WHAT SHANNON REALLY SAID ABOUT COMMUNICATIONS AND ITS IMPLICATIONS TO CATV.

Robert S. Burroughs

Panasonic Technologies
Communication Systems Technology Laboratory

ABSTRACT

HDTV and the application of fiber optics is causing engineers to rethink the way television pictures are being delivered to the home. In discussions about new delivery methods engineers occasionally will quote Shannon, on what the theoretical limitations are for communication systems.

This paper reviews Shannon's fundamental theorems and premise, and postulates a method for evaluating communication systems. Based on the analysis, future digital CATV distribution systems are contemplated.

INTRODUCTION

The U.S. has had the same basic television transmission system for the past 50 years. It has served the U.S. well, but in the meantime modern communications theory has matured. We are now at a transition point where the television system, for possibly the next 50 years, is about to be decided. It would be a shame if the system of the future was constrained by the technologies of the past.

Several proponents are now proposing ATV systems for testing by the FCC's Advanced Television Test Center, (ATTC). Most of the proponent's systems are extensions of the existing NTSC system. In November of 1988 the System Analysis Working Party of the FCC, Systems Subcommittee for Advanced Television Systems, met for a week with the proponents of ATV systems, at a Days Inn, just outside of Washington D.C.. Each proponent presented their system to the committee. During one of the presentations the committee was not able to evaluate or analyze one of the proponent's system. The system did not look like a typical NTSC system, it was more digital in appearance. It was even remarked by committee members that they thought the system violated "Shannon's Limit". This comment struck me as odd, until I realized that I was in the company of primarily analog engineers.

After the meeting was over I contemplated the comment about "Shannon's Limit". If you want move from a classical communication approach to a modern approach you need to go back and review the fundamentals. And the fundamentals started with Shannon. It was then that I decided to go back to Communication Systems 101, and review "What Shannon Really Said About Communications ....".

SHANNON'S CONTRIBUTION

Shannon's fundamental theorem for a discrete channel with noise, (Theorem 11), is the basis by which all systems should be judged - it is the *ideal.*

Prior to Shannon's classic 1948 paper [1], "A Mathematical Theory of Communications", it was universally accepted that the accuracy of a transmitted signal was irrevocably altered by noise. This thinking was only natural. If random noise, n(t) is added to a signal, s(t), the result is a new signal r(t) = s(t)+n(t), which is also a random signal, for which an accurate replica of s(t) can not obtained.

Shannon, however, proved the contrary; a signal s(t) can be recovered to any desired accuracy, in the presence of noise N, if the bandwidth of the signal W is constrained and the signal magnitude S is restricted. Then the effects of noise can be combined with S and W in a parameter called "Channel Capacity" C, in the following form :

$$C = W \text{ LOG } (S+N/N) \quad (1)$$

This shows that the rate,

$$W \text{ LOG } (S+N)/N \quad (2)$$

measures the capacity of a channel for transmitting information. Shannon defined capacity C of a noisy channel as the maximum possible rate of transmission when the source is properly matched to the channel. He used a new measure of information which he called entropy, H to define Channel Capacity :

$$C = \text{Max } \{H(x) - H(y|x)\}, \quad (3)$$

where the maximum is averaged over all possible information sources.

The implications of Shannon's Channel Capacity theorem were quite revolutionary to communication theory. Consider the situation where a number of message possibilities M increases as a function of the signal duration T, slowly enough so that;

$$M < 2^{(CT)} \quad (4)$$

then, although perfect accuracy can not be attained, one can get arbitrarily as accurate as one wishes by choosing T large enough, by using sufficiently long signals. Shannon also showed the converse was true - reliable communications is not possible, regardless of signal-processing schemes, when

$$M > 2^{(CT)} \quad (5)$$

For a source rate R < C it is possible to make the probability of an error in transmission as small as desired by properly choosing the set of :

$$M = 2^{(RT)} \text{ signals.} \quad (6)$$

Or conversely, for a source rate R > C it is not possible to make the probability of an error arbitrarily small with any choice of T or any choice of signals.

The theorem is extremely general and is not restricted to Gaussian or discrete channels. Note that the theory does not say what form the transmitted signal should have or how one should go about finding signals which will achieve "Capacity" C.

