Berthold Wedding, and Gustav Veith, 10-GB/s Transmission and Beyond, Proceedings of the IEEE, Vol. 81, No. 11, November 1993.

(8) Koji Koshima, Using Equalizers to Offset the Deterioration in SCM Video Transmission Due to Fiber Dispersion and EDFA Gain Tilt, Journal of Lightwave Technology, Vol. 10, No. 10, October 1992.

Oleh J. Sniezko
Vice President of Engineering
TCI-Communications Inc.
5619 DTC Parkway
Englewood, CO, 80111, USA
(303) 267-6959

Oleh in his capacity of the Vice President of Engineering for TCI Communications, Inc is responsible for networking technologies and network architectures, equipment selection and approval, engineering policies and standards, and overall engineering activities related to inside and outside plant.

For his achievements in implementing fiber optic technology, Oleh received the 1997 Polaris Award sponsored by SCTE, Corning, and CED.

He authored a series of documents, papers and presentations published in industry publications.

Tony E. Werner
Executive Vice President of Engineering and Technical Operations
TCI Communications Inc.
5619 DTC Parkway
Englewood, CO, 80111, USA
(303) 267-6950

Tony Werner in his capacity of the Executive Vice President of Engineering and Technical Operations for TCI-Communications Inc. has overall responsibility for network architecture, product testing and approval, network upgrades, purchasing and the introduction of new services.

Tony has published numerous technical papers and is a frequent speaker for CableLabs, the NCTA and the SCTE.

Low-Cost Implementation of Ethernet-based Services over CATV Interconnect Networks

Robert W. Harris
ADC Broadband Communications

Michael Coden
ADC Codenoll

Abstract

A low cost solution for implementing Fast Ethernet data transport over local backbone and long distance regional CATV networks is presented. The approach allows switching and transmission of standard Ethernet and Fast Ethernet data in its native mode over any size ring network with no limitation to distance. Furthermore, this method is independent of both data format and protocol allowing service providers to deliver fully transparent Ethernet-based services to their customers. This approach, in effect, extends the functionality and low cost of the local area network (LAN) into the Wide Area Network (WAN).

This paper describes the theory of operation of a new "ring-based" Fast Ethernet switch and transport solution known as EtherRing™. Implementation of EtherRing in cable data modem applications is presented as an alternative to current data transport solutions. EtherRing makes use of the data networking concept known as "Route Once-Switch Many" and is extended into the CATV network environment as a method for centralizing network routers. This solution, when used in conjunction with EtherRing transport technology, is shown to provide a highly scaleable and efficient approach to data modem service delivery resulting in network cost savings between 50% to 80% over alternative data transport methods. Additional CATV data transport applications are presented including; advertisement insertion transport, set-top box access and control transport and general purpose LAN service delivery.

INTRODUCTION

Computer networks have become commonplace in large and small businesses, universities, government facilities and other organizations. With the advent of Internet access, advertisement insertion, video on demand, IP-video, IP-telephony and other emerging CATV applications, computer networking principles and techniques are fast becoming an integral part of the CATV network. By definition, computer networks allow a number of users or devices to share data and resources such as file servers, data storage systems, printers, switches, routers, modems and other peripherals. The CATV industry will need to interface to these and other IP-based devices such as cable modems, MPEG encoders and network management systems over their networks.

During the past 20 twenty years, a number of methods for connecting computer devices within a network have been devised. The IEEE 802 and ANSI committees have developed a number of standards used for various computer networks. These same standards and practices will be followed as CATV networks deploy data-based services.

Computer networks are often classified in two categories - the Local Area Network (LAN) and the Wide Area Network (WAN). The LAN is characterized as a shared medium where all devices share the network bandwidth. The shared medium can consist of coax, fiber, twisted pair (Category 3 and 5) or any combination of the three. LAN bandwidths typically range from 10 Mb/s for Ethernet to 100 Mb/s for Fast Ethernet and FDDI (Fiber Distributed Data Interface).

The LAN is typically used to transport data between network devices over relatively short distances such as within a building or groups of buildings as in a campus. The maximum distance between devices on a LAN is generally (but not always) limited to several km due to the protocols used within the LAN.

The WAN is used to connect two or more LANs separated by much larger distances - several km to thousands of km. WAN network devices have the function of re-packetizing LAN data packets and provide routing protocols to determine the destination of the transmitted data. The WAN can employ standard telecommunication trunking such as T1 circuits, fractional T1, DS3, ATM, etc. The WAN circuit may be a private link or supported on the publicly switched network. The main advantage of WAN functionality is its support of long distance transport. This, however, comes at the expense of the use of complex protocols, potential traffic congestion and high costs.

There are three basic types of protocols used for transmitting data in LAN networks: Ethernet, Token Ring (or FDDI) and encapsulation. Each of these network protocols have specific advantages and disadvantages and are described below. Following this is a discussion on packet switching and its advantages over routing.

This paper will then introduce a new technology concept, called EtherRing™, presented as a solution for significantly reducing the cost and complexity of data transport within the CATV network. Following this discussion, the concept known as Route Once - Switch Many is presented as a method for consolidating network routers at a single headend location leading to additional savings in network costs. Several CATV network applications incorporating EtherRing are also presented.

DISCUSSION OF CURRENT NETWORK PROTOCOLS

To understand the functional capabilities and limitations of existing network protocols, a brief description of three popular protocol solutions is provided below.

Ethernet

Ethernet is basically a broadcast protocol where its main advantage lies in it simplicity. This allows Ethernet to be implemented with less costly hardware and software. The main drawback with conventional Ethernet is that there are limitations on the physical distance that the network can cover. 10Base-T is limited to 4.5 km while 100Base-TX is limited to 1 km.

Despite its distance limitations, Ethernet has become the most common protocol for LANs due in large part to its low cost, ease of use, low complexity and support of relatively high bandwidths (10 Mb/s for 10Base-T and 100 Mb/s for 100Base-TX, commonly referred to as Fast Ethernet). According to one market report, [1] more than 80% of all networked devices are connected via Ethernet.

For the purpose of this paper, the term Ethernet includes the entire class of Carrier Sense Multiple Access with Collision Detection (CSMA/CD) protocols covered by the family of computer industry standards known as IEEE 802.3. A number of good tutorials covering on the operation and functionality of Ethernet are available from a variety of sources [1,2,3].

Token Ring and FDDI

Token Ring and FDDI, as described in computer industry standards IEEE 802.5 and ANSI XT3.9, respectively, provide the distinct advantage that data can be transported over greater distances relative to conventional Ethernet. Further, Token Ring and FDDI provide virtually equal bandwidth throughout the network. The main disadvantage of both is their use of complex protocols. As a result, Token Ring and FDDI hardware and software costs are significantly higher relative to Ethernet. Further, the loss of a token - which determines what network device can send its data - can cause significant network transport delays as token recovery routines are performed leading to loss potential of critical data.

Encapsulation

Encapsulation protocols have been developed to allow Ethernet packets to be transported over the longer distances covered by the WAN. In such protocols, the entire Ethernet packet is placed within another type of packet with its own header including additional addressing information, protocol information, etc.

These protocols typically suffer from the problem that they may require special higher level protocol information to be included in the data field of the Ethernet packets for the purpose of directing routers in the network. This has the effect of limiting the types of data packets that can be handled and places a significant processing burden on both the network devices generating the packets and the routers used to transmit and receive the packets between the various Ethernet network segments.

These additional protocol elements and restrictions typically require expensive hardware and software be added to an otherwise inexpensive Ethernet network. Further, such protocols typically require the use of manually created address tables for the routers.

PACKET SWITCHING

A packet switch, often referred to as an Ethernet switch, or just a “switch”, is a multiport bridge that simply forwards packets from a device connected on one port to a device connected on another port. The forwarding decision is based on the destination MAC (Media Access Control) address at the head of the packet. A switch will ignore a packet that is destined for a device located on the same port as the source device. Forwarding decisions of the switch are based on link-layer (layer 2 of the open systems interconnection, or OSI, model) information. Routers, on the other hand, forward packets based on network-layer (layer 3) information. Key here is that switches do not modify packets as they pass through, whereas routers must change the packet to include the MAC address of the router at the next-hop and may also increment a hop count field [4].

Each port on a switch may be connected directly to network devices such as servers, routers, printers, PCs, etc. or be connected to completely different LANs operating independently and simultaneously. Only those packets that need to pass from one LAN segment to another are forwarded by the switch. As a result, a multiport switch will often increase the overall bandwidth of a single shared LAN by many times.

Layer 2 switches provide a simple and elegant way to increase the aggregate bandwidth of the network and are often less expensive and faster (higher throughput) than routers. Switches are simpler because they operate at layer 2, do not modify packets and do not require complex routing protocols like Routing Information Protocol (RIP) or Open Shortest Path First (OSPF). Routers often use proprietary and complex routing protocols which typically have to segment and reassemble packets on the fly. Moreover, routers are generally protocol specific requiring different software for each protocol used and require constant maintenance of the routing tables as network parameters change. As a result, the cost to maintain routers on a monthly and yearly basis can be significant.

Switching, on the other hand, is entirely self-learning. This means that the switch automatically “learns” which port (or LAN segment associated with a port) each device is connected to - even if the device is moved to a different port or LAN segment. Each switch port records the source address of every packet, as it receives the packet, in a memory table for that port.

Further, when a packet is received at a port of the switch, the destination address of the packet is compared to the memory tables for the other ports of the switch. When a match is found for the destination address in the tables for one of the ports, the packet is switched to and sent out that port. With broadcast packets, the packet is broadcast to the other ports on the switch but never back to the original receiving port.

Likewise, “multicast” packets using specially reserved destination addresses will be broadcast to a selected group of devices. Therefore, the switch does not require special “management” of memory tables since the process is performed automatically. As a result, the switch is effectively maintenance free compared to router functionality.

Layer 2 switches offer key advantages over routers in other ways that affect network performance and cost. Because they operate at layer 2 (switching by MAC address) instead of routing by network address at layer 3, switches can operate at higher speeds contributing to lower latency and higher throughput. Further, routers are generally 3 to 10 times the price per port of layer 2 switches [5].

One reason why switches provide higher throughput, lower latency and are lower in cost than routers is because the switching functions are often implemented entirely in VLSI (very large scale integration) rather than performed by software running on an expensive high performance processor. Consequently, both the initial cost and the maintenance cost of layer 2 switching will always be less expensive than routing.

Given the above attributes of layer 2 switching, it’s not surprising that organizations often deploy layer 2 switches to “front-end” their routers, off-loading some of the traffic from the router and utilizing much less expensive switch solutions for switching traffic between LAN segments [5]. This concept, used often in the traditional LAN environment, is commonly referred to as "Route Once - Switch Many". In the CATV environment, an extension of this concept can be employed leading to a powerful and economical solution for centralizing most or even all network routers at a main server or headend location and distributing the data service(s) via layer 2 switch technology. Route Once - Switch Many as related to CATV-based data transport applications is discussed in further detail later in this paper.

However, before one can implement Route Once - Switch Many techniques within the CATV network, a number of significant challenges need to be overcome. This is because CATV networks and LANs have significant differences in network topology, network distances and necessary protocols and applications to support. Overcoming these challenges led directly to developing a new technique to deliver layer 2 functionality - with its low cost, low maintenance and higher throughput - over the wide area CATV network while maintaining complete compliance with data networking standards. The solution is known as EtherRing.

EtherRing - A NEW CONCEPT FOR ETHERNET TRANSPORT

Because of the clear advantage of Ethernet switching, there is a strong to desire to extend this functionality and flexibility into larger scale applications. However, transmission over the regional CATV network effectively means transmission over a WAN. As a result a number of challenges exist in delivering native-mode Ethernet over the WAN.

Challenges to Overcome

Where Ethernet topology is basically broadcast in nature, telecommunication networks, such as those used in CATV regional headend consolidation, are typically configured in a ring. In standard Ethernet if a packet is not used, or is a broadcast packet, it travels once through the LAN then fades away into the "Ether". On a ring, the packet will come back to the originating point and can continue circulating the ring indefinitely. Figure 1 summarizes the challenges of transporting native mode Ethernet over telecommunications networks.

Challenge	**10BaseT Ethernet**	**100BaseTX Fast Ethernet**	**Telecom Systems**
Topology	Broadcast	Broadcast	Ring
Maximum Time Delays	51.2 µs	5.12 µs	1,000's of µs
Maximum Distances	4.5 km	1 km	1,000's of km
Protocols	Any and All	Any and All	Any and All

Figure 1. *Challenges for Long Distance Ethernet Transport*

Standard Ethernet is unable to cover the long distances (potentially >1,000's of km) associated with ring telecommunication networks - this has been the job of the WAN. Because of the collision detection scheme in Ethernet, the maximum time an Ethernet packet can take to traverse an entire network is 5.12 µsec for 100Base-TX (51.2 µsec for 10Base-T). This maximum time restriction translates into a distance restriction (based on velocity of propagation through the medium) of 1 km for Fast Ethernet and 4.5 km for 10 Mb/s Ethernet.

The alternative currently in use is routing, which, although powerful, can have a number of limitations listed earlier such as; lower throughput, protocol dependence and higher hardware and maintenance costs relative to Ethernet switching. Ideally, a data transport system would be protocol independent so that the operator can connect Novell Netware LANs, TCP/IP LANs, Netbeui LANs, digital set-top box controllers, network management systems and any other device(s) from any manufacturer using any other protocol all on the same transport platform without requiring expensive software and routing table maintenance.

EtherRing Functional Overview

Through simple modifications of the Ethernet standard [6] (while still maintaining IEEE 802.3 compliance on the local ports) EtherRing allows native mode Ethernet packets to be transported via a ring rather than the typical star configuration.

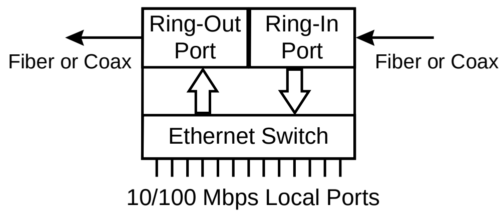


Figure 2. *Functional Block Diagram of EtherRing*

The EtherRing platform consists of three primary components; a Ring-In port, a Ring-Out port and an Ethernet switch. The Ethernet switch is just that - a standard 10Base-T/100Base-TX switch that possesses all of the normal features and functionality found in Ethernet switches. Connected to the switch are Ring-In and Ring-Out ports that may have either a fiber optic or coaxial interface.

Distance Solution. The maximum-distance / collision-time-domain problem is resolved by eliminating collisions altogether on special “Ring-In” and “Ring-Out” ports. By connecting only one Ring-Out port to each Ring-In port, (and vice versa) we ensure that each Ring-In port will only receive packets from one Ring-Out port (see figure 3). Therefore, there are no other possible transmissions to “collide” with the transmissions occurring (simultaneously all around the ring) on each connected pair of Ring-out and Ring-in ports.

The key point is that there can be no collisions on a Ring-In or Ring-Out port since the Ring-In port on any switch will only receive packets from the Ring-Out port of one and only one switch. Therefore, the only time limit on data transmission is that of the higher level software protocols, whatever that may be, but typically it ranges from half a second to several seconds.

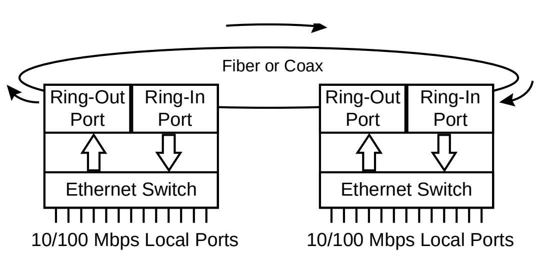


Figure 3. *EtherRing Distance Solution to Eliminate Transport Collision Domain via Ring-In/Ring-Out Ports*

An added benefit resulting from a Ring-In/Ring Out design is a doubling effect of data throughput. Because there is no collision domain on the ring, a full 100 Mb/s is available on *each* In/Out port. Consequently, an effective throughput of 200 Mb/s is available between any two devices.

Protocol Solution. The protocol and unique device software problems are entirely avoided by using MAC address switching. The MAC address is a unique 48-bit number that is built into the Ethernet hardware by every manufacturer, as the device is manufactured, and is completely protocol independent. Therefore, any protocol, such as TCP/IP, IPX, etc., will be switched correctly. Users and system manufacturers are able to install a wide variety of software products on the hardware without concern for incompatibility.

Self-learning Solution. As the packet enters a port of a standard Ethernet Switch, the switch reads the Destination Address and if it finds a matching address in its internal address table, switches the packet to the port on which that device is located. The switch learns the addresses by then reading the Source Address as each packet comes in each port and making entries in the address table that store that MAC address and associate it with the port on which the packet was received. This standard process is applied to the local Ethernet ports in the EtherRing switch and, with simple modification, on the Ring-In and Ring-Out ports. When a packet is received at the Ring-In port, the Source Address is read and entered into the address table. But in this case it is associated with the Ring-Out port instead of the Ring-In port at which it arrived. This teaches the switch that while all packets arrive at the Ring-In port, all devices on the ring must be reached by transmitting on the Ring-Out port. The result is a very simple, low cost and self-learning solution that is, again, maintenance free.

Topology Solution. To solve the problem of the Ring Topology, i.e. that a packet may come full circle around the ring and could continue going around the ring forever, another simple modification is performed on the standard switching technique. This “problem” is a normal part of Ethernet. A packet will not be switched off the ring by a Destination Address match at a local port if it is a broadcast packet, initial packet or packet that does not find a Destination Address match for any other reason at a local port. That packet will come full circle around the ring, back to the switch which has the originating device, and if not dealt with properly that packet would continue to travel around the ring indefinitely.

We solve the Ring Topology problem by performing packet filtering on the Source Address at the Ring-In port instead of the usual filtering on the Destination Address done by a standard Ethernet Switch and done at all the local ports of the EtherRing Switch. In a normal switch, and on the local ports of the EtherRing Switch, if the Destination Address of a packet is already in the table for that same port, we know that the packet is going from one device on a hub connected to the switch port to another device connected to a hub on the same switch port.

Therefore, we filter the packet, i.e. we don't let the packet into the switch because we "know" that the source and destination devices are both on the same port and there is no reason to use up switch bandwidth. In the case of the Ring-In port, if the Source Address of the packet entering the Ring-In port is the same as a source address that has already been entered into the table for a local port, then we know the packet has been full cycle around the ring and we filter it, i.e. we don't let the packet into the switch, thereby removing it from the ring. We know that we can catch all such packets, because the switches (unlike routers) create no packets or data. All packets have to be created originally by one of the Ethernet devices attached to a local port. Therefore, before the packet can get started on the ring, we have learned its Source Address, put that address in the table and associated it with the appropriate local port. When the packet comes around it gets caught.

NETWORK CONFIGURATIONS USING THE FAST EtherRing SWITCH

This section describes the basic building blocks used to configure data networks based on the EtherRing standard.

Stand-alone EtherRing

To build the most elementary EtherRing-based network, a series of Fast EtherRing switches are connected together to create a simple ring as shown in figure 3. Customers can then be served at each location via the local 10Base-T and 100Base-TX ports. The maximum distance between each EtherRing switch is determined only by the optical components being used. In this case, either 1310 nm or 1550 nm optics can be used with a loss budget up to 25 dB.

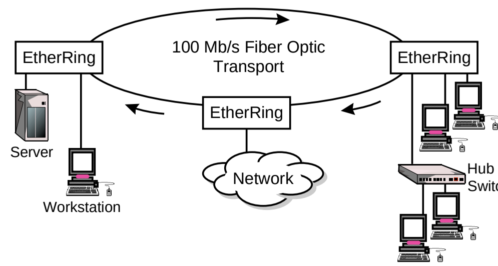


Figure 3. *EtherRing as a Stand-alone Solution*

Multiple EtherRing Elements within a High-speed Transport

The Fast EtherRing Switch has also been designed to interface directly to a high-speed multichannel digital fiber optic transport platform. This type of high-speed digital platform is used extensively [7,8] within the CATV industry for the purposes of distributing video, QAM and data services to local hub sites from a regionalized headend.

This design allows up to 16 separate and independent 100 Mb/s EtherRing channels to be multiplexed onto a 2.4 Gb/s transmission system (see figure 4). Further, using Dense Wavelength Division Multiplexing (DWDM) techniques, up to 144 separate and independent Fast EtherRing channels may be combined on a single fiber. This is accomplished by using 8 wavelengths within the 1550 nm window along with another wavelength at 1310 nm where each wavelength carries a 2.4 Gb/s aggregate data rate. This solution addresses the significant issue of scalability as network demand and subsequent traffic load increases.

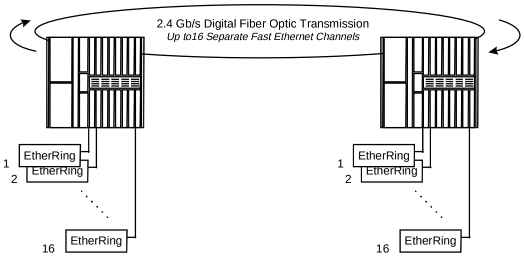


Figure 4. *16 EtherRing Channels as Part of a Higher Speed Multiplex*

The trunking method also allows multiple Fast EtherRing channels to be delivered to hub site locations where individual EtherRing channels are broken out as single channel tributaries (via direct fiber optics). This allows data delivery to remote hub sites that may not have the subscriber density to support the cost of a multichannel Ethernet distribution system - let alone support its own router. Further cost savings are realized as the EtherRing bandwidth is shared over the several sites. As shown in figure 5, EtherRing channel number 16 is not only distributed via fiber from the multichannel shelf but its bandwidth is also shared over several remote sites.

As a final note, these types of high-speed digital trunking systems allow for complete opto-electronic and fiber path redundancy to achieve maximum reliability. And, the individual EtherRing units may also be configured with optical path and terminal redundancy when used as a stand-alone solution or when integrated into the multichannel digital trunk.

With the development of EtherRing, and its use in both stand-alone and multichannel trunking architectures, the low cost, low maintenance, high throughput, high bandwidth, native-mode Ethernet technology normally associated with the LAN environment is now extended into the WAN. Armed with this powerful solution, CATV operators are able to take advantage of cost saving techniques performed within a LAN. Consequently, the CATV network can now realize significant cost savings through the application of "Route Once - Switch Many" thereby consolidating network routers at a single location and allowing network bandwidth to be shared over multiple hub sites.

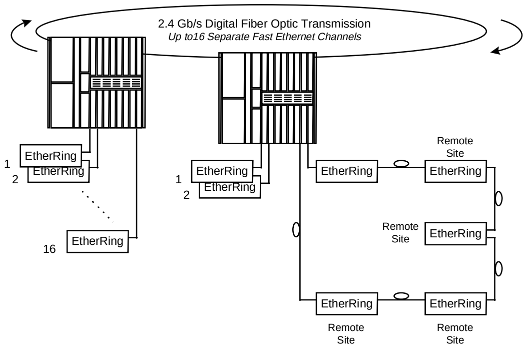


Figure 5. *Multichannel EtherRing Transport with Single Channel Fiber Tributaries*

ROUTE ONCE - SWITCH MANY

A common phase heard within the data networking industry is "switch when you can, route when you must." In other words, the deployment of layer 2 switches is the preferred mechanism for linking LANs, servers and workstations while deploying routers only when necessary. Prior to the advent of Ethernet switches, the generally accepted practice was to use routers to segment congested LAN segments or to link LANs to one another in a building or campus environment [5].

While standard Ethernet switches provide the level of sophistication (self-learning and protocol independence), these devices have been limited in distance due to the Ethernet definition. EtherRing effectively extends the functionality and flexibility associated with the LAN into the wide area network. Therefore, the Route Once - Switch Many technique, traditionally used only in the LAN, can now be extended into the WAN resulting in the consolidation of most, if not all, network routers in a single location. See figure 6a and 6b.

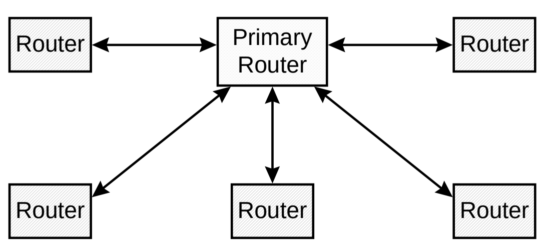


Figure 6a. *Routers Distributed using Current WAN Solutions*

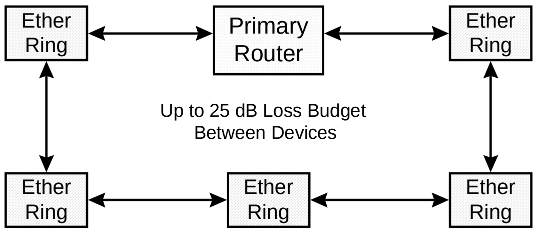


Figure 6b. *Routers Centralized using EtherRing for WAN Transport*

Several key outcomes of this approach follow. First, an initial cost savings is incurred directly through the replacement of higher cost routers with lower cost Ethernet transport leading to an overall lower initial installed cost. Bandwidth utilization is optimized as well by sharing the EtherRing bandwidth among several or more sites.

As network demand increases, the EtherRing network can be easily reconfigured at the headend to allow more or less bandwidth at each hub site. Further, with centralized routers and servers, all key personnel with the necessary expertise can be located at a single main headend site.

Current Data Transport Solutions over the WAN for Cable Modem Services

Data has traditionally been transported over the WAN using conventional routers along with some type network interface which may include; T1, DS3, OC-3c and others. In these cases the network circuits are dedicated point-to-point links. Depending on the type of network and network interface being used to transport the data service, an operator may have too much (OC-3c) or not enough (T1) bandwidth directed to the destination site.

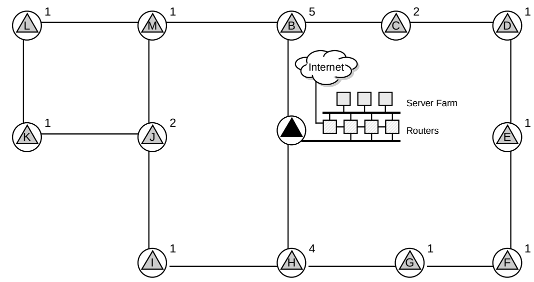


Figure 7. *Regional Headend Interconnect with Required Number of CMTS Units*

Figure 7 shows a typical regional headend interconnect. Based on subscriber count, homes passed, demographics, etc., each hub site is allocated a certain number of Cable Modem Termination Systems (CMTS). This value is labeled near the hub site symbol. In this example, each CMTS requires a 30 Mb/s data channel. The CMTS performs the digital to RF conversion (and vice versa) for access to and from the HFC network.

While some sites are expected to support multiple CMTS units, others will require only one. In fact, none of the multi-CMTS hub sites will likely require and deploy all CMTS units as data modem services are initially being implemented.

Figure 8 shows a possible data transport solution using current methods. The data signal for each CMTS unit is carried in a single OC-3c channel. As a result, an OC-48-based network is used to support multiple OC-3c circuits. Distributed routing must also be used because of the transport of encapsulated data over the network requiring subsequent re-packetization and assembly of the data channel prior to hand-off to the CMTS unit.

The network shown in figure 8 has an initial installed cost of approximately $1.1M. Note that the annual maintenance costs associated with supporting the functions of each router is not included.

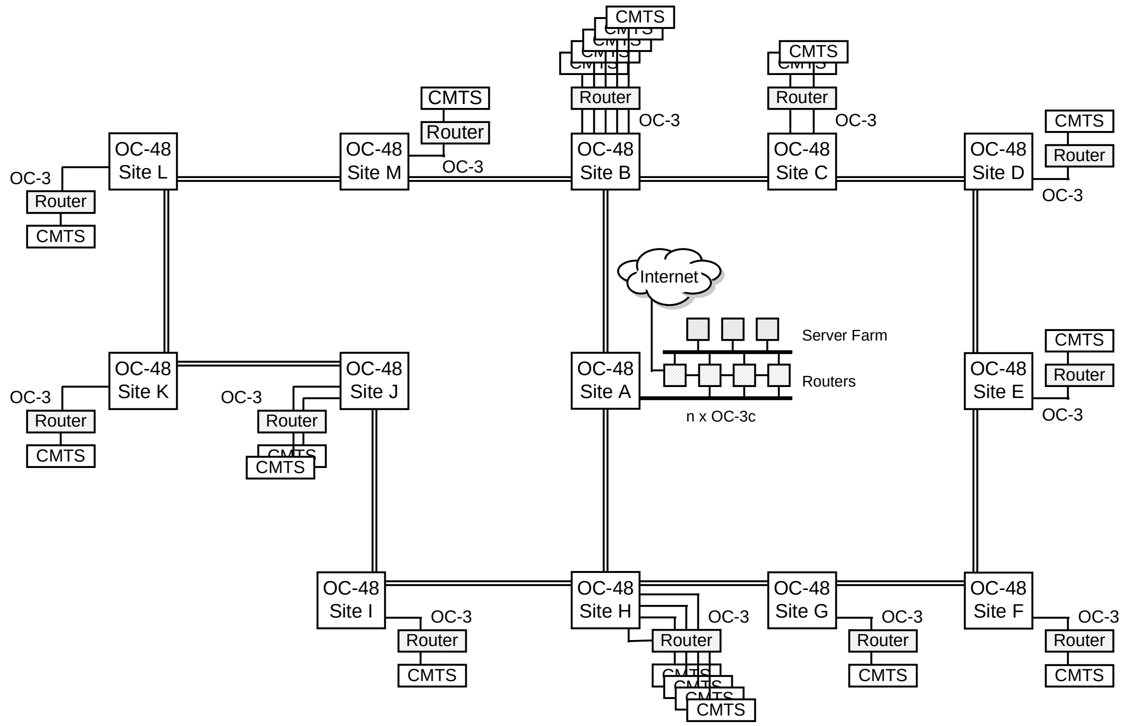


Figure 8. *Current Approach to Data Delivery over the Regional CATV Network*

As this design requires each hub be interconnected via OC-48, an extraordinary waste of network bandwidth results. For example, 21 CMTS units are served - each by a dedicated OC-3c on the OC-48 network. This results in a total of 630 Mb/s delivered on a 2.4 Gb/s platform.

