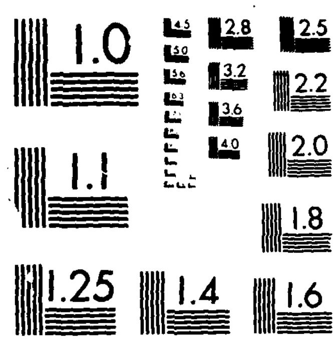
AD-A171 666 A CATV (CABLE TELEVISION)-BASED HIGH-SPEED 1/2
PACKET-SWITCHING NETWORK DESIG.. (U) MASSACHUSETTS INST
OF TECH CAMBRIDGE LAB FOR COMPUTER SCIENCE..
UNCLASSIFIED D C FELDMEIER APR 86 MIT/LCS/TR-359 F/G 17/2 NL



ROCOPY RESOLUTION TEST

MIT/LCS/TR-359

A CATV-BASED HIGH-SPEED PACKET-SWITCHING NETWORK DESIGN

David Charles Feldmeier

April 1986

This research was supported (in part) by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by the Office of Naval Research under contract number N00014-83-K-0125.

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER MIT/LCS/TR-359	2. GOVT ACCESSION NO. ADA171666	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (*and Subtitle*) A CATV-Based High-Speed Packet-Switching NETWORK DESIGN		5. TYPE OF REPORT & PERIOD COVERED M.S. Thesis April 1986
		6. PERFORMING ORG. REPORT NUMBER MIT/LCS/TR-359
7. AUTHOR(s) David Charles Feldmeier		8. CONTRACT OR GRANT NUMBER(s) DARPA/DOD N00014-83-K-0125
9. PERFORMING ORGANIZATION NAME AND ADDRESS MIT Laboratory for Computer Science 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DARPA/DOD 1400 Wilson Boulevard Arlington, VA 22209		12. REPORT DATE April, 1986
		13. NUMBER OF PAGES 137
14. MONITORING AGENCY NAME & ADDRESS(*if different from Controlling Office*) ONR/Department of the Navy Information Systems Program Arlington, VA 22217		15. SECURITY CLASS. (*of this report*) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (*of this Report*)

Approved for Public Release, distribution is unlimited

17. DISTRIBUTION STATEMENT (*of the abstract entered in Block 20, if different from Report*)

Unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (*Continue on reverse side if necessary and identify by block number*)

cable television, metropolitan area networks, broadband networks, access control techniques

20. ABSTRACT (*Continue on reverse side if necessary and identify by block number*)

A high-speed packet-switching data network to the home can be built on an existing, unmodified, residential cable television (CATV) system. This thesis considers the theoretical and practical technical aspects of providing such a service and suggests a possible system design. All network data must pass through the CATV hub, so the network design is divided into three major parts: upstream transmission, downstream transmission, and access scheme.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (*When Data Entered*)

Upstream transmission is difficult because of the high noise level on the upstream channel caused by ingress of shortwave signals and impulse noise. The noise level is increased by the CATV system topology that funnels all system noise to the headend. Several noise-reduction techniques must be used simultaneously for robust upstream transmission. The downstream channel has low noise, but the data signal must be compatible with the CATV system, video signals and television receivers. Vistigial sideband data modulation is suggested for total system compatibility. Existing access schemes are suggested as possible solutions. The best techniques for upstream transmission, downstream transmission and access scheme are combined into a single proposed system.

A CATV-Based High-Speed Packet-Switching Network Design

David Charles Feldmeier

April 1986

This research was supported in part by the Defense Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under contract number N00014-83-K-0125

Massachusetts Institute of Technology
Laboratory for Computer Science
Cambridge, Massachusetts 02139

A CATV-Based High-Speed Packet-Switching Network Design

by
David Charles Feldmeier

Submitted to the
Department of Electrical Engineering and Computer Science
on April 24, 1986 in partial fulfillment of the requirements
for the Degree of Master of Science

Abstract

A high-speed packet-switching data network to the home can be built on an existing, unmodified, residential cable television (CATV) system. This thesis considers the theoretical and practical technical aspects of providing such a service and suggests a possible system design. All network data must pass through the CATV hub, so the network design is divided into three major parts: upstream transmission, downstream transmission, and access scheme.

Upstream transmission is difficult because of the high noise level on the upstream channel caused by ingress of shortwave signals and impulse noise. The noise level is increased by the CATV system topology that funnels all system noise to the headend. Several noise-reduction techniques must be used simultaneously for robust upstream transmission. The downstream channel has low noise, but the data signal must be compatible with the CATV system, video signals and television receivers. Vestigial sideband data modulation is suggested for total system compatibility. Existing access schemes, such as those for local area networks and satellite networks, are unsuitable for a high-speed CATV-based network. Modified versions of two satellite access schemes are suggested as possible solutions. The best techniques for upstream transmission, downstream transmission and access scheme are combined into a single proposed system.

Key Words: cable television, metropolitan area networks, broadband networks, access control techniques

Thesis supervisor: Jerome H. Saltzer.

Acknowledgments

I would like to thank my thesis advisor *Jerry Saltzer* for proposing a CATV-based computer network as a research project and for his thoughts on various aspects of communication via residential CATV systems.

John Cafarella of Micrilor spent many hours with me, discussing aspects of modulation and error-correction coding for noisy channels, both of which are important for the design of the upstream transmission system. His help and expertise are greatly appreciated.

George Papadopoulos of the University of Patras, Greece, spent a few weeks considering possible access schemes for CATV networks. He and I had several meetings discussing the pros and cons of various access schemes.

Stanly Reible of Micrilor helped with measurement and interpretation of noise on some local CATV systems.

My special thanks to *Patrick Mock*, *Katy Isaacs*, and *Corine Bickley* for their helpful comments on my thesis.

Table of Contents

Chapter One: Introduction **9**

1.1 Background 9
1.2 Goals 11
1.3 Communication Systems to the Home 13
1.4 Previous Work 15

Chapter Two: An Introduction to Cable Television **18**

2.1 Community Antenna Television 18
2.2 Distribution Plant 19
2.3 Residential versus Institutional Cable 21
2.4 Constraints on Network Design 22
2.4.1 Central Clock 22
2.4.2 Economics 23
2.4.3 CATV Topology 24
2.4.4 Noise 24
2.5 Problem Division 26
2.6 Summary 26

Chapter Three: Upstream Data Transmission **28**

3.1 Power Measurements on CATV Systems 28
3.2 Phase Distortion 29
3.2.1 The Cause of Phase Distortion 29
3.2.2 The Effects of Delay Distortion 30
3.2.3 Overcoming Delay Distortion 32
3.3 System Characteristics 32
3.3.1 CATV Topology 33
3.3.2 Digital Regenerators 33
3.4 Upstream Noise 33
3.4.1 White Noise 34
3.4.2 Narrow Band Noise 35
3.4.3 Impulse Noise 36
3.5 Effective Channel Utilization 37
3.5.1 Modulation Techniques 37
3.5.1.1 Noise and Signal Power Considerations 38
3.5.1.2 Delay and Interference Considerations 44
3.5.2 Selective Spectrum Utilization 44
3.5.3 Coding 46
3.6 System Noise Budget 48

3.7 Summary 51

Chapter Four: Downstream Data Transmission **52**

4.1 CATV System Considerations 52
4.2 Frequency Shift Keying 53
4.3 Vestigial Sideband Modulation 54
4.3.1 Transmitter Baseband Processing 55
4.3.1.1 Number of Signal Levels 55
4.3.1.2 Low-pass Filtering 56
4.3.2 Amplitude Modulation 57
4.3.3 Transmitter Bandwidth Restriction Filter 57
4.3.4 Vestigial Sideband Filtering 59
4.3.5 Signal Detection 61
4.3.5.1 Coherent Detection 62
4.3.5.2 Envelope Detection 62
4.4 A Sony VSB Data Transmission System 65
4.4.1 Transmitter 65
4.4.1.1 Nyquist Filtering 65
4.4.1.2 Modulation 66
4.4.1.3 Transmitter Bandpass Filter 66
4.4.2 Receiver 67
4.5 Channel Capacity 68
4.6 Master Clock Reception 70
4.7 Summary 72

Chapter Five: Network Access Schemes **73**

5.1 Introduction 73
5.2 CATV Access Schemes Considerations 73
5.2.1 Packet-Switching Broadcast Networks 73
5.2.2 Preliminary Access Scheme Evaluation 74
5.2.3 Monitoring of the Upstream Channel 75
5.2.4 An Example CATV System 76
5.2.5 Expected Traffic 77
5.2.6 Traffic Modeling 78
5.2.7 The Comparison Metric 79
5.3 Deterministic Access Schemes 82
5.3.1 Polling 82
5.3.2 Token Passing 83
5.3.3 Fixed Reservation 84
5.4 Contention Access Schemes 88
5.4.1 Aloha 88
5.4.1.1 Collision Detection 89
5.4.1.2 P-Persistent CSMA 91
5.4.2 Contention Reservation 92

5.5 Propagation Delay Determination 95
5.6 Summary 97

Chapter Six: System Design **98**

6.1 Downstream Transmission 98
6.2 Upstream Transmission 99
6.3 Proposed Access Scheme 102
6.4 The Network Center 104
6.5 The Network Interface 105
6.6 Summary 106

Chapter Seven: Conclusion **108**

7.1 Major Points 108
7.2 The CATV System as a Metropolitan Area Network 109
7.3 Future Research 110
7.4 Conclusion 111

Appendix A: Filter Power Loss **117**

1 Pre-Modulation Filter Loss 117
2 Transmitter Bandpass Filter Loss 120
3 Determination of VSB Filter Loss 121
4 Calculation 125
4.1 Transmitter Bandpass Filter 125
4.2 VSB Calculations 126

Appendix B: Bit Error Rate of an Envelope Detector **127**

Appendix C: Contention Reservation Slot Calculation **134**

Table of Figures

Figure 3-1: CATV Network Delay Errors as a Function of Signal Frequency and Bandwidth (from [30]) 30

Figure 3-2: Measurement Method for Linear Delay Distortion (from [28]) 31

Figure 3-3: Maximum Signal Impairment as a Function of Delay for Phase and Frequency Modulation (from [28]) 31

Figure 3-4: Measured Upstream Noise Spectrum from 0 to 30 MHz with a 20 dBmV Reference for an Example CATV System (from [6]) 36

Figure 3-5: Probability of Bit Error as a Function of E_b/N_o for Several Modulation Systems (from [7]) 39

Figure 3-6: Error Rates for Several Modulation Systems (from [27]) 41

Figure 3-7: Bit Error Probability as a Function of C/N Ratio for BPSK (from [23]) 45

Figure 4-1: Frequency Responses for the Receiver's Vestigial Sideband Filter and the Transmitter's Bandpass Filter (from [25]) 58

Figure 4-2: Bit Error Rate versus E_b/N_o for BPSK (from [7]) 63

Figure 4-3: Frequency Characteristics of the Sony Pre-modulation Bandpass Filter (from [18]) 66

Figure 4-4: Bit Error Rate as a Function of C/N Ratio for the Sony Digital Transmission System (from [18]) 69

Figure 1: *Vestigial Sideband Filter Frequency Response (from [25])* 122

Table of Tables

Table 3-1: Table of Concatenated Codes With a Bit Error Rate Below 10^{-12}(from [12]) 50
Table 4-1: Table of Error Rate versus Sony System Carrier Power 68
Table 5-1: Table of k and E_w versus s 95

Chapter One

Introduction

The demand for high-speed communication to the home is increasing as the economy becomes information-oriented. Although no suitable system exists to meet this demand, a communication network to the home can be built quickly and inexpensively on an existing residential cable television system. The goal of this research is to propose the design of a system suitable for providing high-speed packet-switched communication to the home.

1.1 Background

As the United States shifts from a manufacturing economy to an information-oriented economy, data processing and computers play an increasingly important role in our work. For people who work with information, it is feasible to work one or more days a week at an office in the home. Working at home has several advantages, including reducing the time and cost of commuting. Work during non-business hours becomes convenient, so that personal schedules may easily be shifted. Work at home also allows one to work during the day without interruptions.

The cost of duplicating office equipment, such as a personal computer, at the office in the home becomes economically competitive with commuting as the price of electronics declines. For work at home to be productive, a worker must have access to the resources available in the office; for example, the use of a mainframe computer or shared data-storage system. The worker at home must not become isolated from the office work environment.

A modern office environment consists of computers and peripherals connected together with a *local area network*. A local area network, or LAN, provides inexpensive, high-speed communication over a limited area. Many LANs have a data rate of 10 megabits per second and reside within a building, although network diameters of over a kilometer are possible. Resources that are too expensive to provide to each user, such as high-quality printers or mass-storage devices, are shared using such a network. The network allows the user of a personal computer (PC) to access those services that are available on a mainframe.

An example of an application that requires high-speed communication is remote disk access. The disk drives in a PC have a small capacity and a low throughput, and do not allow data to be accessed by more than a single computer. An alternative is to provide a computer with a large storage system on the network to act as a remote disk that can be accessed by other computers on the network. With a high-speed network, data can be retrieved from a remote disk as quickly as from the local hard disk drive. If data are stored on a remote disk, a worker at home can use the data that are available at work without having to keep an extra copy at home.

Another application for high-speed networks is the transmission of pictures, such as circuit diagrams. *Often, the ability to show something to a person rather than describe* it saves time, but only if a picture can be transmitted reasonably quickly. To transmit the bitmap display of an engineering workstation[1] takes a tenth of a second at 10 Mbps. If the worker at home is restricted to a conventional 1200 bps modem on a phone line, the same transmission takes almost 14 minutes. With a high-speed network, screen displays can be transmitted to a workstation as quickly as turning the pages of a book. A new generation of computers with faster processors, graphic displays and larger memories increases the need for high-speed communications. High-speed, low-cost communications must be brought to the home to support the services provided at the office by a local area network.

[1]A Symbolics 3600 workstation has an 1100x900 pixel bitmap display.

1.2 Goals

The office in the home can be efficient if the home worker has access to the same computational resources that are available at work. Because distributed computing is becoming common and computer prices are decreasing, data communication from the home to the office increasingly will be among computers, rather than between a computer and a home terminal. Computer to computer communication is bursty, with a high peak-to-mean ratio of network use. Unlike computer to terminal communication, intercomputer communication traffic is symmetric, with comparable amounts of data in both directions.

The goal of this thesis is to describe a system that provides computer communication to the home, with performance similar to that of a local area network. As with a LAN, the system transports packets on a single, shared broadcast channel, rather than using point-to-point links. This network provides high-speed service on demand to a single user at a time and any one user occupies the channel for a very short time, which allows many users to share the same channel. Although a high-speed network is needed to handle the expected traffic bursts, each individual station on the network uses this capacity only a small percentage of the time. This type of operation is compatible with computer communication, and data transmission and reception rates are identical because of the expected traffic symmetry.

Computer communication requires high-speed bursts for short periods of time, but what is the minimum necessary speed for acceptable performance? A network to the home is a metropolitan area network because of the large area that it covers. The Institute of Electrical and Electronics Engineers (IEEE) 802.6 committee is developing metropolitan area network standards and the committee is mainly interested in LAN interconnection, and high-speed voice, video and data service. The 802.6 committee charter stipulates that 1 Mbps is the minimum speed for a proposed network standard. Because the network to the home connects computers to LANs, the traffic should be similar to that of LAN interconnection, so 1 Mbps to the home is the minimal acceptable data rate.