Note the remarkable aspects of this theorem. If for any value of S, (signal power), greater than 0, a value of W can be picked such that one can transmit virtually error free messages at a rate, R < C. Or, theoretically , we can recover any reasonable signal buried in noise, given the proper code sequence. In actuality, close to these conditions exist in communication with deep space probes. Compare this situation with a typical S/N that is used to send TV pictures over a Cable system to the home. Current Cable targets are to get approximately a 50 dB S/N to the home vs. a 0 dB S/N used in space communications.

The price we pay for getting arbitrarily close to zero errors is long durations T. The consequences of this will be discussed in the section titled, "Geometrical interpretation of signals".

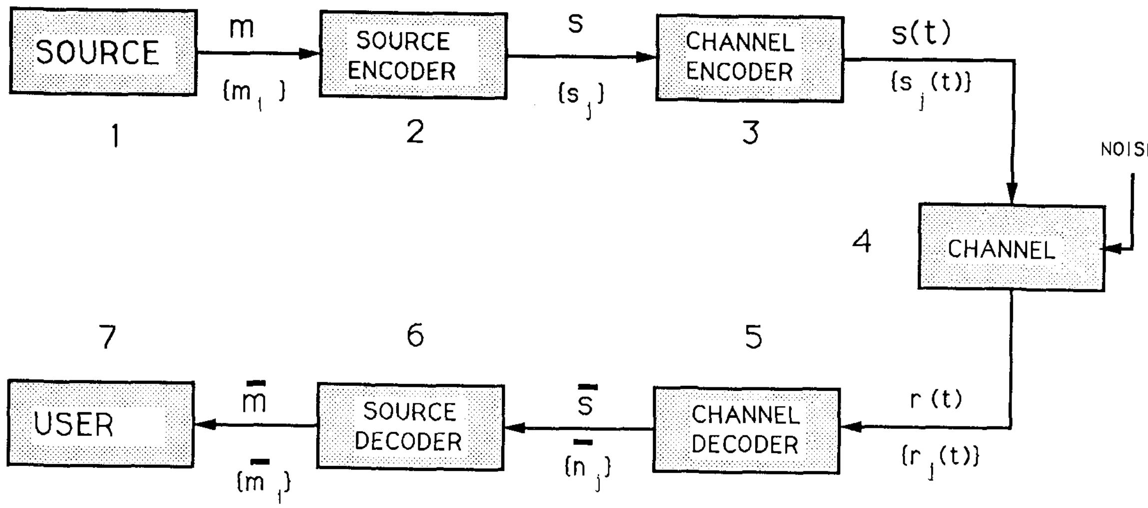
Even though most of the theory presumes discrete signals, Shannon showed that any continuous signal can be represented by a discrete source and the Channel Capacity theorem holds.

System block diagram.

A block diagram of a general communication system is shown in Figure 1. The first element is the 1. information source. The output of this source may be a sequence of discrete symbols, letters, or numbers in which case the source is referred to as a *digital source*. If the source output is a waveform or sequence of continuous valued variables, the source is referred to as a *continuous source*. In this paper, *digital sources* are the focus, since the concepts generally apply to *continuous sources* as well. An example of a *digital source* might be a sequence of "ones" and "zeros", or the text of this paper which is stored on the disk of my computer. These symbols could be expressed by the 128 symbols of the ASCII code. The essential feature of any source is that their output is generated by a *random* or *probabilistic* mechanism. This randomness is required, for if the output was known before the source generated it,there would be no need to communicate the source output to anyone.

The next block in the system is the encoder which is broken into the, 2. source encoder and the, 3. channel encoder. The reason for this separation is because of the different coding requirements required for the source and the channel.

The 5. channel decoder, 6. source decoder, and 7. destination perform the inverse operations of 1,2, and 3. The difference is that the received signals are only approximations to what was sent. Block 4. Channel is the particular medium used such as fiber, wire, free-space, etc. It is also the point where external noise is introduced to the system.



Block diagram of communications system with encoder and decoder each split into two parts.

Figure 1.

INFORMATION AND ENTROPY.

Although the Channel Capacity theorem in the presence of noise, is Shannon's main contribution, he is also responsible for his insight and pioneering work into the definition of Information and its subsequent application to the communications problem. Many of Shannon's concepts were not totally new, but he brought a fresh approach to explaining the fundamental concept of communication - "what is information and how best can one communicate it" ?

Information can have at least three levels of meaning :

1. Technical: how accurate can symbols be communicated ?

2. Semantic: how precise is meaning of symbols communicated ?

3. Effectiveness: how effectively does received meaning affect conduct in desired way?

Shannon concentrated on the technical level, even though the generality of his results also apply to levels 2 and 3.