Route Once-Switch Many as Applied to Cable Modem Data Transport

When EtherRing is applied to the same network (figure 9) dramatic cost savings and operational efficiencies are achieved. First, all routers previously located at the distribution hub sites have been eliminated as router functionality in now centralized within the main headend. In place of routers, Fast EtherRing switches are used. This leads to a direct hardware cost savings. Rather than an OC-n transport, a combination high-speed digital trunk along with single channel tributaries are employed. The EtherRing-based network as shown in figure 9 has an approximate cost of $500k.

This implementation takes advantage of both the statistical nature of Ethernet as well as the increased throughput (200 Mb/s) of EtherRing. Consequently, hub sites initially requiring only one or two CMTS units can now share the same EtherRing bandwidth over several sites resulting in a significant cost savings. Further, this provides an optimal solution for network scaleability where initial service demand may be low.

OTHER CATV TRANSPORT APPLICATIONS USING EtherRing

Since it’s based on the Ethernet standard, an EtherRing solution may also be extended to other data transport uses within the CATV environment. These applications may include both high bandwidth as well as low bandwidth usage. A brief review of these applications follow.

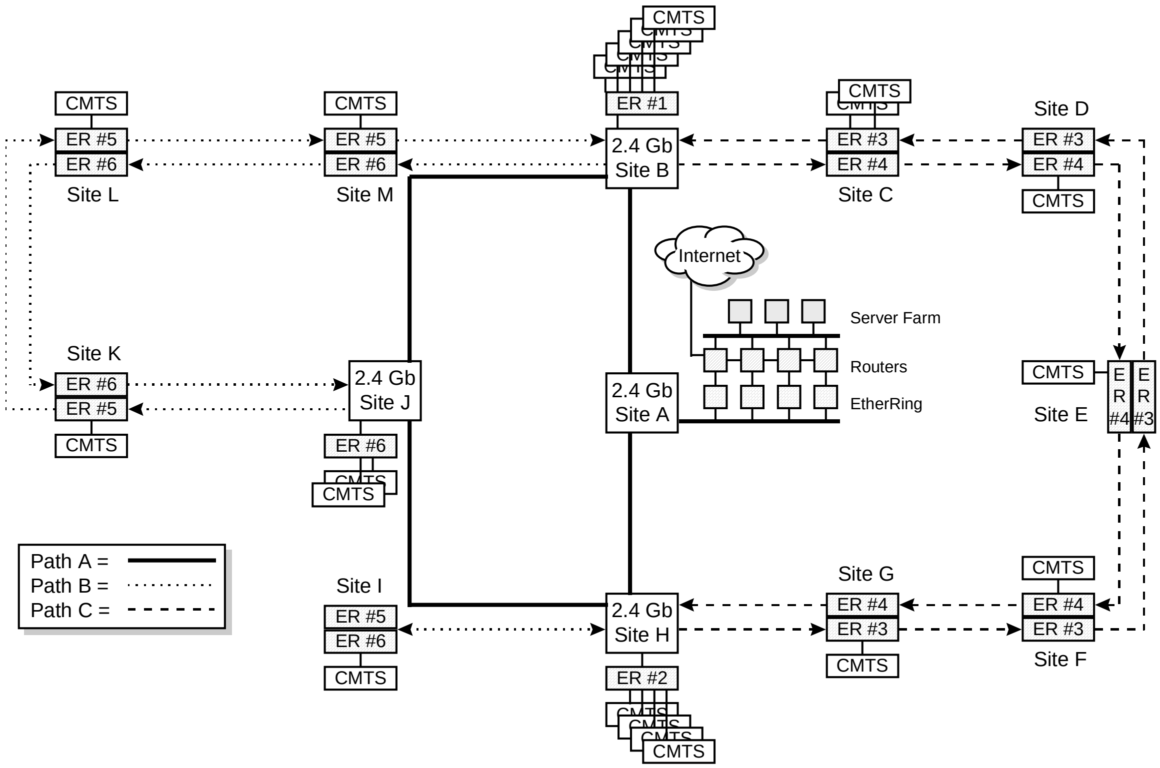


Figure 9. *EtherRing Approach to Data Delivery over the Regional CATV Network*

Advertisement Insertion

Advertisement insertion systems store MPEG compressed video files on a primary server. These files are typically delivered to local servers located within Hub sites over low-speed circuit connections such as T1. Multiple Hub sites require multiple leased circuits costing thousands of dollars per year in usage charges. Further, T1 speeds (1.544 Mb/s) are a potential bottleneck when multiple large (~10 Mb/s) files are transferred.

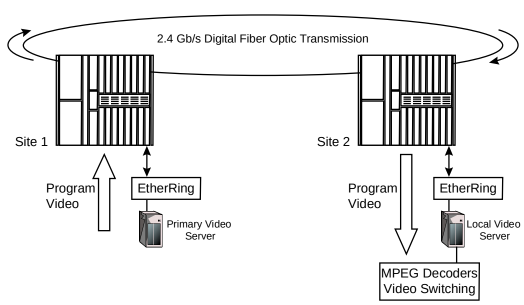


Figure 10. *Advertisement Insertion using EtherRing Transport*

A solution based on EtherRing (see figure 10) offers a simple, low-cost, high bandwidth method for transferring MPEG data files to local Hub site servers. Program channels slated for ad insertion (CNN, MTV, etc.,) may be transported in *uncompressed* format on the same platform as the compressed MPEG files allowing further network cost and operational efficiencies. A native mode Ethernet-based delivery system also offers a significant advantage not offered in current solutions - the delivery of an acknowledgment back to the Headend as the advertisement is inserted. This acknowledgment is logged into a central database thereby giving the operator an accurate record of all ad insertion activity.

LAN Interconnects - For Internal and Revenue Generating Networks

Hub sites often serve as local CATV offices that may support technical, dispatch and CSR staffs. Subsequently, office computers, servers, printers, etc., may be part of an internal LAN that is connected to the main office via a telco WAN connection (56 Kb/s circuit, fractional-T1, etc.). An EtherRing solution (figure 3) allows multiple internal LANs to be interconnected to the central office LAN - completely by-passing the telco and the associated monthly access charges. The same approach may be extended to revenue generating applications for businesses, schools, libraries, etc.

Set-Top Box Access & Control and Network Management Transport

Set-top box manufactures are currently modifying their addressable network interface controllers (ANIC) to interface directly to an IEEE 802.3 10Base-T connection. Likewise, network management platforms are migrating to network control over Ethernet connections. These applications are not as bandwidth intensive as cable modem, data and ad insertion transport. Subsequently, their Ethernet data may be directed to spare ports within an EtherRing network supporting another application.

FUTURE DIRECTIONS FOR EtherRing

Since EtherRing technology is based on industry standard definitions and practices, future products based on this technology will make use of inevitable enhancements and developments within the Ethernet standard. Of particular note are the emerging developments in Virtual LAN (VLAN), Broadcast/Multicast IP, quality of service and Gigabit Ethernet. These developments will extend the functionality and versatility of Ethernet-based transport systems while still operating within the defined standards allowing complete backwards compatibility.

CONCLUSION

This paper has described a method for delivering the functionality, low cost and flexibility (normally associated with the LAN) into the WAN through the use of a new technique known as EtherRing. Consequently, native mode Ethernet and layer 2 packet switching may be employed throughout a larger network. Several key advantages of Ethernet packet switching over routing were presented. These advantages include lower cost, higher throughput, lower latency, independence of network and application protocols and, MAC address switching with automatic self-learning.

When applied to cable data modem transport applications, EtherRing allows the Route Once - Switch Many concept to be employed over the large scale regional CATV interconnect. This permits the centralization of network routers as well as the sharing of Ethernet bandwidth over multiple hub sites resulting in several key outcomes:

- *Lower network cost*
- *Increased operational efficiencies (centralize key technical staff in a common location)*
- *Reduction of network complexity within the hub site*
- *Easier network administration (through elimination of router maintenance functions at each hub site)*
- *Increased bandwidth efficiency and utilization*
- *Simple network re-configuration for increased service demand*
- *Scaleable for service growth*
- *Support of multiple applications based on the Ethernet standard*

EtherRing™ is a registered trademark of ADC/Codenoll

REFERENCES

(1) Bay Networks White Paper, “Networking: A Primer. 1996. http://www.baynetworks.com/Products/Papers/wp-primer.html

(2) http://www.lantronix.com/htmlfiles/mrkg/catalog/et.htm

(3) http://pclt.cis.yale.edu/pclt/comm/ether.htm

(4) Molle M., Watson G., “100Base-T / IEEE 802.12 / Packet Switching”, IEEE Communications Magazine, Vol. 34 No. 8 pp 64-71.

(5) Passmore L., Bransky J., “Route Once, Switch Many” White Paper from Decisys, July, 1997.

(6) Coden M., “Circuits and Methods for a Ring Network”, United Stated States Patent Application, August 1997.

(7) Harris R.W., “Digital Backbone Network Applications for Inter-City and Intra-City Regional CATV Networks” NCTA Technical Papers, May 1994.

(8) Holobinko J., “An Analysis of Interoperability and Cost Factors for Regional Digital Backbone Networks”, NCTA Technical Papers, May 1994.

Robert W. Harris
Market Program Manager, Digital Transmission
ADC Broadband Communications
999 Research Parkway
Meriden, CT 06450
203.630.5732

Mr. Harris serves as Marketing Program Manager for the Broadband Communications Division of ADC responsible for new market, product and business development of digital transmission solutions for video and data applications. Mr. Harris is an 18 year veteran of the CATV and telecommunications industry. During the 1980's he held positions in construction, field service and system engineering with Viacom and Storer Cable. He later worked as a LAN/WAN Network Engineer at NASA's Langley Research Center. Since 1990, he has held engineering and product management positions with C-COR Electronics, Nortel and ADC working with analog and digital fiber transmission, computer network and MPEG based products. He completed his undergraduate training in physics at Old Dominion University and graduate training in engineering science and business at Penn State University and the University of Phoenix, respectively.

Michael Coden
President
ADC Codenoll
200 Corporate Boulevard South
Yonkers, New York 10701
914.965.6300

Michael Coden is President of ADC Codenoll (a subsidiary of ADC Telecommunications Inc.) which he founded in 1980. ADC Codenoll develops, manufactures, installs and services LAN hardware and software products for Ethernet, Switched Ethernet, Fast Ethernet and FDDI LANs. Mr. Coden has also served in various engineering, management and executive level positions for Exxon Enterprises, Maher Terminals Inc., Digital Equipment Corporation and Hewlett-Packard. Mr. Coden holds a Bachelors Degree in Electrical Engineering form MIT, an MBA from Columbia University and a Masters Degree in Mathematics from the Courant Institute of Mathematical Sciences at New York University. He has obtained nine patents in LAN technology and fiber optics, some of which are used in IEEE and ISO international standards.

MULTIPLE APPLICATION CO-ORDINATION AND DEMAND-BASED APPLICATION RETRIEVAL VIA INBAND FOR SET-TOPS

Abhijit Chatterjee and Navneeth Kannan
General Instrument Corporation

Abstract

The advanced set-top terminals of today are characterized by extensive software content. They provide a multitude of features to the subscriber at home via a variety of applications. The demand for new and enhanced applications is growing as cable operators and subscribers are discovering new facets of usability of the Set-top terminal. Smooth co-ordination between applications and the need to have more applications than a cost-effective memory model can support are two problems that face set-top software designers.

This paper describes one implementation of an Application Manager based on user-input to solve the first problem and a Demand-based application retrieval system to solve the second problem.

The Application Manager and Transient Application Server have turned out to be successful tools around which many applications have been built in the product-line of Advanced Network Systems group in GI.

PART-1: APPLICATION MANAGEMENT IN SET-TOPS

Software Architecture

The advanced set-top terminal of today is characterized by extensive software content. A real time operating system is at the heart of many of the advanced set-tops. The use of a real-time operating system allows flexibility to add applications and at the same time to be capable of addressing any real-time communication needs for a set-top. In a set-top, typically, there are drivers and servers for providing various platform services. Applications are the clients of these services. This view of the architecture is presented in Fig.1

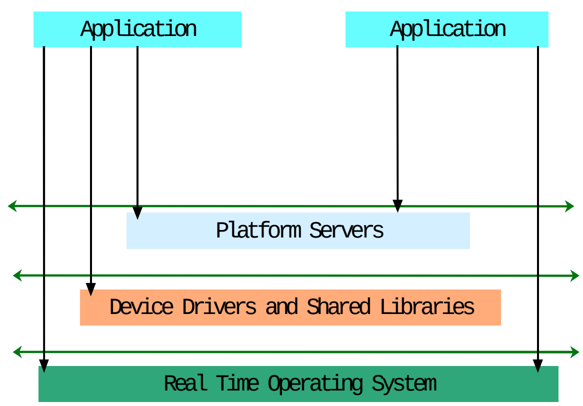


Fig.1 Traditional Embedded Software Architecture

The above traditional embedded architecture is acceptable only as long as all the applications are written by the same vendor and the applications all have a smooth, predetermined way of transferring control from one to another.

In a typical advanced set-top system, there are different applications provided by potentially different groups. There are also applications written by Independent Software Vendors (ISVs) on the advertised, open platform.

The problem then faced by different application writers immediately becomes one of "Acquisition and Transfer of Control". How is an application supposed to take control, and when and to which other application is it supposed to relinquish control, if it chooses to, are questions that need to be answered.

We can try to make an analogy to the Personal computer world and suggest a solution similar to the Program Manager of the Windows paradigm.

However, the advanced set-top terminal is still characterized by the need to drive down the cost to the cable operator. This often translates to a limited amount of memory that is made available in the terminal. With stringent memory constraints, it is difficult to suggest a solution based on complex Operating Systems. Using a traditional real-time kernel is the best way for reducing the memory requirements.

Token based approach for Application Management

In this paper, we propose a unique way of application management using the concept of tokens.

Tokens and Token Management:

The term token is used to represent the unit of communication exchange between applications. The Token based Application Manager (also called the Token Manager) passes tokens amongst applications. The Manager receives all the keys from the remote control, and passes them as tokens to the current active application. Any unprocessed tokens from that application are also obtained and passed to other applications that may need them.

In the set-top scenario, the IR (InfraRed) Remote Control and the Front panel keys on the terminal are the standard devices using which a user interacts with the set-top to invoke applications.

The approach presented here for application management takes into account the essential nature of applications characterized below:

1. An application that is active usually controls the standard devices for user interaction (the remote control and front panel keys for input and On Screen Display (OSD) for output).
2. Applications are typically invoked upon a special key or because of a specific user selection from a displayed menu.
3. Applications handle a set of keys; interpreting some of them for specific actions; dropping other keys that they don't know how to handle.
4. Applications exit upon some specific user action, whereupon they are not expected to handle any more keys.

Receiving Application:

We define the Receiving Application as an application that is currently active (or <u>in-focus</u>). The Receiving Application is the one that receives all the tokens from the Application Manager.

Token Registration:

Applications that need to participate in this process register with the Application Manager. As a part of the registration, an application would specify the token upon which it would like to become the "in-focus" application. This could be either a direct key or a token generated by a menu program. A direct key is a key on the remote control (like the GUIDE key invoking the electronic program guide) directly invokes an application. While registering, the application specifies whether it wants to see the token before the current in-focus application or after the current in-focus application has had a chance to look at it.

Default Application:

The Application Manager needs a 'default application' to send those tokens that have not been registered as input tokens by any other application. Typically, in a set-top scenario, the application that is active when the viewer is watching TV is the default application (which handles channel surfing, volume control and power-on off conditions)

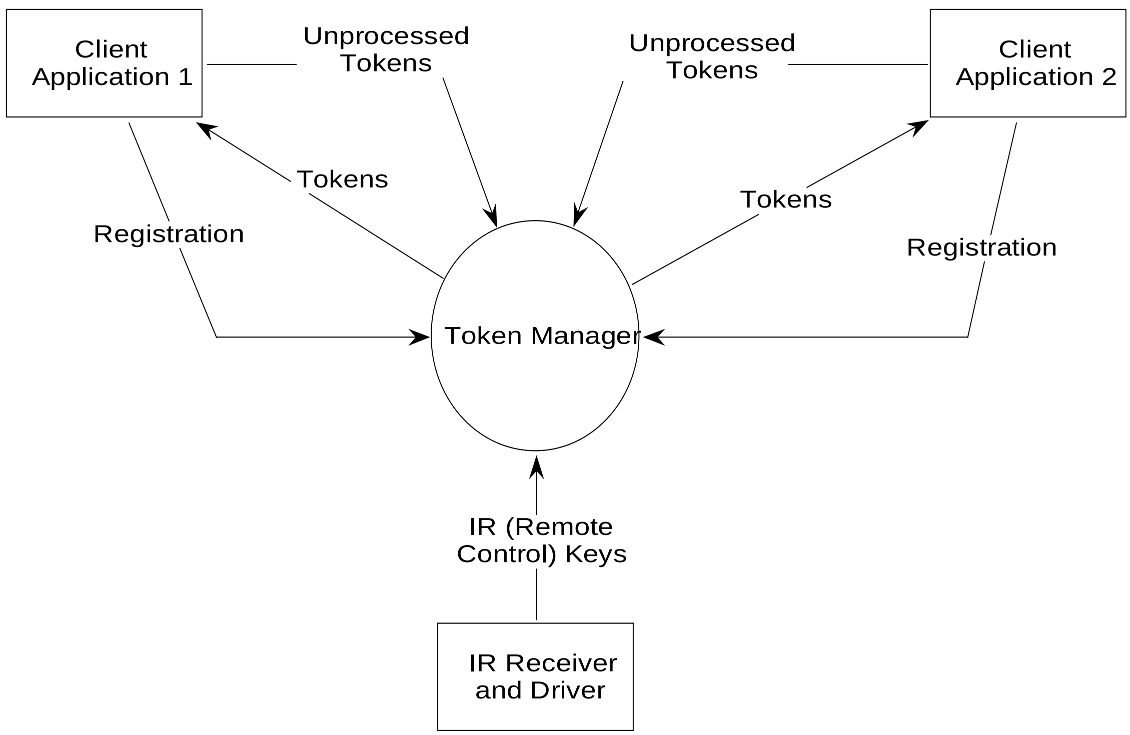


Fig.2: Data Flows amongst Applications and the Application (Token) Manager

Application Management:

The following are the main aspects of the Token based Application Manager.

- Key Forwarding
- Application Flow Control
- Hyper Token Generation
- Application Switching

Key Forwarding

The Application Manager receives all the keys from the hardware. A predefined map establishes the relationship between a key and a token. A menu key pressed on the remote control and a menu key pressed from the front panel of the set-top both map to the same essential token (while retaining source information which may be relevant to the application). Usually, it is the in-focus application, termed the Receiving Application that receives all the tokens.

Application Flow Control:

Flow Control consists of metering out tokens to the applications at the rate at which the applications process the tokens. This ensures that applications don't have to worry about handing over any buffered tokens after their decision to relinquish control. There is just a single buffer for user input and it is maintained by the Application Manager.

Hyper Token Generation:

Sometimes, the user presses a remote control key and keeps it pressed, expecting the same action to be taken multiple times. This is common in channel surfing, browsing through an electronic program guide etc. The Application Manager handles multiple repeat sequences from the key, and directs additional key tokens to the application at a regular interval that can be configured. Applications may choose to turn this feature off, if they cannot handle the held-down key.

Application Switching:

When the Application Manager gets a key, it examines the registry to see if there was any application that had registered for looking at this token before the current in-focus application has had a chance to do so. If it finds one, then it is time for doing an application switch. The current in-focus application is informed that it is no more the in-focus application. It is the responsibility of the in-focus application to give up any substantial or relevant resource (like the On Screen Display, etc.), clear up the display and inform the Application Manager that he is done winding up. The Application Manager then invokes the new application by sending a message to it indicating that it has indeed become the new in-focus application.

More often than not, applications register for a specific token, but do allow the current in-focus application to look at it before processing it. This allows two applications to process the same key, and interpret it differently.

Consider the following example of two applications in the set-top. First, an EPG (Electronic Program Guide) which is activated by the GUIDE token. The other application is the one that handles the Digital Audio features, and also incorporates an electronic guide for the Digital Audio Music Channels. When the Digital Audio Application is the one that is being used by the subscriber, pressing the GUIDE key should present the Music Channel Guide (as opposed to the Video Program guide).

This is accomplished by the two applications by following simple rules. The following sequence of events describes briefly as to how this works.

1. The Electronic Program Guide registers with the Application Manager for the 'GUIDE' key. (It also specifies Post Processing, which would allow the current in-focus application to look at the key before sending it back to the Application Manager)
2. The Digital Audio Application registers with the Application Manager for the 'MUSIC' key.

3. The Tuning Application registers with the Application Manager as the default application.

Scenario 1:

When there is no user interaction, the Tuning Application is the one that is active and is the 'Receiving Application'.

1. The User presses the 'GUIDE' key. Now, since he is in a mode viewing TV, the expected EPG is the Video Channel EPG.
2. The Tuning Application receives the GUIDE key, and realizing that it does not know how to interpret it, passes it back to the Application Manager.
3. The Application Manager examines the registry to see who has registered to post-process the GUIDE key, finds the EPG application and forwards the key to it, also making it the in-focus application.
4. The Application Manager also informs the Tuning Application that it should clear up.
5. The Tuning Application clears up, and the EPG application takes Control.

Scenario 2:

1. The User Presses the Music Key, while watching the EPG application.
2. The EPG application does not handle the MUSIC key, and therefore hands it back to the Application Manager.
3. The Digital Audio Application takes over control.
4. User Presses the GUIDE key now. Application Manager gives it to the in-focus application (which is the Digital Audio Application)
5. The Digital Audio Application (unlike the Tuning Application) knows how to handle this key, and puts up the Music Channel Guide.

In some cases, the current application may relinquish control because of user action (like pressing the EXIT key). In this case, the current in-focus application, upon receiving the EXIT token, relinquishes control by itself, and responds back to the Application Manager to indicate that it not only processed the token, but also wishes to relinquish control.

In certain conditions, asynchronous events could necessitate switching of 'in-focus' status between applications. The following sequence serves as an example of this scenario:

1. The viewer sets up a VCR timer to record an event occurring in the near future and goes back to watching video.
2. A few seconds before the clock reaches the start time of the chosen event, the viewer presses GUIDE key to bring up the Electronic Program Guide and starts surfing through it.
3. The VCR timer indicates that it is time to start recording. The 'Record a Program' application informs the Application Manager that it has to become the 'in-focus' application right now.
4. The Application Manager informs the EPG application that it has to relinquish control due to an external stimulus.
5. The EPG application clears up and informs Application Manager likewise which then makes the 'Record a Program' application the in-focus one.
6. The recording of the event is started.

Conclusion:

In our experience, the concept and use of the token-based application management scheme greatly enhanced the ability of independent application development groups (both within the organization and the Independent Software vendors) to work in unison. The problems associated with integration of multiple applications especially with respect to their smooth co-existence were minimal.

We also have been able to utilize the principles of application management and coordination that we learnt and applied them to more difficult problems like the one described in Part-II of this paper.

PART-II: DEMAND-BASED APPLICATION RETRIEVAL VIA INBAND FOR SET-TOPS

Currently the software for the set-tops is stored either in ROM or programmable Flash memory. The size of the storage area is limited in Set-top terminals and cannot be upgraded easily as demand for newer applications grow. Hence alternate ways for making the Set-top terminal capable of executing multiple applications within the ROM/Flash size constraints had to be identified. The goal was to maximize the usage of the limited storage space and create a notion of a larger 'virtual' executable space. Demand-based Transient Application Retrieval via Inband was designed to meet this requirement.

Overview of Demand-based Transient Application Retrieval

The various applications that are invoked and executed by subscribers can be categorized based on the frequency of usage and expected response time. If an application is not a frequently chosen one or if it does not impose any restriction on the response from the set-top terminal, it is a candidate for demand-based retrieval. Some examples of such applications are Set-top configuration, Favorite Channel setup, Parental Control Password setup etc. These applications can be classified as 'transient' and need not be stored in the ROM/Flash. They could potentially free up some more of the limited storage space for applications that are called upon more frequently or demand faster responses. The 'transient' applications are loaded into the set-top only when demanded and they relinquish all resources once the execution is completed. Essentially, they don't occupy precious storage space permanently.

Instead, the executable code for the 'transient' applications is continually broadcast on one or more dedicated inband channels and the set-top will be able to pick up the code for the particular application in demand and store it in dynamic memory for subsequent execution. Thus, the set-top uses the inband channel as a secondary storage medium. In order to improve the cycle time of the inband carousel, the executable code for the 'transient' applications can be transmitted in a compressed form. In such a case, the set-top inflates the code before execution. Since this process of code acquisition and conditional decompression causes additional latency, a ‘transient’ application should be chosen selectively.

The set-top has a Transient Application Server that controls the acquisition and execution of all ‘transient’ applications. During the start-up sequence of the terminal, the Transient Application Server acts as a ‘proxy’ for all ‘transient’ applications and takes care of registration with Application Manager.

The Server reserves a portion of the dynamic memory as the execution space for transient applications. The size of this dedicated area (called TRansient Application Code Execution area or TRACE) is determined based on size of the largest ‘transient’ application defined for the system. This size also acts as a guideline for determining future candidates for ‘transient’ label as well as in the design of future transient applications. Based on the dynamic memory configuration of the terminal, space could be reserved for more than one transient application thereby improving throughput.

When the subscriber selects to run any ‘transient’ application (by user menu selection or otherwise), the Application Manager switches the ‘in-focus’ application status to the chosen transient application and sends a special API message to inform it of the new status.

The Transient Application Server intercepts the API message and examines its contents to find the target ‘transient’ application name. Thereafter it examines the TRACE area to ascertain if the application is already available. If so, the application starts executing immediately. Otherwise, the Server off-tunes the terminal to the appropriate inband channel and starts acquiring the executable code for the application. Upon completion of acquisition, control is transferred to the ‘transient’ application for execution.

The transient application continues to execute till the user action requires another application to be invoked. After the Application Manager switches the ‘in-focus’ status to the new application, the Transient Application Server instructs the transient application to relinquish all resources.

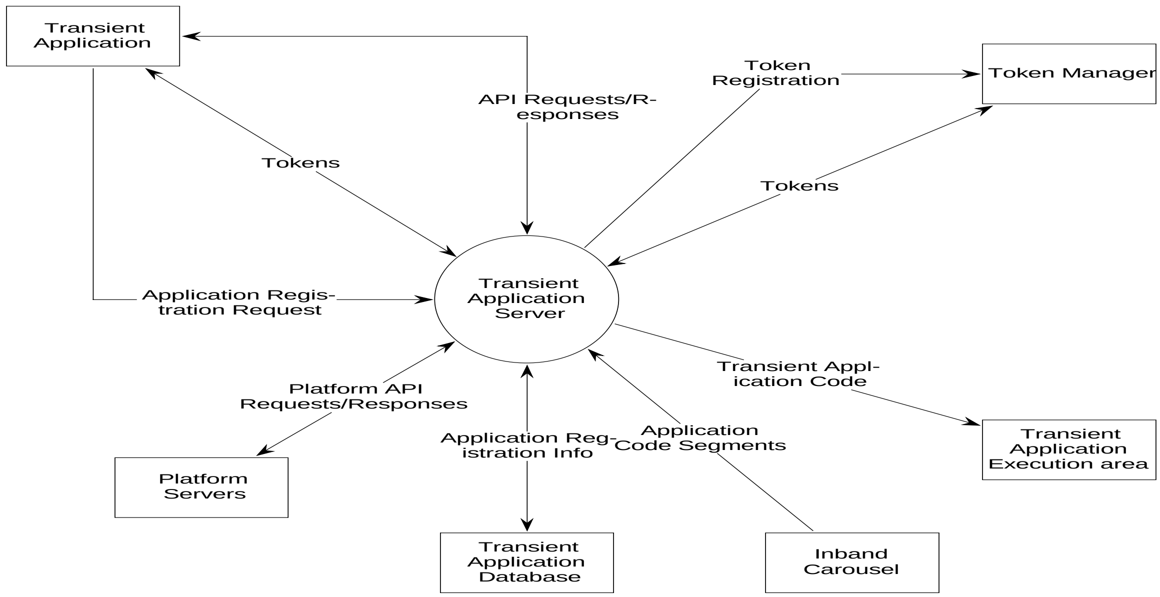


Fig 3: Data flow among Transient Application Server, applications and Inband carousel.

Transient Application Management

The primary aspects of interaction between 'transient' applications and the Transient Application Server are as follows:

- Registration with the Server
- API Message handling
- Transfer of Control
- Management of TRACE area

Registration with Server

All the 'transient' applications defined for the system have a small proxy agent defined in the software resident on the set-top terminal's ROM/Flash. During system start-up, these agents provide registration information to the Server regarding the associated application:

- application name
- 'compressed' vs 'uncompressed' status
- dynamic memory requirements
- Execution space requirements
- Inband Channel information

The Transient Application Server saves the registration information for all applications in the Transient Application Database. It also allocates DRAM for run-time requirements at this time so that the transient application is not starved for DRAM when it is loaded off the inband channel for execution.

API Message Handling

The Transient Application Server handles the API interface with the Platform Servers on behalf of the applications. In order to send an API request, the application uses a set of library functions provided by the Server that identifies the requesting application uniquely. The Manager receives all API responses and forwards them to the appropriate application based on unique routing information embedded in the response.

Transfer of Control

Based on user actions, when Application Manager decides to switch the 'in-focus' status to a transient application, it sends the special 'start-up' token meant for the application. The Transient Application Server intercepts the message and performs the following steps:

1. First, it examines the message contents to find the target 'transient' application name and looks up the Registration Database to locate the record for the application.
2. Thereafter it examines the TRACE area to ascertain if the application is already available. If so, the 'start-up' token is forwarded to the application and it starts executing immediately.
3. Otherwise, the Server off-tunes the terminal to the appropriate inband channel and starts acquiring the executable code for the application.
4. Upon completion of acquisition, the 'start-up' token is transferred to the 'transient' application for execution.