Although the data rate for the entire network is at least 1 Mbps, another important consideration is the minimum throughput for a single user. The bursty nature of computer communication demands a high throughput for a single user because at any given time, only a few stations have data to transmit. The hard disk drive on a personal computer has a raw bit rate of about 500 Kbps. Since remote virtual disk access is a desired application, and one of the most demanding, 500 Kbps should be the network throughput for a single user. Since the data rate is still 1 Mbps, the minimum network efficiency is 50%.

Throughput and delay on a network are related, so although throughput may be increased at the expense of delay, delay must not be excessive. Interactive activities, such as remote login, have a maximum tolerable delay between the queueing of a packet and its transmission. A decent typist will have bursts of over 100 words per minute, and remote login requires two packets per character (one of which is an acknowledgment). This is about twenty packets per second, ten packets in each direction, which implies a maximum delay of 50 milliseconds between packet queueing and transmission.

The computer in the home will be connected to a local area network at the office, and it is simplest if both networks use similar protocols. Specifically, LAN protocols correct data errors by retransmitting damaged packets, so the network error rate should be low enough for most packets to be delivered undamaged. Local area network research at MIT suggests that the correct error rate to control is the frequency of corrupted packets, rather than the bit error rate, because the system overhead of detecting and retransmitting damaged packets dominates the resulting performance. The performance of automatic repeat request error correction depends critically on the packet error rate of the channel. If the error rate is too high, the retransmission overhead dominates the system. Experience also suggests a network design target of less than one corrupted packet per 1000 for acceptable performance.

A communication network to the home must be affordable if it is to be used, so the system must be comparable in cost to systems that provide similar services. The two closest systems are local area networks and dial-up telephone modems, both of which cost $500 to $700 per interface in 1985 dollars.

The parameters above define a system to provide high-speed communication to the home. There are several existing technologies that could be used to build this network, and these are discussed in the next section.

1.3 Communication Systems to the Home

To provide a communications network to the home at a reasonable cost, the system must use either free-space transmission (microwave or light) or previously installed cables. It is impractical to install dedicated cables to each home because of cost and right-of-way considerations.

The telephone system is commonly used for data communication to the home. However, telephone technology limits the data rate to about 9600 bps, which is too slow to provide acceptable packet switching service to the home. Data communication is also an inefficient use of telephone company resources. Telephone switching equipment expects many calls of short (3 minute) duration, but data communication holds a circuit for much longer. Also, the bursty nature of computer communication means that the circuit will often be idle, so system capacity is wasted. In addition, the long economic recovery time for telephone switching equipment precludes its replacement for many years.

A new service to be offered soon by the telephone company is the Integrated Services Digital Network (ISDN). This service provides two 64 Kbps circuit-switched channels and a 16 Kbps packet-switched channel to the home for a total of 144 Kbps. Although this will be a great improvement over a conventional telephone line, it is still slower than desired and does not exist yet.

Another method of data transmission to the home is the FM Subsidiary Communications Authority (SCA) channel. Commercial FM stations can broadcast data on an SCA channel simultaneously with normal broadcasts. The Boston Community Information System uses this type of broadcast channel to provide data to the home [13]. The system broadcasts data to homes in a metropolitan area and is directed toward residential users. The system is also distributed - all data is broadcast to all receivers and it is up to each station to selectively receive the information of interest, store and present the information to the user in a suitable manner. A disadvantage of this system is that the data rate is only 4800 bps, and therefore too slow for the applications that we are interested in.

An alternative is a satellite-based network. Satellite equipment prices are falling and Direct Broadcast Systems (DBS) are designed to provide video entertainment to the home via satellite to small, inexpensive dish antennas with inexpensive electronics. The major problem with satellite systems is that if the satellite is in geosynchronous orbit, propagation delay to the satellite and back is a quarter of a second. This means at least half a second passes before a reply can arrive from any other computer on the network, which is too long. Propagation delay could be reduced by lowering the satellite orbit, but a low orbit requires more satellites so that one is always within line-of-sight, as each satellite would be in range only a short time. In addition, the dish antennas at each home would require a tracking mechanism, which increases the cost for each station. DBS also does not offer a simple return path from the home. Communication from the home requires a powerful transmitter operating in the Gigahertz range. The antenna beam-width must be narrow and directed well enough to avoid interference with other satellites. These requirements could make the return path quite expensive.

The most straight-forward way of approaching the problem is to use an existing link to the home - the residential cable television system. A residential Community Antenna Television (CATV) system distributes television signals with a network of coaxial cables branching out from the cable headend to homes in a community. Modern CATV systems not only provide video channels to the home but also carry signals from the home back to the CATV headend. The coaxial cable is a high-bandwidth medium that is already installed to the home, meaning that a communication system could be added at a low marginal cost. In addition, CATV has the advantage of a small propagation delay relative to the area covered by the network and it does not require any of the already crowded radio spectrum as a microwave system would. Of the possibilities considered above, a CATV system provides the best basis for the desired data communication network to the home.

1.4 Previous Work

Several existing or proposed systems that are built on residential CATV system or use CATV equipment are mentioned below. Although none of them provide a high-performance channel to the home, they show the current state of the art.

The prototypical residential CATV-based communications system is the Warner-Amex QUBE system [20]. This is the first system to take advantage of the two-way capability of a CATV system. The QUBE system runs a single channel at a 256 Kbps rate and uses an adaptive polling access scheme. A polling system queries each user on the system from a central point one at a time for data traffic. Although polling is simple and inexpensive, it is an inefficient access scheme for a network the size of a CATV system because most of the time, the polling system is waiting for a poll response. Adaptive polling is a better system that polls active stations more often than inactive ones, but still the network utilization is low.

A more recent system developed jointly by General Instruments and Sytek is MetroNet, which is designed to bring videotex service to the home via existing residential CATV systems [15]. A MetroNet is frequency divided into 50 channels, each of which has a data rate of 128 Kbps and uses a Carrier-Sense Multiple Access with Collision Detection (CSMA/CD) access scheme. This data rate is almost that provided by ISDN, and although it is adequate for computer to terminal communication, it is too slow for demanding tasks such as remote-virtual-disk access.

Rather than a single high-speed channel, Metronet has many slower channels. Videotex and terminal traffic is less demanding than computer traffic and the narrower bandwidth of the slower channels allows the system to overcome noise on the cable more easily [8]. The relatively slow channel speed allows MetroNet to efficiently use CSMA/CD as an access scheme, but efficiency declines if the data rate is increased.

Unlike the two systems described above, the following two systems are incompatible with a residential CATV system. The Ethernet system is suited to an industrial network within a building or a company. Homenet is a generalized version of an Ethernet that is designed to cover a large area efficiently.

Digital Equipment Corporation developed an Ethernet system with broadband transceivers that operates on CATV equipment [2]. However, the system diameter is limited to a few kilometers because the system operates at 10 Mbps and uses a CSMA/CD access scheme, which is efficient only over a limited area at high data rates. The system also requires 18 MHz of bandwidth in each direction, which would be difficult to fit onto the upstream channel of a residential CATV system because of noise considerations.

Homenet is a system proposed by Bell Labs designed to be built with CATV components [16]. The system consists of cells of transmitters that contend for the channel using a CSMA/CD system. All transmitters within a cell use the same frequency and receivers shift to the frequency of the transmitter's cell. This system allows the total system to be very large, but since all contention is within a cell of a kilometer in diameter, the access scheme remains efficient.

Without modification, the structure of a residential CATV system could accommodate only one cell, because each cell needs a frequency translator and only the headend could have a translator. As a consequence, the access scheme would be inefficient and Homenet would not work well on a residential CATV system.

Little research has been done concerning high-speed communications to the home on a residential CATV system. Some systems are low speed for providing videotex services; some are high-speed systems that require extensive modification to the CATV system. The thesis of D. L. Estrin discusses the technical and regulatory issues surrounding the use of a residential cable television system for data communications [9].

This thesis is a study of the technical and system issues of providing high-speed, packet-switched computer communication to the home on an existing residential CATV system. The purpose of this thesis is to review the type of communications necessary for work at home and to describe a system that provides this communication. This thesis outlines a system that provides LAN-like communications to the home, and is specifically designed for peer communication among computers on an unmodified residential CATV system.

Chapter 2 contains a tutorial on residential cable television. It describes the design of a CATV system and points how it might be used for communication to the home. An analysis of the system leads to the problem being broken into three parts: access scheme, upstream transmission, and downstream transmission.

Chapter 3 discusses transmission from the home to the CATV headend. The tree-shaped CATV system acts as a noise funnel that channels the amplified sum of the noise over the entire system to the headend. A system is proposed for high-speed transmission in this adverse environment that includes a robust modulation scheme, selective spectrum utilization, and error-correction coding.

Chapter 4 covers communication from the CATV headend to the home. Because the CATV system is designed for communication in this direction, the design is relatively simple. The main considerations are compatibility with the current CATV system and equipment including commercial television receivers, and subscriber receiver cost.

Chapter 5 discusses access schemes for a CATV-based data network. It begins with the CATV system constraints and the network expectations, and explores possible access schemes. Suggestions for improved access protocols include a modification of a satellite reservation scheme to accept variable instead of fixed slot lengths and a method of increasing network efficiency by reducing contention on the system.

Chapter 6 proposes the design of a complete system for providing high-speed packet-switched communication to the home. The results of chapters 3, 4 and 5 are used to provide the pieces of the system. The chapter also describes network controller design and subscriber interface design.

Chapter 7 concludes the thesis by considering the strengths and shortcomings of the presented analysis. It also suggests directions for the continuation of this research.

Chapter Two

An Introduction to Cable Television

Since the high-speed data network to the home is to be built on a CATV system, the design of a CATV system must be understood. The technical characteristics of a CATV system will determine how a high-speed data network can be built on the system.

2.1 Community Antenna Television

A Community Antenna Television system is a tree-shaped structure of coaxial cables and amplifiers designed to distribute television signals to homes in a region. At the base of this tree is the system *headend*, from which the video signals originate. Although distribution of video entertainment signals from the system headend to the subscriber is the predominant design consideration, most modern CATV systems are two-way systems that carry signals both from the headend to the subscriber (downstream), and from the subscriber to the headend (upstream). This two-way capability of the cable system allows it to be the basis of a data communication network to the home.

A CATV system is frequency-divided into adjacent 6 MHz channels. The most common, single-cable design used for residential areas divides frequencies into three bands: downstream channels, upstream channels and the guard band. Downstream channels carry signals, typically entertainment video, from the cable system headend to the home. Downstream (or forward path) channels begin at 54 MHz and continue up to the highest frequencies carried by the system - as high as 450 MHz, depending on the system design and technology employed. The upstream (or return path) channels occupy 5.75 to 29.75 MHz and carry signals from the home to the cable headend. The frequencies from 29.75 to 54 MHz form a guard band between the upstream and downstream channels and no signals are transmitted at these frequencies. The cable spectrum below 54 MHz is referred to as the *subband*, and a cable system that allocates that spectrum to upstream channels is referred to as a *subsplit* system; residential cable systems are typical subsplit systems.

Each downstream channel normally carries a single vestigial sideband (VSB) video signal. VSB is the standard modulation for broadcast video signals in the United States, and it is an amplitude modulation system followed by a vestigial sideband filter that eliminates most of the lower sideband. The audio carrier is processed separately and combined with the video signal before transmission. Video carriers on modern cable systems are phase-locked to a master oscillator so that intermodulation distortion products of the carriers are phase-coherent with the carriers themselves. These coherent products are removed along with the carriers during demodulation at the television receiver, but sideband intermodulation products limit the subjective picture improvement to 4 - 6 dB [30].

2.2 Distribution Plant

The distribution plant of a CATV system is an array of coaxial cables and amplifiers that carry signals from the headend to the home. The amplifiers on such a system are adjusted for unity gain across a cable segment, which means that the gain of an amplifier is exactly enough to overcome the loss in the previous cable segment. Amplifiers maintain unity gain with Automatic Gain Control (AGC)[2], which allows the amplifier to adjust its gain automatically so that the signals leaving the amplifier are at the correct signal level. This dynamic adjustment allows the system to operate correctly as the loss characteristics of the cable change with time and temperature, but AGC alone is insufficient because signal loss is also a function of frequency. Signal loss is roughly proportional to the square-root of frequency, a characteristic called *slope*. Amplifiers have Automatic Slope Control (ASC) to correct for system slope by increasing the amplifier gain at higher frequencies. The control voltages for both of these amplitude controls are derived from either two special pilot tones or two designated video carriers. With two signals at different frequencies, the amplifier has enough information to

[2]Referred to in some literature as Automatic Level Control (ALC)

determine the proper gain and slope.

CATV amplifiers are unidirectional, but two-way systems need amplification in both directions. Two-way amplifiers are really two amplifiers in a single box, one for each direction. Since both upstream and downstream signals travel on the same cable, the upstream and downstream channels are separated at each amplifier by *duplexing filters*[3].

The CATV headend feeds into the *trunk* cable of the distribution plant, and the amplifiers along the trunk cable maintain the proper signal level throughout the system. Trunk cables span a city but cannot go everywhere - the trunk cable is too expensive and the cable system itself allows only a limited number of amplifiers in cascade. *Feeder cables* branch out from the trunk cable and serve a limited area, such as a neighborhood. A *bridging amplifier* amplifies signals from the trunk and drives the feeder cable. If a feeder cable is long or serves many homes, additional amplifiers called *line extenders* are used. Branching from the feeder cable to the individual home is done with directional splitters called *taps*. Taps are directional and couple power predominantly from signals traveling away from the trunk cable. This is so that mismatch reflections from taps farther down the cable are attenuated and do not cause interference at the television receiver. As distance from the trunk increases, the taps couple a higher fraction of the decreasing power so that the received signal levels at every home are about the same. Service drops are the flexible cables that run from the taps on the feeder cable to the home and are 50 - 250 feet long. Drop cables are the lowest quality cable in the system and are the most susceptible to interference or *ingress noise* from outside the cable system.

Heading upstream from the home, the directional tap couples into the feeder cable heading toward the trunk. The directional taps work well for upstream transmission because the signal is sent towards the trunk cable rather than to other subscribers. The feeder cable is connected to the trunk cable through an addressable, remote-control switch. The upstream channel can be fed into the trunk directly or through a 6 dB

[3] Referred to in some literature as *diplexing filters*

attenuator, or the feeder cable can be terminated at the switch to prevent any signals from entering the trunk. The switches are controllable from the headend, and they are used to reduce the upstream noise by limiting the number of upstream feeder cables that are simultaneously feeding signals and noise into the upstream channel of the trunk cable. Another use for these switches is the location of parts of the system where noise is entering - by selectively disconnecting parts of the system and observing the noise, bad cable segments can be isolated.

Although all CATV systems operate similarly, large urban areas may be covered by multiple hub networks rather than a single large system. These hubs are connected to a master headend by microwave, fiber optic or coaxial cable links. The master headend transmits the programming to all hubs, and each hub is a complete distribution system. This thesis assumes a one hub CATV system, but a system with multiple hubs could run a network on each hub. These networks could then communicate with each other through the master headend. Each hub would have a gateway, and these gateways would communicate with each other through the master headend. By treating each hub as an independent data communication network, the results of this thesis are directly applicable except that a network controller is needed at each hub rather than a single controller at the headend.

2.3 Residential versus Institutional Cable

This thesis is concerned with data communication to the home and focuses on residential cable systems; sometimes a community will also have an Institutional Network (I-Net) broadband cable system. Unlike a residential system, an I-Net is specifically designed for bidirectional use and provides communication among government offices, schools and other institutions. The frequency allocation on an I-Net is known as *mid-split*, which means that the low VHF spectrum is for upstream transmission and the high VHF spectrum is for downstream transmission, with a guard band in between.