The use of the term R, rate, and message possibilities, M were used to define "Channel Capacity". Shannon also uses the word *information* in a very special sense that should not be confused with meaning. And R is the rate at which *information* can be communicated.

Shannon once stated that the "semantic aspects of communication are irrelevant to the engineering problem". Note that the opposite is not necessarily true.

Information

When one message is selected from a set of possible messages, the information produced when this message is chosen can be quantified, (under certain conditions). As suggested by Hartly and Nyquist the logarithmic function is a convenient measure to use.

The meaning of "message" is quite general. Message can be a simple yes or no , (1 or 0), or a message can be a two hour television program. The information contained in a message of two possible choices is one (1), because of our choice of logarithms to measure information:

$$\log_2(2) = 1 \quad (7)$$

If we had 4,8,16,... choices, the information would be 2, 3, 4, ... bits respectively.

The content of information is typically measured in bits, a contraction of "binary digits".

For a typical communications source, we do not make a single choice, but a series of choices, one following the other as letters in a word or words in a sentence.

Shannon used a probability measure to define information. Using probability theory he defined three types of information :

1. Mutual information -

$$I(x,y) = \log P_{x|y}(a_k|b_j)/P_x(a_k) \quad (8)$$

the information provided about the event $x = a_k$ by the occurrence of the event $y = b_j$.

In terms of Figure 2, event $x = a_{1or2}$, $y = b_{1or2}$

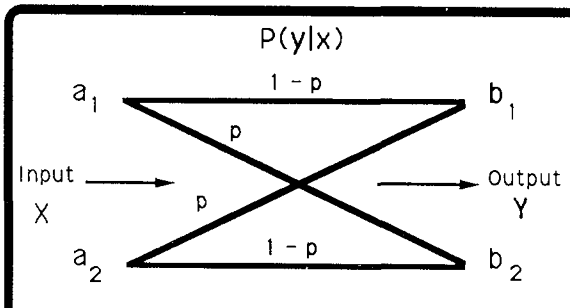


Figure 2 Binary symmetric channel.

1-p is probability that output letter is replica of input, and p it is the opposite of the input letter.

2. Self information -

$$I_X(a_k) = \log [1/P_X(a_k)] \quad (9)$$

the mutual information required to specify $x = a_k$

3. Conditional self-information

$$I(x|y) = \log 1/P_{x|y}(a_k|b_j) \quad (10)$$

the self-information of an event $x = a_k$, given the occurrence of $y = b_j$.

Self information, mutual information, and conditional self-information are all random variables.

Entropy

The entropy of an ensemble (x,y) is defined to be the average value of the information, or in the case of Self-information :

$$H(x) = \sum_{k=1}^{K} P_x(a_k) \log 1/P_x(a_k) \quad (11)$$

The average mutual information between x and y is the difference between the entropy of X and the conditional entropy of X given Y or :

$$H(x) - H(X|Y), \quad (3)$$

the form used in Shannon's coding theorem. Where H(x) is the average information of the source x and H(x|y) is the average information required to specify x,(input), after y, (output), is known. Or H(X|Y) is the uncertainty in y as to which x was transmitted. Shannon refers to this uncertainty as equivocation.

Shannon defined "entropy", similar to the thermodynamic definition which connotes the random character of nature. (Shannon once said that the mathematician John von Neumann urged him to use the term entropy, since no one really knows what it means, Shannon would have an advantage in debates about his theory.)

Although an understanding of the mathematics of entropy is not essential to the purpose of this paper, a simple explanation is warranted. If from the, 1. <u>source</u> , we have a set of n independent symbols or messages, whose probabilities of occurrence are $p_1, p_2, \ldots, p_n$, then the corresponding entropy is:

$$H = -[p_1 \log p_1 + p_2 \log p_2 + \ldots + p_n \log p_n]$$

or

$$H = -\sum_{\text{all } i} p_i \log p_i .$$

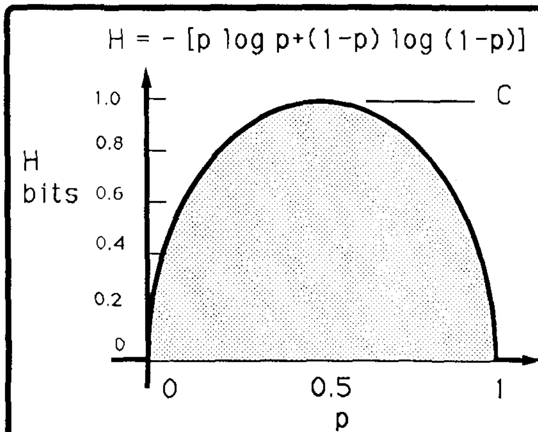


Figure 3.