Similarly, when the transient application is ready to relinquish control in response to some user action, it sends a special 'exit' token to the Application Manager via the Server. The Application Manager switches the 'in-focus' status and sends a notification to the erstwhile active transient application too. The Server intercepts this message and instructs the transient application to release all system resources acquired during execution.

Management of TRACE area

Based on the dynamic memory configuration of the set-top terminal, the Transient Application Server may reserve a larger area for storage and execution purposes of more than one transient application. This improves the throughput of the terminal since inband acquisition may not be required for all 'transient' application accesses. The TRACE area is managed based on a Least-Recently-Used criterion. For each application currently resident in this area, the Server maintains information about its size and the time at which the application was loaded.

When need arises to accommodate a new transient application in the TRACE area, the Server compares the application executable space requirements with the available free space. If enough space is available, the portion of TRACE area is marked for the new application and inband acquisition begins. On the other hand, if the free space is not enough, the Server looks for application(s) that has not been used recently and is of adequate size to accommodate the new application. The storage space allotted to this application(s) is acquired, gets marked for the new application and inband acquisition begins.

Conclusion

In our opinion, the concept of Demand-based Transient Application Retrieval will enhance the capabilities of the advanced set-top terminals immensely. If applications are categorized appropriately, this implementation will allow the set-top to execute multiple applications within the constraints of a limited storage space without impacting the performance. It will provide cable operators with a mechanism to add on new features without enhancing the hardware configuration of the terminals.

Author Information:

Abhijit Chatterjee
Staff Engineer
Advanced Network Systems,
General Instrument,
2200, Byberry Road,
Hatboro, PA, 19040
(215-957-6749)

Abhijit Chatterjee has been in the software engineering field for 15 years and has been with the Embedded Software Engineering group of GI for nearly 4 years. He got is BSEE from IIT Kharagpur, India.

Navneeth Kannan
Senior Staff Engineer
Advanced Network Systems,
General Instrument,
2200, Byberry Road,
Hatboro, PA, 19040
(215-957-8391)

Navneeth Kannan has been in the software engineering field for 15 years and has been with the Embedded Software Group of GI for the past nearly 5 years. He got his BSEE from Madras University and MSCS from IIT Kanpur India.

Multiwavelength Analog Video Transport Network

Dogan A. Atlas, John J. Kenny
ANTEC Corporation
Norcross, GA 30092

Abstract

We report an experimental 8-channel dense wavelength division multiplexed 1550-nm analog video transmission using directly modulated lasers over 40 km of standard fiber, through two cascaded Erbium-Doped Fiber Amplifiers with adequate carrier-to-noise and distortion performance. Multiwavelength analog video transport networks provide an end-to-end transparent information pipe with increased service penetration.

I. INTRODUCTION

Progress in 1550-nm linear fiber optic technology is continuing to move forward rapidly. Long-distance fiber transmission of broadcast channels, narrowcast (i.e., targeted) services and high-speed data are possible by increasing the channel capacity and power budget. The use of dense wavelength division multiplexing (DWDM) technology combined with Erbium-Doped Fiber Amplifiers (EDFAs) provides a simple and powerful method of increasing channel capacity in Hybrid Fiber/Coax (HFC) networks [1]. The installation of a DWDM CATV system may initially require higher capital cost. However, service upgrading will be significantly cheaper as more wavelengths are added to increase service penetration [2].

In designing a transparent DWDM analog transport network many system parameters must be considered carefully. These include transmitter source characteristics, multiplex and demultiplex optical filter requirements, fiber effects, optical amplifier performance, and optical receiver design. The fundamental issue is the accumulation of capacity limiting impairments in the optical transparent network [3]. This paper reports an experimental demonstration of eight wavelength division multiplexed 1550-nm analog lightwave system and investigates linear and nonlinear amplifier, fiber, and demultiplex optical filter effects.

II. TRADITIONAL CATV OPTICAL NETWORKS

A. Broadcast 1550-nm network:

Figure 1a shows a high-power 1550-nm CATV network. 1550-nm externally modulated transmitters combined with optical amplifiers provide 750 MHz bandwidth of broadcast services. Although the high-power 1550-nm network provides low cost broadcast services, the network is limited in that it can not deliver targeted services. Therefore, the one service it provides is shared among optical nodes (8-16 typically) with thousands of homes passed.

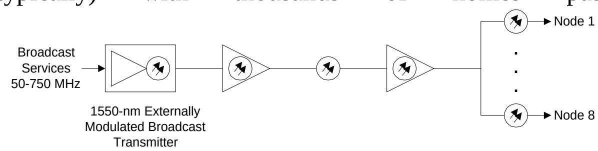


Figure 1a Schematic of a broadcast 1550-nm CATV network.

B. Scalable 1310-nm or 1550-nm network:

Figure 1b shows a scalable CATV network. The headend is composed of several directly modulated 1310-nm or externally modulated 1550-nm transmitters each carrying dedicated 750 MHz bandwidth of broadcast and targeted services on dedicated fibers. The scalable network offers future targeted service capacity expansion [4]. However, targeted service penetration is only possible by increasing dedicated fibers that run from the headend to the optical nodes.

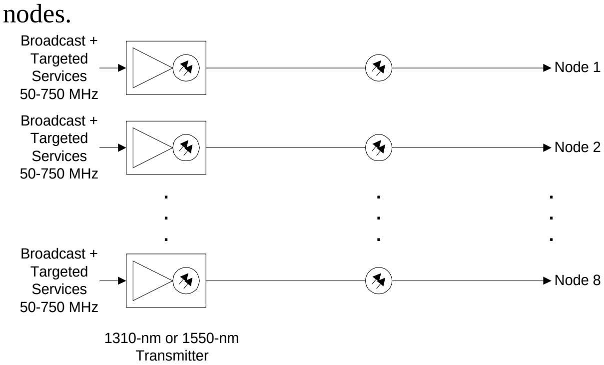


Figure 1b Schematic of a scalable 1310-nm or 1550-nm CATV network.

C. Broadcast 1550-nm and narrowcast 1310-nm overlay network:

Figure 1c shows the 1550-nm broadcast, 1310-nm narrowcast overlay CATV network. These networks place SONET interconnect equipment, service-enabling digital terminals, modulators, optoelectronic converters and 1310-nm narrowcast transmitters in a large multi-cabinet hub [5]. Low-power 1310-nm DFB laser transmitters carry targeted services in the 550-750 MHz frequency range to optical nodes. The targeted services on a 1310-nm optical carrier are overlaid on the same fiber as the broadcast services on a 1550-nm optical carrier through a WDM or coupler. Such a CATV optical network is opaque in the sense that services cannot be targeted to a particular node by routing from the headend.

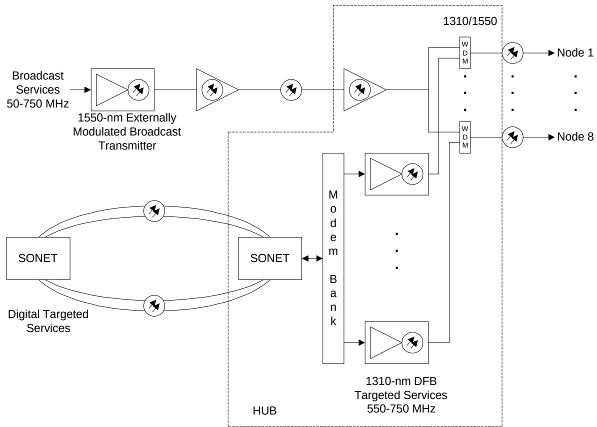


Figure 1c Broadcast 1550-nm and narrowcast 1310-nm overlay network.

III. DWDM 1550-nm CATV OPTICAL NETWORK

Figure 2 shows the DWDM CATV optical network. The headend consists of a 1550-nm broadcast transmitter transporting broadcast services through the first optical fiber and with two erbium-doped fiber amplifier cascades to the hub. A second optical fiber link combines eight narrowcast optical signal beams operating at wavelengths which follow the International Telecommunication Union (ITU) wavelength standard [6]. The combined optical beam passes through two EDFA cascades before entering to the hub. The network is flexible such that each narrowcast transmitter is capable of accepting a unique service. For capacity expansion with unique services additional transmitters operating at a different ITU wavelengths can be deployed. At the hub, the narrowcast signals are demultiplexed such that each unique service is combined with the broadcast signal and transmitted to their respective nodes.

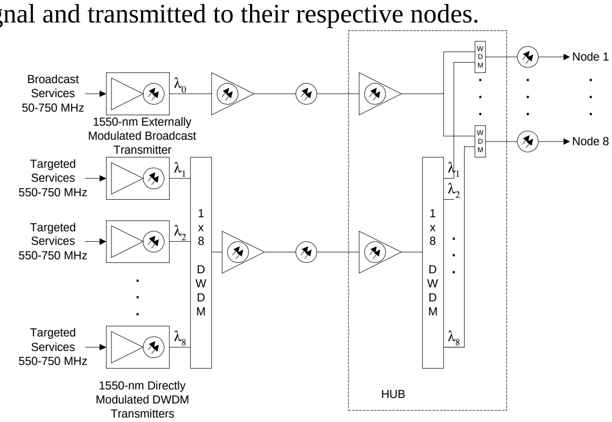


Figure 2. DWDM CATV optical transport network.

The use of DWDM technology enables electro-optical processing for targeted services to be routed from the headend rather than at the hub, as is the case shown in Fig. 1c. This network flexibility provides transparency through a much smaller hub site [5]. DWDM CATV architecture offers a cost effective solution compared to installing additional optical fiber. With DWDM system installed, the secondary hub may initially service tens of thousands of homes passed with one wavelength targeting about ten thousand homes. As service penetration is increased by adding more wavelengths, one wavelength could service much smaller home segments.

IV. TECHNICAL ISSUES

Several system parameters require consideration in designing 1550-nm DWDM CATV optical transport networks. These design and technical issues are related to transmitter source characteristics, optical amplifier performance, fiber effects, and multiplex-demultiplex optical filter performance. Amplifiers can degrade CNR by noise accumulation due to cascaded EDFAs with low input power per optical carrier and degrade distortion due to amplifier gain-slope. Fiber effects can cause distortion, crosstalk, and CNR performance degradation. Fiber effects are stimulated Brillouin scattering (SBS), stimulated Raman scattering (SRS), Self phase modulation (SPM), cross-phase modulation (XPM), and chromatic dispersion (CD). Multiplex-demultiplex optical filters can also degrade performance due to crosstalk generated by adjacent channel isolation and distortion induced by non-uniform passband flatness.

A. 1550-nm transmitter source design:
Broadcast transmitter

In the design of 1550-nm analog broadcast systems, the choice of source must be an externally modulated transmitter. A 1550-nm externally modulated transmitter is attractive and offers several advantages such as very low chirp, and high power with the ability to be amplified further by optical amplifiers to service remote locations.

DWDM narrowcast transmitters

However, the CW DFB laser source, external intensity modulator, and associated electronics and RF circuitry that constitute the transmitter are altogether too expensive to be used as narrowcast transmitters. Two alternatives to an externally modulated transmitter are directly modulated DFB laser and electro-absorption modulated DFB laser (EAML) source. Current EAML sources are moderately expensive, have low output power and have very nonlinear light-electrical transfer characteristics. Directly modulated DFB laser sources are lower cost, have high output power, and have linear light-current transfer characteristics.

However, the magnitude of the laser chirp of directly modulated DFB lasers may limit the number of subcarriers to be used by each narrowcast optical channel. The operating center wavelengths of the DFB sources follow the ITU wavelength standard.

When two optical signal beams share a receiver at the node, mixing takes place. The mixing of the two optical signal beams is unavoidable and generates optical beat interference noise. To minimize optical beat interference noise between broadcast and narrowcast optical channels at the node, channel separation must be more than about 2-nm. However, actual required minimum separation depends on the optical line shape of the two optical channels.

A. Optical Amplifier effects:

Noise

If more than one optical beam is passed through the amplifier the intensity noise increases as a function of the number of optical channels M. The accumulated amplifier RIN per optical channel in an amplified DWDM analog transport system is proportional to $M{\cdot}F$ where F is the noise figure of the amplifier cascades in a single optical channel system. Here we have assumed that the optical power delivered by the multiple optical signal beams to each amplifier input is equal to the optical power delivered by a single optical signal beam in a single channel system. With DWDM CATV systems, to reduce the impact of noise cascades per channel we must increase the total power to the input of the amplifier.

Gain flattening

The gain of the EDFA across the waveband 1530- to 1560-nm is not flat. The need for improved flatness across the whole wavelength window is to ensure that all narrowcast optical channels are equally amplified, separated, and sent to their respective nodes. To flatten the EDFA gain profile sophisticated filter technologies are used [7]. With these enabling filter technologies it is possible to achieve a 1 dB peak-to-peak flatness over the narrowcast waveband 1540- to 1560-nm.

Gain-slope

Amplifier gain-slope coupled with laser chirp, $\Delta\lambda_{chirp}$ causes composite second-order intermodulation distortion (CSO). The amplifier gain-slope induced distortion is flat, i.e., not dependent on modulating frequencies, and is proportional to $\Delta g{\cdot}\Delta\lambda_{chirp}$ where Δg is overall gain-slope of the amplifier cascade. To keep distortions to an acceptable level, the magnitude of total gain-slope should be less than 0.2 dB/nm when the chirp is less than 0.0017 nm.

B. Fiber effects:

Laser phase noise to intensity conversion

As the optical signal beam travels through the fiber, fiber chromatic dispersion converts laser phase noise to intensity noise. The intensity noise is proportional to $\Delta\nu{\cdot}L^2{\cdot}f^2$ where $\Delta\nu$ is the laser linewidth.

Stimulated Brillouin scattering

SBS is a nonlinear fiber effect that limits the amount of optical power per channel that can be launched into the fiber. In standard fiber, for narrow linewidth sources the SBS threshold at 1550-nm is 6 dBm. SBS scatters light into the backward direction and reduces received power as well as increases noise in the forward direction, thereby degrading the received CNR. The SBS launch power threshold can be increased by 10 dB (i.e., 16 dBm fiber launch power) by spreading the optical spectrum. However, to reduce fiber nonlinear effects such as SRS and XPM to an acceptable level, the launch power per optical channel in analog DWDM systems should be maintained to be less than 10 dBm.

Stimulated Raman scattering

SRS is a phenomenon similar to SBS. SRS in optical fiber is a nonlinear process that depletes the lower wavelength optical channels and amplifies the higher wavelength optical channels. In analog DWDM systems, each optical channel is intensity modulated by subcarriers. The intensity of the optical channel with the lowest wavelength modulates the intensity of optical channels at higher wavelengths. Therefore, SRS leads to crosstalk (XT) between optical channels. The magnitude of XT depends on launched power per channel, polarization of each optical signal beam, channel spacing, number of optical channels, chromatic dispersion, and fiber length. Fiber chromatic dispersion reduces the magnitude of crosstalk due to walk-off where the group velocity of adjacent channels are slightly different. For fiber distances greater than 20 km SRS crosstalk is almost fiber length independent. To reduce SRS crosstalk, the launched power per optical channel should be kept to less than 10 dBm.

Self- and cross-phase modulation

SPM is caused by the change in the optical index of refraction with intensity. The intensity dependent nonlinear index causes phase shift of the signal which is dependent on the launch power and fiber distance. XPM is similar to SPM where intensity of one optical beam modulates the phase of the other optical beam. The phase shifts combined with fiber CD causes frequency shift which introduces a relative delay as the intensity of the optical beam is modulated. The intensity modulated signal is distorted as it travels through the fiber. The distortion increase with modulating frequency.

XPM generated second-order distortion is proportional to $(2M\text{-}1)\cdot L^2\cdot f^4$ where M is the number of optical channels, L is the length of fiber and f is the frequency of modulation. Phase modulation (generated by SPM or XPM) occurs mostly within the first 20 km and becomes less dependent on fiber length at longer lengths. Frequency deviation (generated by SPM or XPM and CD interaction) increases as a function of fiber length. When amplifiers are cascaded, after each amplifier the power of the optical signal beam is at its highest before launched into the fiber. In amplifier-fiber cascades, to reduce distortion, XPM and SPM must be kept to a minimum by launching low powers into the fiber.

Chromatic dispersion

When fiber chromatic dispersion is combined with laser chirp, frequency deviation is converted to intensity modulation. The product of the induced intensity modulation and the original intensity modulation generates distortion. The intensity modulated optical signal is distorted as it travels through the fiber, and the distortion increases with fiber length and modulation frequency. Chirp-dispersion induced second-order distortion is proportional to $\Delta\lambda^2_{chirp}\cdot L^2\cdot f^2$ where $\Delta\lambda_{chirp}$ is the laser chirp, L is the length of fiber and f is the frequency of modulation. To eliminate CSO distortion products falling within the broadcast band, narrowcast channels can be arranged to occupy frequency bands such that CSO distortion products fall out-off band (OOB). Other methods are also being considered to reduce CSO and CTB distortion generated by laser chirp combined with fiber CD.

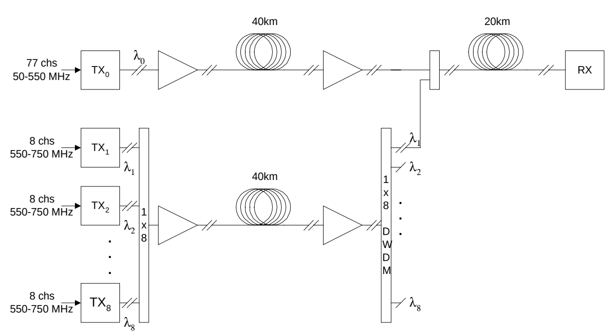


Figure 3. Experimental set-up of a DWDM 1550-nm analog lightwave transmission system.

V. EXPERIMENTAL DWDM SYSTEM

Figure 3 illustrates the experimental DWDM system configuration. The system consists of a broadcast transmitter operating at 1544.54-nm with an output power of 7 dBm and a RIN of less than -164 dB/Hz. The broadcast transmitter is modulated by 77 NTSC channels. The eight narrowcast transmitters operate at 200 GHz spacing, with the lowest wavelength laser operating at 1549.32-nm. The narrowcast transmitters are modulated by 8 channels spaced consecutively above the broadcast channels. The output signal power of each narrowcast transmitters is adjusted to 10 dBm. The transmitters have relatively low RIN of less than -160 dB/Hz and linewidth of about 2 MHz. The eight narrowcast optical carriers are modulated by eight analog channels. Figure 4. shows the optical spectrum of the DWDM narrowcast channels after being combined at the headend. The wavelength of the broadcast channel is about 5-nm below the lowest wavelength narrowcast channel. Broadcast and narrowcast optical channels are transmitted through separate dedicated fibers of 40 km length (see Fig. 3). At the end of the 40 km narrowcast fiber link a dielectric filter type demultiplexer is used to separate the eight wavelengths at the hub. One of the narrowcast wavelengths is combined with the broadcast wavelength and launched through 20 km of standard fiber. The broadcast and narrowcast optical channels are then mixed on a single optical receiver.

TABLE I
Measured CNR and Distortion Performance of Broadcast and Narrowcast (1550.92-nm) Optical Channels mixed on a Single Receiver.

Frequency	Broadcast Band			Narrowcast Band		
(MHz)	CNR	CTB	CSO	CNR	CTB	XT
61.25	51.2	66.8	66.1			
211.25	51.9	68.4	67.4			
319.25	51.3	67.0	68.9			
445.25	51.3	66.7	67.0			
547.25	52.0	57.2	66.5			
553.25				41.4	50.2	
577.25				40.8	47.2	
595.25				41.6	46.5	55*

* XT is measured at 598 MHz.

Table I shows a summary of the experimental results. The received optical power for the broadcast and narrowcast optical signals were 2.0 and -8.0 dBm, respectively. The RF level difference was about 10 dB. The SRS and optical filter induced XT is measured to be about -55.0 dBc at 1550.92-nm. Polarization controllers were used on four narrowcast transmitters to measure worst case XT numbers. We have observed that the narrowcast OOB CTB distortion degrades the broadcast channel CTB performance at frequencies 545.25 and 547.25 MHz. Conversely, the broadcast OOB CTB distortion degrades the narrowcast CTB performance at frequencies 553.25 to 595.25 MHz. Narrowcast OOB CSO distortion products were below 54 MHz.

VI. CONCLUSIONS

These architecture feasibility tests demonstrated the enhanced narrowcast capability that DWDM can add to the more conventional 1550-nm broadcast network. Excellent 77 channel analog AM-VSB transmission performance from the head end, through the hub, to the node was observed. The carrier-to-noise ratio and distortion performance of the narrowcast channels permit transmission of at least 8 digitally modulated RF subcarriers to carry targeted forward path services such as cable telephony and cable modem. The transparency of the DWDM approach allows the targeting of a service to a particular node only, it does not require routing activities at the head end and none at the hub site.

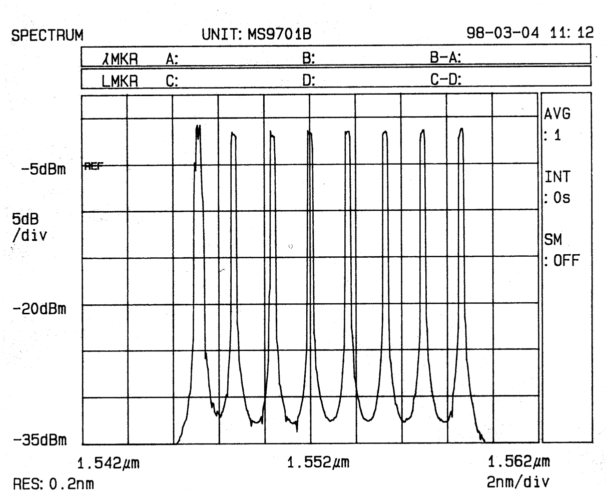


Figure 4. Spectrum of eight wavelength division multiplexed narrowcast optical signal beams.

ACKNOWLEDGMENTS

We would like to thank O. Sniezko of TCI for his encouragement and expert advice. The contributions of M. Linford, M. Ellis, and H. Ghuman of ANTEC are gratefully acknowledged.

REFERENCES

[1] W. I. Way et al, " Simultaneous distribution of multichannel analog and digital video channels to multiple terminals using high-density WDM and a broadband in-line erbium doped fiber amplifier, *IEEE Photon. Technol. Lett*., pp. 665-668, September 1990.
[2] M. Lafferty, D. Cervenka, " Redefining state of the art," *CED*, pp. 25-42, February 1998.
[3] R. W. Tkach et al, "Fundamental limits of optical transparency," in *Proceeding OFC'98*, San Jose, CA, February 1998, pp. 161-162.
[4] J. Holobinko, " Scalability: the newest HFC network weapon," *CED*, pp. 30-35, January 1996.
[5] O. Sniezko, "Video and data transmission in evolving HFC network," in *Proceeding OFC'98*, San Jose, CA, February 1998, pp. 131-132.
[6] ITU-T draft Recommendation G.mcs, (study not finalized).
[7] S. Turley, " Technology choices abound for the optical network," *Lightwave*, pp. 51-60, February 1998.

New Developments in System Information and Program Guide Standards

Mark K. Eyer
General Instrument Corporation

In anticipation of North American broadcasters commencing digital transmissions in 1998, digital television standards work continued in earnest in 1997. In December, the Advanced Television System Committee (ATSC) adopted its newest standard, A/65, Program and System Information Protocol for Terrestrial Broadcast and Cable [(4)]. The ballot approval culminated over twelve months of work in the ATSC's T3/S8 Transport Documentation Specialists subcommittee with help from the Society of Cable Telecommunications Engineers (SCTE) Digital Video Subcommittee (DVS), where the standard was also approved by letter ballot in January. The acceptance of A/65 for use with both terrestrial broadcast and cable transmission media is a triumph for cooperative efforts in standards-setting and will facilitate interoperability between broadcast and cable as we move into the age of digital transmission.

This paper describes the new A/65 standard, beginning with the system functional requirements upon which it was based. An overview of the protocol is presented to give the reader a flavor of the features it provides. The relationship of A/65 to the prior ATSC standards for program guide and System Information, A/55 and A/56 respectively, is discussed. Finally, the use of A/65 on cable is explored.

OVERVIEW OF PROGRAM AND SYSTEM INFORMATION PROTOCOL (PSIP)

The ATSC A/65 *Program and System Information Protocol (PSIP) for Terrestrial Broadcast and Cable* is a standard method for delivery of program guide and system data tables carried in any compliant MPEG-2 multiplex. The primary purpose of PSIP is to facilitate acquisition and navigation among analog and digital services available to a particular receiver or set-top box, but it also serves as a support platform for applications such as data broadcasting. Use of A/65 is mandatory for digital terrestrial broadcasts in North America, and it is expected that cable operators will support it as well to support cable-ready digital televisions and VCRs.

In analog broadcast or cable television, if a user selects channel "4," the receiver knows to tune the frequency of channel 4 as standardized by the FCC and EIA—the 66-72 MHz band. The situation changes, however, with advent of digital transmission—digital broadcasters will now have the freedom to define channel numbers independently of the RF frequency used to carry the signal. PSIP (like the original A/56) is architected around the concept of "virtual channels." A virtual channel is called "virtual" because its definition is given by indirect reference through a data structure called a Virtual Channel Table (VCT). So, when an end-user selects "channel 4" she is actually selecting the channel record associated with user channel number 4. The definition of the channel as given in the VCT includes its frequency and method of modulation, textual name, channel type (analog audio/video, digital a/v, audio only, data), and the channel number the user may use to access it.

The A/65 protocol introduces a new navigational concept, the "two-part" channel number. Broadcasters declared the need, as new digital services are introduced, to retain the brand-identity they currently have as a result of years of marketing and advertising. For broadcasters, the first part of the two-part number (called the "major" channel number) is required to be the same as the EIA/FCC channel number already in use for the analog service. The second part of the number (called the "minor" channel number) identifies one service within the group of services defined by the major number. From the point of view of the user, where before there was just "channel 4," now there may also be channels 4.1, 4.2, 4.3, and so on.

A/65 also standardizes the digital equivalent of the V-chip content advisory data now included in analog broadcasts in EIA-608A XDS packets. PSIP delivers a Rating Region Table (RRT) that defines the structure of a multi-dimensional content advisory system for a specific region (e.g., country), and a Content Advisory Descriptor that can be used to associate specific program events with rating levels defined in the RRT.

SYSTEM FUNCTIONAL REQUIREMENTS

US broadcasters were aware in early 1997 that the ATSC *Digital Television Standard* as written did not completely meet all of their needs. While they recognized the importance of providing program guide data along with digital broadcast programming, the standard indicated that use of the ATSC program guide standard (A/55) was "optional." Likewise, though the ATSC standard suite included the A/56 *System Information* standard that defined network data tables for various transmission media, use of A/56 for broadcast was also optional. In the case of SI, broadcasters were not even sure they had any use at all for the network data it provided. After all, they felt, receivers already know the standard 6 MHz frequency plan, and the MPEG-2 *Systems* standard itself defines ways of labeling the various parts of a digital multiplex that make up each service.

Early in 1997, the problem of finalizing program guide and navigation issues was assigned by ATSC to its T3/S8 Transport Documentation Specialists group. In May, the ATSC T3/S8 group began focusing on system functional requirements for what they coined the "Naming, Numbering and Navigation," or "N^3," problem.

The following presents the basic set of system functional requirements T3/S8 adopted as their starting point:

1. **Must support direct access to any channel.** The navigational model must support the ability to access any analog or digital channel by direct entry of its channel number or call letters.

R1. Must support grouping of selected digital services with an existing analog service, or with digital services on other multiplexes. There will be a period of time in which broadcasters operate an analog channel in addition to a digital multiplex. The navigation model must include a grouping concept to support channel surfing within a set of related analog and digital channels.

R2. Must preserve channel branding. When a broadcaster begins a digital service, the system must allow him to associate the new programming with the channel label he has used in past years of advertising.

R3. Must be harmonized with cable standards. The chosen solution for broadcast N^3 must recognize that cable set-tops and cable-ready receivers will also employ navigational and channel naming methods. These must work in harmony with methods defined for terrestrial broadcast.

R4. Must accommodate the flexibility of digital transport. The MPEG-2 standard provides a great deal of latitude in defining a "service" on the multiplex. The approach must accommodate the wide variety of service structures possible via digital transport.

R5. Must allow a broadcaster to "package" or "market" some services separately from others on the multiplex. For example, as a public service, the owner of a digital broadcasting license may offer a couple of spare megabits of SDTV bandwidth to a college or community access channel, or to a government affairs (city politics) channel. It must be possible for that operator to label that channel separately from the other services he offers on that multiplex.

In addition to these general requirements, an important consideration was identified specific to cable transmission media: the cable system operator must be able to label digital services independently of the EIA RF channel number he uses to transmit them. In other words, cable would use a "virtual channel" scheme. Each cable operator must be free to use whatever frequency slot he chooses to deliver the digital signal and still be able to label it for the user in a meaningful way. Practically speaking, when the cable operator takes an off-air digital broadcast signal and re-modulates for cable, he should be able to label it the same as the broadcaster does for televisions that receive it directly off-air.

At first glance, this arrangement might seem obvious. After all, most cable operators place each local broadcast channel at the same channel number which the broadcaster uses for his terrestrial broadcast. But consider that the cable operator may wish to take advantage of the faster data rate available when the digital signal is delivered on cable: 8-VSB modulation provides an information rate of about 19 Mbps, while 64-QAM modulation on cable allows 43% more bits at 27 Mbps. With 256-QAM modulation, the available data rate is double that of the terrestrial rate: 38.8 Mbps. The operator may wish to add together two broadcast multiplexes into a high-rate transport stream QAM modulated in a single 6 MHz frequency band. Despite this remultiplexing, any standard receiver connected to the cable must be able to determine how to present all the services to the user, and to promote ease of use, it should be able to label them consistently with their broadcast versions.