An I-Net is a better system for data transmission because it is designed to provide bidirectional communications to targeted areas. The I-Net has fewer and higher quality drop cables than a residential system, thus reducing the total noise on the system. Because the I-Net is mid-split, upstream channels exist at higher frequencies than on sub-split systems. These frequencies are above the short-wave bands, and therefore a major source of ingress noise for lower frequencies is absent. The upstream channels are in the same spectrum as broadcast television, whose standard frequency allocation practices insure empty channels. Because an I-Net has much more hospitable upstream channels, a system designed to work on an I-Net will not necessarily work on a residential CATV system.

2.4 Constraints on Network Design

2.4.1 Central Clock

An advantage of communication on a residential CATV system over most other communication systems is the existence of a central clock. All video carriers on the downstream channels of a modern CATV system are usually phase-locked to a master oscillator at the headend, and the clock can be recovered from any of these carriers. Even on older systems that are not phase-locked, any single video carrier can be used as a central clock.

The master clock can be recovered with little error because of the high signal-to-noise ratio of the downstream channel. FCC minimum standards, which most CATV systems exceed by a wide margin, require that the downstream channel maintain a minimum of 36 dB carrier-to-noise ratio for its downstream video signals. The smallest that the video signal may be is at the reference white level, where the signal is 12.5% +/- 2.5% of peak envelope level [7]. Since signal power is proportional to the square of the amplitude, if the minimum amplitude of the signal is 10% of the unmodulated carrier amplitude, then the power is 1% of the unmodulated carrier power. If the carrier is at 36 dB above noise, then the minimum signal power of 1% of this is 16 dB above noise for a 4 MHz bandwidth (a standard bandwidth width for CATV measurements). If a clock receiver bandwidth is 100 kHz, then the signal-to-noise ratio is increased by

$$10\log\frac{4\times10^6}{10^5}$$

or 16 dB, for a minimum signal-to-noise ratio of 32 dB. This is a very high signal level, and the recovered clock signal should be very accurate.

2.4.2 Economics

It is important to keep in mind the concerns of the cable operator when designing the network. Most of the income from a residential cable system is derived from entertainment video, and this will continue to be true even with the advent of data communication. Consequently, the operator is unlikely to install any new system that could affect his major source of revenue. The network to the home must therefore be compatible with the existing CATV system and television receivers. An already installed CATV system has little frequency flexibility because channel assignments are reflected in filters installed on telephone poles; the network must be able to operate in whatever spectrum happens to be available.

Another problem is the maintenance of the physical plant. The amount of maintenance necessary for good picture quality on the downstream channel is inadequate to provide a low-noise upstream channel. Presumably, income from data services will be relatively low and the operator may be unwilling or unable to provide the aggressive maintenance that a low-noise upstream channel requires. If the upstream channel is maintained, it is because the operator already has another service running on the upstream channel. The network to the home should not conflict with any such existing services.

The operator may be unwilling to invest in an expensive high-speed network because of the uncertainty of an adequate return on the investment. The initial system must be inexpensive and easy to install, which suggests that all equipment be installed either at the headend or in the home. The best approach to these problems is to change the CATV system as little as possible. This implies that the network to the home can have no packet-switching devices within the CATV system itself, for reasons of expense and maintenance, although packet switches at the branching points of the network could make a communications system work more efficiently [24].

2.4.3 CATV Topology

The tree-shaped structure of a residential CATV system constrains the network design because the headend of the system is the only point on the system that can receive from all points on the system. As a result, a broadcast network requires that stations transmit data to the headend on the upstream channel to a network repeater that retransmits the data to all stations on the downstream channel. Although data passes through the CATV headend, the headend simply repeats data - all data origination and reception takes place at the network nodes distributed across the system. As with local area networks, all stations use this single, shared broadcast channel to communicate, rather than point-to-point links between stations. A single station transmits at a time, but it broadcasts to all stations on the network, and each station must selectively receive the data addressed to it.

2.4.4 Noise

Channels in both directions are susceptible to many types of noise, such as thermal noise from amplifiers, ingress of broadcast (off-the-air) signals, intermodulation products caused by amplifier non-linearity, rectification by corroded connectors and insertion noise from equipment connected to the cable [9, 30]. CATV topology and frequency assignment, however, make the upstream channel noisier. The upstream channel is between 5 and 30 MHz and is susceptible to ingress noise from short-wave signals and impulse noise from electrical arcs. Not only are the upstream frequencies more affected by noise, but the CATV system topology drastically worsens the situation. The tree-shaped structure of the CATV system acts as a noise funnel for the upstream channels. As branches of the cable system combine, the upstream noise on the branches is summed and amplified, so that the headend receives the amplified sum of all the noise on the system.

The downstream channel, on the other hand, is designed for distributing video with good picture quality. Consequently, the downstream channel suffers few of the problems associated with the upstream channel and is thus better suited to high-speed data transmission. As mentioned earlier, the FCC requires that the carrier-to-noise ratio on the downstream channel be at least 36 dB, and most cable systems exceed this.

Because the upstream channel is so noisy and the downstream channel is so quiet, simply translating the frequency of the upstream channel to that of the downstream channel has the effect of rebroadcasting the accumulated noise. The headend instead should receive the upstream channel, demodulate the data and remodulate the data on the downstream channel. The differences in the environments of the two channels are so great as to suggest strongly that the upstream and downstream channels should use different transmission techniques. On most local area networks, all stations receive and transmit directly one to another, and therefore both the transmitter and receiver must use identical modulation techniques. On this CATV network, we may optimize the upstream and downstream transmission methods separately. In addition, network interfaces contain an upstream transmitter and a downstream receiver. Independent selection of modulation techniques in the two directions may allow a lower cost subscriber interface in exchange for a more costly headend receiver and transmitter. Because there may be thousands of subscriber interfaces for each headend, the overall savings could be substantial.

Two additional methods exist of reducing the amount of noise on the upstream channel. The first is to replace some of the upstream amplifiers with digital signal regenerators. These regenerators would receive a signal, demodulate and remodulate it rather than merely amplify it, thereby eliminating noise. Since most noise enters through the relatively low-quality drop cables, one need only replace the upstream amplifier in the trunk-bridger with a repeater to keep most of the noise out of the trunk cable. However, for reasons of cost and maintenance, repeaters should be used only if upstream noise cannot be overcome by other methods.

The second method is to take advantage of the remotely addressable switches that connect the upstream channels on the feeder cables to the upstream channel on the trunk. By feeding only a single feeder cable into the trunk at a time, the upstream noise is greatly reduced. Unfortunately, correct operation of the switches requires that the headend know which station is going to be active at any given time so that the correct switch is closed while the others remain open. Such a strategy implies a deterministic access scheme, which may not be optimum.

2.5 Problem Division

The design of a high-speed data network to the home is easier if the problem is divided into several smaller pieces. Each piece should be independent of the others so that it may be worked on separately. The network design naturally divides into three major pieces: upstream transmission, downstream transmission and access scheme (which of several stations that have to send data, goes first). These problems are addressed individually in the next three chapters, and the solutions to each are used to propose a system for high-speed data communication to the home in Chapter 6.

2.6 Summary

A residential cable system not only *distributes television signals to the home*, but also provides a path from the home to the CATV headend. This two-way channel provides the basis of a data communication network to the home. The system is broadband and divides the available spectrum into 6 MHz channels, most from headend to subscriber (downstream), but some from subscriber to headend (upstream). A tree-shaped structure of cables and amplifiers distributes signals from the headend on the downstream channel and receives signals from subscribers on the upstream channel.

An inexpensive way to provide data communications to the home is to use the existing CATV physical plant and change it as little as possible, perhaps adding a network controller at the headend and inexpensive network nodes at the subscriber end. Because the headend is the only point on the system that can receive from all other points, the minimum network consists of an upstream channel from the subscriber to the headend, a headend repeater, and a downstream channel from the headend to the subscriber. This forms the single, shared channel that is used by all stations for communications.

Unlike the downstream channel, the upstream channel has high noise because of the funneling effect of the CATV system's tree-shaped structure - the headend receives the amplified sum of all the noise in the system. The system design must take this noise into account.

The network design divides into three major pieces: upstream transmission, downstream transmission and access scheme. Each of these three may be solved individually and the solutions to each combined into the total solution.

Chapter Three

Upstream Data Transmission

A network built on a CATV system must have a repeater at the headend to which all stations transmit and from which all stations receive. Therefore, data transmission on a CATV network consists of two parts: transmission from the home to the headend (upstream) and transmission from the headend to the home (downstream). Upstream transmission is the more difficult of the two because high noise levels exist on the upstream channel. Several noise reduction techniques must be used simultaneously to combat upstream noise. This chapter discusses the upstream transmission problems on a CATV system, the expected noise on the channel and the techniques required for robust data communication.

3.1 Power Measurements on CATV Systems

Upstream transmission quality depends on the signal power and noise levels on the upstream channel. Information about power and noise levels on a CATV system is mostly available in CATV literature and CATV-specific power measurements are used. Most power measurements are made in units of dBm (decibel milliwatts), a measure of power relative to a milliwatt:

$$dBm = 10 \log \left(\frac{P}{10^{-3}} \right)$$

where P is the power being measured. Power levels on CATV systems are measured in dBmV (decibel millivolt), the amount of power necessary for a one millivolt signal to flow across a 75 ohm resistor, because cable television systems are built with 75 ohm impedance cable and impedance-matching termination. To convert from dBmV to dBm, we must calculate the power necessary to produce a 1 millivolt voltage across a 75 ohm resistor. The power across a resistor is

$$P = \frac{V^2}{R}.$$

With V = 10^{-3} volts and R = 75 ohms, a dBmV is 13.3 nanowatts. The conversion factor is

$$dBm = dBmV + 10\log\left(\frac{13.3\times10^{-9}}{10^{-3}}\right) = dBmV - 48.75.$$

Some literature refers to the S/N ratio at the detector input as the carrier-to-noise (C/N) ratio, to distinguish this S/N from the S/N of the detector output. For modulation schemes like Phase-Shift Keying (PSK) or Frequency-Shift Keying (FSK), all of the power in the carrier is used to transmit data, thus the signal-to-carrier ratio is unity, and C/N is equivalent to S/N. Modulation schemes like Amplitude Modulation (AM) and Vestigial Sideband (VSB) have a signal-to-carrier ratio of less than one because some transmitter power is used to transmit a reference carrier. In this case, the C/N ratio refers to the carrier power without modulation, and the S/N ratio is lower.

3.2 Phase Distortion

Although high noise levels are the main difficulty for upstream transmission, another factor unique to the upstream channel contributes as well. Filters associated with the upstream amplifiers induce phase distortion on the channel, which impairs data transmission.

3.2.1 The Cause of Phase Distortion

The phase delay on the channel is caused by duplexing filters, equalizers, impedance matching transformers and cable power filters associated with the system amplifiers. Figure 3-1 shows the maximum group delay error versus signal bandwidth at several frequencies [30]. The signal bandwidth scale is logarithmic, and the delay is roughly proportional to bandwidth.

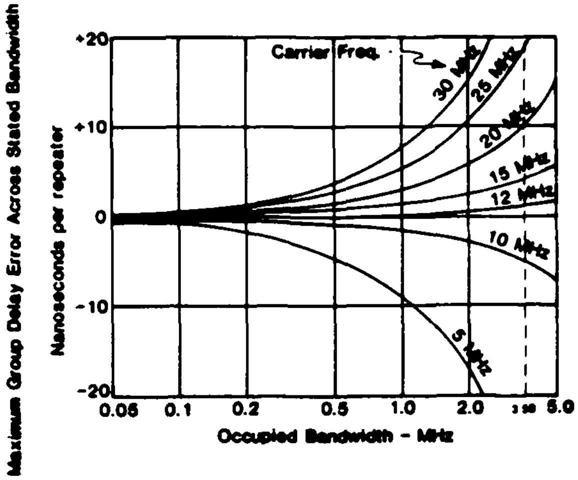


Figure 3-1:CATV Network Delay Errors as a Function of Signal Frequency and Bandwidth (from [30])

If the upstream signal frequency and bandwidth are known, then the amount of delay distortion per amplifier is shown in Figure 3-1. The total delay on this channel depends not only on the delay per amplifier, but also on the number of amplifiers in cascade. For example, assume a 1 MHz wide signal at a 5 MHz center frequency. The signal suffers a 9 nanosecond per repeater delay difference across its bandwidth. In a cascade of 25 repeaters, the total time delay across the bandwidth is 225 nanoseconds.

3.2.2 The Effects of Delay Distortion

As the delay distortion on a channel increases, transmission is impaired. The amount of impairment depends on the difference in delay between the midband frequency and the maximum frequency, f_{max}, from midband. Figure 3-2 shows how linear delay distortion is defined. The important characteristic is the ratio of delay time (d) to pulse interval (T). Here $T = 1/f_{\mathrm{max}}$, so $d/T = df_{\mathrm{max}}$.

As an example, assume that we intend to use Binary Phase-Shift Keying (BPSK) on the upstream channel of a system with 25 amplifiers. Looking at the graphs in Figure 3-3,

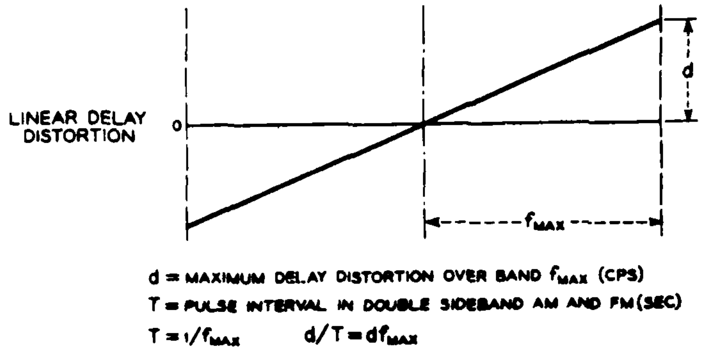


Figure 3-2: Measurement Method for Linear Delay Distortion (from [28])

we see that as long as the d/T ratio is less than 1, the signal impairment is less than 1.5 dB [28].

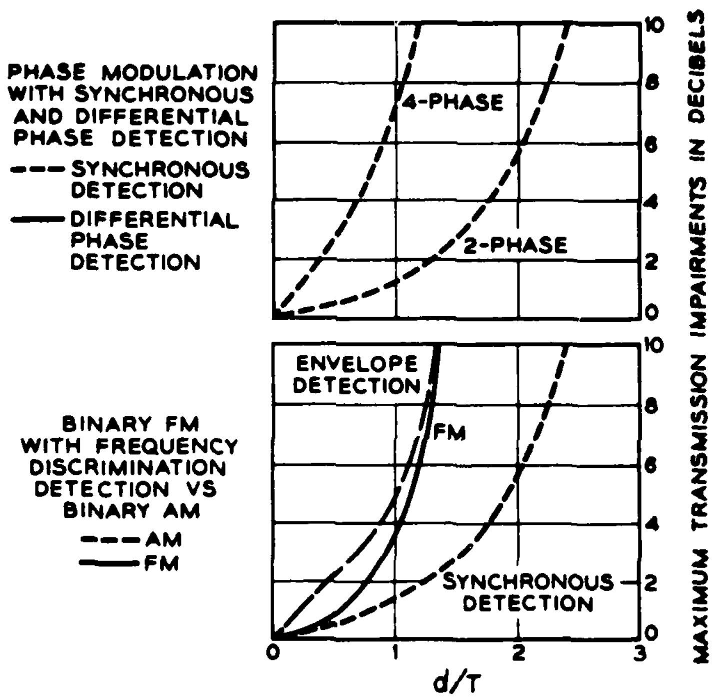


Figure 3-3: Maximum Signal Impairment as a Function of Delay for Phase and Frequency Modulation (from [28])

For the worst case delay of 225 nanoseconds, the bandwidth of the signal such that $df_{\max} = 1$ is 4.5 MHz. The bit rate that a 4.5 MHz bandwidth can support is determined by the modulation *spectral efficiency*, which is calculated in terms of the ratio bits/Hertz. Although BPSK has a theoretical spectral efficiency of 1, a rate that performs well on a real channel is 0.8 [21], which gives a bit rate of 3.5 Mbps. Even on a system with 54 amplifiers, the signal bandwidth is 2 MHz, more than enough for the 1 Mbps that we desire. If faster transmission is desired, then the delay distortion on the channel must be compensated for.