Entropy in the case of two possibilities with probabilities p and (1-p)

Figure 3 is an example of the entropy of the binary symmetric channel of Figure 2. The maximum entropy can be seen to occur for the case where the input probabilities p_i are all equal, or the most uncertain condition. The condition of zero entropy is when one of the two p_i is certain. The implications are that when the inputs are know for certain, then no information is communicated, or that when the inputs are equally likely then the maximum transfer of information occurs and the channel achieves capacity, C. The first condition is obvious, if we know the inputs with probability 1, then there is no need to transmit them. The second situation is paradoxical. The maximum information is transferred when the inputs are equally likely, or completely random, but this is similar to the definition for noise. So, when is the transmitted signal, information and when is it noise ? I leave this paradox to a later paper, after I have had a chance to by some books newer than 1968.

DISCRETE SOURCES.

Although Shannon showed that the "Channel Capacity" theorem applies equally well to continuous time functions, it is the concept of a discrete source which is fundamental to the development of the theorem. The source outputs individual messages, each message a point in the message space of all possible messages. The transmitter, from a geometrical standpoint, maps the message space into the signal space. Many of the possible messages in the message space are redundant and do not convey any more information, than if they were not transmitted.

If we want to represent a continuous signal s(t) as a discrete message, we can represent this signal, if it is limited in frequency, to no frequencies greater than W_o hz.; then s(t) can be exactly represented by taking 2 W_o independent samples. This is the well know sampling theorem. These independent samples are discrete in time, such that a finite number of sample values are needed to define s(t) over a period T seconds. To each sample, if we can assign a number, we have an ensemble of numbers which are sequentially tied together. We now have a discrete digital source. But, unless we are very lucky, we can not always assign the exact value to the sample. We can, however, assign a value which is as close to the actual value of s(t) as we wish. This is the concept of quantization. Typically the accuracy we assign to this value is some multiple K of the RMS noise voltage N_o at the input to the decoder, Figure 1, 5. Channel decoder. This defines the resolution to which we have approximated the signal s(t) to a value:

$$q = KN_o. \qquad (12)$$

The important point is that we have defined the source s(t) to be a sequence of numbers

$$s_i(1), s_i(2), \ldots s_i(n). \qquad (13)$$

We also have defined the rate at which these discrete symbols occur, 2W. We now have a set of numbers and the communication problem comes down to transmitting this series of numbers to the user as close to the original sequence as possible. Since these are just numbers, and not a fixed single signal, s(t), we can perform almost any mathematical operation on these numbers we desire, as long as from the resulting sequence of numbers we can decode the original sequence. This is the basis of modern communications theory. Note that in Figure 1, we are not restricted from using conventional "analog" modulation methods such as FM, AM, PM, etc. One possible use of coding would be for more effectively use of spectrum. Consider normal TV signals where most of the energy is located near zero frequency. We could develop a coding scheme which would select codes such that the frequency characteristics of the sequence would spread the energy more evenly across the band. This transformed set of signals would fit within the same bandwidth as the original s(t), but would distribute power more efficiently such that less peak power would be required by the transmitter and less distortion products in the channel. This technique would reduce any benefits gained by companding or pre and de- emphasis schemes such as are typically used for FM systems. This scheme would be a natural scrambling scheme for CATV signals with the added advantage that it could improve system performance rather than reduce performance.

It should be emphasized that digital does not mean binary as is generally assumed. Digital, in communications theory, means that the source is a discrete source.

GEOMETRICAL INTERPRETATION OF SIGNALS.

A set of three numbers can always be used as the coordinates of a point in three dimensional space. In mathematics the concept of n-dimensional space is common. Similarly we can use the 2WT sample values, from the sampling theorem, to be the coordinates of a 2WT dimensional space. All of the points in this 2WT space represent all of the possible messages of length 2WT samples.

The size of this space is quite large. For a typical television program lasting an hour, with a bandwidth of 6 mhz. this space will have about

4.3×10^{10} dimensions. And the total possible messages which can be transmitted in this space will fill all of the possible points in the 2WT dimensional space.