Lastly, two requirements were identified specific to terrestrial broadcasting. First, the system must accommodate terrestrial broadcast translators. Broadcasters don't want to have to re-process SI or the program guide data if they transmit the same MPEG-2 Transport Stream (TS) on an alternate frequency. Second, the receiver should be able to deal with a movable antenna, either steerable by the user or because the receiver itself is movable. For example, the receiver may be in the back of an RV traveling across the country.

THE PSIP SOLUTION

The A/65 PSIP standard addresses all these system concerns for both terrestrial broadcasting and cable. Specifically:

- **Direct access**— The user can access any service, whether it is an audio/video service, an audio only service, or a data broadcast service, by specifying either its channel number or its service name. The channel number method is compatible with numeric-keypad RCUs.
- **Channel grouping—** Channel numbers are assigned through the Virtual Channel Table. The two-part channel number scheme provides a grouping function in that the major channel number identifies the group and the minor number the member of the group. Broadcasters are required to use the RF channel for the current analog NTSC broadcast channel as the major channel numbers for all the new digital services.
- **Channel branding**— As described, the broadcasters keep the "brand identity" associated with their analog channels because the new digital channels use that same number for their major channel number.
- **Harmonized with cable—** PSIP was designed with the needs of cable in mind. The Digital Video Subcommittee (DVS) of SCTE was invited to participate in the development of the standard from the beginning and contributed important technical input. DVS voted and approved PSIP (as its document DVS-097) in January, 1998.

  The overall design of System Information tables in PSIP achieves nearly complete parallelism between the methods used for cable and terrestrial broadcast. There are only a few places where the specification differentiates cable and terrestrial broadcast. One example is that major channel numbers on cable can range from 1 to 999 for video services, where on terrestrial broadcast the range is limited to 2 to 99.
- **Accommodates flexibility of digital transport**— PSIP builds upon the MPEG-2 *Systems* standard (Ref. [(5)]) without imposing constraints on its use. Any service that can be represented by the MPEG-2 standard method can be referenced by the PSIP VCT. Note that this is in contrast to the original A/53 *ATSC Digital Television Standard* [(1)] which mandated the use of the so-called "program paradigm" for broadcast television. (See *The "Program Paradigm"* below for a discussion of the program paradigm and why ATSC is moving to delete it from the ATSC standard).
- **Separate packaging of services—**PSIP states that for the US, in addition to using the major channel number that matches his analog broadcasting license for some services, a broadcaster may label one or more virtual channels on his multiplex with major channel numbers in the 70-99 range. Certain conditions must be met, however: assignment of major channel numbers must be coordinated such that they are unique within a region. Otherwise, one receiver could access two different digital services labeled with the same channel number, and that would cause user confusion. Once the value of this feature is recognized, one might guess that the FCC would be called upon to coordinate such regional number assignments.
- **Cable virtual channels**—The requirement is met because PSIP is based on the virtual channel concept. The more significant aspect is that digital terrestrial broadcasting is now also based on virtual channels.
- **Broadcast translators—**If a digital broadcast is shifted in frequency at a translator, all SI and EPG information it carries remains correct except for frequency references. As it happens, even though carrier frequency information is included in the SI tables, a receiver can operate without it (or, more likely, compensate for frequency errors it might find). The techniques involved are discussed in *Identifying the Signals* below.
- **Movable broadcast receivers and antennas—** PSIP tables are delivered repetitively. Broadcast Virtual Channel Tables, for example, are repeated no less often than once each 400 milliseconds. A receiver may therefore quickly learn the labels to use for navigation. If a receiver itself is moving, even as one signal fades out and another is acquired, navigational and channel label data will always be readily accessible.

THE PROGRAM AND SYSTEM INFORMATION PROTOCOL—USER'S VIEWPOINT

Looking at it from a broad perspective, PSIP is a standard method for delivering the data that a digital TV or cable box needs to support basic navigation among digital services. The term "navigation" is used in this context to mean the process of discovering where one is; moving in the desired direction; and then reaching the desired destination within what could be an ocean of digital and analog service offerings.

The “Where am I?” question is answered in two ways by PSIP: a channel number and channel name. PSIP provides labels, with both a name and a two-part channel number, for each digital service offering. The answer might then be “I am watching WXYZ on channel 7.3.” PSIP also defines the name of the program currently on-air, so a further answer can be “I am watching *Geraldo*.”

The “Where am I going?” question in the context of TV watching might be re-phrased “where do I go when I hit the CHANNEL UP button?” PSIP organizes channels into groups by means of the two-part virtual channel numbering scheme: channels are sorted for navigation first by the group associated with their common major channel number and, within each group, the minor channel number. With this sorting, CHANNEL UP is expected to take a user to the next (higher) channel number within the sorted list of all channel numbers known to the receiver.

PSIP helps with the “How can I get where I want to go?” question also, by providing two mechanisms for travel to a known destination. In some cases, the user will know the destination by its channel number. Printed program guides such as what might be found in the newspaper will identify programs by their local channel number. If the channel number is known, the DTV or set-top box will likely support direct entry of the two-part number via the RCU as is done today with the single-part channel number.

Perhaps the most user-friendly way of getting to a destination is to interact with an Electronic Program Guide (EPG) application. PSIP provides the basic program title and schedule database the EPG uses to construct and display a program guide. A user can interact with the EPG to scroll through all the channels to find and select a program of interest.

THE TWO-PART CHANNEL NUMBER

We are all familiar with “TV channel numbers.” For broadcast TV, we know VHF channels 2-13 and UHF channels 14-69. On cable, the numbers can range higher. Because it hasn’t been seen before, the two-part channel number concept introduced by PSIP is one that might take some getting used to by consumers. PSIP does not say anything about how consumer equipment will operate from the point of view of a user interface. It doesn’t even suggest what punctuation character should be used as a delimiter between the major and minor channel numbers when they are displayed or printed. All this is left for the marketplace, or popular opinion, to decide.

One might conclude that the Remote Control Unit (RCU) or TV front panel might require a new key: the “dash” key, if that is the preferred delimiter. In fact, options are possible that don’t involve adding a new key. For example, the ENTER key found on many RCUs could be used to switch modes between entry of the major and minor numbers.

Another possibility worth considering was suggested to the ATSC T3/S8 committee by TeleTV: the use of a cluster of arrow keys on the RCU: “↑” “↓” “←” and “→”. The up and down arrows would take the user from one channel group to the next (increment/decrement the major channel number), while the left/right arrows would allow navigation within the group (increment/decrement the minor channel number.

With this scheme, a mental model is formed of a two-dimensional “space” in which one navigates. In the following figure, three channel groups are shown, for major channels 6, 7 and 9. If the current channel is 7.2, an UP arrow would take the user to a channel in the 9 group (probably the most recently viewed channel in that group, though other methods are possible).

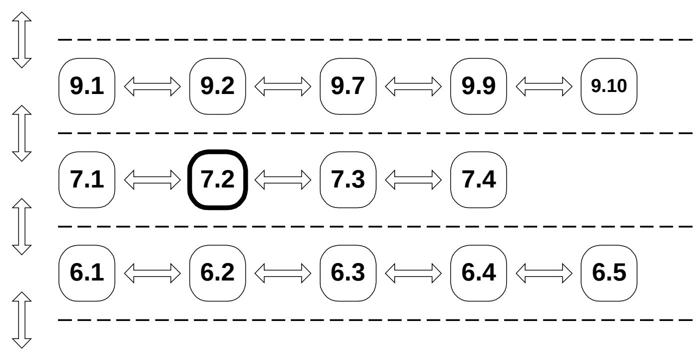


Figure 1. Example of Two-Dimensional Navigation

Entry of a left arrow from 7.2 would take the user down one minor channel to 7.1. A second left arrow would take the user either around to 7.4, or down to the highest minor channel in the next-lowest group (channel 6.5 in this case). A mental model that supports the latter interpretation is that all the channels are sorted like chapters and pages within a book, where each chapter starts numbering again from one. Use of the left/right arrows is then analogous to paging through the book, while use of up/down arrows is analogous to jumping from chapter to chapter. User testing will be needed to determine the most acceptable user interface among the many possibilities.

WHY ARE CHANNEL MAPS NEEDED?

Within the broadcaster community, the initial mindset was that channel mapping was not needed, and the extra complexity wasn’t worth the effort. When considering a 6 MHz channel carrying a digital multiplex with just one

HDTV program, the channel numbering paradigm in use for analog TV seemed to work just fine: tell the TV "47" and it will tune to the channel band assigned to channel 47 and acquire the program.

When the ATSC recognized that Standard Definition (SD) formats were possible and desirable, they acknowledged that a method was needed for selection of just one program in the multiplex. MPEG-2 *Systems* provides data called Program Specific Information (PSI) that included a parameter called a "program number" that seemed to fit this purpose. They felt that the user could specify or enter the MPEG-2 program number directly.

Those in the committees representing the cable community objected to this use of the MPEG program number as a user "sub-channel number" because of considerations related to the re-multiplexing that will be routinely done at cable headends. Cable operators need to be able to decouple the physical channel used to deliver a programming service from the user's method of access (the channel number). With this kind of decoupling, they are free to move a service to a new carrier frequency, or rearrange the delivery method of a group of services by remultiplexing, without causing user confusion.

The following is a concrete example. Let's say one broadcast multiplex is transmitted terrestrially on physical channel 27, and it consists of four SD program sources. The consumer equipment uses the FCC channel number and the MPEG program number to identify and select these sources. If MPEG program numbers 1, 2, 3 and 4 were used, to the consumer, the four sources would be identified as 27.1, 27.2, 27.3, and 27.4.

Now consider what could happen at a cable headend. The off-air signal from channel 27 could be received, and the cable operator might want to combine it with another off-air multiplex to make most efficient use of cable bandwidth. Perhaps he wishes to combine channel 27 with an off-air channel 39 that is also available. Channel 39 also has four SD program sources, and these want to be labeled 39.1, 39.2, 39.3, and 39.4. Let's say the cable operator will put the combined multiplex on cable channel 27.

If consumer equipment connected to the cable used the physical channel/MPEG program number method for labeling and identifying services, the channel 39 services would have to be accessed in some awkward way, such as at 27.10, 27.11, 27.12, and 27.13. Much consumer confusion would result.

So, since cable *must* use a virtual channel scheme, this scheme was promoted to ATSC for PSIP.

The argument for virtual channels that held the most sway with broadcasters, though, was related to the branding issue. Broadcasters were asked to consider the investment they had in brand recognition—nearly all TV channel logos in media and print advertising feature the local broadcast channel number. The channel number is recognized by the public as the way to access the service (where to get it). Now consider what would happen when a local broadcaster is given a high-numbered UHF channel to use for his new digital broadcasts. Maybe the broadcaster is known locally as "Channel 8," but now would have to use "channel 41" for the added digital services. Will he feel good about changing his logo to say "8/41"? Probably not. But what is worse, when a user sits in front of the TV and surfs channels, there are lots of channels between the old "8" and the new group of "41" channels. A virtual channel numbering scheme solves this problem by allowing the channels which are broadcast on UHF channel 41 to be labeled so they appear to the consumer as "parts of" Channel 8.

THE "PROGRAM PARADIGM"

The broadcasters also wished to use the "program paradigm," a method whereby the PID values used for audio and video were related algorithmically to the MPEG program number. Using the paradigm limits flexibility, however. If one is using the program paradigm, for example, it is not possible to define a channel that shares a video component with another channel and at the same time offers a different audio track (e.g. a different language or language rating). PSIP references the MPEG-2 service directly and thus doesn't incur such limitations. The ATSC Digital Television Standard (A/53) and the guide to its use (A/54) specify the use of the program paradigm, but ATSC now recognizes that it should be removed from the standard.

THE PSIP TABLES

The Program and System Information Protocol consists of six different kinds of tables: Master Guide Table (MGT), System Time Table (STT), Virtual Channel Table (VCT), Rating Region Table (RRT), Event Information Table (EIT), and Extended Text Table (ETT). The VCT comes in two flavors, one specifically for terrestrial broadcast called the Terrestrial Virtual Channel Table (TVCT) and the other for cable, the Cable VCT (CVCT). The following sections describe the functions of these tables in more detail.

All PSIP tables follow the same basic structure for data transport, conforming to the MPEG-2 Program Specific Information (PSI) private section data format. The "long form" syntax provides sectioning and versioning information, as well as a 32-bit CRC for robust error detection.

Like the predecessor A/56 SI protocol, all PSIP tables are extensible via the descriptor method popularized by

Europe's Digital Video Broadcasting (DVB) SI protocol. In PSIP, some descriptors are mandatory, while others are optional. A receiver that does not recognize a descriptor of a certain type is required to simply ignore it, so addition of new features via new descriptor definitions is a powerful way to add new features to the protocol. This extensibility is analogous to the extensions that have been made recently to the EIA-608 Vertical Blanking Interval protocol to define new Extended Data Services (XDS) packets for various uses.

One hard-coded PID value is chosen for the transport packets that carry all PSIP data except the program guide data and extended text. PID 0x1FFB was chosen so as not to collide with other known fixed-assigned PID values. The A/55 and A/56 protocols used PIDs 0x1FFD and 0x1FFC, respectively. Choice of a base PID that does not conflict with those used in other protocols makes it possible for dual-carriage of SI or program data, if it is ever necessary.

With ATSC's adoption of A/65, all ATSC-compatible digital broadcast television multiplexes must carry at least the following instances of PSIP tables:

- the Master Guide Table, repeated at a minimum rate of once every 150 msec;
- the System Time Table, repeated at a minimum rate of once per second;
- the Rating Region Table, repeated at a minimum rate of once each minute;
- the Terrestrial Virtual Channel Table repeated at a minimum rate of once each 400 msec.; and
- the first four Event Information Tables (representing twelve hours of program schedules).

Further EITs may be transmitted if a broadcaster wishes to provide program schedules beyond half a day. The recommended repetition rate of the EITs is twice per second.

A receiver is required to handle data rates of PSIP data in PID 0x1FFB (base PID) of up to 250 kbps. Each EIT and text PID can also be sent at a rate of up to 250 kbps.

The A/65 standard defers to SCTE for specification of data rates on cable, but does state that the MGT, STT, RRT, and cable VCT are required. Cable systems may elect to supply program guide data in a format other than PSIP's EIT/ETT.

Master Guide Table

The Master Guide Table (MGT) provides three important functions in PSIP. For each instance of a PSIP table, the MGT provides:

1. its physical location in the Transport Stream (PID);
2. the current version of the table instance; and
3. the length, in bytes, of each.

If a receiver monitors the MGT, it can look for the availability of updates to any table it may have stored. The MGT also indicates the amount of memory that must be allocated to store the updated table.

MGT is extensible such that it can, in the future, describe other types of tables that may be standardized. For example, the ATSC T3/S13 data broadcast standards group is working to define a Data Information Table (DIT) that provides schedules for data events analogous to program events found in the EIT. The MGT can easily describe location, version, and length of any instances of the DIT.

Virtual Channel Tables

The virtual channel tables in PSIP function identically to the VCTs found in the original A/56 SI standard. A Virtual Channel Table consists of one or more virtual channel definitions, where each channel is characterized by:

- the two-part (major/minor) channel number the user will use to access the service;
- its textual name (up to seven characters)[1];
- how the service is physically delivered (carrier frequency and modulation mode);
- its MPEG-2 program_number;
- its "source ID" (see below);
- the type of service (analog TV, digital TV, audio only, data);

Other data specific to each terrestrial virtual channel includes a flag that tells whether the service is access controlled or not, and an indication as to whether or not "extended text" is available to provide a textual description of the service.

The Cable Virtual Channel data structure is identical to the terrestrial version, but it defines a few bit fields that are reserved in the other. A cable virtual channel can be labeled "out of band," or "hidden." Hidden channels may be used by a cable operator for system tests—they are unavailable to a consumer unit but are accessible to test equipment. Some cable systems use out of band channels that are structured like MPEG-2 Transport Streams. The

[1] PSIP provides a descriptor mechanism to define longer channel names as needed.

Cable VCT can define a service physically carried on an out of band channel, if desired.

Descriptors

Any VCT record can include descriptors to further describe the service. PSIP defines an "extended channel name" descriptor that allows a broadcaster or cable system operator to give any channel a name exceeding the seven characters offered by the basic VCT record. The seven-character limit for the basic name label was chosen to support on-screen program guides in which a limited amount of screen real estate is available for the name text.

Another descriptor, usable only for terrestrial broadcast, quotes the PIDs used by each elementary stream component, saving the receiver from having to get this information from the MPEG Program Map Table (PMT). A third type of descriptor can be used to indicate that the channel carries programming identical to another channel, except time-shifted by a given amount.

Source ID

PSIP also makes use of the "source ID" concept introduced in the A/56 SI standard. A source ID is defined as a number that uniquely identifies a source of scheduled programming. PSIP introduces a new level of flexibility into the definition of source ID by stating that the scope of uniqueness is local to the Transport Stream for values in the range zero to 0x0FFF, and the scope is network-wide for values 0x1000 or above.

A national database has already been set up to assign unique source ID values for program sources in the US, so that in this case the "network" is nationwide. When using network-scoped source IDs, a supplier of program guide data can create EPG data in EIT or other formats that can be used as-is throughout the network. Program title and schedule data records are tagged with source ID which is linked to whatever the Virtual Channel Table defines as services available to a given receiver.

The source ID concept may find other uses as well, for example as part of a Uniform Resource Locator (URL) scheme that would be used to target a programming service. Much like Internet domain names in regular Internet URLs, such an URL does not need to concern itself with the physical location of the referenced service.

Event Information Tables

The Event Information Table (EIT) is the PSIP table that carries program schedule information for each virtual channel. Each instance of an EIT covers a three-hour time span, and provides the following information for each programming source:

- event start time
- event duration
- event title
- a pointer to optional descriptive text for the event
- program content advisory data (optional)
- caption service data (optional)
- audio service descriptor, which can list available languages (optional)

Note that descriptive text blocks are not included in the EIT itself, but instead are delivered in a separate data structure, the Extended Text Table. Separating the text from the basic schedule data allows an operator to send text at a slower rate, if desired, to make more efficient use of bandwidth.

Extended Text Tables

Each instance of an Extended Text Table carries one text block. Fields in the EIT and VCT link a program event or virtual channel to an ETT instance. As with all text delivered with PSIP, multiple languages are supported.

System Time Table

The System Time Table is the simplest and smallest of the PSIP tables. Its function is also simple: to provide a reference for time of day to receivers. In addition, the STT provides daylight savings time information.

Like A/56, the STT in A/65 bases its reference for time of day on Global Positioning Satellite (GPS) time, which is measured in terms of seconds since the first week of January, 1980. This count increments monotonically, and hence can be used as a reliable and predictable timebase for specification of future times of action.

A receiver needs one other piece of information to derive Coordinated Universal Time (UTC)[2]: the current count of the number of leap seconds that have occurred since the beginning of GPS time. The STT delivers this data as well. Leap seconds account for the difference between time based on atomic clocks (as is GPS time) and time based on astronomical events such as the earth's rotation.

The STT also provides daylight savings time status (whether or not daylight savings time is in effect), and indicates the day of the month and the hour that the next transition will occur.

---

[2] The international standards bodies couldn't agree on an acronym that reflected either English or French word order, so they chose one that reflected neither!

A receiver needs two pieces of additional information before it can use the STT data to track local time of day: 1) the offset in hours from UTC (the time zone); and 2) whether or not daylight savings time is observed locally. For a digital television, this information may be entered directly by the consumer via a unit setup function. For a cable set-top box, this information can be delivered by the system operator.

The System Time data is required to be no less accurate than plus or minus four seconds, which will make a DTV receiver one of the most accurate timepieces in the household.

The Rating Region Table—V-chip for Digital Television

The last type of PSIP table is the Rating Region Table (RRT). The function of the RRT is to define a "rating system" for a given region, where the rating system is characterized by a number of "rating dimensions," each of which is composed of two or more "rating levels." An example of a typical rating dimension used on cable is the Motion Picture Association of America (MPAA) system. The levels within the MPAA dimension include "G," "PG," "PG-13," etc.

Once a receiver learns the dimensions and levels of a rating system it can do two things: 1) provide a user interface to allow the user to set limits on program content and; 2) interpret content advisory data on individual program events. Based on a user's preference for certain program content, the receiver can block programming that exceeds a desired threshold (the V-chip function).

PSIP does not define the actual dimensions and levels of any rating region; it provides the transport mechanism to deliver the table. Specification of the actual systems for each region will be left to future standardization efforts to decide. One would hope that a standard definition for the US region can be reached in a spirit of cooperation among all interested parties.

The RRT concept originated in A/56, in almost the identical format. PSIP provides two important improvements, however:

- in PSIP, the number of rating dimensions for any region can be very large, whereas in A/56 it was limited to six;
- for added flexibility, the values that define a dimension can be declared to be based on a graduated scale, or not, as appropriate.

The table structure in PSIP allows one or more instances of the RRT to be sent, as needed, where each instance defines one region. For terrestrial broadcast, for many parts of the country, only the US Rating Region will be applicable. For areas close to the national border, however, a Canadian or Mexican rating table may be sent in addition.

In the committee meetings, broadcasters in the US repeatedly voiced their concern that the US rating system should not change often, if ever. Many felt that the system was too flexible in supporting dynamic changes to the content advisory system. A compromise was reached in that the document states that, for the US region, a "next version" of the RRT cannot be sent. The implication is that once a decision is reached by US policymakers, it will not be rescinded or altered.

With regard to the content advisory system, the architects of PSIP envisioned unity among delivery systems as a clear goal. Consider that one program sent down the cable might have originated at a local station or network affiliate, while another might have been delivered by satellite to the headend. Each of these programs should carry program content advisory data using a common rating system so as to present a consistent system view to the cable subscriber.

The same point can also be made another way. Consider what would happen if broadcasters rated TV programs using one scheme (which included some information about maturity level and sexual, violence, language, and dialog content), while cable-originated programming was rated with a similar but different scheme. (This is the case today!) If the broadcaster and cable community can't agree on the meaning of "mild violence" for example, then PSIP would have to have two "Violence" dimensions, one for network TV programming and another for cable. Users would likely be confused and annoyed by the redundancy. The fear is that if a system is too difficult to use, it isn't used at all.

EXAMPLE OF PSIP TABLES

Figure 2 shows a very simplified view of the relationship between the PSIP tables (the ETT is not shown).

As indicated, the MGT, Terrestrial or Cable VCT, RRT, and STT are transported in the "base PID," 0x1FFB. The MGT provides location (PID), version, and length (not shown) of all tables except STT. In the example, the two EITs are transported in PIDs 0x3E00 and 0x3E01.

The Virtual Channel Table defines a set of services available, including (at minimum) those on the Transport Stream carrying the VCT itself. In this case, two channels named KVGN-S and KVGN-M are shown. The VCT gives the channel numbers associated with the channels (10-1 and 10-2), the MPEG-2 program numbers the receiver would use to extract the elementary streams from the multiplex, and the source ID values for each.

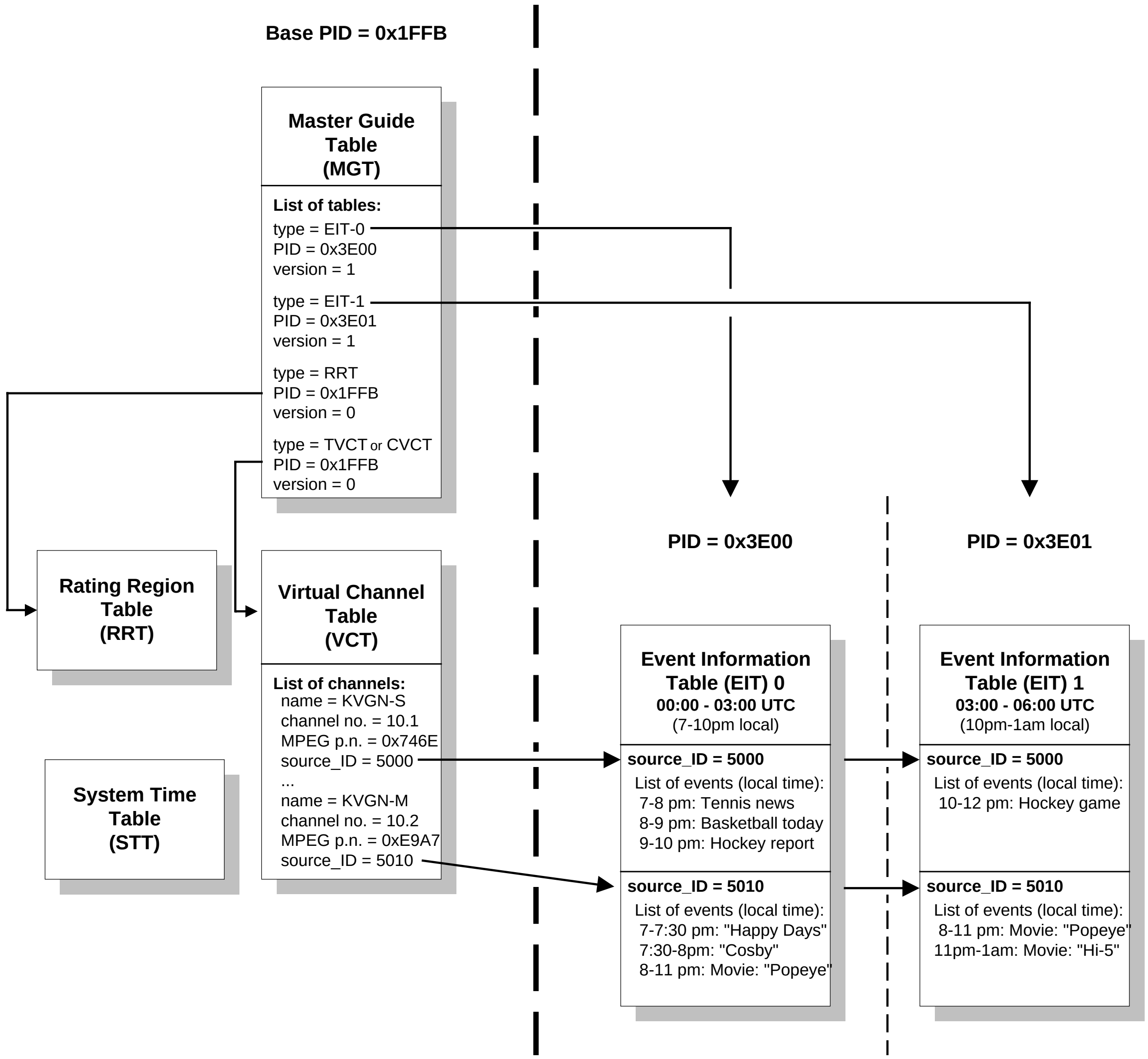


Figure 2. Relationship between PSIP tables

As indicated, source ID is used to link the VCT with the EPG data delivered in the EITs. Each PSIP EIT describes program events for a three-hour time slot, and lists, for each source ID, start times, durations, and program titles. Pointers are provided to Extended Text Tables which can provide further descriptive text.

For cable use, the EIT and ETT are optional. Figure 3 shows PSIP used on cable in a case in which the VCT links to a proprietary EPG database. A cable operator may want to offer users a program guide function that provides extended features not available with PSIP.

TEXT REPRESENTATION

Like the A/56 SI standard, PSIP uses Unicode character coding, and offers a similar method of Unicode character code page selection. But unlike A/55 or A/56, PSIP includes methods for text compression. Two Huffman encode/decode tables for English text are included in A/65. One is optimized for title text, where the first characters of words are often capitalized, and the other is optimized for the event description text. These Huffman tables were provided to ATSC by General Instrument Corporation on an unlimited-use royalty-free basis. A compression efficiency approaching 2:1 can be expected.

Another difference between A/65 and A/56 relates to multilingual text support. In A/56, one would deliver separate instances of an SI table for each supported language. In PSIP, by contrast, the language selection occurs inside the text object itself. Any text block (channel description, program title, or program description) may be defined in one or more languages. Within a single EIT, a Spanish-language television channel, for example, might have all its program titles and descriptions given only in Spanish while the other channels in that same EIT are described only in English.