3.2.3 Overcoming Delay Distortion

There are several ways to overcome delay distortion, should it become a problem. One is to restrict transmissions to frequencies near 12 MHz, where the delay distortion is smallest. Preferably, the data communication system should be able to utilize any portion of the upstream spectrum, so that it does not conflict with other services running on the upstream channel.

Another way to reduce the effects of phase-distortion is to use a narrower transmission bandwidth. To retain a high bit rate, this implies that multiple subchannels are needed. Each channel would be received separately, so that each channel is independent of the others. Therefore, the important characteristic becomes the bandwidth of each subchannel rather than the total bandwidth of all channels. By limiting the upstream data rate to 1 Mbps, phase distortion becomes a minor consideration and we can concentrate on other characteristics of the upstream channel that make transmission difficult.

3.3 System Characteristics

Two characteristics increase the noise on the upstream channel of a CATV system: CATV topology and the lack of digital regenerators.

3.3.1 CATV Topology

One of the causes of high noise levels on a CATV system is system topology. The system is designed for the inexpensive distribution of video signals on the downstream channel, with less regard for the upstream channel characteristics. The upstream noise problem is intensified by the tree-shaped structure of the CATV system. As the upstream branches combine, the noise on the channel is summed and amplified. Therefore, the headend receives the amplified sum of all the noise over the entire system. Ingress noise and common-mode noise are particular problems because they sum coherently.

A way to reduce this effect is to divide the cable system into segments and to allow only a single segment to transmit to the headend at a time. All segments but the selected one are isolated with the remote control switches in the bridging amplifiers. By allowing a single segment to contribute noise, the noise level at the headend is reduced. The main drawback of this scheme is that it constrains the access scheme, since only those stations on the selected segment may transmit to the headend. Another problem is that some systems do not have remotely operated switches.

3.3.2 Digital Regenerators

Another reason for the high noise level is that all repeaters in the system are linear amplifiers that amplify the noise as well as the signal. Upstream noise could be reduced by replacing some of the linear amplifiers with digital signal regenerators. Each regenerator would demodulate and remodulate the signal before passing it on, thus eliminating the noise. The disadvantage of this scheme is the need to replace some upstream amplifiers, making the system more difficult and costly to install and maintain.

3.4 Upstream Noise

Four types of noise on the upstream channel are white noise, ingress noise, common-mode distortion and impulse noise [6]. Because these types of noise on the upstream channel have such different characteristics, several noise-reduction techniques must be used simultaneously to combat the noise.

3.4.1 White Noise

White noise is caused by thermal molecular noise in the cable and repeater amplifiers. The amount of white noise is proportional to both the receiver bandwidth and the system temperature. The power spectrum of white noise is kTB, where k = Boltzman's constant, T = temperature in degrees Kelvin and B = noise power bandwidth. An appropriate temperature for calculation of thermal noise is 290 Kelvin (17 Celcius, about room temperature), and the noise power bandwidth is the receiver bandwidth.

Amplifiers increase the white noise by an amount determined by the amplifier's *noise figure*. The noise figure of an amplifier is the amount of white noise at the output of an amplifier above the amount that results from amplifying the white noise at the input. The noise figures for CATV trunk amplifiers is 9 - 11 dB above kTB, and for feeder amplifiers is 10 - 14 dB above kTB [30]. Noise figures are independent of bandwidth.

To determine whether upstream transmission is possible, a S/N budget needs to be computed. An example system, taken from a paper by Citta and Mutzabaugh, has the following characteristics [6]:

1. 31,000 subscribers
2. 2312 amplifiers
3. 9 dB noise figure for all amplifiers
4. unity gain system
5. 100 kHz receiver bandwidth

The white noise at the headend receiver in this system is:

$$WN\ floor\ (dBmV) = -59 + 10\log\left(\frac{B}{4\times 10^6}\right) + NF + 10\log(N)$$

where B is the noise power bandwidth, NF is the amplifier noise figure and N is the number of amplifiers. The -59 dBmV is the white noise power for a 4 MHz bandwidth, a standard bandwidth for television receiver measurements[4].

For the parameters above, the noise floor of this system is about -32 dBmV. A typical upstream operating level for closely-spaced 100 kHz carriers might be 4 dBmV, so the S/N is then 4 - -32 = 36 dB. If we assume that the carrier power is proportional to bandwidth, the signal-to-noise ratio does not depend on system bandwidth because the bandwidth-dependent factors cancel. This is a good signal-to-noise ratio, so white noise alone is not the problem. As we shall see, other noise sources raise the noise floor considerably.

3.4.2 Narrow Band Noise

Narrow band noise is caused by common-mode distortion and the ingress of short-wave radio signals. Common-mode distortion originates in non-linear passive elements of the distribution plant. Sometimes corrosion causes a connector to form a point-contact diode, and common-mode distortion is caused by signals on the system mixing in these diodes. The difference products caused by common-mode distortion occur at 6 MHz intervals on the upstream channel because the downstream carriers are spaced at 6 MHz intervals.

Ingress noise is caused by short-wave signals that leak into the cable system through imperfect shielding. Much ingress noise in the subband is radio signals that originate in Europe and the Far East that reach the United States by reflecting off the ionosphere. Since the reflective characteristics of the ionosphere vary with the time of day and the season, ingress noise is also time-variable. Other sources of ingress include amateur radio and citizens band transmissions.

Ingress noise has a second effect aside from introducing narrow band interference; it also raises the noise floor of the system significantly above the white noise floor. For the system described above, the white noise floor is at -32 dBmV, but observed ingress noise

[4]Although a television channel bandwidth is 6 MHz wide, only 4 MHz of this bandwidth contains useful video

raises the noise floor to between -5 dBmV at 5 MHz and -12 dBmV at 30 MHz, for a 100 kHz noise power bandwidth. The carrier power for a 100 kHz channel is 4 dBmV, so that the observed S/N ratio on the channel ranges from 9 dB to 16 dB. The observed noise on the upstream channel is shown in Figure 3-4.

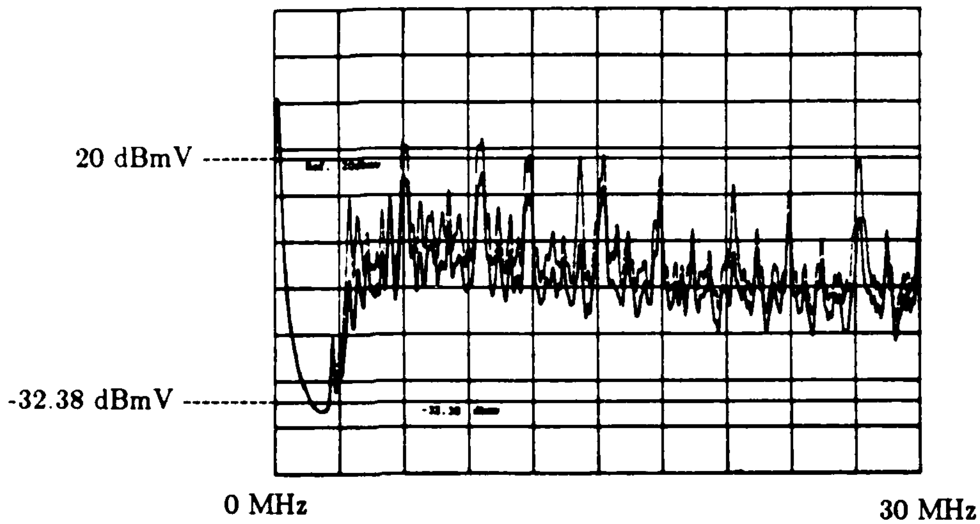


Figure 3-4: Measured Upstream Noise Spectrum from 0 to 30 MHz with a 20 dBmV Reference for an Example CATV System (from [6])

3.4.3 Impulse Noise

Impulse noise is caused by both internal and external sources. Internally, impulse noise is caused by the power on the cable for the remote amplifiers. This power is between 30 and 60 volts at 60 Hz, and it arcs across intermittent center connections in the transmission path [1]. High-voltage power lines are the primary external source of impulse noise, specifically corona and gap noise. Corona noise is caused by electrical discharge into the air, and it occurs randomly. Gap noise is caused by an electrical discharge through the discontinuity of a cracked insulator. The discharge occurs as the voltage reaches a peak, and since each cycle of a 60 Hz line has two peaks, discharges occur at 8.33 millisecond intervals. The discharge has a sharp rise and fall time with a duration in the microseconds, but neither the duration nor the interval is exact. Impulse noise enters the cable system in the same way as ingress noise, or else through insufficiently grounded cable shields [6].

Impulse noise affects the entire 5 - 30 MHz band, but it is most severe at low frequencies. This is because as frequency increases, the amplitude of the impulse harmonics decrease. If the impulse is modeled as a square pulse of a few microseconds duration, in the frequency domain its amplitude is

$$N(f) = \frac{\sin \pi f d}{\pi f d}$$

where f is the frequency of interest and d is the impulse duration. For a 1 microsecond pulse, the impulse amplitude at 29.5 MHz is down 7.3 dB from its amplitude at 5.5 MHz. Since any real discharge has a finite rise-time, its high-frequency components are even lower.

Peaks as high as 30 to 40 dBmV have been observed but peaks of 0 to 10 dBmV are typical [1, 6]. Since a data carrier might be of a 4 dBmV level, an impulse could cause bit errors on the channel. By monitoring the channel with an amplitude detector, the receiver can determine which bits have been corrupted by impulse noise.

3.5 Effective Channel Utilization

To effectively use the upstream channel, several noise reduction techniques, such as noise-resistant modulation techniques, selective spectrum utilization, and error-correction coding, must be used simultaneously.

3.5.1 Modulation Techniques

The modulation scheme should provide a small error rate at low S/N ratios, and tolerate moderate delay distortion and interfering carriers on the channel.

3.5.1.1 Noise and Signal Power Considerations

For modulation techniques that exhibit the same probability of error, the one that requires the lowest bit-energy-to-noise (E_b/N_o) ratio is preferable. On a CATV system, the transmitter power and bit duration (and thus the total bit energy) have a fixed ceiling. Therefore, a modulation scheme that requires a lower E_b/N_o ratio tolerates a higher noise level.

To compute the necessary E_b/N_o ratio for the system, the desired bit error rate must be known. Assume that the noise has a Gaussian distribution, and therefore the bit errors on the channel are independent. Also assume a binary symmetric channel, a channel on which the probability of a bit error is independent of the bit value. Further assume that 1 and 0 are equally likely to be transmitted on the channel. One of the system requirements is a 10^{-3} packet error rate on the channel; if the errors are independent and the average packet is 10^4 bits in length, this gives a 10^{-7} raw bit error rate for acceptable network performance.

Figure 3-5 shows the E_b/N_o ratio necessary for a given bit error rate for several modulation schemes. Since we are interested in the schemes that have the lowest E_b/N_o ratio, the best modulation scheme is BPSK.

Any BPSK scheme requires coherent detection at the receiver; a coherent detector compares the received signal with a phase reference to extract the data from the received signal phase. Storing a phase reference at the headend receiver for each transmitter is impractical, so somehow the transmitter and receiver must communicate the phase reference.

Usually phase reference information is transmitted along with the data on the same channel. Any system that transmits additional information over the channel such as phase information for detection, must use some of the channel power to transmit this information. To overcome upstream noise, the maximum possible power should be applied to the channel. Because the total channel power is bounded, the transmission of phase information must necessarily lower the power available for data transmission and

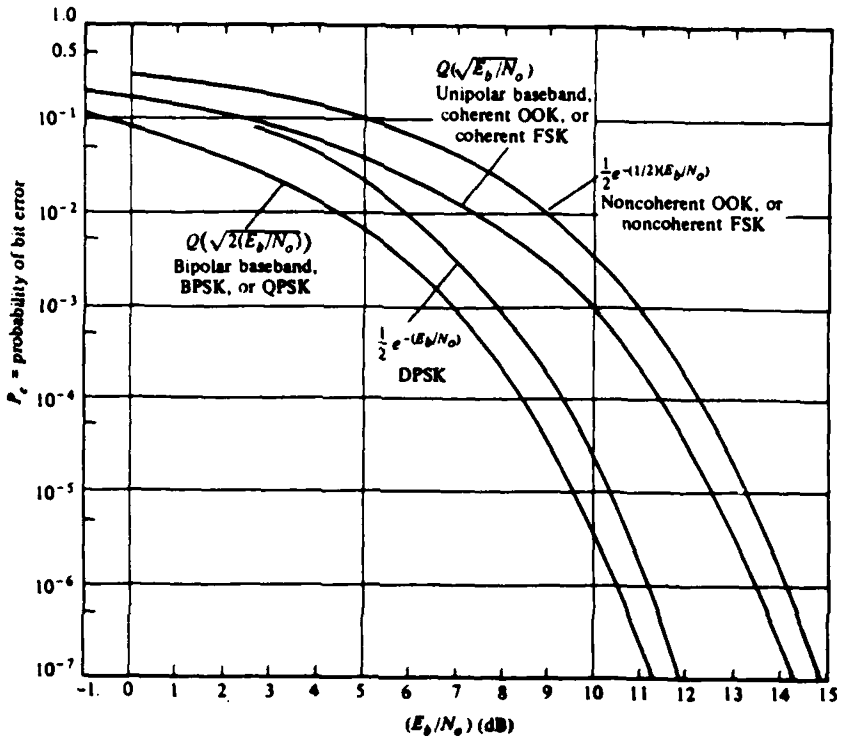


Figure 3-5:Probability of Bit Error as a Function of E_b/N_o for Several Modulation Systems (from [7])

therefore require a better S/N ratio for the same bit error rate as pure BPSK.

The probability of bit error depends not only on the E_b/N_o ratio on the channel, but also on the phase error between the transmitter and the receiver. Clock jitter between the transmitter and receiver increases the bit error rate of the channel. Since the power on the channel must be divided between the data and the phase information, increasing the amount of power used to send the phase information decreases the clock jitter but increases the susceptible of data error due to channel noise. The problem of dividing the power use on a channel optimally for the lowest error rate is known as the power allocation problem. The optimal assignment of power depends on the technique used to communicate the phase reference. Three methods of obtaining a phase reference are pilot tone systems, squaring loops, and Differential Phase-Shift Keying (DPSK).

A *pilot tone* system transmits an unmodulated carrier for a phase reference. This tone is received by a phase-locked loop (PLL) at the receiver and the output of the PLL is applied to the detector. The pilot tone frequency is near the the data signal frequency so that the channel impairments are the same to both, and usually the pilot tone is a residual carrier of the data signal. The amount of power assigned to the pilot tone must be subtracted from the data signal power and a typical system might assign 30% of the system power to the pilot tone for optimum performance.

The PLL must lock to the pilot tone before detection can occur. Ideally, the PLL should have low jitter and fast lock time, but these two desires are contradictory. If the loop bandwidth is wide, then the loop locks quickly but has a large phase jitter because more noise enters the loop. A narrow bandwidth PLL has less jitter, but the lock time increases.