When considering the length of a sample T, simple PCM systems do not use more than one word of 7 to 10 bits. The previous example of T equal to one hour would make for an extremely complicated and slow system. But, Ts which encompass several symbols are quite common in communication systems which work with very low signal to noise ratios.

The importance of this representation is that the mathematics of geometry can be used in discussing and solving communications problems.

If the co-ordinates of this space are at right angles, (orthogonal), then the distance from the origin to one of the points can be interpreted as 2W times the energy of the signal,

$$d^2 = 2WE \qquad (14)$$

$$= 2WTP$$

where P is the average power over the time T.

When noise is added to a signal, this corresponds to a new point in the space which is proportional to the RMS value of the noise.

Different co-ordinate systems can be used. A specific co-ordinate system which is used in many communications problems uses sines and cosines, such as used in the Fourier series expansion.

In modern communication theory, the vector representation of signals is typically used, [4]. In the theory, a set of orthonormal functions is selected. Each waveform $\{s_i(t)\}$ will be completely

determined by a vector and its coefficients :

$$s_i = (s_{i1}, s_{i2}, \ldots s_{iN}); \quad i=0,1,\ldots,M-1 \tag{15}$$

We now have M vectors $\{s_i\}$ defining M points in an N dimensional vector space, called the signal space, with N mutually perpendicular axes. If the set of unit vectors defining the space are

$x_1, x_2, \ldots, x_N$, then the signal

can represented as:

$$s_i = s_{i1}x_1 + s_{i2}x_2 + \ldots + s_{iN}x_N. \tag{16}$$

The key benefit to being able to visualize transmitter signals geometrically is illustrated in Figure 3, which shows four signals in a two-dimensional signal space.

The points s_0, s_1, s_2, s_3,

are all a distance:

$$d = E_s^{1/2} \tag{17}$$

from the origin, where

$$d = \int s_i^2 \, dt \quad i=0,1,2,3 \tag{18}$$

is the energy dissipated in a 1-ohm resister if the voltage is $s_i(t)$.

Note that X_1 and X_2 are:

$$X_1 = (2/T)^{1/2} \sin 2pi\ f_o t \tag{19}$$

$$X_2 = (2/T)^{1/2} \cos 2pi\ f_o t \tag{20}$$

respectively for

$0 < t < T$, and 0 elsewhere.

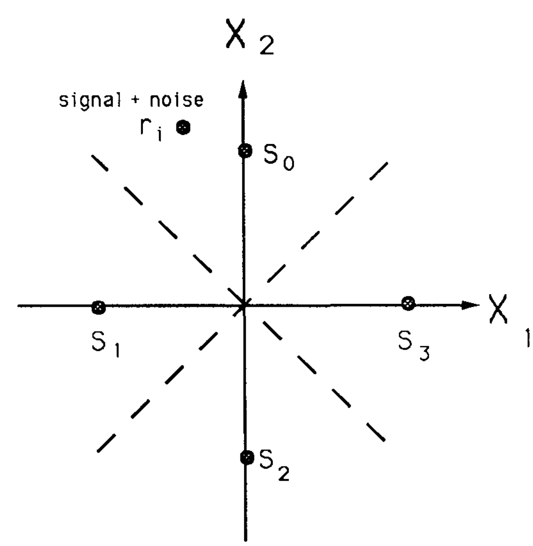


Figure 4 : Bi-orthogonal signal set

$d = \sqrt{E_S}$

(distance of signal point from origin)

Vector diagrams, of the type shown, are convenient ways of modeling signals for design-O*p+90Xmitted. ing various signal sets for transmission. From these diagrams the optimum decoding algorithm can be developed. Point r_i, in Figure 4, represents a signal s_i to which noise has been added. Equally spaced regions have been defined by the dashed lines. The signal points enclosed within these dashed lines define the respective optimum decision regions for each signal, when the a priori probabilities of the input signals are not know. This is known as a *maximum-likelyhood* receiver. If the received signal vector r_i is in the region associated with a particular s_i, then the s_i in that region is selected as the signal transmitted. It can be shown that choosing in this manner is optimum in the sense that the probability of an error is minimized.

In the general case, if the a priori probabilities are known, then the shape of the decision regions are shaped such that the optimum receiver, on observing the received vector r, sets the estimate of the received signal m = m_k whenever the *decision function* :

$$P[m_i]\ p_r(r|s=s_i)$$

is maximum for all i = k.