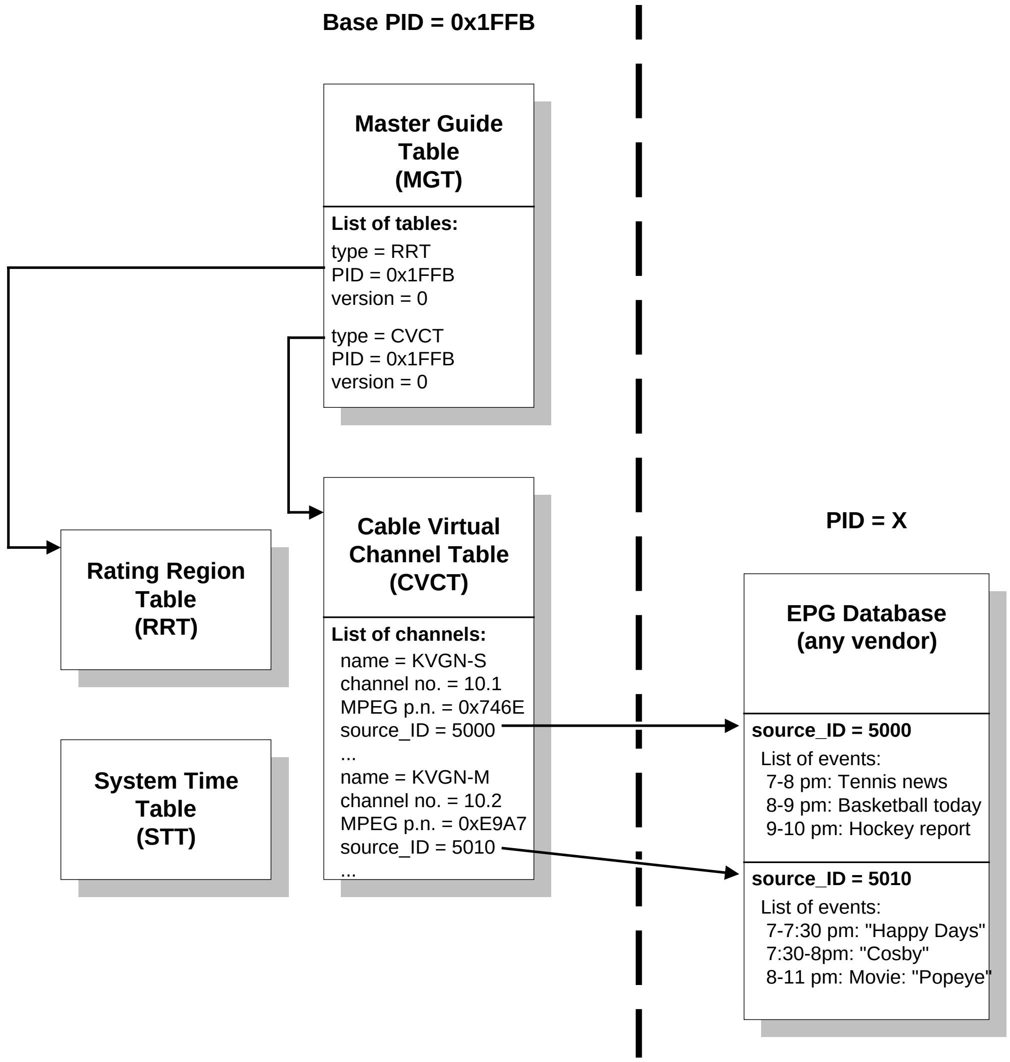


Figure 3. VCT for Cable Linked to a Proprietary EPG Database

IDENTIFYING THE SIGNALS

Transport Stream and Transmission Signal IDs

PSIP, like all current System/Service Information standards for digital television, is based on the MPEG-2 *Systems* standard, ISO/IEC 13818-1. Most digital television delivered over cable or terrestrial broadcast today uses a transport protocol defined by the MPEG-2 standard. Digital data is divided into a sequence of 188-byte transport packets. Each packet is associated with a data stream such as an audio or video service by means of a tag called the Packet Identifier (PID). Special PIDs identify packets carrying the SI and program guide data in the multiplex. The combination of service streams comprising the services and control data (SI and PSI) is called the MPEG-2 Transport Stream (TS).

The MPEG-2 standard defines a way to identify the multiplex itself so that it can be referenced within a larger network of digital multiplexes: each Transport Stream is identified by a 16-bit number called the Transport Stream ID (TSID). PSIP indicates that, in the US, the FCC will allocate values of TSID to the broadcasters to ensure uniqueness. It is expected that Canada and Mexico will cooperate to use TSID values distinct from those assigned in the US. North America can be considered a "network" in the sense that TSID values must be unique within a network.

Analog signals, until recently, had no analogous identifying tag. As a result of the ATSC PSIP work, however, the EIA has now approved EIA-752 [(6)], an extension to the EIA-608 standard for NTSC VBI data which defines an analog Transmission Signal ID (the acronym is also TSID). This 16-bit number is also expected to be assigned by the FCC.

Either analog or digital TSID values can appear in Virtual Channel Tables. In the normal case for terrestrial broadcast, the VCT will contain channel definitions for digital channels carried on the same multiplex that carries the VCT. It will also very likely carry channel data and program guide information for the analog channel associated with the broadcast digital services. The analog TSID is important for PSIP and digital receivers because it allows the receiver to verify that a received analog signal is actually the one referenced by the PSIP data.

In virtually all cases a receiver won't be able to receive a signal other than the one referenced. That's because both the analog and digital transmitters are intended to serve the same geographic area. Receivers on the fringe areas or those with directional or movable antennas, however, may be able to pick up signals other than the one expected to be found at a given frequency. A check of TSID will allow the receiver to avoid incorrect displays of channel name and/or program guide data.

Use of TSID data in the receiver actually makes it possible for the receiver to correctly display channel data and perform navigation even if the frequency data given in PSIP is incorrect. For example, a broadcast translator will shift the frequency of a transmitted signal. A receiver will find the signal when it "learns" the channel lineup, however, and it can take note of the frequency at which this particular TSID was found. The same logic applies for analog signals that have been moved to different carrier frequencies.

COMPARISON OF A/65 AND A/55-56

Some of the similarities and differences between the new A/65 standard and the predecessor SI and Program Guide standards A/55 and A/56 have already been mentioned. The following table summarizes similarities and differences.

Table 1. Comparision of A/55-56 with A/65 PSIP

Feature	A/55-56 SI/PG	A/65 PSIP
Table syntax	Mixture of MPEG long- and short-form syntax	Consistent use of long-form syntax; all tables use MPEG-2 sectioning and version control
PIDs used	A/55: 0x1FFD for MGT, plus others A/56: 0x1FFC for all Network data	PSIP uses 0x1FFB for MGT, VCT, RRT, and STT, plus others for EIT/ETT data
Text representation	A/55: ISO Latin-1, no compression A/56: Unicode based, no compression	Unicode, with two standard English language Huffman compression tables
Program rating system	A/55: hard-coded MPAA, sex, violence, and language ratings A/56: based on downloadable Rating Text Table (RTT) to define at most six dimensions	Rating system is downloadable via Rating Region Table (RRT); supports effectively unlimited number of rating dimensions
Content advisory data	A/55: included some content advisory data, not regionalized; no descriptor defined for content advisory data A/56: program rating descriptor was standardized in SCTE DVS-011	PSIP defines a descriptor usable in either the PMT or EIT; PSIP integrates content advisory data for individual programs (via PMT) with data for future programs (in EPG data)
System time	A/55: included time of day with daylight savings time indicated; A/56: included system time in GPS time format	Unified method for time representation throughout (GPS seconds); daylight savings time indication
Virtual channels	A/55: concept not present A/56: VCT is defined and utilized, with source ID linkage to EPG data	PSIP uses the A/56 VCT concept, including the source ID linkage to EPG data.

A/65 ON CABLE

Hundreds of thousands of General Instrument digital cable decoders have been deployed in North America by various cable operators conforming to the A/56 System Information standard. These decoders use an out-of-band (OOB) signaling channel to deliver network data (SI). At the current time, none of the in-band digital Transport Streams carry SI data or a Network PID.

Typical cable boxes use an out of band control channel for addressable control. The cable operator is thus afforded guaranteed access at all times to control or to update data tables in each box. SI or EPG data delivered on the OOB channel can flow at low data rates since each cable terminal will store the received data in RAM for instant access. In-band PSIP data, on the other hand, is repeated at a very high rate on the assumption that a receiver may not have had recent access to the data and needs to be refreshed as soon as possible.

PSIP was designed to allow the owner of a single digital terrestrial multiplex to include SI and EPG data describing services on that same multiplex (plus EPG data for an associated analog NTSC service). A cable-ready receiver can use PSIP on cable in just the same way: it can collect SI data from each multiplex as it is acquired and aggregate all the data into a larger channel map and EPG database. Digital cable-ready receivers and VCRs will most likely not be equipped to process out-of-band data, and will rely on in-band PSIP data for navigation. It appears likely that cable systems will move to include in-band PSIP data to support cable-ready consumer devices.

THE FUTURE

ATSC has recently finalized A/65 and already proposals for extending it are coming into the committees. As mentioned, the ATSC T3/S13 committee on data broadcasting standards is proposing a Data Information Table (DIT) analogous to the EIT, but for data "program events." Another proposal to extend PSIP's capabilities came in to the T3/S8 Transport Specialists group. The proponent wished to be able to target channel and program definitions to specific receivers, perhaps by their geographic location (via postal zip code, for example). This way, a broadcaster could better tailor programming to a target audience. Of course many variations are possible, and ATSC will have to find a procedural method to handle the many proposals that will be put on the table.

ACRONYMS

ATSC	Advanced Television Systems Committee
CVCT	Cable Virtual Channel Table
DIT	Data Information Table
DTV	Digital Television
DVB	Digital Video Broadcasting
DVS	Digital Video Subcommittee
EIA	Electronic Industry Association
EIT	Event Information Table
EPG	Electronic Program Guide
ETT	Extended Text Table
GPS	Global Positioning Satellite
MGT	Master Guide Table
MPAA	Motion Picture Assoc. of America
MPEG	Motion Picture Experts Group
PID	Packet Identifier
PMT	Program Map Table
PSI	Program Specific Information
PSIP	Program and System Information Protocol
QAM	Quadrature Amplitude Modulation
RCU	Remote Control Unit
RRT	Ratings Region Table
RTT	Ratings Text Table
SCTE	Society of Cable Telecommunications Engineers
SDTV	Standard Definition Television
SI	System or Service Information
SD	Standard Definition
STT	System Time Table
TSID	Transport Stream ID or Transmission Signal ID
TS	Transport Stream
TVCT	Terrestrial Virtual Channel Table
URL	Uniform Resource Locator
VBI	Vertical Blanking Interval
VCT	Virtual Channel Table
VSB	Vestigial Sideband

DOCUMENT REFERENCE

(1) ATSC A/53, ATSC Digital Television Standard, September 1995.

(2) ATSC A/55, Program Guide for Digital Television, January 1996.

(3) ATSC A/56, System Information for Digital Television, January 1996.

(4) ATSC A/65, Program and System Information Protocol for Terrestrial Broadcast and Cable, December 1997.

(5) ITU-T Rec. H.222.0 | ISO/IEC 13818-1:1996, Information Technology Generic coding of moving pictures and associated audio; Part 1: Systems.

(6) EIA-752, Transport of Transmission Signal Identifier (TSID) Using Extended Data Service (XDS).

* * *

Mark Eyer
Principal Staff Engineer
Advanced Technology
General Instrument Corporation
6262 Lusk Blvd.
San Diego, CA 92121
(619)404-2929
meyer@gi.com

Mr. Eyer is the author of the ATSC A/56 SI standard and one of the primary architects of the A/65 PSIP standard. He has performed system design and standards development in the field of secure and digital television delivery systems for satellite and cable for over fifteen years. He received an MSEE degree from the University of Washington in 1977.

ON PLANT RENEWAL STRATEGIES

Dan Pike
Senior Vice President
Science and Technology
Prime Cable
One American Center
600 Congress Avenue
Austin, TX 78701

Tony E. Werner
Executive Vice President
Engineering and Technical Services
TCI Communications, Inc.
Terrace Tower II
5619 DTC Parkway
Englewood, CO 80111-3000

Abstract

Just as many plants were built originally in a short time in the early 1980's, many now face renewal issues. There are many new variables beyond the traditional technological ones to consider in plant renewal strategy, including competitive risk factors and financial risk factors. This work integrates the concerns known to be influential in the process, and suggests ways to balance between them in the overall adjudication of available technology interventions for plant renewals.

This paper focuses on macro trends as well as specific considerations relating to economics, performance and network design. The Appendix includes extensive work relating specifically to node size determinants, based on most recent theory and experience in traffic engineering, noise accumulation, service levels and network availability.

Introduction

The duty of the cable industry to its constituency has always been to provide entertainment to the residential consumer. The form of technology used for this purpose has changed regularly, from the original tall towers and tube type amplifiers, to technologies such as microwave relays, satellites, fiber and computer controlled terminals, but the purpose has always remained the same--to provide more and better.

The satellite platform of the later 70's was a watershed technological platform. It enabled many new services and engendered urban viability not possible with only the previous terrestrial microwave platforms. Throughout the 80's the industry made optimum use of the technological opportunities, and today all the nation's homes are substantially passed by cable service. Now, in the 90's we contemplate the new platform with which to renew the cable facilities to continue to satisfy the fundamental duty of providing more and better information and entertainment.

This paper discusses factors important to the task, and strategies useful for the purpose.

Today's Environment

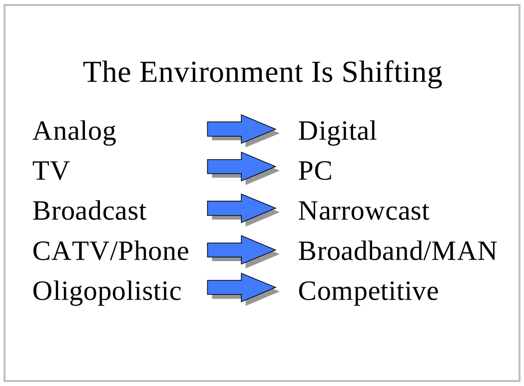


Figure 1

As today's plant renewals are being planned there are some shifts happening in the telecom environment that must be taken into consideration. (*Figure 1*) In the environment of today, traffic moves more to digital from analog form, and information and entertainment previously viewed on a television set is being viewed on a home computer. Broadcasted information is moving more to narrowcasted information. Traditional industries such as the cable industry or the telephone industry are migrating to more of a wide area data structure of a metropolitan area network and are becoming more broadband in nature, and the public policy of the nation today encourages industries that were once isolated by rules or economics to become competitive.

The Impact

These shifts will have an impact as television moves more to a digital format and information becomes indistinguish-able between text, imagery, graphics and moving video. (*Figure 2*) The social trend seen with audio will move quickly to video in the form of individual consumption. Market expansion in the form of product and packaging will follow the differentiation and stratification.

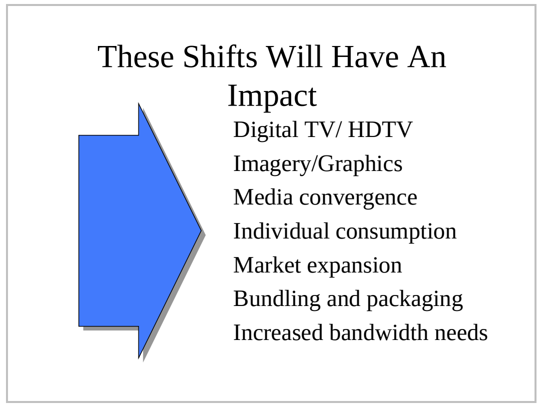


Figure 2

The common denominator of this generally manifests itself in the need for more bandwidth, but provided in ways that do not necessarily have a concomitant cost structure. Today's consumer wants more, but doesn't necessarily want to pay more.

The consumer of today is perhaps best characterized by the 3C's. They desire to be in control of their decision spending, they desire to be able to select among choices for their best value fit, and they desire to bring convenience to a busy life. Selection of plant renewal strategy must be done in consideration of these consumer attributes.

Though it has been discussed for a decade, nearly all would agree that the industry is at the verge of some substantial digital television usage. The implications of that along with the broadcast issues of must-carry and multicasting will have an impact on plant renewal strategies. Up till now, channel bandwidth used for television distribution had serious linearity constraints necessary for transmission of VSB-AM television signals. As more digital service is contemplated, the concern shifts to one of a unity gain bandwidth solution rather than one constrained by linearity. Better silicon transistors with higher F_T products, and better device technologies, such as GaAs, make unity gain easier and cheaper to attain for digital service. [1, 2] Similarly, there is little debate about the rate of increased usage of the internet and as people use the internet with its imagery and graphics in a more interactive form, the demands of the transport facility to achieve that will be more digital in nature.

Mentioned earlier, but expanded here is the concept of individual consumption. In the last twenty years audio devices have gone from one or less per person per household to many, and as video trends follow audio trends, the individual consumption of video services is an attribute that must be considered with new plant initiatives. Rather than one or two it is likely that several devices will be in use simultaneously per household for the asset life contemplated with today's renewal strategies.

Some will require the lowest cost, most ubiquitous distribution through the household, for example the breakfast area television that only requires some news and information sources fed by an additional analog outlet. Others will be more demanding of choices, as for example, the entertainment devices in the media room where pay per view, near video on demand services, and internet response features are enjoyed. All of this has a common denominator requiring more bandwidth, yet as the digital trend moves forward, bandwidth can be thought of as more than one kind, and with more than one cost structure.

Legacy Approach

In the cable industry the previous practice to consider plant upgrades was to decide at what point along one axis one wanted to go. (*Figure 3*) One simply increased the bandwidth of the cable plant as governed by economics or technology. Generally, the performance of the fundamental amplifier component was the limiting consideration, particularly the third-order distortion characteristics.

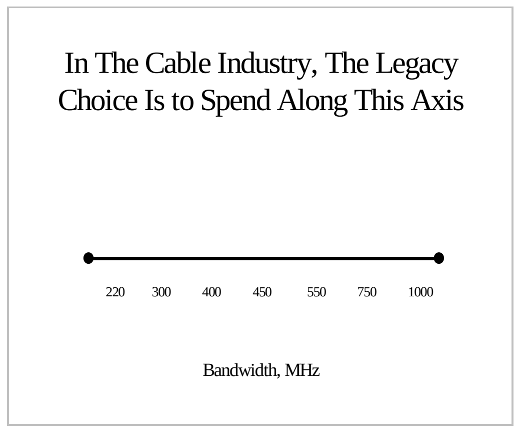


Figure 3

Now There's Another Dimension

Now there is another, digital, dimension to consider. Digital bandwidth may be considered and expressed in megabits per second, and contrasted to analog bandwidth which can be expressed in megahertz. One can equate the two, for purposes of this discussion, though not rigorously, with the aid of a modulation conversion terminal or expression of efficiency(bits/Hz).

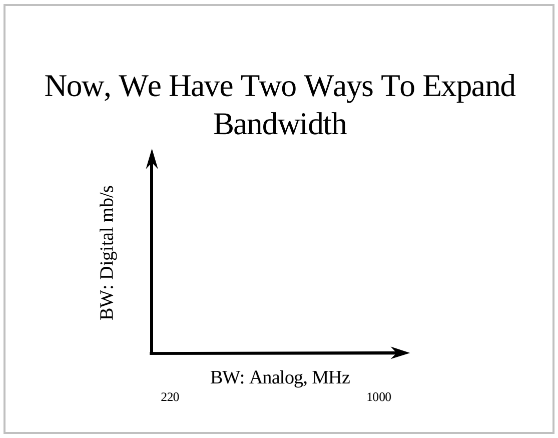


Figure 4

This means there are two axes now to consider in capital spending decisions. (*Figure 4*) We must arbitrate whether to allocate capital for digital bandwidth accomplishment or plant analog bandwidth accomplishment. *Figure 5* is an illustration of the many varieties of ways that capital can be deployed both today and in the future to achieve the situational best answer for any given set of circumstances, such as plant renewal timeframes driven by asset useful life concerns, demographics, financial considerations, and product availability.

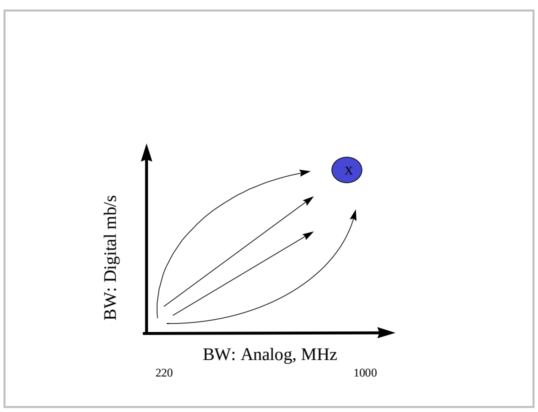


Figure 5

One could take a more digital oriented approach with the purchase of more terminals that achieve digital bandwidth, or one might take a more bandwidth oriented approach to deliver more analog oriented services, or any mixture in between. In any case it all adds up to total capital spent on total bandwidth capacity and each market consideration will determine the best trajectory for that particular location and market.

Risk Elements

Between the historical comfort of known services and plant platforms and the trepidation of new technology and services and the uncertainty it brings lies the concept of decision making under risk. This is today's case far more than in the original decisions for the cable plant platform. Compensation for imperfect knowledge may take the form of safety factors, overcapacity, or capacity built too early.

Figures 6 and 7 illustrate risk factors necessary to consider in upgrade strategy. Risk of too little and too much capital is illustrated as well as the combined risk levels of service acceptance and technological life cycles

.

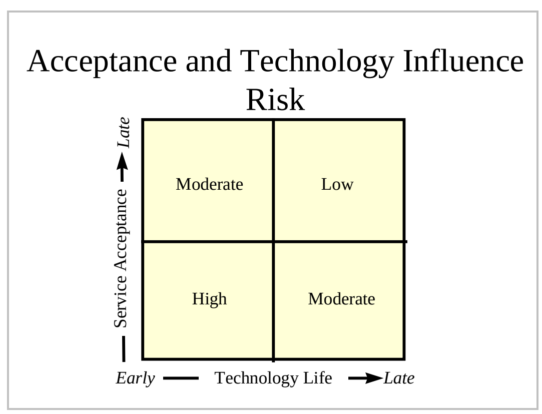


Figure 6

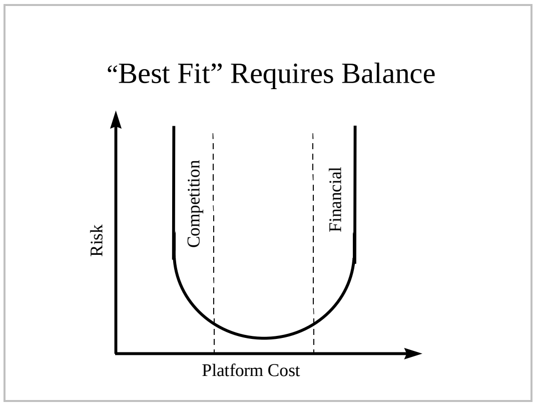


Figure 7

Conceptually illustrated here is the risk tolerance to be observed and realized in plant renewal strategy. Too much capital spending produces risk of financial nonviability; too little implies risk from a noncompetitive service offering. A broad "sweet spot" of lowest risk is the desired operating region but many different approaches are possible within this region. A possible strategy of this risk region is a digital intervention, making early use of digital transport structure enabled by digital terminals, with options of digital-grade additional bandwidth.

Digital Encoding Advantages

Figure 8 illustrates the resultant efficiency from digital modulation, or encoding, and transport, which as stated earlier can be considered casually as the numerical bridge between analog bandwidth in megahertz and digital bandwidth and megabits per second.

Judicious use of digital technology can provide several methods to mine the embedded value from existing coaxial networks. Thorough testing across several systems reveals incremental bandwidth that can be obtained for the fraction of the cost required for a full 750 MHz. This combined with the capacity offered from digital technology provides for robust service offering while reducing cost burdens.

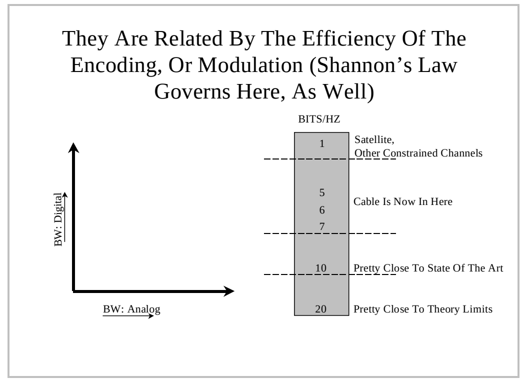


Figure 8

The efficiency of the encoding is treated in the body of science called information theory[3] which began with the publication of Claude Shannon's work about the time the transistor was invented. This very lengthy body of work describes how information can be encoded and transported in any communications channel and discusses the relative theoretical efficiencies. It is not easily summarized[4,5] into a single simple equation, but a useful concept for this discussion can be drawn to express the relationship of digital capacity (C) to bandwidth and channel quality as

$$\boxed{C = Bw\ Log_2\,(1 + {}^{S}/_{N})}$$

The satisfactory transport of the traditional VSB television signal in recent years has required a pristine and very linear channel for consumer satisfaction as described in previous investigations.[6, 7] Typical urban-grade carrier to noise ratio (typically) in the high 40's and distortion requirements in the low to mid 50's leave a very linear, very noise free channel for conversion to digital bandwidth. The total conversion range that is possible, considering the Shannon capacity limits plus the efficacy of concatenated coding[8], can be expressed as a linear range of bits per hertz multipliers beginning at about one and ending at about twenty. Our industry's linearity advantage over propagated or more channel constrained narrowband services means that the cable industry can operate in the five to ten b/Hz range while satellite and other generally more noisy channels must operate below this area. Above this area is another doubling of payload capacity before the theory limit is matched in practice as digital technology and software algorithms improve. This would be analogous to the Moore's Law progression of computer chips and the increasing speed of computer modems of recent years. Thus one can expect cost effective advantages along the digital axis brought on by fundamental Silicon Valley improvements while on the analog axis the linearity of amplifiers remains the governing criteria.

Cable vs Competitive Capacity

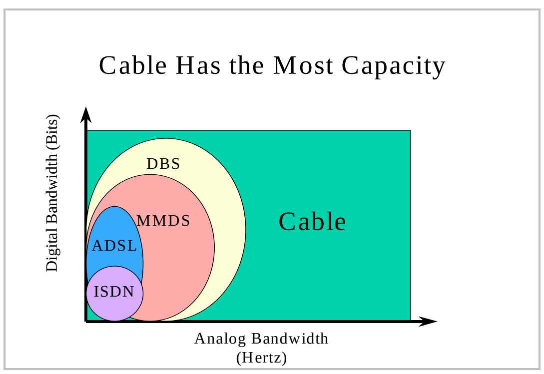


Figure 9

Many of the industry's broadband competitors have made early use of the digital axis available to them. The purpose of the conceptual illustration in *Figure 9* is to show that the cable industry is more advantaged by its intrinsic linearity for both the digital axis bandwidth conversion and the traditional analog bandwidth usage. The cable industry can go further and do better.

Best Fit Engineering

The "Best Fit" engineering approach is used to optimize the balance between network capabilities/performance, market requirements and cost. The process involves a thorough market evaluation, detailed system diagnostics and network modeling. Once completed the options are catalogued so that business and strategic decisions can be made.

This process provides several benefits. The first as alluded to above is an optimized cost structure. The second is a well defined scope of work that provides for an accurate budgeting process. The third is assurance that bad components get identified and replaced during the plant renewal and that good components do not.

Drivers on the market requirement side include current and future must-carry services, competitive offerings, personal computer penetrations and market expansion potential for new services.

There is embedded value in most of the existing networks. The objective of this engineering approach is to maximize this embedded value. Most operators for example re-use the majority of the existing coaxial cable during network renewals. The cable that is added (typically 5-20%) is to make the new design more efficient, not because the existing cable is defective or incapable of the new bandwidth.

There are several additional leverage points that can provide significant value above and beyond the coaxial cable. These largely result from the differential specification performance of passive devices and the initial design. When these are combined with new technologies including fiber optics, digital video and advanced hybrid designs, the result is often dramatic as the following cases illustrate.

Case Study 1

This case studies an existing 300 MHz system without fiber technology in place.

After significant testing of the embedded cable, passive devices and active devices it was determined that the cable provided no practical limitation to bandwidth, that the taps limited bandwidth to just past 450 MHz, the splitters limited the bandwidth to 400 MHz and the amplifiers limited the bandwidth to 300 MHz.

Distortions were analyzed based upon several different scenarios for loading and cascade length. Based upon these findings a network design was developed that installed fiber to limit the cascade lengths and a digital loading of a minimum of 50 MHz was assumed. All of the splitters and couplers will be replaced as will the amplifiers with 750 MHz equipment. The fiber architecture was designed so that if required to later move to 750 MHz, this portion of the network would not have to be redesigned. The network included two way service activation. The costs for this case, in $/home passed, are outlined below and contrasted to a 750 MHz industry standard renewal.

	450 MHz	**750 MHz**	**Diff**
Fiber Cost	$31.00	$31.00	$0.00
Coax	$33.00	$212.00	$179.00
Upstream	$18.00	$28.00	$10.00
Total	$82.00	$271.00	$189.00

Case Study 2

This case studies an existing 450 MHz system without fiber technology in place.

After significant testing of the embedded cable, passive devices and active devices it was determined that the cable provided no practical limitation to bandwidth, that the amplifiers, taps and splitters limited bandwidth to just past 450 MHz. The spacing and design of the system, however, allowed for 550 MHz operation with 50 MHz of digital loading with all of the devices remaining in their exact same location. This meant that amplifier modules and tap plates could be simply changed out which gave least time to completion and least customer disruption. Again fiber was installed to the same architecture as required for a 750 MHz system and two way was activated from day one.

The cost comparison is similarly outlined as follows:

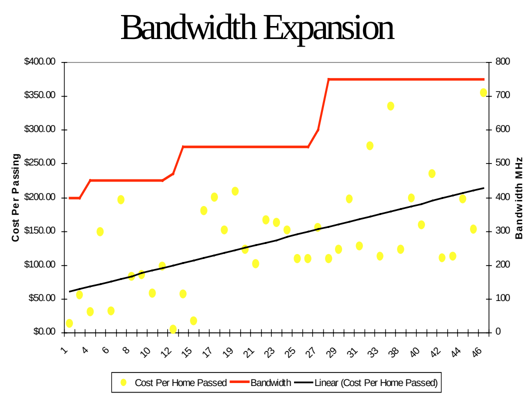
	550 MHz	750 MHz	Diff
Fiber Cost	$39.00	$39.00	$0.00
Coax	$42.00	$124.00	$82.00
Upstream	$22.00	$26.00	$4.00
Total	$103.00	$189.00	$86.00

As stated earlier, the public policy of the nation is to encourage competition among broadband providers, thus any capital spending decision must be considered with all the salient inputs of that environment. There are several ways to do this, but one of the more common is the Internal Rate of Return, which is sensitive to the expected sales success enabled by the capital platform. Using like assumptions, and a Hurwicz criterion[9] of 0.5, three cable systems were considered for three renewal strategies, with cost and IRR information detailed in Table 1. Careful study will show that each case is optimized differently, where IRR is considered as a significant indicator of competitive comparison. Moreover, the standard deviation of renewal efforts is large. Significant bandwidth improvements can be found for as low as $7/home passed to values between one and two orders of magnitude beyond. Illustrated in Table 2 are recent results from 46 projects studied. Clearly no single simple solution platform comes of this disparity, which is why the Best Fit solution is advocated.