A *squaring loop* system extracts the phase reference from the data signal itself. Part of the signal is processed with a frequency doubler (squarer) to produce an unmodulated carrier at twice the carrier frequency. A PLL is used to track and smooth the doubler output and the PLL output is halved and used as a phase reference. Because the PLL operates at twice the carrier frequency, there is a possibility of a 180 degree phase ambiguity in the system, so the absolute phase is determined by transmitting known reference symbols for synchronization. Although this system requires no separate signal to obtain a phase reference, the phase reference is noisier than that in a pilot tone system.

DPSK is a form of PSK that obtains its phase reference by comparing each received bit with the previously received bit. Comparison of adjacent bits must recreate the original bit stream, therefore differential coding is used. Because DPSK is a form of PSK, the error rate of DPSK is as good as PSK on a single-bit basis, but because of the differential encoding, DPSK bit error tend to occur in pairs. To bring the bit rate down to that of BPSK requires about 1 dB more power than pure BPSK.

Figure 3-6 shows the error rate for several systems that transmit a phase reference. DPSK and squaring loop system are the most efficient, followed by pilot tone systems.

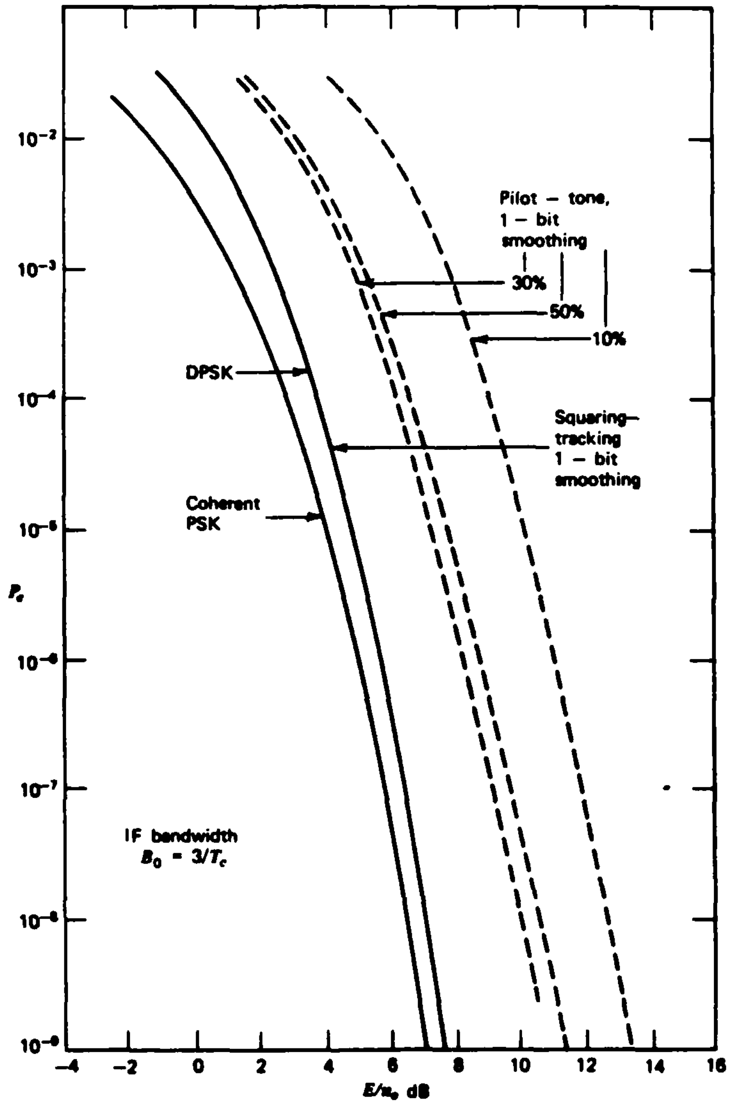


Figure 3-6: Error Rates for Several Modulation Systems (from [27])

For the optimum bit error rate, it is preferable to avoid the power allocation problem altogether, which implies that the phase information is transmitted somewhere other than the data channel. A separate channel for the phase information retains only the relative phase of the clock because of the delay distortion on the channel. A phase matching system is needed to correct for the phase distortion between the signal and the pilot tone on the channel. The receiver must match the phase exactly at the headend, but once the phase is matched, it stays matched because the relative phase is constant.

If we can obtain only a constant relative phase anyway, there is no need to transmit the phase information on the upstream channel. If the master clock is steady enough that its phase remains relatively constant over the maximum round-trip propagation delay of the network and the channel jitter is small, then the phase information can be carried on the downstream channel. An advantage is that all stations are coordinated with the clock at the headend, and this coordinated clock is important for the access scheme.

The downstream channel has a higher S/N ratio than the upstream channel, so the received clock signal will have less jitter than a similar system on the upstream channel. Another advantage is that there is a single transmitter on the downstream channel for all PLLs to lock to. Because the pilot tone does not change with each new transmitter, the PLLs can have very narrow bandwidths. Although the lock time could become long, each PLL need only lock to the pilot tone once when the station is turned on, so a long lock time is no disadvantage.

The phase reference needs low jitter to avoid raising the error rate on the channel. The error probability when taking phase jitter into account is [27]

$$P_e = \frac{1}{\pi\sigma_\phi}\int_0^\infty exp\left(-\frac{\phi^2}{2\sigma_\phi{}^2}\right)\int_{\sqrt{2E_b/N_o}\cos\phi}^\infty exp\left(-\frac{y^2}{2}\right)\,dy\,d\phi$$

where $\sigma_\phi{}^2$ is the variance of ϕ. The above equation assumes that ϕ is random and has a gaussian distribution. Since BPSK has about a 1 dB advantage over other modulation schemes at the E_b/N_o ratio considered, the tolerable amount of jitter is the amount that negates the advantage of BPSK over other modulation schemes.

The solution to the equation above does not exist in closed form, and numerical evaluation of the equation is difficult. Fortunately, an approximation exists for large E_b/N_o [27]

$$P_e \approx 1 - erf\left(\frac{\pi}{2}\frac{\sqrt{E_b/N_o}}{\sqrt{(1 + 2\sigma_\phi{}^2 E_b/N_o)}}\right)$$

The superiority of coherent BPSK over other schemes is least for high E_b/N_o, thus we shall use a bit error rate of 10^{-7} to obtain the maximum allowable phase variance. If we let E_b/N_o increase by 1 dB from the value necessary assuming no jitter, the allowable jitter is $\sigma_\phi = 0.24$ radians^2. As will be shown in Chapter 4, the jitter for the received clock can be much smaller than 0.24; small enough that the jitter will have no appreciable effect on the bit error rate of the channel. Of course, this assumes that the jitter of the master clock and the upstream and downstream channels is small. If the jitter caused by other sources is too high, then DPSK is probably the optimum modulation scheme because it performs well and can be synchronized to quickly because the transmitter and receiver still have the same clock frequency.

The signal-to-noise ratio required for a 10^{-7} bit error rate depends not only on the E_b/N_o ratio, but also on the bit rate and the bandwidth of the signal. The relationship is

$$\frac{S}{N} = \frac{E_b R}{N_o B}$$

where R is the bit rate and B is the signal bandwidth. The expression above is only approximately true because it does not take into account the amount of signal power lost as the signal is filtered to a finite bandwidth. A practical R/B ratio for BPSK is 0.8 [21] and extrapolating from a table in reference [31], the signal loss due to filtering is probably less than 2 dB.

$$\frac{S}{N}(dB) = 11.3\ dB + 10 \log 0.8 + 2\ dB,$$

giving a signal-to-noise ratio of approximately 12.3 dB for BPSK with an error rate of 10^{-7}.

3.5.1.2 Delay and Interference Considerations

The modulation must also perform well in the presence of delay distortion and carrier interference, and BPSK is one of the best performers with both [21]. For a 1 Mbps rate, the bandwidth must be $10^6/0.8$ or 1.25 MHz. The worst case delay is 9 nanoseconds per repeater, for a total delay of 225 nanoseconds on a system with 25 repeaters in cascade. From this, we compute the loss due to delay distortion.

$$d/T \;=\; df_{max} \;=\; (225\times10^{-9})\times(1.25\times10^{6}) \;=\; 0.28$$

Looking at Figure 3-3, we see that the impairment is 0.5 dB, which increases the minimum necessary S/N ratio to 12.8 dB.

The S/N ratio needed on the channel is increased if there are interfering carriers. Typical peak levels of ingress interference are observed to be from 0 to 15 dBmV, depending on the size of the system [23]. Figure 3-7 shows the S/N ratio necessary for a given BPSK error rate in the presence of interfering carriers. Carrier-to-noise ratio for BPSK is equivalent to S/N ratio as used in this thesis. Even interference 10 dB below the data carrier power increases the minimum S/N ratio to 15.5 dB, and since the carrier is only 4 dBmV, this is a -6 dBmV signal. Although BPSK is tolerant of low ingress level, some other mechanism must reduce the ingress noise on the data channel.

3.5.2 Selective Spectrum Utilization

Although BPSK can handle low-level ingress interference, high-level ingress must be eliminated in some other manner. Peak ingress noise can be 30 to 45 dBmV, far above the normal carrier power of 20 dBmV on the upstream channel. If the upstream signals are transmitted only where the channel is clear, then the overall signal-to-noise ratio for data transmission is improved. Abbondi suggests that all of the frequencies used for short-wave, citizens band and amateur radio be avoided until all other frequencies are used [1]. Citta and Mutzabaugh say that long-term reliability may be possible by using narrow (less than 100 kHz) bandwidths and squeezing signals between areas of interference [6]. MetroNet assigns 50 channels in the 16 - 34 MHz range, the width of each channel being 300 kHz. The system simply abandons those channels with too much

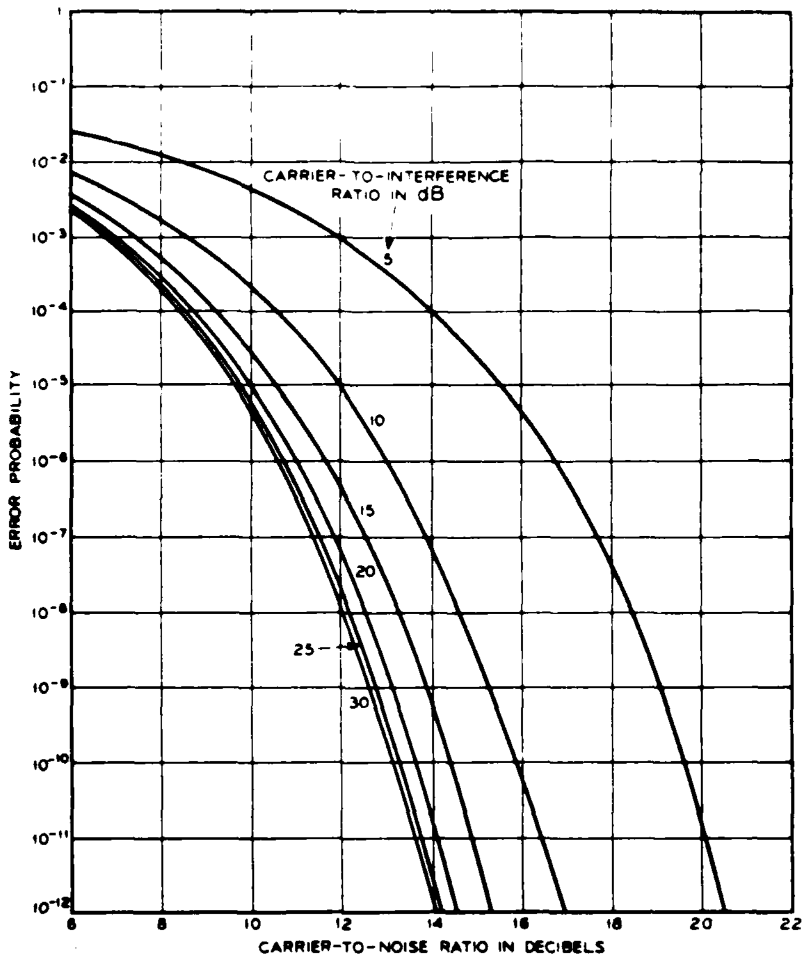


Figure 3-7: Bit Error Probability as a Function of C/N Ratio for BPSK (from [23])

noise, and Ennis estimates that a typical hub could support 38 of these channels [8].

The best approach is to use several smaller subchannels whose total data rate is as high as desired. For cost reasons, we would like to minimize the number of channels, but interference is easier to avoid with narrow channels. Suppose each station has M transmitters. By selecting the operating frequencies of the M transmitters carefully, the amount of interference from ingress noise and common mode distortion is reduced. As an additional benefit, the reduced bandwidth needed for a subchannel reduces the delay distortion problem.

Because ingress is a time-varying signal, the subchannel frequencies for the upstream modems should be assigned dynamically. In this way, the system will have the lowest error rate given the current conditions on the upstream channel. Most of these signals will vary in terms of hours or minutes rather than seconds, and the frequencies of the modems could change as often as every packet.

The data rate of the upstream channel is limited by noise, but not so with the downstream channel. Therefore, the downstream channel has the capacity not only to carry data repeated from the upstream channel but also to carry additional data from the headend to the network nodes simultaneously. A network control channel could inform the transmitters of the clearest frequencies and report the received signal strength at the headend. The upstream modems then use this information to adjust both operating frequency and transmitter output power.

3.5.3 Coding

Unlike common mode distortion and ingress noise, impulse noise is impossible to avoid. Impulses arrive about every 8.33 milliseconds, and since the network data rate is 1 Mbps, this means that a bit error may occur at least every 8333 bits or 1042 bytes. A packet longer than this is sure to be received with at least one error. On local area networks, if a packet arrives with errors, the receiver simply asks the transmitter to repeat the damaged packet. Although repeat-request error correction is relatively inefficient, the error rate of a local area network is low enough that this scheme is effective. On a CATV system, however, impulse noise could cause an error-free packet transmission to be almost impossible. In this case, a repeat-request form of error correction could not work.

A way to reduce the channel error rate is to use error-correcting codes. These codes introduce redundancy into the transmitted signal so that bit errors in the transmitted data can be detected and corrected using this redundant information. The number of errors that a code can correct is determined by the type of code and the code rate. The rate of a code is the number of data bits communicated per number of bits transmitted on the channel. For example, if 8 bits of code are needed to transmit 7 bits of data, the

the rate of the code is 7/8.

Coding can correct not only burst errors, but it also can increase the S/N ratio on the channel. If the S/N ratio of the channel is not high enough for the desired error rate, then coding allows data rate to be exchanged for error rate. Most of the noise on the upstream channel is caused by ingress noise, which varies with the time of day and the season of year and consequently, the S/N ratio of the channel changes with time. Since coding can compensate for a low S/N ratio, and the S/N ratio changes with time, then the coding rate should be able to change with time also to provide the maximum data rate with an acceptable bit error rate.

A code that has the desired properties is a Reed-Solomon (RS) code. An RS code is a special case of the BCH block codes, and it encodes k data symbols into n transmitted symbols; therefore its code rate is k/n. RS coding is well-known for burst-error correction [32], and it can correct multiple bursts per code block [22]. Although RS codes do not work well with binary alphabets, they can be easily converted to a binary representation. The code is usually built over a finite field of 2^n, but an element over a field of 2^n can be represented as n elements over a field of 2, so mapping an RS code to a binary code is trivial.

Since the upstream modem has M separate transmitters, the coding can either be for M binary channels or for a single 2^M-ary channel; the best choice is the system that minimizes the symbol error rate of the channel. For random bit errors due to white noise, the symbol error rate is identical for either case, because the probability of error on any of the M transmitters is independent of the others. This is not true for impulse noise, because an impulse will affect all frequencies simultaneously and each of the M subchannels will have simultaneous errors. If M binary channels are used, then M symbols are corrupted; but if a single 2^M-ary channel is used, then only a single symbol is affected. Therefore, a single 2^M-ary channel gives superior results.