In the case where the input probability functions are known, the symmetrical placement of the signal vectors, s_i, are not optimum. This could imply that a specific receiver might readjust the *decision function* depending on the specific television program to be transmitted.

COMPARISON OF TELEVISION TRANSMISSION SYSTEMS FOR CATV.

We are now armed with tools to evaluate various systems to be used for transmitting A/V signals over a CATV network.

Motivation

Why should we not just keep with the current methods ? Primarily , as was stated previously, we have had essentially the same television transmission system for the last 50 years. The new ATV and HDTV television systems being proposed are changing the requirements for transmission systems. Unlike the early days of television, where there was only the broadcast channels, there are now several television delivery methods capable of delivering the new ATV pictures - DBS, fiber telephone systems, video tapes, CATV, etc.. The media which is capable of providing the best possible pictures, with the most consumer convenience, at competitive prices, will have an edge in the competitive market. But, if traditional analog approaches are used, for new ATV systems, compression of a 30 mhz. HDTV baseband signal to a 6 mhz. band could be stretching the capabilities of the channel. This in itself is not bad, because there are several other channels available with considerably more bandwidth available, and the consumer will choose the one they like best. The near term dilemma is the fixation on making the new television systems compatible with existing receivers, as opposed to the opposite, making the old NTSC system compatible with the new HDTV receivers. Cable has a long tradition of making interface boxes to consumer electronics equipment, for the purpose of bringing more services to the consumer.

SYSTEM EVALUATION.

How can different television transmission systems be compared on an equal basis ? How efficiently is the transmission channel utilized ? What is the net effect at the television receiver, in terms of picture quality as measured by S/N and cost ?

What are the physical limitations of the channel and what is an ideal model by which we can use to measure all systems? Of course, Shannon's "Channel Capacity" comes to mind.

To test this method, a QPSK system and a typical FM system, used for "Super Trunking", are compared to illustrate how the concepts of information theory and "Channel Capacity" can be used to rate the two systems.

Analog systems do not lend themselves well to discrete analysis, primarily because, once a particular discrete model is selected for the analog system, changing parameters of the discrete system can dramatically alter the fundamental characteristics of the modeled analog system, such that it may no longer be the same system. To get around this difficulty, the analog systems will be compared to an equivalent digital system of the same bandwidth and information transmission characteristics at the receiver. The comparison will then be made between CNR required in the channel, with the *channel encoder* of the digital system being selected to give approximately the same bandwidth in the channel as the analog system. The difference in required CNR will be an indication of how well each system utilizes the channel. Or more specifically how much power we need in order to get a specific level of performance.

An example is given to illustrate the method. Most of the calculations are approximations. The important aspects are the trends and relative magnitudes, not the exact numbers.

FM System:

This system is designed to use 40 mhz. channel bandwidth, 65 dB video SNR, 35 dB channel CNR, and it also uses pre- and de-emphasis which accounts for 12 dB increase in video SNR.

In the comparison, 12 dB is subtracted from the video SNR because the digital system used, did not use this technique, but it could achieve the same effect by using variable sample sizes. This would complicate the calculations and would not give any more insight into the comparison.

Digital System:

A digital system is selected to match the parameters of the FM system.

Sampling rate of 12 Mhz. Although this is lower than what is typically used for digitizing video signals, much lower rates can be achieved by pre and post processing and using statistical sampling methods.

A 7 bit word was selected to give a video SNR of 53 dB. This is about the same as the FM system with 53 dB, when the 12 dB of pre-emphasis is subtracted.

This then gives a channel bit rate of:

7x12 = 84 mhz.

We can use a bandwidth equal to the bit rate.

We then select a Bi-Orthogonal signal set, (QPSK) to reduce the channel bandwidth to 42 mhz. If the bandwidth of the FM system were greater, giving more video S/N, we would use more bits in our word. If the FM system used less bandwidth, with the same S/N, then we would use a higher dimensional signal set, such as an 8-phase system.

Information comparison.

The information communicated by both systems is the same since the video S/N and the video bandwidth were chosen to be the same. The Channel Capacity is determined by the dimensionality W, and the maximum CNR of the channel.

If the FM parameters are used, it is seen that the Capacity for this channel is :

$$C = W \log_2(1 + CNR)$$

$$= 40 \text{ mhz. } \log_2(1 + 56)$$

= 233 million bits.