Table 1

		Option 1 450 MHz	Option 2 550 MHz	Option 3 750 MHz
System A	Cost Per Passing	$131	$183	$233
	IRR	27%	26.6%	20.1%
System B	Cost Per Passing	$132	$152	$171
	IRR	28%	36%	32%
System C	Cost Per Passing	$151	$167	$184
	IRR	19%	25.7%	23.7%

Table 2



These studies identify situations where a reduced bandwidth approach met the capacity requirements of the local markets and provided substantial savings in implementation time and in cost.

This is not always the case though. Sometimes there is only a modest savings that can be obtained because of the existing network, density or other factors. Or in other cases there are revenue potentials that can only be realized through the additional analog bandwidth.

Figure 10 provides a histogram of selected architectures by bandwidth. These were selected based upon internal rate of returns. It should be noted that the 750 MHz architectures rarely had a more favorable rate of return, but they were close enough to the network with the best IRR, that the marginal difference was overlooked for the extra capacity.

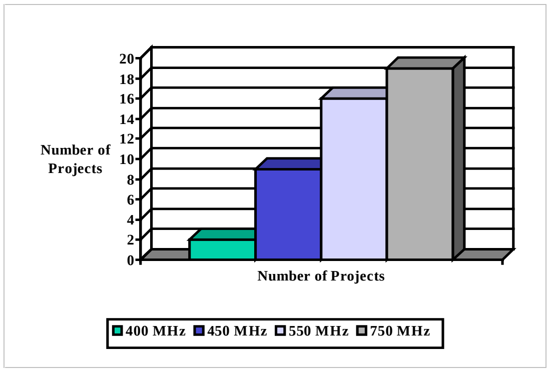


Figure 10

OPTIMAL NODE SIZE STRATEGY

The node size employed in these projects ranged from 600 homes passed to 1500 homes passed per node. Again it was determined by economics and the specific layout of the particular network including housing density.

Discussion of node size requires in-depth examination of trafficking, noise accumulation, and other assumptions. Due to its considerable length this discussion is presented in Appendix 1.0. Latest theory and experiential data is used.

The Appendix provides a substantial review of node size calculations. *Figure 11* provides an overview of the bandwidth required for various node sizes, even when modem penetration reaches 100%.

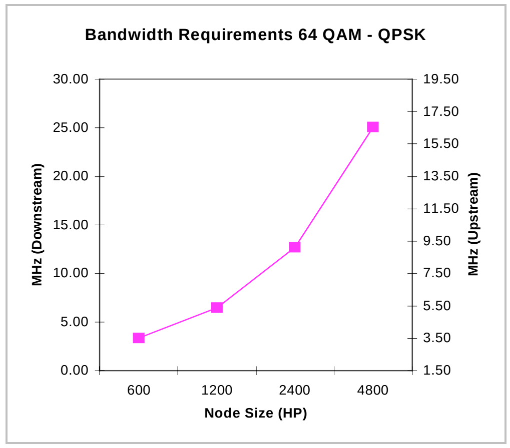


Figure 11

Other graphs are included in the Appendix report that identify the requirements when 256 QAM/16 QAM are employed in the downstream and upstream respectively.

The Appendix report concludes that HFC networks have significant capacity for data communications, so much capacity that even large nodes can support a model based upon 100% modem penetration. This, combined with technology advances in Dense Wave Division Multiplexing (DWDM) and Dense Frequency Division Multiplexing (DFDM) technologies provide for almost unlimited bandwidth capabilities.

Neither noise contributions nor service level performance limit node size designs. High performance can be achieved even with large nodes. Cascade length should be a consideration and must be defined based upon the long loop AGC capabilities of the modem technology combined with the thermal characteristics of the network.

There is no imperative from traffic, performance, maintenance or economics to design nodes below 600 homes passed. In fact, in many cases there are several arguments to design nodes to larger sizes especially in high density areas where reasonable cascade lengths are still being maintained.

Allocating two fibers for every 600 homes passed is more than ample to support the residential business applications. Because of advancements in DWDM technology, these counts may even be reduced in the future.

Conclusion

This paper integrates a range of strategies into a renewal calculus useful to achieve both digital and analog bandwidth additions to existing plants considering the latest theoretical traffic models and empirical evidence, and today's competitive environment. General conclusions supported in the discussions are that Best Fit Analysis produces the best IRR performance, node size is not a significant design variable in plant renewal, and traffic modeling for broadband facilities is highly nonlinear and not yet well understood.

APPENDIX 1.0

Considerations For Determining Optimal Node Sizes in HFC Networks

Node Size Considerations

Communication companies have been deploying Hybrid Fiber Coax (HFC) networks for the last ten years. Construction of these networks started to gain momentum in the early 90's and today, they are the network of choice for video and multi-media communications in the local loop. These networks have been designed and constructed with node sizes varying from 200 to 5,000 homes passed. Until recently, the industry has not had solid theoretical models or empirical data required to determine optimum node sizes. This paper integrates the current theoretical information that is available with as empirical data that has been derived from CableLabs experiments and actual field experience.

Factors that must be considered to determine optimum node sizes include:

- Services
- Traffic Engineering
  - ✓ Today's Requirements
  - ✓ Future Requirements
- Noise Accumulation
- Service Levels
- Technological Trends
- Maintenance Considerations

1.0 Services

Network designs in the late 1980's and early 1990's anticipated that a wide array of technologies and payloads would be utilized to provide advanced communications services over an HFC network. This included different protocols as well as different modulation and multiplex schemes for applications such as telephony, high-speed data, teleconferencing and games. It was envisioned that these different services would be frequency division multiplexed onto the HFC network.

Due of the broad acceptance of the IP protocol, several of these services can now be provided over a single, common, time division multiplexed network employing the IP protocol. This results in several advantages, including;

reduced CPE and office terminal equipment, ubiquity through national IP networks, reduced requirements for additional gateways, significantly more efficient use of bandwidth and a consolidated traffic model.

While most of the envisioned residential services have stayed the same, the method of providing these services has converged to either the cable modem or the OpenCable[10] set top device.

2.0 Traffic Engineering

This section focuses on IP traffic modeling for today's services and extrapolates the bandwidth requirements for emerging and future services. While certain discussions will relate to the global Internet backbone, the predominant focus of this paper is on the HFC portion of this architecture.

Simple mathematics do not apply to shared access mediums. If they did, 10 Mbps Ethernet LANs with 50 to 100 users would perform only marginally faster than dial-up modems. Yet Web pages representing several million bits of information are retrieved and displayed in a few seconds. The reason for this lies in the bursty nature of the data and in the manner that traffic aggregates.

IP Traffic engineering is influenced by several factors, including user behavior patterns, applications, home platforms and quality of service requirements. Residential traffic today, consists mainly of news groups and web surfing. Electronic mail, file transfers and Telnet are popular applications, but they do not account for much traffic. Traffic types will change though, as high-speed networks become more ubiquitous, promoting new services.

Several challenges exist in predicting IP based traffic over the HFC network. These challenges result from the HFC network providing an extension to the global Internet. The Internet itself is largely heterogeneous in both applications and behaviors. While the IP architecture unifies many different network technologies, it does not unify behavior in network administration or by the end user.

Adding to the challenge is the fact that the Internet is big. It was most recently estimated at more than 16 million computers (January 1997, Lottor-97). This large scale means that rare events now occur routinely in certain parts of the network.

Change is another challenge in determining Internet traffic. Change in size, with the Internet nearly doubling on an annual basis. Change in traffic characteristics including median packet length and protocol mix. But, the largest change that pales all others is the change in applications and use.

This is a moving target that is fueled by several factors. The following lists parameters that can significantly affect traffic characteristics and are causing continual change:

1. Continuous increases in processor and communication speeds;
2. Pricing structure;
3. Equitable distribution of resources instead of FCFS (first-come-first-served) scheduling algorithms (fair queuing);

4. Native (source) multicasting over the network (instead of predominantly unicasting;
5. QoS (quality of service) protocols for different classes of traffic (a must for IP telephony) and other services requiring better than "best-effort" performance;
6. Local Traffic caching, and
7. New killer applications

While change has to be accounted for in any forward looking traffic model, we have to start with a basis for how we aggregate traffic from various sources. Today there are two common methods for modeling traffic arrivals. The first is using traditional Poisson processes. The second is based upon statistically self similar modeling.

Both techniques have their appropriate application, largely based upon traffic type.

2.1 Network Characteristics

You must first analyze the characteristics of the network itself. This includes characteristics of the users, the protocols and the network medium.

2.1.1 Users

Traffic may be generated by the aggregate of several small connections (Web "mice"). Or it may be dominated by a few extremely large, one way, rate adapting bulk transfers ("elephants") or long-running, high volume data streams containing audio and video that are "multicasted" from one sender to multiple destinations. It can also consist of bidirectional multimedia traffic generated by interactive gaming.

2.1.2 Protocols

The IP (fundamental Internet protocol) is responsible for routing packets through the network. Other protocols such as TCP - transmission control protocol are built on top of the IP layer and based upon implementation differ significantly. These protocols define how to divide streams of data into packets such that the original data can be delivered to the receiver. Certain protocols will deliver this data even if some of the packets are dropped or damaged.

On top of these two layers there are application specific protocols providing such network services as email (SMTP), WWW (HTTP), file transfer (FTP), and others. Traffic modeling and simulation must take into account interactions between all the protocols in a stack.

2.1.3 Network Medium

Both the network (comprised of several, differing, networks) and the services continue to evolve, adding to the complexity and dynamics. The IP protocol allows for uniform connectivity but not uniform behavior. Moreover, it interconnects millions of users (computers) and even if some of them behave atypically, the Internet can still include thousands of these atypical users. The phenomenon, of course, changes over time from the growth in the number of users and from the growth in service types.

Additionally, new technologies and topologies add to the growing heterogeneity of the network. An example is satellite links (geostationary satellites with constant, large latency or LEO satellites with continually varying latency). New protocols will emerge for these technologies to compensate for the issues resulting from their latency and other disadvantages, this again will change the network characteristics. Other examples include fast modems such as ISDN, XDSL and ultra fast modems such as DOCSIS modems connected over HFC networks.

2.2 Traffic Characteristics

Traditionally, network arrivals have been modeled as Poisson processes. If this model is applied consistently to all network processes, it would lead to the conclusion that packet interarrivals are exponentially distributed. However, traffic studies showed this not to be true. Especially, in LANs and WANs, the distribution of packet interarrivals clearly differs from the exponential (no natural length for a burst, traffic bursts appear on a wide range of time scales).

The network can be a perfect aggregate of traffic for different applications or can be dominated by one or a couple of applications. The following are the most popular protocols and applications:

- TELNET: interactive client server communications traffic,
- FTP: file transfer traffic,
- SMTP: email is a small machine generated and sometimes timer driven, bulk transfers of data,
- NNTP: network news transport is a small machine generated and sometimes timer driven bulk transfers of data,
- HTTP: World Wide Web traffic,
- IP Telephony

There are many other processes (RLOGIN and X11, for example) that also contribute to the network traffic.

2.2.1 Aggregation Levels

The traffic will show different characteristics at different levels of aggregations. The empirical data indicate that some models used for traffic on LANs (for example self-similar process models) are not applicable to WANs. It is not known precisely what level of aggregation can be applied to the traffic on HFC networks with different levels of segmentation.

2.2.2 Distribution of Traffic Processes

2.2.2.1 TELNET

TELNET connection arrivals within one hour intervals are well described by Poisson model with fixed hourly rates. This is valid after neutralizing the dominating 24-hour pattern. Each arrival represents an individual user starting a new session.

Packet arrivals are well described by an empirical, heavy-tailed Tcplib process. This process preserves the burstiness of the packet arrivals over many time scales. This distribution is primarily network invariant and is determined by human typing characteristics. Packet inter-arrivals are far from an exponential distribution and are more accurately described by a Pareto distribution with infinite mean and variance.

Connection sizes in bytes have log-extreme distribution.

Connection sizes in packets have log-normal distribution.

2.2.2.2 FTP

User generated FTP session arrivals within one hour intervals are well described by a homogeneous Poisson model with fixed hourly rates. This is valid after neutralizing the dominating 24-hour pattern. Each arrival represents an individual user starting a new session.

Data connection arrivals, on some aggregation levels, behave as self-similar processes. These are initiated within a session, whenever the user lists a directory or transfers a file. This traffic has a long-range dependency and comes in bursts.

Byte-number distribution in every burst has a heavy upper tail. A small fraction of the largest bursts carry almost all FTP bytes during data transfer. Therefore, all that matters in analyzing the network behavior is the behavior of a few large bursts. A 5% tail for these bursts fits well into a Pareto distribution. Arrivals of the upper tail bursts are not well described by a Poisson model. Generally speaking, the FTP connection arrival processes show large-scale correlation indicating long-range dependency for data connections.

2.2.2.3 SMTP and NNTP

Connection arrivals for NNTP are not well modeled using Poisson processes. This is explained by the periodicity of machine generated IP traffic that can result in traffic synchronization which is not characterized by the Poisson model.

The SMTP connection arrivals within 10 minute periods, are well described by the Poisson model.

2.2.2.4 HTTP

Connection arrivals are decidedly not well modeled with Poisson processes.

2.2.2.5 IP TELEPHONY

This traffic will most likely follow Poisson models for telecommunication traffic.

2.3 Traffic Simulation

Traffic simulation is important, as testing with collected traffic traces does not take into account the adaptive congestion control used by the network and the sources attached. Traffic simulations and protocol simulations also allow us to test the network under different scenarios. Moreover, simulations protect the network from an unintentional overload of the network by a very successful application (so-called success disaster).

Simulation processes must be based on network invariance and judicious exploration of the parameter space. Invariance include long-term correlation's in the packet arrivals in aggregated (LAN) traffic. Internet traffic is well described in terms of "self-similar" (fractal) processes instead of traditional approach (Poisson or Markovian modeling). Long term here is defined as hundreds of milliseconds to tens of minutes. Short term behavior is affected by network transport protocols which affect short-term correlations, and longer term is affected by non-stationary effects such as diurnal patterns.

Network users session arrivals are well described using Poisson processes. Examples are remote logins and initiations of a file transfer (FTP) dialog. Unlike the packet arrivals, session arrivals are at a much higher level (exchange of hundreds of packets). This is true after accounting for diurnal patterns in session arrivals. Individual network connections within a session are not well described by Poisson processes.

Connection sizes or duration can be described by log-normal distribution with mean and variance based on empirical measurements. These, however, are not generally useful due to variations in connection characteristics from site to site.

Network activity can be characterized by a distribution with heavy tails (for example, Pareto distribution with shape parameter $\alpha<2$ with infinite variance). This conclusion is supported by such examples as Unix processes (CPU time consumed), sizes of Unix files, compressed video frames, WWW items, bursts of Ethernet and FTP activity.

The pattern of network packets generated by a user typing (TELNET) has an invariant distribution with a Pareto upper tail and a Pareto body.

Parameter space must identify all parameters that have to be exercised during simulation. This includes selecting one and varying it over several orders of magnitude and in very small steps, to account for non-linear feedback, during the simulation process. Additionally, one must exercise all parameters to determine network sensitivity to each of them independently.

2.4 CableLabs Experiment

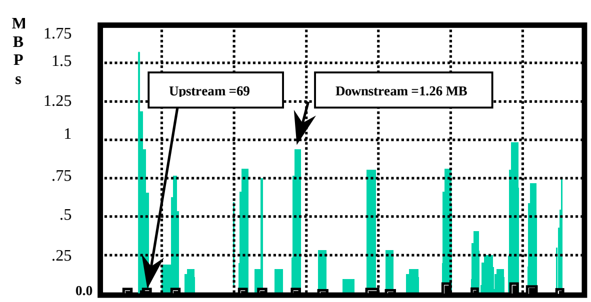
In 1996 CableLabs conducted research into high-speed data transmission over HFC networks and concluded the following:

- Large node sizes (~2,000 homes passed) work well for data services;
- Nodes can be subdivided in the frequency domain long before a physical reconfiguration is necessary, and
- That the quality of service does not degrade noticeably as up to several hundred users are added

CableLabs designed a high-speed surfing experiment to provide an understanding of WWW impacts on HFC based, shared mediums. In a controlled laboratory setting containing both the PC performing the surf and the server holding the requested pages, the engineers performed a graphically intense web surf. During this surf, the engineers analyzed the data transmissions over the broadband pipe that connected the machines. *Figure 1* provides a representative, single user web surf derived from this work.

Figure 1

Binned Throughput (Mbps)

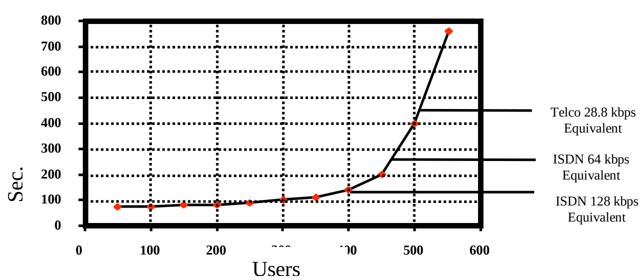


Time 60 Seconds

The solid bars indicate downstream data bursts and the dashed lines indicate upstream data. In a surf, this is usually a request or a search that creates the upstream data. The downstream burst of data is a web page being downloaded to the client. In this case, the web pages were graphicaly rich and included significant spacial and temporal information. As can be seen in *Figure 2*, bursts hit in the range of 1.6 Mb/s in the downstream and less than 50 Kb/s in the upstream. The total data downloaded over a 60 second period was 1.26 Mbytes. Concurrently the upstream data over this 60 second period was 69 Kbytes providing for an 18:1 assymetry.

Based upon this data and the MAC layer behavior of the MCNS protocol, CableLabs created a predicted target architecture performance based upon the previous WWW surf. This result was based upon a sophisticated computer model and then validated with real modems. It is illustrated in *Figure 2*.

Figure 2



Based upon *Figure 2*, the knee in the performance curve starts at approximately 400 simultaneous users on a single MCNS modem channel.

It should be noted that this experiment identifies the possible upper limit of Web surfing, not typical behavior. The pages were downloaded faster than what could be absorbed by the user.

2.5 Empirical Data

While a number of comprehensive studies have been published over the years, there are relatively few recent studies to on internet traffic and characteristics. Some of the early traffic studies include CBP93 which investigated detailed NNstat and SNMP NSFNET data in the summer of 1992. At that time, application usage was reported as being dominated by FTP, SNMP, NNTP, DNS, Telnet and a growing amount of "other" traffic.

Merit's final NSFNET report summarized statistics on the backbone from 1988 to 1995 where it was noted that information retrieval services such as Gopher and Web were beginning to overtake mail and file transfers in their share of network traffic. Traffic make up in packet counts in the spring of 1995 are outlined in Table 1.

Table 1

Other	27%
WWW	21%
FTP-data	14%
NNTP	8%
Telnet	8%
SMTP	6%
IP	6%
Domain Name Server	6%
IRC	2%
Gopher	2%

This distribution is based upon packet count and would change significantly based upon byte count because of the distribution of packet sizes.

Paxson published a comprehensive report in 1997 that analyzed 20,000 TCP transfers among 35 sites and over 1,000 internet paths. His findings examine such areas as route pathologies, loss characteristics, specific TCP implementation behaviors and packet queuing and delay differences. His research is the basis of several of the preceding sections that describe the nature of different traffic sources.

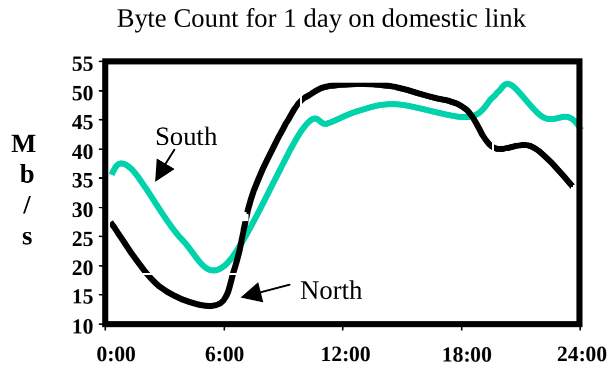
2.5.1 MCI (vBNS)

In the fall of 1997, Thompson – Miller - Wilder published results from monitoring the MCI very-high speed Backbone Network Service (vBNS) for the National Science Foundation (NSF).

The MCI paper provides daily and weekly traffic patterns and composition for two traffic points on the commercial internet backbone. In this section we will highlight some of the key findings from point one, which was a domestic traffic point versus point two which was an international traffic point. Point one was within a node that serves as a junction point for several backbone trunks as well as an access point for local customer traffic near a major U.S. East Coast city.

Figure 3 below is an approximate recreation of the byte volume for a one day record on the domestic link. Note that there are some differences in communication symmetry, which also changes over time.

Figure 3



These curves show consistency with other patterns whereby the traffic volume nearly triples from 6:00 am to noon and then holds a moderately steady pattern until near midnight when it starts to trail off.

Weekly traffic patterns indicated that weekday traffic was much heavier than weekend traffic, but the daily patterns were similar, especially when time zone impacts are considered. An HFC network will differ as it will be much more asymmetrical as it relates to byte counts and it will likely not diminish as much during the weekends.

Again the MCI study showed significant modal distribution of packet sizes based upon protocols. Over 50% of the packets were smaller than 45 bytes and nearly all were smaller than 1500 bytes. This modality again brings into question the value of average packet lengths when discussing traffic characteristics. This conclusion may be different for HFC networks, where by definition they will have packet sizes that will range from a minimum of 64 bytes to a maximum of 1500 bytes.

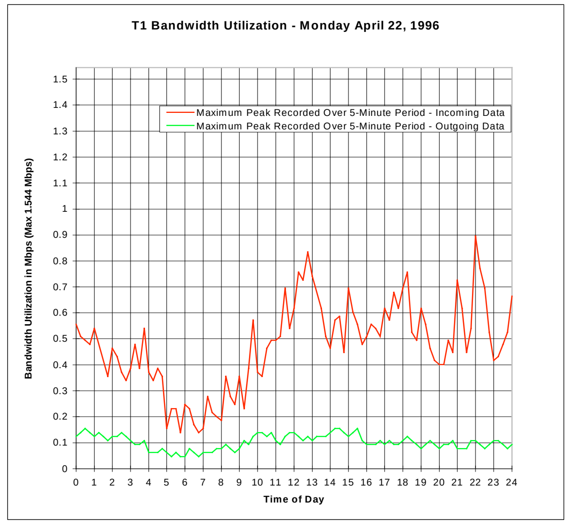
Traffic composition for the MCI study revealed that for IP protocols, TCP was the predominant protocol accounting for over 75% of the flows, 90% of the packets and over 95% of the bytes. UDP accounted for the majority of the remainder accounting for 20% of the flows, 10% of the packets and 5% of the bytes. The remainder was accounted for by emerging protocols, such as IPv6, encapsulated IP, ICMP and others.

Within TCP, the most predominant application is WWW, accounting for 75% of the bytes. FTP accounted for 5% of the bytes, NNTP accounted for 2% of the bytes and Telnet accounted for 1% of the bytes. The remaining traffic is distributed over several other types of traffic. UDP traffic mainly consisted of DNS and RealAudio, RealVideo apps.

2.5.2 Rogers Engineering

Rogers Engineering monitored HFC and ISP gateway utilization during the Spring and Summer of 1996. During this time, they had 471 early adopter data customers connected to Zenith high speed modems over an HFC network. This network was connected to the internet via a T-1 connection and during this time traffic caching was not performed. The T-1 utilization is graphed in *Figure 4* for a typical day out of the analysis. You will notice that even with 471 customers that the peak aggregate bit rate never exceeded 900 kb/s. The pattern is, however, quite similar to that identified in the backbone study by MCI where by dramatic traffic increases occurred between 6:00 AM and Noon. The bottom trace indicates the outgoing data. Asymmetry in byte count, increases as traffic increases, largely based upon the traffic types. It starts out as low as two to one and grows to as high as ten to one during the peak usage. The CableLabs WWW experiment showed that this asymmetry could grow as high as eighteen to one during exaggerated peak usage, especially if it was based upon pure WWW traffic.

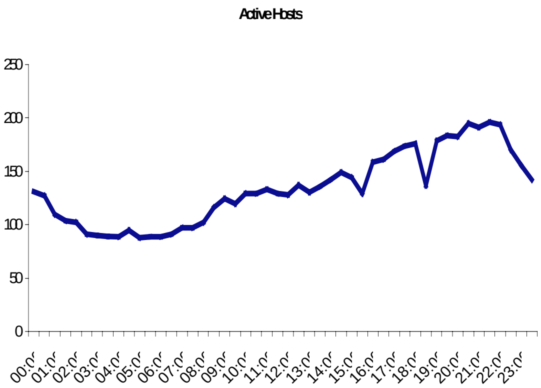
Figure 4



Active users are plotted for a typical day in *Figure 5*. In it, you will notice that the curve differs from byte count graphs shown previously. This is largely accounted for by users that leave their computers on all evening, even though they are not being actively used.

This graph depicts the highest simultaneously active hosts from the study, whereby 194 of the 471 (41%) hosts were active. This peak occurs at 9:30 PM and represents a growth from 5:00 AM when 19% were active. It must be noted that active terminals do not represent simultaneous use as it is common practice to have active, but idle computer terminals. This pattern and its growth supports the typical industry standard for a 25% peak simultaneous usage.

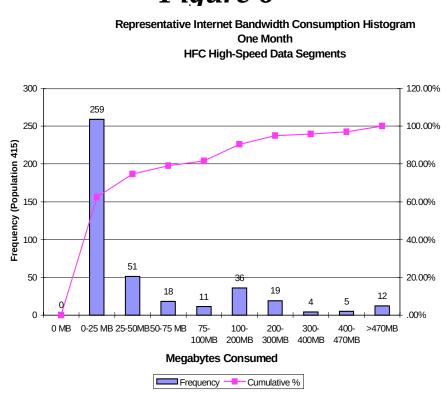
Figure 5



Traffic composition based upon the Rogers experience indicated that WWW (HTTP) and News (NNTP) accounted for 98% of the traffic volume in bytes. News (NNTP) traffic accounted for a higher percentage of the HFC traffic than it did on the backbone. This became especially true as the operation implemented local caching of the News group data. This combined with the early adopter profiles is the predominant reason for this characteristic.

User profiles indicate that the majority of the users are similar and in the case of Rogers, 62% of the customers consumed less than 25 MB of data over a one month period. The remaining heavy users form a fairly long tail. This characteristic becomes even more predictable as larger numbers of users are present and the early adopter impact has less of an effect on the total usage.

Figure 6



2.6 HFC Capacity

The preceding sections have illustrated that it is nearly impossible to develop a traffic model that can account for all of the dynamics that will affect traffic and user behavior as we move into the future. One can, however, model an upper bound that will demonstrate the ample capacity of HFC networks even with relatively large node sizes. Operators have typically taken a significantly conservative approach up until this time in traffic provisioning. This has resulted in many under utilized ISP connections and the practice of node combining in the HFC networks themselves.

As can be seen from the previous studies, even though peak burst rates are often high, the aggregation of this traffic is quite modest. This is what allows the backbone to operate with relatively small circuits and the 471 customers in the Rogers early experience to only require 900 kb/s.

The remainder of this section demonstrates the significant capacity of HFC networks for two-way data-communications. This section demonstrates that even large nodes can support 100% modem penetration and incredibly high consumption and peak characteristics while allocating only a modest portion of the up and downstream bandwidth.

The following provides assumptions based upon a 100% modem penetration as a result of the OpenCable initiative. It should be noted that while there may be 100% modem penetration, only a percentage will be attached to stand alone computers. The remainder of the modems will provide communications to the OpenCable terminal, which will operate as a network computer. This device will have a powerful processor, but it will not contain substantial local storage or rapid input mechanisms. Because of this architecture, the upstream communications will mainly consist of short packets used to request information. It will also serve a much different need and application than that of the personal computer.

Table 2 provides assumptions baselines for the calculations. We start with the assumption that 100% of the cable households have a high-speed modem installed and operational. We use today's penetration of 62% for cable penetration. We base PC penetration at 50% which is much higher than average penetrations today.

We have divided the users into power users and casual users. Power users would be those who spend considerable time on the PC and who have the latest generation processor, etc. The baseline uses a 40%-60% split respectively. In the Rogers experience, 62% of the customers fell into the bin of 0 to 25 MB per month. From there, the usage distribution had a long tail, with 80% of the users consuming less than 75 MB per month.

For simultaneous peak usage, we have used 25%. This number is based upon today's low penetration levels and early adopter make up. This number should actually go down as penetration increases. With the move to OpenCable types of services, TV viewing habits will drive a substantial amount of the usage. Typically the maximum TV on rate is in the high 60% range. This usage is split based upon market share, which today, usually breaks this down into small segments. Major interest items will typically use a multi-cast approach and also reduce peak data requirements.

Telephone penetration is calculated at a 30% penetration and a 3 CCS traffic load. This is the average residential usage today. This is likely conservative as these homes will typically have two phone lines and high-speed data access through the DOCSIS modem.

IP video telephony is estimated at a 20% penetration and a 1 CCS utilization. In other words, consumers will use video connections one minute for every three minutes that they are using the standard POTS line. Again this is likely extremely conservative as these services will have a rate differential and they will require that you connect to another customer who has this video capability. Just like wireless communications, customers will likely use wireline POTS if it will satisfy the business or personal need, before they will use the more expensive service.

Table 2

Penetration Baseline

Customer Penetration	62%
PC Penetration of Customers	50%
Power User	40%
Casual User	60%
Simultaneous Peak	25%
OpenCable Penetration	100%
Simultaneous Peak	25%
IP Telephony Penetration	30%
Usage CCS	3
IP Video Telephony Penetration	20%
Usage CCS	1

Table 3 provides estimates for peak average data rates. For WWW, FTP and NNTP applications, we have used the 168 kbps downstream and 8 kbps upstream results that were derived from the CableLabs high-speed surfing experiment. Again this is not burst speed but aggregate speeds that are then averaged. We have used 80 kbps and 1,544 kbps for real-player audio and video respectively. Both of these are typical of today, but likely conservative for the future. IP telephony is based upon 32 kbps for voice and 128 kbps for video. This should likely be conservative as well. Today the majority of the IP voice is compressed to levels 50% to 60% lower than this 32 kbps rate.