Another consideration is the optimal block length. At 1 Mbps, burst errors occur at 8333 bit intervals and ideally the block length should be long enough to average the burst error over the longest possible period, so that the block length should be a multiple of 8333 bits long. Unfortunately, this is already far too long. As will be seen in Chapter 5, the majority of packets expected for the network are short packets of 50 bytes. This is 400 bits and if the coding rate for short packets is to be 0.8, then the maximum block length is 500 bits. If the block length is longer, then short packet transmission becomes increasingly inefficient because the ratio of useful data bits to code bits decreases.

The main reason for shortening the block length is to reduce the delay at the headend repeater. Because the entire block must be received before it can be decoded, the headend repeater delay is increased by at least n/R (the block length divided by the channel rate). For a fixed coding rate, the error rate on the channel is determined by the block length. For low noise, the block length is shortened for minimum repeater delay. For high noise, the block length is extended for maximum error correction.

3.6 System Noise Budget

The number of subchannels should be minimized to reduce the number of transmitters on each network interface, so the widest possible bandwidth should be used for each. For simplicity, all subchannels have the same bandwidth. MetroNet has been successful with 300 kHz bandwidths, so this looks like a good first choice. More noise measurements of upstream channels need to be made to determine if 300 kHz is the correct size. Another advantage of fewer subchannels is that the carrier on each subchannel may use more power. If the white noise covers all frequencies equally, then the noise increases along with the carrier power, and the S/N ratio remains the same. The higher carrier power increases the signal to interfering-carrier ratio; thus narrow-band noise causes less interference.

Since BPSK is a relatively efficient modulation scheme, a 1 Mbps data rate could be achieved with four 300 kHz wide channels. To account for coding overhead, five transmitters could be used. This allows a 1 Mbps data rate with a coding rate of 0.8. Also, Reed-Solomon coders on a chip that have symbols over a field of 2^5 should be available soon, which also points toward having five transmitters [5].

In the example above, the S/N ratio on the channel is between 9 and 16 dB. The total budget is:

1. 11.3 dB = E_b/N_o necessary for a 10^{-7} bit error rate

2. -1 dB = 10log (0.8) R/B ratio

3. 2 dB loss due to finite transmission bandwidth

4. 0 dB delay distortion

5. 0 dB carrier interference loss

The overall S/N ratio needed is 12.3 dB. The multiple subchannels not only reduce the transmission bandwidth, but also increase the bit duration. This works out to 3 nanoseconds of delay per repeater, and 75 nanoseconds for 25 repeaters. The $d/T = df_{max}$ ratio becomes $(75\text{x}10^{-9})(300\text{x}10^{3}) = 0.0225$, for a negligible delay distortion loss. Because it is difficult to tell how well selective spectrum utilization works, it is difficult to estimate the signal loss due to carrier interference. For now, assume that there is no loss so that the channel bit error rate can be estimated.

There are several ways to increase the signal-to-noise ratio on the channel. The carrier power has been calculated assuming that each subchannel has a carrier of the same power level. If, for example, we assume that half of the channels will have too much noise for data communication, then the carrier power on the remaining channels may be doubled without overloading the return amplifiers, adding 3 dB to the S/N ratio. This raises the S/N ratio to between 12 and 19 dB, almost enough for the required bit error rate in the worst case. In fact, since the system uses only five of twenty subbands on a 6 MHz channel, the carrier level can be increased up to 6 dB. Raising the carrier level increases the carrier-to-interference ratio on the channel, which also decreases the bit error rate.

We could also use coding to lower the bit error rate. If the maximum S/N ratio is 9 dB, then E_b/N_o is 8 dB. The bit error rate for an 8 dB E_b/N_o ratio is between 10^{-3} and 10^{-4}. An estimate of the block length necessary to drop the error rate below 10^{-7} can be found in a table of concatenated codes [12]

Concatenated codes are really two codes in sequence. The outer code is often a Reed-Solomon code, and its output is encoded with the inner code. The decoder reverses the procedure. The advantage of concatenated codes is that only the inner coder runs at the data transmission rate, while the outer coder runs at a lower speed.

Forney has calculated tables for many concatenated codes of a specified error rate, all of which use a Reed-Solomon code for the outer code [12]. However, his conclusion is that the only purpose of the inner code is to drop the error rate below 10^{-3} and that the RS coder produces all further improvement. Since our system already has an error rate below 10^{-3}, we can drop the inner code entirely. Therefore, we can use the reported performance for concatenated codes to estimate the performance of the RS coder for the network.

Codes of Rate 0.5 Which Achieve $Pr(\mathcal{E}) \leq 10^{-12}$ on a Binary Symmetric Channel with Crossover Probability $p = 0.01$

(N,K)	D	T_0	(n,k)	d	t_0	nN	Comment
(414,207)	51	25	—			414	S–S
(15,11)	3	1	(76,52)	25	12	1140	E–O
(31,21)	5	2	(69,51)	19	9	2139	E–O
(63,36)	11	5	(48,42)	7	3	3024	"Best" E–O
(63,39)	9	4	(52,42)	11	5	3276	E–O
(63,45)	7	3	(54,38)	17	8	3402	E–O
(127,71)	19	9	(38,34)	5	2	4826	E–O
(127,78)	15	7	(33,27)	7	3	4191	E–O
(127,85)	13	6	(32,24)	9	4	4064	E–O
(127,92)	11	5	(46,32)	15	7	5842	E–O
(127,99)	9	4	(62,40)	23	11	7874	E–O
(31,20)	6	2	(45,35)	11	5	1364	E & E
(31,21)	5	1	(77,57)	21	4	2387	E & E
(63,36)	11	4	(40,35)	6	2	2520	E & E
(63,36)	11	3	(72,63)	10	1	4536	E & E
(63,38)	10	4	(41,34)	8	3	2583	E & E
(63,38)	10	3	(47,39)	9	2	2961	E & E
(63,39)	9	3	(42,34)	9	4	2646	E & E

Table 3-1: Table of Concatenated Codes With a Bit Error Rate Below 10^{-12}(from [12])

From Table 3-1, we are interested in the outer RS codes (denoted by lower case n and k), and the EO (Error Only) codes, because we cannot use erasure coding for random bit errors as we could for impulse burst errors. Codes that have an error rate of less than

10^{-12} have a block length of about 50 bits. Although none of the codes in the table is exactly what is needed, they give an estimate of the necessary block length. The code circled in the table has a rate of just slightly over 0.8 and a block length of 52 bits for a 10^{-12} bit error rate. The error rate is lower than we need, so for an error rate of 10^{-7}, the block length can be shorter.

3.7 Summary

The upstream transmission is difficult because of the high level of upstream noise and delay distortion on the channel. The four types of upstream noise are white noise, ingress noise, common mode distortion and impulse noise. White noise and ingress noise raise the noise floor of an example system to between 9 and 16 dB. Coherent BPSK works well with the high noise floor, but other techniques are needed to overcome other types of upstream noise. Ingress noise and common mode distortion are avoided by selectively transmitting on clear parts of the channel. Error correction coding overcomes impulse noise and increases the S/N ratio of the channel.

For the example system that was studied in this chapter, it looks like we can just barely meet the network requirements because of the high noise levels on the channel. Other techniques can reduce the noise on the upstream channel. Bridger switches reduce upstream noise by selectively allowing only certain feeder cables to feed signals into the trunk, but this places constraints on the access scheme. Digital data regenerators can be used to reduce upstream noise if all else fails, but are to be avoided because of cost and maintenance problems.

The upstream channel of a CATV system is extremely noisy, but a combination of good modulation technique, careful selection of operating frequency, and coding allows the upstream channel to be used for data transmission.

Chapter Four

Downstream Data Transmission

Once data has reached the headend, it must be broadcast to all stations on the downstream channel. A CATV system is designed for video distribution, and the downstream data channel must be compatible with the CATV system, video signals and television receivers. Although a simple modulation scheme might be cheaper, vestigial sideband (VSB) data transmission is completely compatible with video transmission and reception. The technical aspects of VSB transmission are discussed, and a system designed by Sony for digital audio transmission over CATV systems is described.

4.1 CATV System Considerations

CATV systems are designed for video signal distribution with good picture quality, which implies that that the downstream noise is low enough for reliable data transmission. The wide bandwidth and high signal-to-noise ratio make compatibility with the existing cable television system the main consideration for the downstream modulation scheme. Despite the need for data communication to the home, video entertainment distribution will remain the main source of income for the cable operator. Consequently, the cable operator will be reluctant to introduce a new system onto the CATV system if it will affect revenue by interfering with video picture quality.

Although compatibility is important, another important factor is receiver cost. Because all stations require a receiver but only a single transmitter exists at the headend, economics suggest that the overall system cost may be reduced by decreasing the receiver cost in exchange for increasing the transmitter cost. For example, with a single transmitter and 1000 receivers, if the receiver cost can be reduced by a dollar each in exchange for a transmitter that costs $500 more, the total system cost is reduced.

The criteria above suggest either that an inexpensive system should be made compatible with CATV systems or that a compatible system should be made inexpensive. Frequency Shift Keying (FSK) is an inexpensive modulation technique often used for data communication, but as described below, it may not be compatible with all CATV systems. Vestigial sideband (VSB) data transmission is completely compatible because it is modulated identically to signals for which the CATV system is designed. Although VSB requires a more complex receiver than FSK, VLSI technology and mass production of television receivers could make the VSB receiver quite inexpensive.

4.2 Frequency Shift Keying

A simple, low-cost approach to downstream transmission would be use of a Frequency-Shift Keying (FSK) modulation system. MetroNet is a residential CATV-based network that uses FSK on the downstream channel because of its comparatively low cost. The FSK carrier is normally run 10 - 15 dB below the level of the video carriers in order to reduce interference with adjacent channels. Since the carrier-to-noise ratio for video carriers on the CATV system is at least 36 dB and the data carrier is at most 15 dB lower, the signal-to-noise ratio for FSK must be at least 21 dB. For reasons of cost, an inexpensive non-coherent FSK receiver should be used because it requires only a 1-dB-higher S/N ratio than does an optimal coherent receiver for the same bit error rate and the 1 dB can be spared because a non-coherent FSK signal with a 10^{-7} bit error rate requires less than a 15 dB S/N ratio.

Although FSK is simple and low-cost, the CATV distribution plant is designed to carry VSB-modulated video signals, and it may not be suited to other modulation schemes. Some hubs may use a hetrodyne signal processor for RF or IF switching or patching. A hetrodyne signal processor down-converts the video and audio signals to an intermediate frequency of 41 - 47 MHz, where the video and audio signals are amplified separately with separate AGC. The two signals are then recombined and up-converted to the appropriate frequency. The CATV hetrodyne processor requires a VSB signal for correct AGC operation; for another type of signal, the amplifier might not maintain the proper signal level [30].

The downstream modulation technique must also be compatible with television receivers. Television receivers have adjacent-channel rejection filters that notch out the video carrier on the next higher channel and the audio carrier on the next lower channel. If the FSK carrier is on a frequency other than one normally occupied by the video and audio carriers, it could pass through the adjacent channel rejection filter with enough power to interfere with the video reception for adjacent channels. If the data carriers are not phase-locked with the video carriers on the CATV system, the intermodulation distortion products of video and data carriers could lower the subjective video quality because the distortion products are not coherent with the video carriers. Also, the traps used on CATV systems are designed assuming a VSB spectral profile. It is desirable to be able to use these traps on the data channels. All of these possible incompatibilities make the use of FSK on the downstream channel undesirable.

4.3 Vestigial Sideband Modulation

Television video signals are transmitted with VSB modulation, thus a system that transmits data with VSB modulation is completely compatible with CATV systems. The disadvantage is that the receiver is more complex, but the use of mass-produced television parts keeps the receiver cost low.

VSB modulation is a special case of AM modulation. Technically, the modulation used for television transmission is VSB plus carrier, because a carrier is transmitted also to simplify demodulation. Video signals are bandlimited to 4.2 MHz before amplitude modulation. This AM signal is 8.4 MHz wide, too wide for a 6-MHz channel, but since both sidebands carry the same information, it is sufficient to transmit one of them. A single sideband receiver requires a filter with a sharp cutoff at the carrier frequency, and this filter introduces severe phase-distortion. VSB is a compromise system that requires less bandwidth than an AM signal, and yet has more reasonable filter requirements than SSB.

The bandwidth of the AM signal is reduced by passing it through a VSB filter. In an actual broadcast system, the VSB filter is at the receiver because the precise filtering necessary is done more easily at low power levels in the receiver's IF stage. The transmitter also has a filter, but its only purpose is to assure that the transmitted signal is less than 6 MHz wide.

A video signal has a large bandwidth and significant low-frequency content. Digital data modulation has more power at higher frequencies than does video because of the fast rise-time of the digital signals, so the sideband power level may be raised above that of video signals. As with video signals, the baseband digital signal must be restricted to less than 4.2 MHz of bandwidth before modulation. To determine the bit error rate of the downstream channel, the first thing we must know is the S/N ratio of the channel.

4.3.1 Transmitter Baseband Processing

The two baseband processes that occur at the transmitter are binary symbol to M-ary symbol conversion, and low-pass filtering to limit the data signal bandwidth.

4.3.1.1 Number of Signal Levels

Although the data to be transmitted is binary, the transmission system itself may use M-ary symbols. For practical reasons, the number of levels M is assumed to be 2^n, because n bits of a binary signal are easily mapped into each channel symbol. The number of levels used by the transmission system depend on the required data rate, the channel bandwidth available, and the receiver cost.

The data requirements for the downstream channel are to repeat upstream data and to carry network control information from the headend to the stations on the network. The downstream packet data channel needs the same data rate as the upstream channel, or 1 Mbps. The network control channel is unlikely to need more than 1 Mbps, so the total downstream data rate is 2 Mbps.

Television channel bandwidth restrictions limit the bandwidth of the baseband signal to below 4.2 MHz, but this is enough to transmit about 8×10^6 channel symbols per second with minimal intersymbol interference. Therefore, the desired 2 Mbps is easily obtained on such a channel with binary modulation.

Much more information could be transmitted with a multi-level system because of the high S/N ratio of the downstream channel. Binary reception is the least expensive and the channel capacity with binary modulation already exceeds the required data rate, so the extra cost associated with multi-level signalling is unjustified.

4.3.1.2 Low-pass Filtering

The baseband signal entering the low-pass filter consists of rectangular pulses of time duration T_b and infinite bandwidth. Before the signal can be modulated for transmission, its bandwidth must be limited. Although the bandwidth must be limited, intersymbol interference should be avoided. To eliminate intersymbol interference, the low-pass filter must meet the Nyquist criterion, which states that the filter cutoff frequency must be $1/2T_b$. This assures that the tails of symbols other than e symbol of interest are zero at sampling times.

The amount of signal power lost in the process is a function of the Power Spectral Density (PSD) of the baseband signal. The original input power is

$$\int_{-\infty}^{\infty} P_i(f)\, df$$

where P_i is the PSD of the input baseband signal. Assuming an ideal bandpass filter with a cutoff frequency at the Nyquist frequency, the output baseband signal power is

$$\int_{-1/2T_b}^{1/2T_b} P_i(f)\, df$$

If the modulation is binary data, and the probabilities for 1 and 0 are equal, then the PSD of the baseband signal is [7]:

$$P_i(f) = T_b\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 \qquad -\infty < f < \infty$$

The ratio of output power to input power is derived in Appendix A. No matter what the bit rate of the system, if the signal is ideally filtered to eliminate intersymbol interference, the power loss is -1.1 dB. A non-ideal filter will have slightly more power loss.