(35 dB = 56)

At the receiver :

R (rate of information) =

$$4.2 \text{ mhz. } \log_2(1+ 128)$$

= 29 million bits

An efficiency can be calculated:

R/C = 29/233
= 12.4%

For the QPSK case selected, R is the same as in the FM case, since it was chosen that way. The C, however, is different since less CNR is required in the channel.

$$C = 42 \text{ mhz. } \log_2 (1+25)$$

= 197 million bits.

R/C = 29/197
= 14.7 %

The percentage difference is slight, but the digital system was constructed to match the equivalent FM system with a corresponding 7 dB_v less power required in the transmitter. This means that lower cost lasers can be used, or signals can go twice as far, or one more level of splitters deeper into the fiber system can be accommodated.

	FM	QPSK
CHANNEL CHARACTERISTICS		
Carrier to Noise (CNR)	35 dB	28 dB (assumes P{e} = $1x\ 10^{-7}$)
Band Width	40 mhz.	42 mhz.
Capacity	233 m bits	197 m bits
RECEIVER CHARACTERISTICS		
Video signal to noise (SNR)	53 dB	53 dB
Video bandwidth	4.2 mhz.	4.2 mhz.
Information rate of video - (R)	29 m bit	29 m bit

The key point is that if the 7-bit system is changed to an 8-bit system, the bandwidth will increase by 1/7 th., but the video S/N will double. This is the exponential tradeoff between bandwidth and SNR that is inherent in PCM and digital systems that is not present in analog FM systems, whose tradeoff is only linearly related. To get an equivalent performance increase in the FM system would require doubling the bandwidth.

S/N_{FM} = function of (log n)

S/N_{PCM} = function of n

(where n is key parameter which gives S/N improvement, deviation in FM, more quantization levels in PCM).

Note, that no compression or coding schemes were used to reduce bandwidth requirements or to improve P(e) of receiving the digital messages. both systems presumed white Gaussian noise. Typical CATV systems have higher levels of coherent noise such as cross-mod and inter-mod. The effects of coherent noise can be substantially eliminated in digital systems by coherent detection and the proper design of the encoders and decoders.

The sampling method can also be improved such that both the input sampling and output recovery are processed digitally.

CONCLUSION.

It is presumed that the entertainment delivery system of the future will have more bandwidth capability than coaxial systems, such as fiber or satellite.

Although it can be shown that certain FM systems can outperform certain digital systems, the choice for the future is clear. Even if cost is the primary reason for not going the digital path today it does not make sense not to make the investment in digital for the future. If the costs and performance of analog and digital systems are almost equal today, the greatest return on investment can be achieved with a digital approach for the future.

The digital CATV transmission system of the future will look much like the block diagram of Figure 1, where

The, 2. source encoder will:

reduce the redundancies in the source.

The 3. channel encoder will:

select transmission schemes, (like Bi-orthogonal signals),

add error correcting codes,

and add program encryption.

Many of the communications problems associated with digital systems have already been solved for the telecommunications, military, and aerospace industries.

The key to providing digital to the home isn't fiber optics to the home, it is digital inputs into the new ATV receivers. And just as in the past, Cable will provide digital to NTSC encoders for NTSC televisions and VCRs. And at the same time getting rid of many of CATVs signal quality problems such as ghosts, and various other modulation effects.

It now remains for the industries which deliver home entertainment to come into the 20 th. Century, before the 21 st. Century is upon us.

REFERENCES

[1] C.E. Shannon "A Mathematical Theory of Communications" Bell Systems Tech. Journal vol. 25 (Jul. 1948), pp 379 to pp 423, vol.(Oct. 1948), pp 623 to pp 656

[2] B.M. Oliver, J.R. Pierce, C.E. Shannon, "The Philosophy of PCM", IRE, 36, 1324-1331, November 1948.

[3] C.E.Shannon, "Communication in the Presence of Noise", Proceedings of the IRE, 37 No.1, 10-21, January 1949.

[4] J.M.Wozencraft&I.M. Jacobs, "Principles of Communication Engineering", John Wiley & Sons, 1967.

[5] R.G. Gallager, "Information Theory and reliable Communication", John Wiley & Sons, 1968.

[6] A.J. Viterbi, "Principles of Coherent Communication", McGraw-Hill, 1966.

[7] M. Schwartz, W.R. Bennett, S. Stein, "Communication Systems and Techniques, McGraw-Hill, 1966.