Table 3

Peak Rate Baseline Kbps

	Peak Rate	
	DN	UP
Protocol/Application	Stream	Stream
WWW	168	8
FTP	168	8
NNTP	168	8
RealPlayer Audio	80	2
RealPlayer Video	1,544	8
IP Phone	32	32
IP Video Phone	128	128

Tables 4 and 5 provide the capacity of HFC networks based upon 64 and 256 QAM downstream modulation schemes and QPSK and 16 QAM upstream modulation schemes.

Equipment is already being shipped with 64 QAM and QPSK schemes today with 256 QAM and 16 QAM equipment likely available in the summer of 1998.

Table 4

Downstream Capacity Baseline

	64 QAM	256 QAM
Downstream Bandwidth MHz	6	6
64 QAM Gross Data Rate Mbps	30	43
Minus Error Correction Mbps	3	5
Minus MAC and Messaging Mbps	1	1.5
64 QAM Net Data Rate Mbps	26	36.5

Table 5

Upstream Capacity Baseline

	QPSK	16 QAM
Upstream Bandwidth MHz	6.4	6.4
Gross Data Rate Mbps	10.2	20.4
Minus Error Correction Mbps	1.4	2.8
Minus MAC efficiency and OH Mbps	3.7	7.5
Net Data Rate Mbps	5.1	10.1

Table 6 derives the number of users for four different node sizes based upon the preceding baselines. This table initially indicates the total number of customers and then extrapolates this for simultaneous use.

Table 6

User Calculations

Node Size (HP)	600	1200	2400	4800
Customers 62%	372	744	1488	2976
OpenCable Customers	372	744	1488	2976
PC Customers	186	372	744	1488
Power Users	74	149	298	595
Casual Users	112	223	446	893
Simultaneous Active *(not including telephony customers)*				
OpenCable Customers	93	186	372	744
PC Customers	47	93	186	372
Power Users	19	37	74	149
Casual Users	28	56	112	223

Tables 7 and 8 calculate the number of lines required to provide a P.01 grade of service based upon the preceding baselines. This is based upon Poisson addition and then discounted the trunking efficiency for small statistical groups. The trunk efficiency typically would be higher as one moved from the 600 home example up to the 4800 home example. Under this IP example, however, blocking will technically be determined by the bit priority that is assigned.

Table 7-IP Phone

Node Size (HP)	600	1200	2400	4800
IP Phone Customers	112	223	446	893
Total CCS	335	670	1339	2678
Lines Required for P.01	26	46	82	152
Trunk Efficiency	0.9	0.9	0.9	0.9
Lines	29	51	91	169

Table 8-IP Video Phone

Node Size (HP)	600	1200	2400	4800
IP Video Phone Customers	74	149	298	595
Total CCS	37	74	149	298
Lines Required for P.01	8	11	16	26
Low Trunk Efficiency	0.9	0.9	0.9	0.9
Lines	9	12	18	29

Table 9 defines the characteristics of Power users, again based upon the preceding baselines. Notice how conservative the model is. We are now up to an average peak usage of 322 kbps in the downstream. If one reinserted this into the previous data on the Rogers customers, the peak usage of 900 kbps that they recorded would have jumped to 38 Mbps or an increase of 42 fold. A customer of this characteristic would consume as more data in ½ hour than the average Rogers customer did in an entire month. This is an extremely aggressive assumption for any user, and contributes to make this model highly conservative.

Table 9

Power User Characteristics (kbps)

Application	DN Stream	UP Stream
WWW/FTP/NNTP	168	8
Real Player Audio 20%	16	2
Real Player Video 1 - 6 Sec. Clip	154	1
Total	322	10

Tables 10 and 11 define user characteristics for casual PC users and couch surfers. Again these are average peak usage, not burst speeds. Again these extrapolate into extremely aggressive usage patterns.

Table 10

Casual User Characteristics (kbps)

Casual User	DN Stream	UP Stream
WWW/FTP/NNTP	84	4
Real Player Audio 20%	8	0.2
Total	92	4.2

Table 11

OpenCable Characteristics (kbps)

Open Cable	DN Stream	UP Stream
Web Retrieval	42	2

Table 12 now uses these preceding baselines and calculations to determine cumulative data requirements for four different node sizes.

Table 12

Cumulative Bandwidth Requirements (Mbps)

Bandwidth Requirements Mb/s								
Node Size (HP)	600		1200		2400		4800	
Stream Direction	DN	UP	DN	Up	DN	UP	DN	UP
Services								
IP Phone	0.9	0.9	1.6	1.6	2.9	2.9	5.4	5.4
IP Video	1.1	1.1	1.6	1.6	2.3	2.3	3.7	3.7
Personal Computer								
Power User	6.0	0.2	12.0	0.4	23.9	0.7	47.9	1.4
Casual User	2.6	0.1	5.1	0.2	10.3	0.5	20.5	0.9
TV Based Web Services								
Open Cable	3.9	0.2	7.8	0.4	15.6	0.7	31.2	1.5
Total	**14.5**	**2.5**	**28.1**	**4.2**	**55.0**	**7.1**	**108.7**	**13.0**

A 600 home node only requires 14.5 Mbps in the downstream and 2.5 Mbps in the upstream to satisfy 100% modem penetration and aggressive and even perhaps unrealistic data use assumptions.

Figure 7 and 8 graph the RF bandwidth requirements for networks based upon 64 QAM/QPSK and 256 QAM/16 QAM modulation schemes.

Figure 7

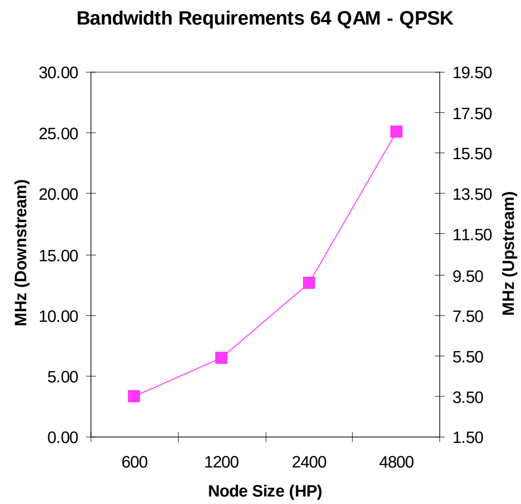
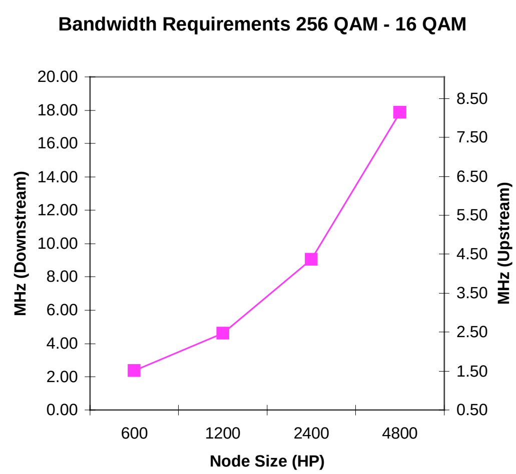


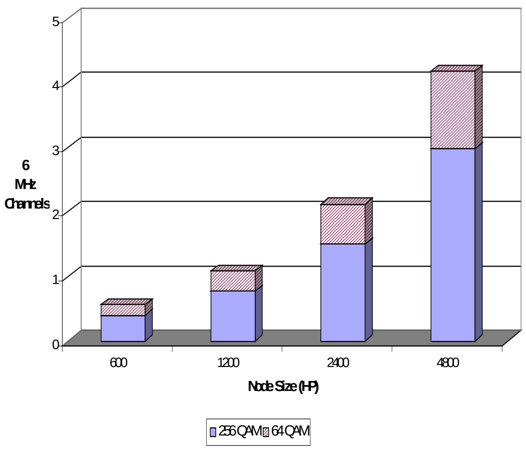
Figure 8



Note that even with QPSK modulation schemes and a 4800 home passed node you can support 100% modem penetration with less than 18 MHz of upstream spectrum.

Figure 9 charts the number of downstream channels that are required to support various sized nodes based upon 256 QAM and 64 QAM modulation schemes.

Figure 9



3.0 Noise Accumulation

Larger nodes do have a higher accumulation of noise from active device contribution. The increase, however, is not a material design factor. Let's first look at the active components, their operational levels and noise contribution as it applies to the upstream network.

The design assumptions for an upstream network are listed in Table 13.

Table 13

Laser Type	DFB
Laser Operating Levels	-56.5 dBmV/Hz
Return Amplifier	Hybrid
Amplifier Operating Levels	-54 dBmV/Hz
Laser Equivalent Noise Factor	27.6 dB
Hybrid Noise Figure	4.5

These design guidelines provide for a 15 dB operating to clipping ratio. Because the dynamic environment of the upstream network and the effect of laser clipping on digital signals, an adequate amount of headroom is critical to return path operation. This allows for level changes and impulse noise effects.

Based upon these design parameters, the carrier to noise resulting from the optical network would be:

Table 14

Noise Floor	-125.1 dBmV/Hz
Minus Input Level	-56.5 dBmV/Hz
Plus Equivalent Noise Figure	27.6 dB
Optical Network CNR	41 dB[11]

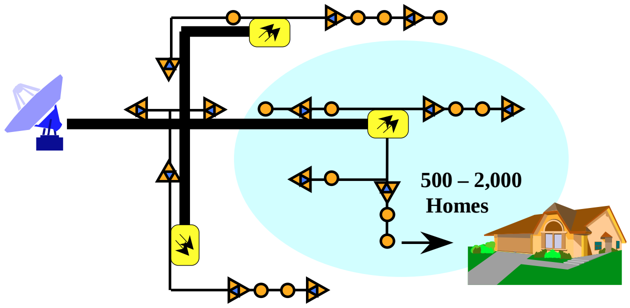
The carrier to noise ratio resulting from a single hybrid would be as follows:

Table 15

Noise Floor	-125.1 dBmV/Hz
Minus Input Level	-54 dBmV/Hz
Plus Noise Figure	4.5 dB
Single Station CNR	66.6 dB

Figure 10 depicts an HFC network indicating nodes that vary from 500 to 2,000 homes passed.

Figure 10



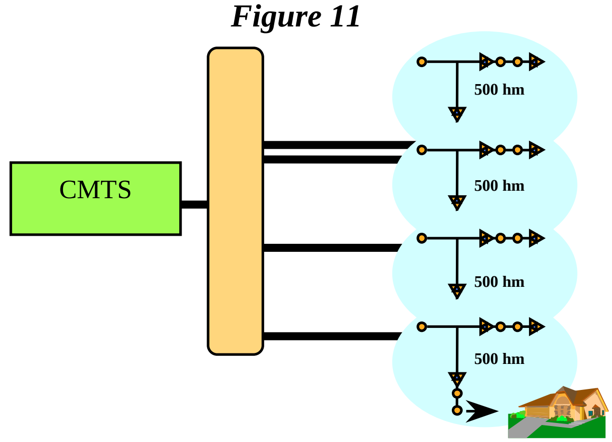
Based upon 5 amplifiers per mile and 100 homes per mile, a 500 home passed node would include 25 amplifiers. The $10\text{x}\log_{10}$ of 25 is 13.97 providing a 52.6 dB CNR for the entire 25 amplifiers. When this is added to the 41 dB CNR from the optical network we arrive at a 40.7 dB CNR for the 500 home node.

A 2,000 home passed node would consist of four times as many amplifiers or 100. The $10\text{x} \log_{10}$ of 100 is 20, providing for a 46.6 dB CNR for the entire 100 amplifiers. When this is added to the 41 dB CNR from the optical link, we arrive at 39.9 dB CNR. Notice that quadrupling the node size provides only a 0.8 dB loss of CNR performance. This analysis is based upon the best laser that is practical to deploy in the upstream network today. Several operators are employing lower cost lasers such as cooled FP lasers. Under these situations there would be practically no improvement in CNR performance of a 500 home passed node over that of a 2,000 home passed node.

Because of the cost and space required for cable modem termination (CMTS) equipment, most operators are serving between 2,000 and 5,000 homes passed from a single channel termination. Some operators are doing this from a single optical node, others have actually combined smaller nodes to arrive at this level. Provisioning a single cable modem termination for a 500 home or even 1,000 home passed node offers poor economics at penetration levels below 15%.

Figure 11 depicts the node combining that is performed to group four 500 home nodes into a 2,000 home passed area to satisfy the CMTS economic and space criteria.

Figure 11



Under this example, the output of four optical receivers is combined at RF and fed into the CMTS. The CNR performance for this example now consists of 100 amplifiers plus four lasers. The four lasers now produce a combined CNR of 41 dB $-10\log_{10}$ of 4 or 35 dB. Note the combined effect of the optical transport and the RF Transport is 34.7 dB CNR. Note that this configuration offers the poorest performance of all of the examples.

Table 16 depicts noise based upon four examples, the three previously illustrated and a 10,000 home passed RF only network.

Table 16
CNR (dB)

	500 HP	2,000 HP	4x500 HP	10,000 HP
RF Network	52.5	46.5	46.5	40
Optical Network	41	41	35	N/A
System	40.7	39.9	34.7	40

Table 16 illustrates the dominant noise contribution from the optical network in upstream applications. This is largely caused by the low optical modulation index requirements for return applications. After laser combining, the four 500 home passed node example provides the poorest noise performance of any of the examples.

4.0 Service Levels

Early on, many network engineers and system designers hypothesized that smaller nodes would yield much higher levels of service availability. This was based upon the lower levels of network contributed noise, the lower level of impulse noise accumulation from home appliances and shorter cascades of RF amplifiers.

Actual implementations, however, have demonstrated that there is minimal if any correlation between node size and service performance of two-way services. *Figure 12* is provided by CableLabs. In it, CableLabs tested and charted the performance of several different networks across North America. After tabulating the results, they categorized the network performance of these systems by node size. As the chart illustrates there is no correlation between service performance and nodes size. In fact in this case the 250 home passed node had an equivalent number of bit errors as did the 2,000 home passed node.

Figure 12

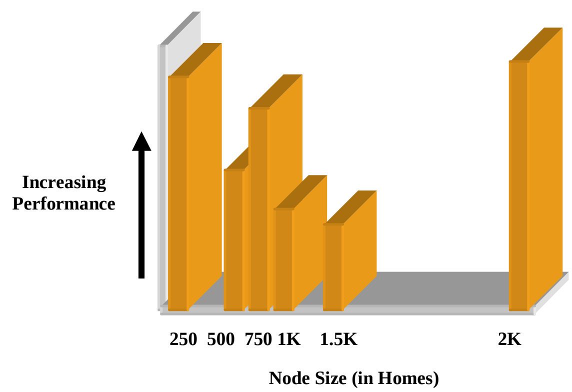


Table 17 is also provided by CableLabs and provides testing results of a 250 home passed node, a 600 home passed node and one 2,000 home passed node. Again you will notice that there is no correlation between node size and bit error rate performance. In these results the 2,000 home passed node exhibited superior performance to the 600 home passed node and nearly identical performance to the 250 home passed node.

Table 17

Node Size Homes Passed	% Performance Above Threshold	
	10^{-4}	10^{-3}
250	99.94	99.95
600	99.87	99.93
2,000	99.93	99.95

The explanation for these results is fairly simple. It isn't that node size doesn't have an impact on two way service performance, it is simply that the node size impact is extremely small when compared to other factors which include maintenance standards and component selection.

System amplifier noise accumulation improvements from a 500 home passed node versus a 2,000 home passed node is only 0.8 db and this is when a high grade, DFB laser is used in the return path. If a Fabry Perot laser is used, this improvement becomes even lower.

Because of the random arrival nature of impulse noise, it does not show statistical addition based upon node size. Either you have impulse noise at a level that is impacting service performance or you don't. Several non service impacting impulse noise sources do not aggregate to cause a service affecting impairment.

Unless RF amplifier cascade lengths exceed the thermal range that can be compensated for by the long loop AGC circuit, they will not materially reduce service level performance. RF amplifiers today typically experience MTBF of 60 years and MTTR of 60 to 90 minutes.[12]

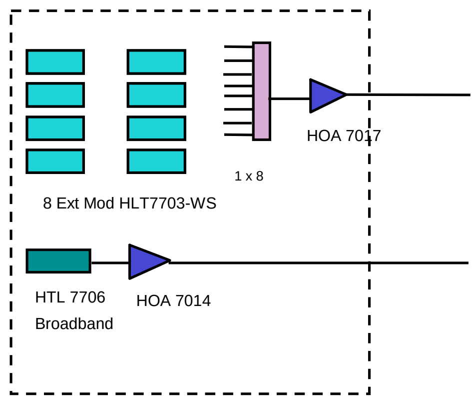
5.0 Technological Trends

Experienced design engineers know that they must understand the current state of technology, but even more importantly the future state of technology. This is extremely applicable to node design as laser prices are falling at a rate of 3% per month and new technologies such as dense wave division multiplexing (DWDM) and small form factor dense frequency division multiplexing (SFF-DFDM) technology is rapidly progressing.

These two latter technologies allow system designers to employ scaleable designs that can exploit the nearly unlimited bandwidth of HFC networks.

Figure 13 illustrates equipment that is available today to provide DWDM for up to 8 separate wavelengths

Figure 13



DFDM technology allows for 8, 5-40 MHz sub-low segments to be multiplexed into each wavelength. If one concatenated the two technologies it would allow for 64 nodes to be carried on a single fiber.

Initially, standard WDM technology can be deployed that will allow for two wavelengths per fiber (1310 nm and 1550 nm). These devices are under $500 today and on a rapidly declining cost curve. Use of these devices is always less expensive than installing additional fiber and usually less expensive than deploying higher fiber counts initially.

6.0 Maintenance Considerations

Maintenance and reliability considerations must be taken into account when determining network architecture and node sizes. Fiber installation reduces maintenance costs and improves reliability. HFC reliability has been demonstrated in theory and also in practice through many commercial launches. Fiber plays an important part in this, but node size correlation has not been identified. Nodes of 300, 600, 1200 and 2500 have been compared with no visible correlation to maintenance costs.

The majority of maintenance improvements result from cascade lengths, not node size. This is true also of reliability. Reliability has been mainly the result of power loss, cut cables and craft error. Each of these items are a function that multiplies with cascade length, not homes passed. It needs to be pointed out that increased cascade length for a given home density increases the total homes. The point is that there is no demonstrated reason to artificially limit cascade length to preserve a low number or homes passed per node.

7.0 Node Size Considerations Summary

HFC networks have significant capacity for data communications, so much capacity that even large nodes can support a model based upon 100% modem penetration. This, combined with technology advances in DWDM and SFF-DFDM technologies provide for almost unlimited bandwidth capabilities.

Neither noise contributions nor service level performance limit node size designs. High performance can be achieved even with large nodes. Cascade length should be a consideration and must be defined based upon the long loop AGC capabilities of the modem technology combined with the thermal characteristics of the network.

There is no justification from traffic, performance, maintenance or economics to design nodes below 600 homes passed. In fact, in many cases there are several arguments to design nodes to larger sizes especially in high density areas where reasonable cascade lengths are still being maintained.

Allocating two fibers for every 600 homes passed is more than ample to support the residential business applications. Because of advancements in DWDM technology, these counts may even be reduced in the future.

Acknowledgments

We would like to thank Rogers Communications for their contributions of empirical user behavior data and CableLabs for their contributions relating to WWW surfing and to service levels based upon node size. We acknowledge helpful discussions with Mark Laubach of Com21 and Tom Moore of CableLabs on the traffic models and Dr. Terry Shaw of CableLabs on the Shannon limits. We appreciate the review and counsel given by colleagues.

We would like to specially thank Oleh Sniezko and Esteban Sandino of TCI for their contributions to the sections on theoretical traffic engineering and empirical data. We note heavy use of the traffic references in that section.

REFERENCES

[1] S. Beauchamp, "Throughput Analysis of Com21 Service Levels and ISDN," Application Note, Number 280-0024-D1, August 4, 1997.

[2] CableLabs, *"Data Over Cable Interface Specification,"*

[3] CableLabs, *"Cable Data Modem Performance Evaluation,"* Summer Conference 1996

[4] Chariot Console 2.1 Specification Sheet

[5] Kathleen M. Nichols and Mark Laubach, "On Quality of Service in an ATM-based HFC Architecture," IEEE ATM Workshop 96, August 27, 1996

[6] Kathleen M. Nichols, "Improving Network Simulation with Feedback," Internal Paper, Bay Networks.

[7] Kevin Thompson, Gregory Miller and Rick Wilder "Wide-Area Internet Traffic Patterns and Characteristics" IEEE November/December 1997

[8] Vern Paxson and Sally Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling," IEEE/ACM Transactions on Networking**,** 3(3) (June 1995): 226-244.

[9] Vern Paxson and Sally Floyd, "Why We Don't Know How To Simulate The Internet," Proceedings of the 1997 Winter Simulating Conference, Lawrence Berkeley National Laboratory, University of California

[10] Vern Paxson,"Fast Approximation of Self-Similar Network Traffic," April 20, 1995, Lawrence Berkeley Laboratory, U.C. Berkeley.

[11] James L. Riggs, Economic Decision Models for Engineers and Managers, McGraw Hill, 1968.

[12] Balakrishnan, V. Padmanabhan, and R. Katz,"The Effects of Asymmetry on TCP Performance", University of California at Berkeley, Berkeley, CA.

[13] Frances Cairncross, The Death of Distance: How the Communications Revolution Will Change Our Lives, Harvard Business School Press, 1997.

ENDNOTES

[1] Phil Miguelez, Gary Picard and Fred Slowik, "The Advantages of Gallium Arsenide Hybrid Power Amplifier Technology in Cable Television Applications," 1998 Conference on Emerging Tech Proceedings Manual: Collected Technical Papers (January 1998) :221-244.

[2] C. Day, "Heterojunction Device Technologies: Characteristics and Advantages for Wideband Linear Hybrid Amplifier Design," 1998 Conference on Emerging Tech Proceedings Manual: Collected Technical Papers (January 1998) :245-254.

[3] Claude Elwood Shannon, Claude Elwood Shannon Collected Papers; ed.. N.J.A. Sloane and A.D. Wyner, IEEE Press, 1993.

[4] Robert S. Burroughs, "What Shannon Really Said About Communications And Its Implication to CATV," 1990 NCTA Technical Papers (May 1990) :23-49

[5] Jerry I. Gibson, ed., The Communications Handbook, Florida: CRC Press and IEEE Press, 1997.

[6] Bronwen L. Jones, "Subjective Assessment of Cable Impairments on Television Picture Quality -- A Preliminary Report," 1991 NCTA Technical Papers (March 1991) :92-94.

[7] Bronwen L. Jones, "Subjective Assessment of Cable Impairments on Television Picture Quality," 1992 NCTA Technical Papers (May 1992) :1-25.

[8] Sergio Benedetto, Dariush Divsalar and Joachim Hagenauer, ed., "Guest Editorial: Concatenated Coding Techniques and Iterative Decoding: Sailing Toward Channel Capacity", IEEE Journal on Selected Areas in Communications, (February 1998) :137-139.

[9] The Hurwicz criteria is a measure sometimes used in modeling decisions under risk. A 0.5 criteria is a neutral case, assuming neither more optimistic nor pessimistic conditions than previous.

[10] OpenCable is a trademark of CableLabs and refers to the generation of digital cable terminals that are based upon open standards.

[11] This based upon a high quality (DFB) return laser.

[12] Werner/Sniezko "Network Availability Requirements and Capabilities of HFC Networks" 1996 Conference on Emerging Tech Proceedings Manual: Collected Technical Papers (January 1996) :123-135.

[13] Werner/Gatseos "Network Availability, Consumer Expectations, Plant Requirements and Capabilities of HFC Networks," 1996 NCTA Technical Papers (April 1996) :313-324.

TV MODULATOR PHASE NOISE - MEANINGFUL PERFORMANCE CRITERIA, SPECIFICATION AND NEW MEASUREMENT METHODS

Ron D. Katznelson
MCSI

---- *abstract only* -----

Over recent years, frequency synthesized TV modulators have gradually replaced many single frequency crystal controlled TV modulators, thereby introducing higher levels of modulator phase noise in CATV systems. While these degradations in many applications are insignificant, they present serious observable signal quality degradation effects in some instances such as reduced Direct-Pick-Up (DPU) rejection in off-air phase-locking and other signal quality degradation effects in coherent head-end applications.

Current practice of specifying phase noise levels in dBc/Hz at some offset frequency (typically 20 Khz) provides only limited value in comparing modulators, much less in providing a meaningful criteria for predicting picture quality or whether Precision Demodulators with synchronous detection can lock on a given modulators' signal. This paper shows that a more relevant criterion of **Total Integrated Phase Noise** (TIP Noise) obtained by integrating the phase noise density over a frequency band centered about the picture carrier frequency should be used to compare performance of TV signal sources. TIP noise is measured in **degrees RMS** and thus provides a meaningful measure of the total phase fluctuations of the RF phase.

As an example of a specific picture quality degradation that can be predicted by the use of the TIP Noise measure, it is shown through analysis that a TIP Noise of more 5° - 10° RMS in an off-air phased-locked modulator produces noticeable degradations in DPU rejection. In fact, it is suggested that these visible degradations may be experienced by an increasing number of subscribers who receive signals from "upgraded" head-ends that now employ new synthesized modulators having TIP Noise between 20° to 150° RMS instead of the crystal controlled modulators with superior performance with less than 5°RMS TIP Noise.

A new measurements method using a Network Analyzer for TIP Noise is presented along with some typical measurement results for single frequency crystal controlled and several synthesized modulators used by the cable industry.

The effects of phase noise in a coherent head-end is also discussed and the TIP Noise criteria is used to explain several known phase-locking degradation effects that have been observed by those who compared picture quality performance between the phase-locked and the phase-unlocked modes. It is shown that the improvements expected in phase-locked coherent head-ends cannot be fully realized unless modulator TIP Noise is better than 5° RMS. Unfortunately, many of the new "phase-locked" modulators are better defined as "frequency locked" rather than "phase-locked" since the excessive TIP Noise reaching instantaneous values of more than 180 degrees frustrates the maintenance of a truly fixed phase relationship with their respective reference signals. Modifications that improve these deficiencies in modulators are presented.

Utilizing Ingress as a Plant Maintenance Strategy

Raymond J. Schneider

A new concept is described which uses a transmitter located in the return path frequency interval to ingress coded signals into the system. A vehicle equipped with a GPS, a leakage detection system and a transmitter sends information, via coded ingression, back to a monitoring device at the head-end creating a centralized leakage and ingress database as a by-product of the maintenance cycle.

We describe a prototype system used to test out these concepts. Data is presented on a small model cable system used to initially prove out the system. More definitive data will be developed on a portion of the Time Warner system in Harrisonburg.

THE PROBLEM SUMMARY

Situation

Ingress is a particularly difficult problem in the return path frequency interval due to the many signal sources that exist below 45 MHz and other factors. This is particularly trouble-some since the return path is becoming increasingly important as the vehicle for various bi-directional interactive services such as Internet access.

Problem Components

Both the leakage and ingress problem have common components. The problem generally arises from weaknesses in the rf shield integrity. These weaknesses create points at which egress, usually called leakage, and ingress can occur. At these weaknesses rf energy can leave or enter the system much more readily than when the shield integrity is intact.

The problem is to detect, classify and localize the leakage or ingress energies so that the weaknesses can be found, diagnosed, and corrected. Leakage is found by driving through the system, however, ingress has been primarily investigated by performing spectral analysis on the return path signals at the head-end. This analysis can be effective for detection and classification, but is of little use for localization.

Current Response

The primary response to both FCC regulation and dealing with system ingress in both the upstream and downstream legs of the system has been a pro-active leakage and system integrity program. This requires systems to perform rf leakage surveys on a continuing basis to detect, classify, localize and correct defects. The detection of leakage energy is accomplished by traversing the system with sensitive leakage detection receivers which listen for emissions from the system.

To ensure that detected emissions are actually from the system, the transmissions use specially modulated coded signals such as the Sniffer transmitter or tagging signals applied to normal downstream traffic.

Deficiencies In the Current Response

This response requires specially trained technical personnel to conduct surveillance of the system on a quasi-continuous basis. The localization of defects is uncertain, due to a number of factors. Such factors include the tendency of rf leakage energy to reach the surveillance receiver by a combination of multi-path and direct path, standing waves on the sheath of the transmission system, intermittency and others.

A *find and repair* sequence is both time consuming and expensive, so it is quite common to separate the functions of surveillance and repair, logging the leakage during surveillance and returning to logged leakage sites to repair the system. Leaks are seldom entirely stable. One must return to the surveyed site and re-detect, localize, diagnose and then repair the defects.

Surveillance Automation

A partial solution is to automate the surveillance process. ComSonics and others have developed leakage detection and logging systems which combine leakage information with GPS (Global Positioning System) information. Figure 1 illustrates this kind of system diagramatically.

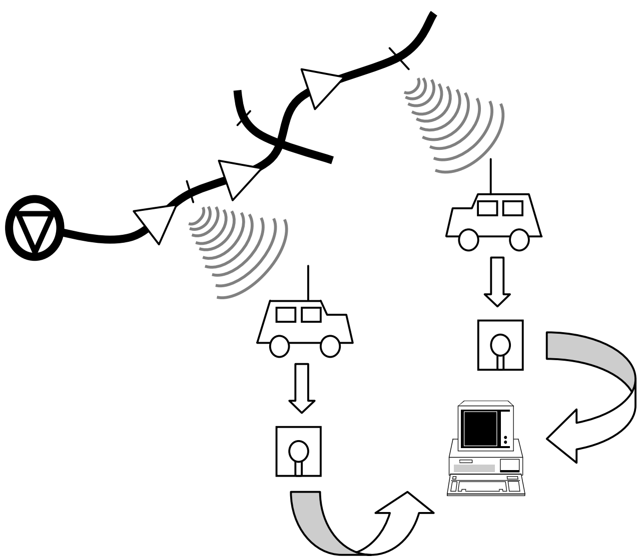


Figure 1. Typical Leakage Surveillance System

The principle advantage of this method is that it provides dense coverage with very little operator intervention. System coverage can be total and nearly continuous if there are sufficient vehicles instrumented and they drive through the system on a regular basis. A disadvantage is the manual process of handling the diskettes generated by several vehicles, and the time required for the associated processing.