4.3.2 Amplitude Modulation

After the baseband signal is bandlimited, it must be amplitude modulated. Data is carried only in the sidebands of an AM signal, and the power in the sidebands is determined by the carrier power and the average power of the modulating signal. The power in the sidebands of an AM signal is

$$\frac{\langle m^2(t)\rangle}{1 + \langle m^2(t)\rangle}$$

Equation 4-1

where $\langle m^2(t)\rangle$ denotes the time average power of the modulating signal relative to the carrier [7]. $\langle m^2(t)\rangle$ is restricted to less than one to allow inexpensive detectors at the receiver. Although square-wave modulation produces the most signal power, even with 100% modulation the power in the AM sidebands represents only half of the total power in the signal.

4.3.3 Transmitter Bandwidth Restriction Filter

The AM signal bandwidth may be as wide as 8.4 MHz because the baseband signal is limited to 4.2 MHz, but the transmitted signal bandwidth must be reduced to 6 MHz to fit on a standard television channel. Although broadcast television uses a a VSB transmission system, the VSB filter is implemented at the receiver for economic reasons. The filter on the transmitter simply eliminates any out-of-channel frequency components of the signal. Figure 4-1 illustrates the frequency response of both the transmitter bandpass filter and the receiver VSB filter.

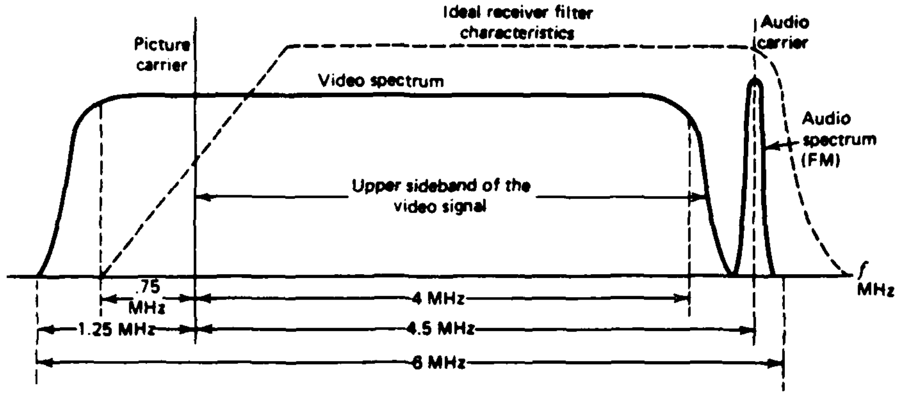


Figure 4-1: Frequency Responses for the Receiver's Vestigial Sideband Filter and the Transmitter's Bandpass Filter (from [25])

Because the baseband signal is purely real, the positive and negative frequency components of the AM signal are symmetrical about $f = 0$. Since we are interested in the ratio of output power to input power, we can work with the positive half of the spectrum only and set the carrier frequency to 0 for mathematical simplicity. The total power entering the filter is then

$$\int_{-4.2\times10^6}^{4.2\times10^6} P_i(f)\,df,$$

where $P_i(f)$ is the input PSD of the AM signal. The signal power at the filter output is

$$\int_{-1.25\times10^6}^{4.2\times10^6} P_i(f)\,df$$

which reduces the transmission bandwidth to 5.5 MHz.

The transmitter bandpass filter loss depends on the PSD of the input to the filter and for binary data, the PSD depends on the bit rate. If the modulation is binary data, and the probabilities for 1 and 0 are equal, then the PSD of the sidebands of the AM signal is:

$$P_i(f) = \frac{T_b}{4}\left(\frac{\sin\ \pi(f-f_c)T_b}{\pi(f-f_c)T_b}\right)^2 + \frac{T_b}{4}\left(\frac{\sin\ \pi(f+f_c)T_b}{\pi(f+f_c)T_b}\right)^2$$

$$f_c - \frac{1}{2T_b} < |f| < f_c + \frac{1}{2T_b}$$

with the frequency restricted by the pre-modulation low-pass filter. The AM sidebands form a Binary Phase-Shift Keying (BPSK) modulated signal, and the total AM signal is a BPSK data signal with carrier. The filter loss is

$$\frac{\int_{-1.25\times 10^6}^{1/2T_b}\left(\frac{\sin\ \pi f T_b}{\pi f T_b}\right)^2}{\int_{-1/2T_b}^{1/2T_b}\left(\frac{\sin\ \pi f T_b}{\pi f T_b}\right)^2}.$$

Appendix A contains a derivation of the filter loss calculations and a specific calculation for an example system.

The visual sideband power of a commercial broadcast is about 20 dB below that of the video carrier [14], and since the minimum C/N specified by the FCC is 36 dB, the video S/N ratio is at least 16 dB. For compatibility with television receiver components, the power in the data signal should not exceed that of a video signal, so the data S/N ratio should be the same as the video S/N ratio, or about 16 dB.

The modulation percentage for the amplitude modulator can be determined from the S/N and the C/N ratios on the channel. The S/N level is fixed at about 16 dB and the C/N level is determined by the detection system that is used and the desired bit error rate.

4.3.4 Vestigial Sideband Filtering

Now the received signal must be passed through a VSB filter to de-emphasize the lower frequencies of the transmitted signal. The amount of power lost in the in the VSB filter depends on the PSD of the received signal and the frequency response of the VSB filter. The frequency response of the VSB filter is shown in figure 4-1.

As we are interested in the ratio of output power to input power, we can work with the positive half of the spectrum only and set the carrier frequency to 0. The power of the received signal is

$$\int_{-1.25\times10^6}^{4.2\times10^6} P_i(f)\, df,$$

where P_i(f) is the input PSD, and 1.25 MHz is the cutoff of the transmitter bandpass filter. The PSD at the output of the filter is

$$\int_{-1.25\times10^6}^{4.2\times10^6} |H(f)|^2\, P_i(f)\, df$$

where |H(f)| is the magnitude of the VSB filter's frequency response. The filter response is divided into three distinct regions that can be calculated separately.

$$\int_{-4.2\times10^6}^{-0.75\times10^6} 0\, df + \int_{-0.75\times10^6}^{0.75\times10^6} \left(\frac{0.75\times10^6 + f}{1.5\times10^6}\right)^2 P_i(f)\, df + \int_{0.75\times10^6}^{4.2\times10^6} P_i(f)\, df$$

The VSB filter loss depends on the PSD of the input to the filter and for binary data, the PSD depends on the bit rate. If the modulation is binary data, and the probabilities for 1 and 0 are equal, then the PSD of the sidebands of the transmitted signal is:

$$P_i(f) = \frac{T_b}{4}\left(\frac{\sin\, \pi(f-f_c)T_b}{\pi(f-f_c)T_b}\right)^2 + \frac{T_b}{4}\left(\frac{\sin\, \pi(f+f_c)T_b}{\pi(f+f_c)T_b}\right)^2$$

$$f_c - 1.25\times10^6 < |f| < f_c + \frac{1}{2T_b}$$

with the frequency restricted by the pre-modulation low-pass filter. Appendix A contains a detailed derivation of the filter loss value and a specific calculation for an example system.

Not only is signal power reduced, but the noise power at the filter is reduced also. Assuming gaussian white noise, the amplitude of the noise is constant across the filter bandwidth. The noise power lost is the filter power out versus the filter power in.

$$\frac{\int_{-0.75\times10^6}^{0.75\times10^6}\left(\frac{0.75\times10^6+f}{1.5\times10^6}\right)^2 df + \int_{0.75\times10^6}^{4.2\times10^6} df}{\int_{-1.25\times10^6}^{4.2\times10^6} df}$$

The power loss of the noise is -1.4 dB. The S/N at the detector is

$$detector\,\frac{S}{N}(dB) \;=\; channel\,\frac{S}{N} + VSB\,filter\,signal\,loss + 1.4dB$$

The C/N ratio at the detector is computed in a similar manner. Since the VSB filter response at the carrier frequency is half of the passband response, the carrier loses 6 dB of power as it passes through the VSB filter.

$$detector\,\frac{C}{N}(dB) \;=\; channel\,\frac{C}{N} - 6dB + 1.4dB$$

4.3.5 Signal Detection

The detector normally used for VSB data transmission is the coherent detector. A coherent detector produces the lowest error rate, but requires a phase reference for correct operation.

Another type of detector is the envelope detector, which is the detector used in television receivers. The envelope detector has a higher error rate than the coherent detector because envelope detectors are susceptible to quadrature distortion and quadrature noise. An advantage of an envelope detector is its low cost. Also, a carrier must be transmitted with the signal to allow demodulation. At high S/N ratios with a large carrier component, the output of the envelope detector approaches that of a coherent detector, but at the expense of placing more carrier power on the channel.

4.3.5.1 Coherent Detection

Coherent detection is usually used for data detection because coherent detectors detect VSB signals without distortion. The disadvantage is that a coherent detector requires a phase reference. On the downstream channel, the phase reference is easily provided as a residual carrier of the data signal. The residual carrier required for a phase reference is much smaller than that needed for incoherent detection.

The S/N ratio for data on the channel is 16 dB, and the S/N of the signal at the receiver is determined by the losses induced by the VSB filter.

To compute the bit error rate, we obtain E_b/N_o from the S/N ratio:

$$\frac{E_b}{N_o}(dB) = \frac{S}{N}(dB) + 10 \log \frac{R}{B}$$

where R is the data rate and B is the receiver bandwidth. With this E_b/N_o ratio, the bit error rate is that of BPSK. Figure 4-2 plots error rate versus E_b/N_o for BPSK.

4.3.5.2 Envelope Detection

Low-cost television receiver chips could reduce the network interface cost, but television receivers use envelope detection rather than coherent detection. The error rate of a receiver with an envelope detector must be calculated to determine the practicality of using envelope detection for data reception. Unfortunately, envelope detection of a VSB signal introduces signal distortion and the exact error rate is difficult to determine.

Quadrature distortion occurs because the sidebands of the received signal are not symmetric. Any asymmetric signal can be resolved into two symmetric parts: one in phase with the carrier and one in quadrature. The original AM signal is in phase with the carrier and the asymmetry introduced by the VSB filter forms a signal in quadrature with the carrier.

Another problem is quadrature noise. The channel noise is shaped by the VSB filter at the receiver and the noise also resolves into in-phase and quadrature noise components.

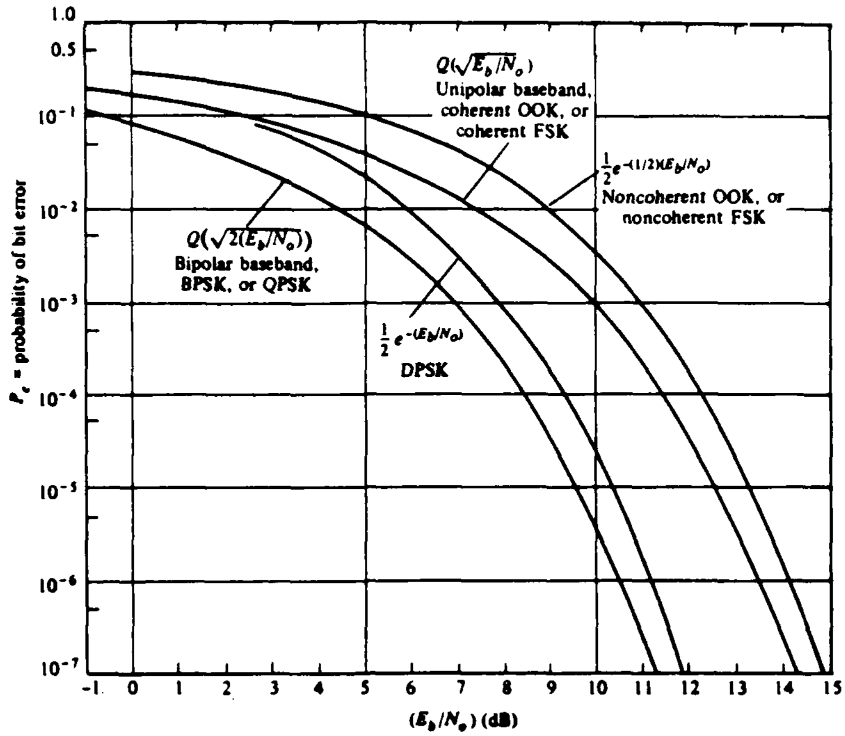


Figure 4-2: Bit Error Rate versus E_b/N_o for BPSK (from [7])

If the channel noise is gaussian white noise, then the in-phase noise is also gaussian white noise, but the quadrature noise is shaped by the VSB filter and generally is not white.

A coherent detector can receive either of the orthogonal signal components without receiving the other. An envelope detector receives the magnitude of the signal and thus receives both. Therefore quadrature distortion and quadrature noise affect only the envelope detector, so the envelope detector has a higher error rate than a coherent detector. The effect of quadrature distortion and quadrature noise is reduced by increasing the amplitude of the carrier relative to the data signal. The desired error rate for the channel is below 10^{-7} and this determines the necessary carrier power.

To calculate the error rate of a VSB signal received by an envelope detector, some simplifying assumptions are made. In general, the signal and noise on the quadrature channel are difficult to calculate with because they are shaped by the response of the VSB filter. The calculations are simplified if the VSB filter cuts off sharply at the carrier frequency, producing an SSB signal with carrier. This is the worst case because it maximize the quadrature distortion and noise, but it has the advantages of simplifying the representation of quadrature distortion and of whitening the quadrature noise.

The error rate for an ASK system with 100% modulation has been determined. The error rate for VSB data transmission is determined by modifying the results for ASK to account for quadrature distortion introduced by the VSB filter and to allow less than 100% modulation. The details of the calculation are in Appendix B.

The resulting probabilities of error for a transmitted zero or a transmitted one are

$$P_0 = Q\left(\sqrt{4(S/N) + 2(C/N) - 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

$$P_1 = 1 - Q\left(\sqrt{4(S/N) + 2(C/N) + 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

where α' is the decision threshold and the S/N and C/N ratios are those at the detector input. Q(a, b) is

$$\int_b^{\infty} t\, e^{-(t^2 + a^2)/2}\, I_o(at)\, dt \equiv Q(a, b)$$

These integrals do not have closed form solutions, but they numerically calculated and tabulated [19].

The probability of error is minimized for α' that satisfies the equation

$$\frac{p_0}{p_1} = \frac{I_o\left(\alpha' \sqrt{4(S/N) + 2(S/N) + 4\sqrt{(S/N)(C/N)}}\right)}{I_o\left(\alpha' \sqrt{4(S/N) + 2(C/N) - 4\sqrt{(S/N)(C/N)}}\right)}\, e^{-2\sqrt{(S/N)(C/N)}}$$

where p_0 and p_1 are the probability of a zero and a one, respectively.

The S/N ratio for the data signal is 16 dB on the channel and the S/N at the detector is determined by the PSD of the signal. Once the S/N ratio at the detector has been calculated, the C/N ratio at the detector is determined by the desired error rate. When the appropriate C/N ratio at the detector is known, the channel C/N ratio is easily calculated.

4.4 A Sony VSB Data Transmission System

The Sony corporation has developed and tested a digital audio and data transmission system for CATV networks [18]. The goals for their system are similar to those for the downstream channel of a CATV-based data communications network:

1. Large transmission capacity
2. Reliable transmission (low error rate)
3. System easily introduced to existing CATV networks
4. TV signals not adversely influenced by the system
5. Low cost receiver.