Once the data has been processed, it is a relatively simple matter to generate work orders to service the hot spots in the system. A difficulty at this point is that the data is collected as latitude/longitude data and must be processed to produce actual street addresses. This processing requires a current map base that is as accurate as the GPS. Then, for each leak listed on the work order, a cable service person must re-localize, diagnose and fix the leak.

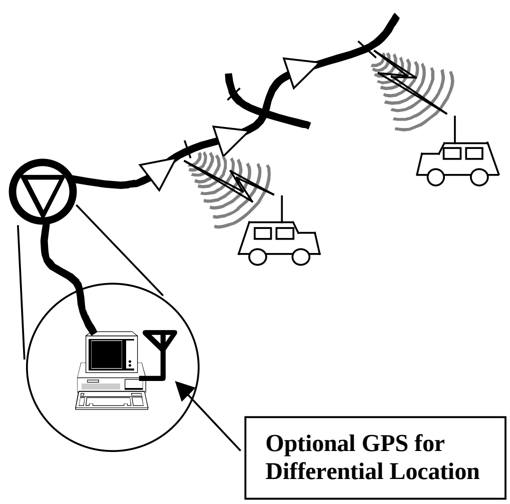


Figure 2. Diagram of the Ingressor Concept

A NEW CONCEPT

We have been working on a new concept which solves the majority of the problems with the mobile data logging approach. In addition, the concept offers new and interesting capabilities.

Figure 2 illustrates what we call the Ingressor Concept. In this concept the leakage monitoring vehicle does not log the data to a disk or other on-board storage media (although this capability would still exist as a backup). Instead, the system transmits position and leakage information which can be gathered at a central point either by direct path reception or, perhaps not as obvious, by receiving the signal via the ingress path.

The Conceptual Components

The Ingressor Concept has several conceptual components:

1. The primary signal is provided by ingress using a coded broadcast in the return path frequency interval of the cable system;

2. The surveillance vehicle transmits the signal which ingresses into the cable system as a function of the path of the vehicle, the location of the cable plant relative to the vehicle, and the relative shielding effectiveness of the portion of the plant closest to the vehicle;

3. On the transmitted signal is impressed both GPS position data for the vehicle and leakage (egress) data detected from the vehicle;

4. At the head-end, a receiver detects the presence of the ingression at the selected frequency, decodes the GPS and leakage information, and measures the signal strength of the received transmission;

5. If a map of the system is available, the system processes the data to produce a running measure of the shielding effectiveness along the plant strand map;

6. Peaks in the received ingress signal are correlated with peaks in the transmitted leakage information and the position of the leak is inferred more accurately than from leakage alone;

Figure 3. The single span model system

7. If differential GPS is used together with a surveyed plant map, the location of points of leakage or weak shielding effectiveness can be located with an accuracy on the order of 5 meters. (Single station GPS accuracy is only 100 meters.)

8. The same system which processes the received ingression signal can also be equipped to perform spectral analysis of the entire return path frequency bandwidth.

Using the GPS timing signals to synchronize the transmissions, a TDMA (Time Division Multiplex Approach) can be implemented. This allows multiple vehicles to share the same frequency and intermix their signals, so that far more coverage efficiency is achieved than if the system were restricted to a single vehicle.

PROTOTYPING THE CONCEPT

To determine if the Ingressor concept was viable, we decided to build a prototype system in two phases. The first phase was to develop a hardware prototype and test it on a miniature cable system built at ComSonics for this purpose. Once the prototype is shown to be functional on this system, it will be migrated to the local Time Warner cable system in Harrisonburg, VA and tested on that system's return path. This second phase will allow both higher fidelity and a much larger range of geometries and leaks to be evaluated.

The Model System

Figure 3 depicts the model system. It consists of a single short 20' span of cable with a trunk bridger, a line extender and a splice block mounted between two support poles. The model system is fed from an equipment rack inside the ComSonics building. For the data shown here, the only signal on the system was a Sniffer transmitter on 108.625 MHz. The ingress transmitter is mounted together with a GeoSniffer system equipped with a ComSonics Sleuth leakage detection receiver in a vehicle. The transmitter is a Motorola Radius low band transceiver transmitting a measured 47 watts into a 5/8 wavelength whip at 27.47 MHz.

Preliminary Data Runs

Figures 4, 5 and 6 depict ingress/leakage runs past the model system under three different conditions:

- a run past the model system with the system totally sealed (Figure 4)
- a run after a small leak (approximately 45 µV/m at 10 feet) has been introduced into the system (Figure 5)

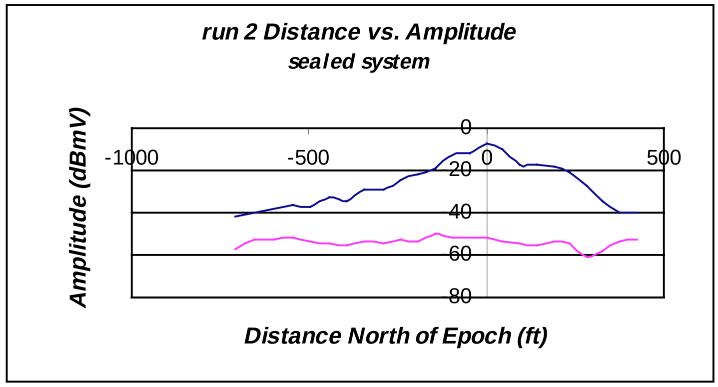


Figure 4. A run past the model system sealed

- a run after a larger leak (approximately 190 µV/m at 10 feet) has been introduced into the system (Figure 6)

The output of the GeoSniffer was connected to a modem and a digital data stream transmitted. The data consisted of the GPS position updates once a second, and the Sleuth leakage data at the time of the GPS update. The data was burst transmitted at a fixed time offset relative to the GPS clock. A receiver connected to the model system looked for the signal and if present synced with it and read the information. In addition to reading the GPS and Sleuth information the system recorded the AGC and RSSI (log output) voltages from the receiver. This data was used to create a calibration table for estimating the amplitude of the incoming ingression signal. Since this is early in our experiment, the errors are not perfectly understood. We believe the data is accurate to about 3 dB.

Data Interpretation

The data depicted in Figures 4, 5, and 6 shows both the ingress and leakage data plotted against distance relative to epoch. The term *epoch* is used to denote that point which has the largest ingress amplitude as the vehicle passes the model system. This point will be near the system, but not necessarily exactly when the vehicle passes closest to the system.

The distance is calculated by differencing the epoch location from the latitude and longitude of the data points and converting the difference to a distance in feet. Distance which is relatively North of the epoch is designated as positive, and distance South of the epoch is designated as negative. If the epoch were exactly abreast of the model system, it would be approximately 100’ from the first pole.

The data is displayed from South to North. In each data plot, the top trace is the ingress data and the bottom trace is the leakage data. The model system is on the South side of the ComSonics building offset about twenty feet.

Fully Sealed System

The first thing to note is that one only gets information from the system when the ingress is present and at a signal to noise level sufficient to read the data reliably. This level is nominally –35 dBmV. In run 2 depicted in Figure 4, the system reliably ingressed from about 700’ South and closer. Notice that the signal intensity grows steadily to the epoch and then declines and there are only slight ripples in the data. The associated leakage data shows no sign of peaking.

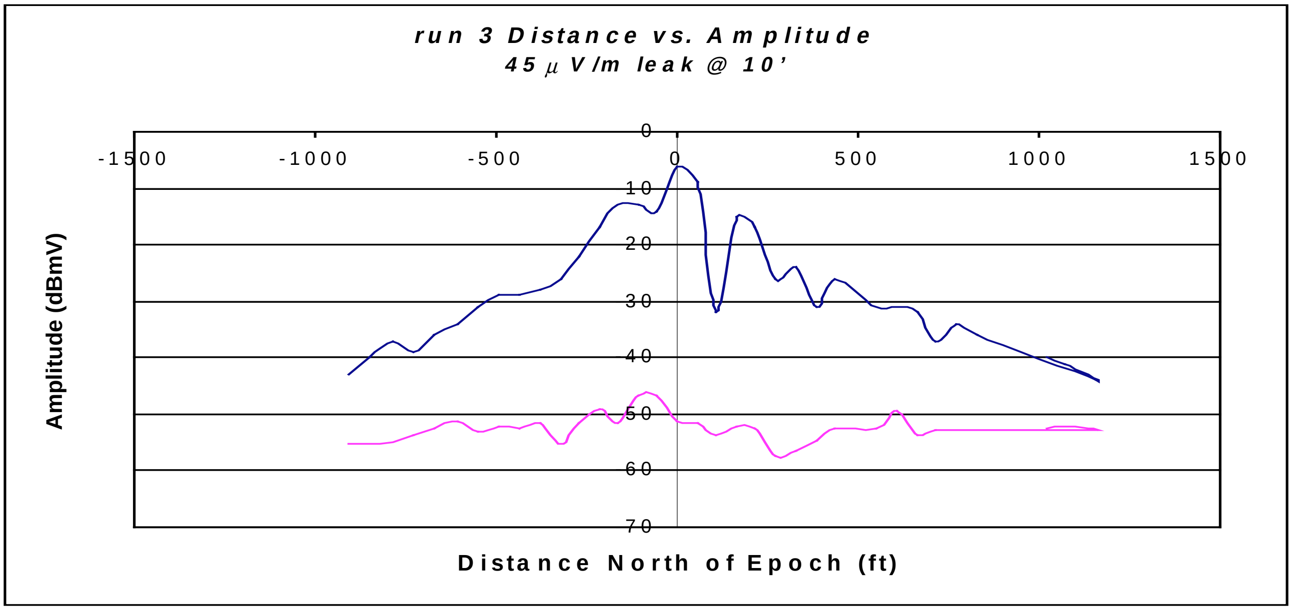


Figure 5. A run past a small (45µV/m) leak

A Small Leak

At least three effects different from the sealed system run are apparent in the data from the small leak run (run 3). These are:

- The signal dropped out at 900’-1000’ instead of 700’ from the epoch.

- The leakage data shows a noticeable peak just prior to the epoch indicating that the leak is being detected.

- There are several major swings in signal amplitude after passing the epoch to the North.

We currently believe that the peak in the leakage data occurs approximately when we are abreast of the model system. The large swings are believed to be due to multipath, an interfering reflection from the metal side of the ComSonics building which goes in and out of phase as the vehicle proceeds North. About three cycles of declining amplitude can be seen in the data of Figure 5.

A Larger Leak

Introduction of larger leaks in the model system created stability problems with the prototype. The initial effort to create a larger leak produced data which contained spikes and oscillations. To correct this problem we inserted a 20 dB attenuator at the input to the receiver. Run 6, depicted in Figure 6 shows the result when we drove past the system with a 190 µV/m leak. Even with a 20 dB attenuation, the receiver began to receive and decode usable data at a range of between 1414’ to 1486’ feet from the model system on this run.

The same kinds of features are observed on this run as on the low leakage run. The first ingression peak correlates with the first leakage peak.

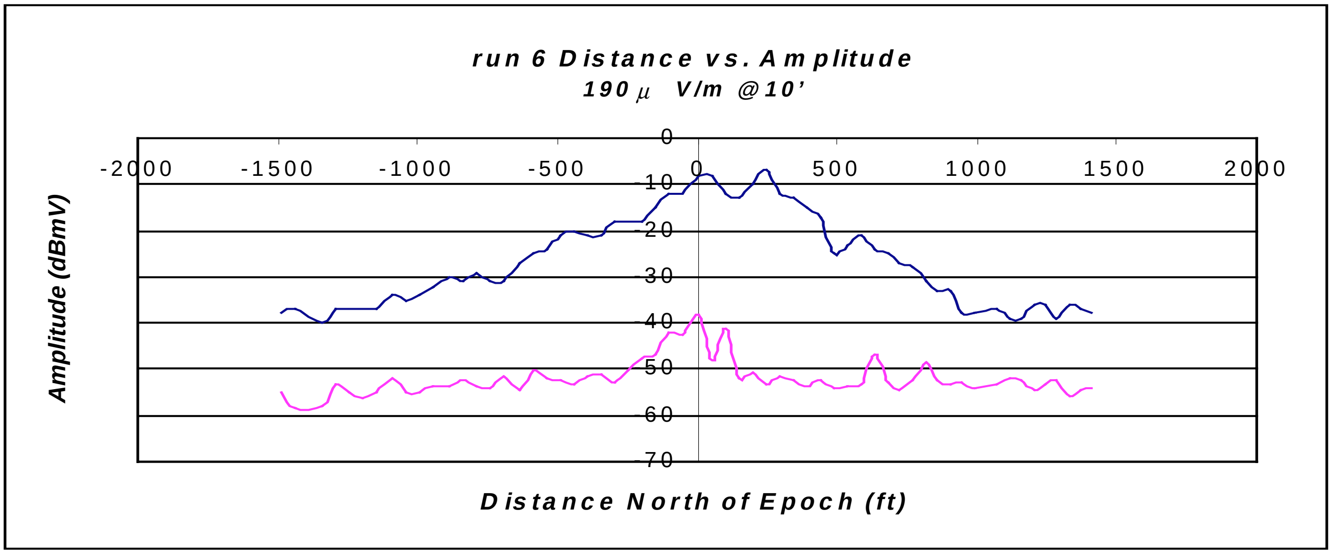


Figure 6. A run past a larger (190µV/m) leak

Conclusions

Ingression occurs into a sealed system from distances on the order of 700’ with 47 Watts of radiated power, such that coded digital data can be extracted. Even a small leak increases this range by about 30%. A larger leak increases this range by over 100% to in excess of 1400’.

The fact that ingress is so readily achieved allows it to be used as an unorthodox but highly diagnostic data path to the headend.

Combining leakage and ingression data gives additional information which supports better leak/ingress position determination.

Structure in the ingress amplitude allows intelligence about the state of the cable system to be extracted.

Lower power than employed in this test will still routinely ingress data into the cable system. Optimum power levels will be studied in phase two of this work.

Acknowledgements

I would like to take this opportunity to thank Dennis Zimmerman, the CEO of ComSonics who had the original idea for the Ingressor Concept. There is a patent currently pending. I would further like to thank Randy Smith, whose hard work made it possible to collect this data.

Raymond J. Schneider
Director of Engineering
ComSonics Inc.
1350 Port Republic Rd.
Harrisonburg, VA 22801
(540) 434-5965
web site: **www.comsonics.com**
e-mail: **rschneid@shentel.net**

Ray Schneider has been the Director of Engineering at ComSonics since 1989 when he joined ComSonics from his position as Chief Scientist at Unisys's Washington Systems Engineering Center. Ray's team of engineers have developed many products for the CATV industry since that time including the Window Lite Plus, the Sniffer ID, the GeoSniffer, the Sleuth and others. Ray has a B.S. in Physics, an M.E. in Engineering Science and is completing his PhD in Information Technology at George Mason University. He is a licensed professional engineer in the state of Virginia.

Will Widescreen (16:9) Work Over Cable?

Ralph W. Brown

Digital video, in both standard definition and high definition, is rapidly setting the standard for the highest quality television viewing experience. Digital Versatile Disc (DVD) is already delivering standard definition digital video in both 4:3 and 16:9 aspect ratios for display on existing NTSC television sets. High Definition Television will extend this experience delivering both of these aspect ratios and significantly higher resolutions. The desire for consumers to watch movies in either letterbox or pan-and-scan format depending on their preferences will push the Cable Industry to support both these aspect ratios. However, there are a number of technical issues that make this problematic. This paper will explore the technical issues surrounding dual aspect ratios and offer possible solutions to them.

INTRODUCTION

With the advent of MPEG video compression the ability to deliver more video content at various resolutions and aspect ratios has become a reality. The MPEG video compression standard [1] allows for both standard definition and high definition resolutions. Standard Definition Television (SDTV) is specifically formatted to work with today's NTSC and PAL television sets. High Definition Television (HDTV) is specifically formatted High Definition television sets that will be commercially available later this year.

SDTV is being delivered today via cable, satellite, and Digital Versatile Disc (DVD) [2]. HDTV broadcast is currently being tested and will begin on a national basis later this year. Aside from the issues of which specific HDTV formats will be used, there is the issue of how the multiple aspect ratios of movies and video material will be presented on the subscriber's television. This is particularly an issue for network operators, as it affects bandwidth efficiency and customer expectation.

The Advanced Television Systems Committee [3] has specified 18 different standard formats for digital television covering frame rate, picture resolution, scan mode, and aspect ratio. These 18 formats specify two aspect ratios, 4:3 and 16:9, and four resolutions covering both SDTV and HDTV display.

Aspect Ratio and Resolution

Aspect ratio is defined as the ratio of the width to the height of the source material or the display device. Typically, this is expressed as "*width*:*height*" or as the fraction *width/height*. Resolution (the number of picture elements, or pixels, displayed on the screen) is typically expressed as "*width* x *height*". The higher the resolution, the sharper the image appears on the display.

Table 1 below provides a list of some of the common aspect ratios and resolutions used in digital television.

Aspect Ratio or Resolution	Fraction	Description
16:9	1.778	HD TVs and widescreen SD TVs
4:3	1.333	Most SD TVs
1920x1080	1.778	HDTV
1280x720	1.778	HDTV
704x480	1.467	SDTV
640x480	1.333	SDTV

Table 1 - Common Digital Television Aspect Ratios and Resolutions

When square pixels are used, the resolution and aspect ratio are the same when expressed as a fraction. Square pixels are of importance to the computer graphics industry, however, in MPEG video there is no requirement to use square pixels.

For the source image to be displayed without distortion on the display device, it is necessary that the source aspect ratio match the display device aspect ratio. A mismatch of the source and display aspect ratios will result in a squeezing or stretching of the image itself. The greater the mismatch in aspect ratios the more noticeable this distortion will be. When one considers the historical evolution of motion pictures, this has become a more significant issue.

The aspect ratios of motion pictures over the years has spanned the range from 1.33:1 to 2.76:1, with 1.85:1 and 2.35:1 being the most common aspect ratios being used today [4,5]. Table 2 below provides examples of common motion picture aspect ratios.

Aspect Ratio	Fraction	Description
16.7:9	1.85	35 mm movies
19.9:9	2.21	70 mm movies
21.2:9	2.35	Panavision, Cinemascope
24.3:9	2.7	Ultra Panavision

Table 2 - Common Movie Aspect Ratios

Most of us are familiar with the tall and skinny pictures from movies not displayed at the proper aspect ratio. Figure 1 shows an example of this distortion. This sometimes occurs as the credits roll at either the start or end of the movie.

Figure 1 - Aspect Ratio Distortion [6]

As can be seen from tables 1 and 2 above, none of the common movie aspect ratios match either the SDTV or HDTV aspect ratios. This presents the problem of how to display an undistorted movie image on either of these types of television sets.

Pan & Scan versus Letterbox

Since aspect ratio must be preserved in order to prevent distortion of the resulting image, we have the problem of presenting movie material on either a 4:3 or 16:9 display. There are two methods used to address this problem when transferring a movie title for television display. One can either use Letterbox or Pan & Scan transfer on a 4:3 television.

In Letterbox transfer the full movie aspect ratio is presented, however, black bars are displayed at the top and bottom of the television to enable it to fit the 4:3 TV aspect ratio. Figure 2 is an example of Letterbox transfer.

In Pan & Scan transfer a 4:3 window is placed over the full movie material. This 4:3 window is moved horizontally or vertically, zoomed in or out, as necessary to capture the "important" portion of each frame of the movie.. The determination of what is the "important" portion of each frame varies from film to film and director to director. Some directors are particularly sensitive to the cinematic effect of the wider aspect ratios of motion pictures and take a more active role in recomposing the film in its 4:3 Pan & Scan version. Figure 3 shows an example of Pan & Scan transfer.

In the case of 16:9 encoded MPEG material the movement of this 4:3 window is captured by the Pan & Scan vector and transmitted along with the MPEG video data, however, the MPEG Pan & Scan vector is limited to horizontal movement only. As we will see shortly, this limitation is an issue for the usage of the MPEG Pan & Scan vector.

Notice that the figure on the right in the Letterbox example is not visible in the Pan & Scan example. This is an objection to Pan & Scan transfer. The black bars above and below the image in the Letterbox example is an objection to Letterbox transfer.

Letterbox display is less objectionable on 16:9 television displays than it is on 4:3 television displays as motion picture aspect ratios more closely match this larger aspect ratio. The black bars above and below the image are much narrower on a 16:9 television. Table 3 below shows the percentage of the display space that is used for the image in Letterbox mode on either a 4:3 or 16:9 display for source material of several aspect ratios. In general, a much greater percentage of the display is used when its aspect ratio is 16:9.

Figure 2 - Example of Letterbox

Figure 3 - Example of Pan & Scan

Display aspect ratio	Source Aspect Ratio		
	1.33	1.85	2.35
4:3	100%	72%	57%
16:9	75%	96%	76%

Table 3 - Display usage for Letterbox

The problems of dual aspect ratios are different for SDTV than it is for HDTV. The remainder of this paper will address the respective problems of transmitting movie material for SDTV and HDTV.

STANDARD DEFINITION TELEVISION

The problem for transmission of SDTV resolution MPEG video relates to having two different aspect ratios for standard definition television sets, either 4:3 or 16:9. A majority of televisions in the US today are of the 4:3 aspect ratio, however, a number of television manufacturers have been producing widescreen televisions that have a 16:9 aspect ratio. These widescreen televisions have a number of operating modes that will be discussed shortly.

For optimal display on both types of televisions you might wish to use Pan & Scan on the 4:3 TV and Letterbox on the 16:9 TV. The viewer may wish to determine which version of the film they watch as well. One viewer may wish to watch a movie in Pan & Scan on his 4:3 television and another viewer may wish to watch it in Letterbox. Ideally, the network operator would prefer to support both of these viewers with a single MPEG video stream. It is more bandwidth efficient to broadcast only one MPEG video stream for both types of displays.

To understand the issues with dual aspect ratio on SDTV it is necessary to understand the behavior of both 4:3 and widescreen televisions having a 16:9 aspect ratio. The behavior of 4:3 televisions is simple, they only have one mode of operation. Widescreen televisions have multiple modes of operation.

The Behavior of Widescreen (16:9) Televisions

Several television manufactures build widescreen (16:9) televisions for the US market and widescreen televisions are already widely accepted in Japan and Europe. Examples of widescreen NTSC televisions include: Pioneer's Elite PRO-119, Toshiba's TW40F80 and TW65G80, JVC's NV55BH6, and Goldstar's WF-32A10S). These TVs all support at least three modes of operation:

1. NORMAL mode - in this mode the TV displays standard 4:3 television images on a 16:9 screen with black side bars to preserve aspect ratio
2. THEATERWIDE or CINEMA WIDE mode - in this mode the TV would provide virtually full-screen images of Letterbox formatted video tapes, laser disks, or DVD material by expanding it vertically to maximize the display. Some 16:9 TVs have multiples of this version to adapt to a variety of Letterbox aspect ratios.
3. FULL or CINEMA FULL mode - in this mode the TV would expect 16:9 material anamorphically squeezed into a 4:3 format and display the full 16:9 material at the proper aspect ratio.

Another issue for SDTV is how the MPEG decoding device (set-top, DVD player, or satellite IRD) signals the widescreen television the appropriate mode (NORMAL or FULL) is to be used. This is less of an issue for DVD players as the user will likely set up the DVD player to operate in one mode and leave it that way. For a set-top or satellite IRD the user will be changing between channels with 4:3 and 16:9 source material, causing a need to signal the TV potentially at each channel change.

In addition to the issues for the television viewer, there are the concerns of movie studios and directors regarding the presentation of their creative material.

Issues for the studios and directors

Ideally, the network operator could broadcast one MPEG program stream of the movie material with the appropriate Pan & Scan vector information for 4:3 display. In this way the owner of a widescreen TV could view the 16:9 movie and the owner of a 4:3 TV could view this either in Letterbox or in Pan & Scan according to their preference. However, there are a number of concerns among the movie studios and directors that make this undesirable from their perspective.

The first of these is the loss of resolution in 4:3 Pan & Scan mode. If the original source material is encoded at a 704x480 resolution at 16:9 aspect ratio, then the 4:3 Pan & Scan image will have a 25% reduction in resolution resulting in a 528x480 display resolution.

The second and more important problem is that the MPEG Pan & Scan vector is limited to horizontal movement only, however, in general when a Pan & Scan transfer is made to video tape, vertical movement and zooming are also used to provide the “best” representation of the cinematic mode the director intended to create. Frequently, the movie director will retain control of the Pan & Scan transfer for this reason.

Solutions for SDTV

Given the studios objections to the use of the MPEG Pan & Scan vector, it really is necessary to transmit two versions of the same movie. One version transmitted in Letterbox for 16:9 televisions and one version transmitted in Pan & Scan for 4:3 televisions. DVD also uses this dual stream solution. Typically a DVD title will come with two sides. The first side will contain the Pan & Scan transfer for 4:3 display and the second side will contain the Letterbox transfer for 16:9 display. The viewer can choose which to view and how to display them. The only realistic option for cable broadcast is to simulcast both the 4:3 Pan & Scan and the 16:9 Letterbox versions of the video together with only one set of audio streams shared between the two video streams.

HIGH DEFINITION TELEVISION

Since High Definition televisions have only one aspect ratio and it is 16:9, they do not share the problems of Standard Definition televisions having two aspect ratios. HDTV presents a different problem for network operators. The memory and MPEG decoder performance required for HDTV decode is substantially greater than for SDTV and consequently more costly. Network operators currently incur the burden of the set-top terminal cost. They purchase the set-top terminals that are installed in subscribers homes and are unlikely to take on the additional cost of HDTV decode without a corresponding increase in revenue.

High Definition televisions will already contain the memory and MPEG decoder performance required to decode HDTV. The subscriber will already have purchased this equipment and it is redundant to have the same capability built into the set-top terminal. This necessitates a digital interconnect between the set-top terminal and the High Definition television to pass the uncompressed MPEG data. The set-top then functions as the tuner, QAM demodulator, and PID filter to direct the proper MPEG program stream to the digital interface. Today, the digital interconnect most commonly being considered is IEEE 1394.

The use of a 1394 interface between the set-top terminal and the High Definition television solves one problem, but introduces two additional issues. The first is tht transmission of HDTV MPEG in the clear across an open interface, such as 1394, permits anyone to make an unlimited number of perfect copies of the original source material. This ability to make unlimited copies is of even greater concern to the studios than the issues for 4:3 Pan & Scan transfer for SDTV. The second issue is presentation of the user interface (UI) generated by the set-top terminal. Since the HDTV is performing the MPEG decompression and not the set-top terminal, the ability for the set-top to composite the UI on top of the video has been lost. A protocol must be defined to transmit the set-top generated UI to the HDTV which must perform the process of combining the UI with the decompressed MPEG video.

Solutions for HDTV

The Motion Picture Association of America (MPAA) has required that there be an acceptable form of digital copy protection placed on common interfaces, such as 1394. There has recently been an agreement between Sony and Intel, the two leading vendors of copy protection technology to merge their copy protection schemes into a single 1394 digital copy protection standard. To date, there has been no well organized effort to define the UI presentation protocol between the set-top terminal and the HDTV over 1394.

Once the digital copy protection compromise and the UI presentation protocol have been implemented, it will be possible to interconnect digital set-top terminals to High Definition televisions using 1394.

CONCLUSIONS

The network operator must consider how do deal with the dual aspect ratio of the 18 ATSC digital TV formats. For SDTV there are two options, either simulcast both a 4:3 and a 16:9 version of the video stream or broadcast only the 4:3 version. For HDTV finalizing the digital copy protection mechanism and UI presentation protocol will enable carriage of HD content over cable.

Given the relative low penetration of standard definition widescreen televisions, it is unlikely that network operators and satellite providers will dual carry movies in both 4:3 Pan & Scan and 16:9 Letterbox format to accommodate both aspect ratios. With the national deployment of HDTV later this year it makes more sense to wait for this deployment to support movie material in its original theatrical aspect ratio.

REFERENCES

(1) ISO-IEC IS 13818-2, International Standard (1994), MPEG-2 - Video.

(2) Jim Taylor, DVD Demystified, McGraw-Hill, New York, NY, 1998.

(3) Advanced Television Systems Committee (ATSC) Standard Document A/53, Digital Television Standard, September 16, 1995.

(4) John Belton, Widescreen Cinema, Harvard University Press, Cambridge, MA, 1992.

(5) Ken Korman, Size Matters, Making Sense of Film and Video Aspect Ratios, Video Magazine, June 1997.

(6) All images for this paper were provided on the web site The Screening Room - Widescreen Movies, http://www.accessone.com/~mco/letterbox.htm

Ralph
Chief Software Architect
Time Warner Cable - Advanced Engineering
160 Inverness Drive West
Englewood, CO 80112, USA
303-799-5622

Ralph W. Brown is currently the Chief Software Architect for Time Warner Cable's Pegasus Program and a member of the Advanced Engineering Group. Ralph joined Time Warner Cable in 1994 to work on the Orlando Full Service Network™ as a software architect and technical program manager. Prior to this Ralph was employed by Solbourne Computer and Bell Laboratories.

Ralph earned a Master's degree in Electrical Engineering from Massachusetts Institute of Technology. His Master's Thesis was in the field of Speech Recognition. He received his Bachelor's degree in Electrical Engineering, Summa Cum Laude, from North Carolina State University.

His current technical interests are in distributed software systems, digital video technology, video servers, computer graphics, and speech recognition.

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