The Sony system transmits data using VSB modulation, and since the data modulation method is identical to that for video modulation, no change to existing CATV equipment is required to accommodate the system. The receiver uses low-cost mass-produced television parts, and the system has been tested on both North American and Japanese CATV systems. This section analyzes the Sony system in terms of the criteria above. The Sony appears to be well suited to data transmission on a CATV system.

4.4.1 Transmitter

4.4.1.1 Nyquist Filtering

The baseband bandlimiting filter in the Sonv system cuts off at 3.7 MHz and $T_b = 1/(7.4\text{x}10^6)$. Figure 4-3 shows the frequency response of the Sony pre-modulation bandpass filter. The power loss in the pre-modulation filter is about -1.1 dB because the

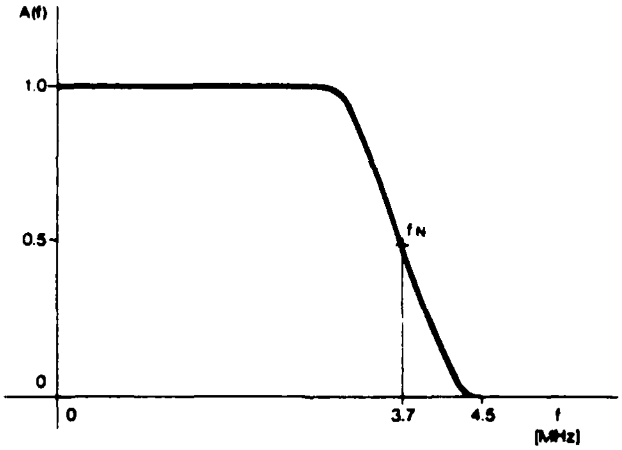


Figure 4-3:Frequency Characteristics of the Sony Pre-modulation Bandpass Filter (from [18])

filter meets the Nyquist criterion for low intersymbol interference.

4.4.1.2 Modulation

The amplitude modulator is operated at a modulation depth of 50%. The time-average power for a 50% modulated binary signal is $<m^2(t)> = p(0.5)^2 + (1\text{-}p)(-0.5)^2 = 0.25$ regardless of the probability of a 0 or 1, p and (1-p) respectively. Therefore, the signal-to-carrier ratio after AM modulation can be calculated with equation 4-1.

$$10 \log \left(\frac{0.25}{1 + 0.25} \right) = -7 \ dB$$

4.4.1.3 Transmitter Bandpass Filter

The transmitter bandpass filter causes a signal power loss of -1.5 dB. See Appendix A for detailed calculations. At this point we can calculate the S/N on the channel.

The Sony data carrier is 10 dB lower than a normal video carrier, and since the minimum C/N ratio on a CATV system is 36 dB, the C/N ratio for the data carrier is at least 26 dB. The S/C ratio is -1.1 dB + -7 dB + -1.5 dB = -9.6 dB, thus the data S/N ratio on the channel is 26 dB - 9.6 dB = 16.4 dB, and the power in the data sidebands is about the same as the power in video sidebands.

4.4.2 Receiver

The signal power loss in the VSB filter is -1.9 dB. See Appendix A for detailed calculations of the signal loss due to VSB filtering for this system.

The S/N ratio at the detector is

$$\frac{S}{N}(dB) \;=\; 16.4\,dB - 1.9\,dB + 1.4\,dB \;=\; 15.9\,db$$

The C/N ratio at the detector is

$$\frac{C}{N}(dB) \;=\; 26\,dB - 6\,dB + 1.4\,dB \;=\; 21.4\,db$$

The detector of the Sony system is an envelope detector. The probability of error for a transmitted zero or one is

$$P_0 \;=\; Q\left(\sqrt{4(S/N) + 2(C/N) - 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

$$P_1 \;=\; 1 - Q\left(\sqrt{4(S/N) + 2(C/N) + 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

where α' is the detection threshold.

To evaluate the error rate of the Sony system for various C/N ratios, the value of Q functions must be determined. A table of the Q function to six digits of precision was prepared by Marcum [19]. Although this is not precise enough to determine error rate below $1\text{x}10^{-6}$, the values in table 4-1 seem to agree with the theoretical values for the Sony system in figure 4-4. Remember that since the Sony system has a modulation level of 50%, as the C/N ratio is reduced, the S/N ratio is reduced also.

The Sony system appears to require about 3 dB more power than theoretically necessary, so the 26 dB C/N ratio on the channel is adequate to produce an error rate of below 10^{-6}.

Channel C/N Ratio	Optimum α'	P_0 Error Rate	P_1 Error Rate	Total Error Rate
20 dB	9.7	0.000094	0.000074	8.4×10^{-5}
21 dB	10.9	0.000011	0.000009	1×10^{-5}
22 dB	12.3	0.000001	0.000001	1×10^{-6}
23 dB	?	< 0.000001	< 0.000001	$< 1 \times 10^{-6}$

Table 4-1: Table of Error Rate versus Sony System Carrier Power

Although a 10^{-7} bit error rate is acceptable, error-correction coding reduces the error rate. The Sony system reduces the error rate on the downstream channel to 10^{-11} with an extended Hamming code of rate 25/32.

The receiver recovers the system clock from the received bit stream. To assure that enough bit transitions occur for clock recovery, the transmitted signal is scrambled with a pseudo-noise sequence that is known to the receiver. The clock is recovered from the scrambled data bits, and the bits are then descrambled to recover the data.

4.5 Channel Capacity

The main purpose of the downstream channel is to rebroadcast data received on the upstream channel. Regardless of the downstream modulation scheme, the downstream channel has a higher capacity than the upstream channel, thus it can not only rebroadcast upstream data, but it also can send network control signals from the network controller to all stations. This network control channel is particularly important for the upstream modems because it carries information about which frequencies on the

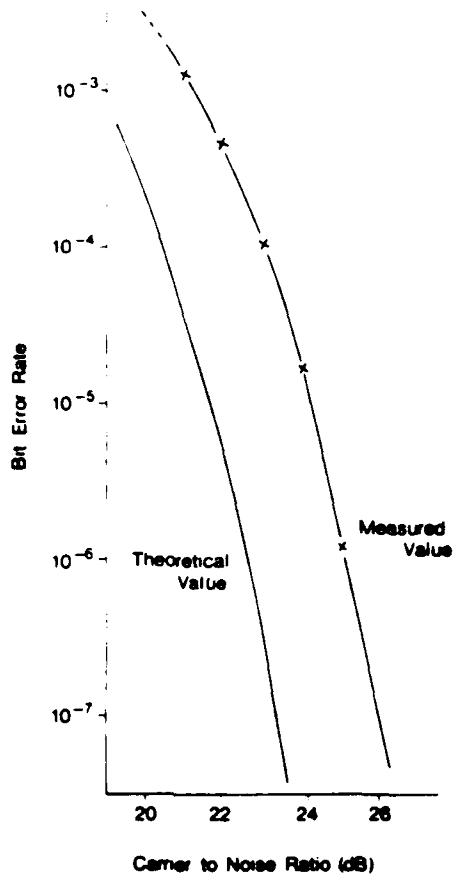


Figure 4-4: Bit Error Rate as a Function of C/N Ratio for the Sony Digital Transmission System (from [18])

upstream channel have the lowest noise. Also, since the amount of noise changes with time, the amount of coding needed to overcome noise changes with time, and the present amount of coding also can be transmitted.

Another use for the network control channel is to provide CATV system parameters to the network interfaces. All interfaces could be mass produced, but for optimum network performance, they could adjust themselves based on the parameters of the specific CATV system. These parameters include the size of the CATV system and the number of stations on the network.

4.6 Master Clock Reception

An advantage that a CATV-based communication system has over many other systems is the existence of a central clock that is used by all stations. Since the downstream channel has excess capacity, some of its bandwidth can be set aside to carry a pilot carrier that is phase-locked to the system clock. All stations receive this signal and use it as a master clock. For a VSB transmission system, the VSB carrier is ideal for use as a clock signal.

The carrier power must be high enough to reduce the clock jitter to an acceptable level. If envelope detectors are used for the receiver, the carrier power may already exceed that necessary for a low-jitter clock. The carrier frequency in the IF stage of the receiver is 45.75 MHz, and this is a low enough frequency that inexpensive PLL and logic chips can be used to implement the clock system.

The variance of PLL phase jitter for an unmodulated carrier with white noise is

$$\sigma^2 = \frac{N_o B_L}{A_o^{\ 2}}$$

where N_o is the noise power bandwidth, B_L is the PLL bandwidth, and A_o is the carrier amplitude. The carrier is modulated, however, and because the modulation is VSB, phase offset is introduced to the signal.

For binary data, the probability density function of the phase will have a one gaussian distribution centered on the phase for a 0-bit and another gaussian distribution centered on the phase of a 1-bit. Making the same assumptions as for envelope detection, we assume that the quadrature signal power is equal to the in-phase signal power. The carrier-to-noise power and phase for one and zero are

$$\frac{{A_1}^2}{N_o} = \sqrt{(C/N+S/N)^2 + (S/N)^2}$$

$$\phi_1 = Tan^{-1}\left(\frac{S/N}{C/N + S/N}\right)$$

$$\frac{{A_0}^2}{N_o} = \sqrt{(C/N-S/N)^2 + (S/N)^2}$$

$$\phi_0 = Tan^{-1}\left(\frac{-S/N}{C/N - S/N}\right)$$

The total variance for equally probable 1's and 0's is

$${\sigma_t}^2 = \frac{{\sigma_0}^2 + {\sigma_1}^2}{2} + \frac{{\phi_0}^2 + {\phi_1}^2 - 2\phi_0\phi_1}{4}$$

where

$${\sigma_1}^2 = \frac{B_L}{\sqrt{(C/N+S/N)^2 + (S/N)^2}}$$

$${\sigma_0}^2 = \frac{B_L}{\sqrt{(C/N-S/N)^2 + (S/N)^2}}$$

Recall from Chapter 3 that the phase variance must be less than 0.24 radians^2. For the S/N and C/N ratio values for the Sony system, the total jitter of the clock PLL is

$$B_L(MHz)\,(6.3\times10^{-4}) + 0.05 \; < \; 0.24$$

which indicates that the loop bandwidth B_L can be up to 300 MHz wide and still produce a low enough clock jitter. In fact, at an 80 MHz loop bandwidth, the dominant cause of phase error is signalling on the channel. Increasing the carrier amplitude decreases the effect of signalling on phase variance.

Because data on the downstream channel is synchronous with the downstream carrier, the master clock can be used for the downstream channel clock also. The receiver simply adjusts the clock phase to match the phase of the bits on the downstream channel. The alternative is to derive the downstream clock from the bit stream as on the Sony system.

4.7 Summary

The structure of a CATV system forces all data to travel through the headend, so all stations transmit to the headend on the upstream channel and a repeater at the headend repeats the data onto the downstream channel. Downstream data transmission must be inexpensive and compatible with CATV systems. FSK is a simple and inexpensive system for data transmission, but it is not completely compatible with CATV systems and television receivers.

Vestigial sideband transmission has the advantage of being totally compatible with CATV because it is identical to the modulation used for video transmission on the system. Downstream transmission with VSB modulation is practical, as demonstrated by the Sony digital audio and data transmission system. Although VSB modulation is more complex, the existence of mass-produced chips for televisions should keep the price low.

Since the downstream channel is not noise-limited, as is the upstream channel, it can carry network data from the upstream channel and other data as well. This extra capacity could provide a network control channel for the headend network controller to provide information to the network nodes. VSB transmission has the additional advantage that its carrier can be used to provide a clock signal.

Chapter Five

Network Access Schemes

5.1 Introduction

An access scheme is an algorithm for allocating network bandwidth to stations on a network. The choice of access scheme depends not only on the network propagation delay, topology and data rate, but also on the expected network traffic. Access schemes are compared with the parameters for a CATV-based network in mind. This chapter discusses the CATV system parameters and the expected data traffic, and with these assumptions, the merits of different access schemes are examined. Changes to some of the access schemes are proposed that would improve their performance on a CATV-based network.

5.2 CATV Access Schemes Considerations

5.2.1 Packet-Switching Broadcast Networks

The CATV network, like the local area network and some satellite networks, is a packet-switching broadcast network that provides datagram service. All stations on the network share a single, high-speed broadcast channel which distributes each transmitted packet to all stations, and each individual station decides whether to receive the packet. This decision is based on the destination address at the beginning of the packet.

By definition, a broadcast network has no internal switching. Switching on a CATV network would involve installing intelligent devices at network branching points, such as the location of bridger amplifiers. These switches would have to be installed and maintained at remote locations along the CATV system and they would add significantly to the initial and maintenance costs of the network. Also, some downstream frequencies must be reserved for use of the switches.

All stations on a broadcast network must receive every packet, but the CATV headend is the only point on the system that can transmit directly to all points on the network. Therefore, stations must transmit data to the headend on the upstream channel, and a repeater at the headend rebroadcasts the data to all stations on the downstream channel.

A datagram network makes its best effort to deliver packets, but does not guarantee delivery. Handling of incorrect or out-of-sequence packets is the responsibility of higher network levels.

5.2.2 Preliminary Access Scheme Evaluation

The first step in determining an access scheme is to examine those networks similar in size and topology to a CATV system. The conventional networks that are the closest in size to a CATV system are local area networks and satellite networks. A CATV network has a propagation delay of ten times that of a local area network, but only a thousandth that of a satellite network. Little work has been done on efficient access schemes for CATV networks, but perhaps access schemes from LANs and satellite networks might be suitable.

The topology of a CATV system is similar to that of a satellite network. Stations on a satellite network transmit data to the satellite, which broadcasts the data to all receiving stations. The CATV network is similar; stations transmit on the upstream channel to the headend, which broadcasts the data on the downstream channel to all receivers. This topology increases the already long packet propagation time because packets travel to their destination via the headend rather than directly.

The data rate of this CATV network is 1 Mbps, the same as that of moderate speed local area networks and the minimum data rate specified for LANs and Metropolitan Area Networks (MANs) by the IEEE 802 network standards committee. In fact, a CATV network is considered a MAN because it covers a large area.

The network connects computers in the home to computers at the office to form a distributed computing environment. Traffic measurements have been made on LANs in distributed computing environments, and the traffic on a CATV network should be similar. Computer data traffic, unlike terminal traffic, is bursty; and bursty traffic is difficult to schedule on a network with the long propagation delay of a CATV network. Bursty traffic requires a dynamic allocation of bandwidth for efficient network utilization; static allocation of time or frequency slots makes network utilization unacceptably low. Current access schemes for LANs or satellite networks do not work well on a CATV network, but perhaps a modified version of one of these access schemes will.

5.2.3 Monitoring of the Upstream Channel

For CATV network access schemes, a major problem is that stations cannot accurately monitor the upstream channel; they can infer the state of the upstream channel only by observing the downstream channel. This means that network state information could be delayed by up to twice the maximum propagation delay of the network.

There are several reasons why monitoring the upstream channel is impractical or impossible. An upstream transmission on one feeder cable cannot be received on the upstream channel of any other feeder cable because of the tree-shaped structure of the CATV system. Even if the transmitter and receiver are on one feeder cable, the directional taps that connect the drop cables to the feeder cable inhibit the detection of another station's transmission. If the transmitter is nearer the trunk cable than the receiver, the transmitter's directional tap will send most of the transmitter power toward the trunk, leaving only a small amount of residual signal traveling towards the receiver due to imperfect coupling. Likewise, if the receiver is nearer the trunk than the transmitter. the receiver's tap is designed to receive signals arriving from the trunk and to reject signals traveling toward the trunk. These taps cause signals traveling toward the trunk to be attenuated by 20 to 40 dB at the network interface.

If two stations on a feeder cable begin transmission simultaneously, even if the other station's signal could be detected, the collision cannot be detected. Collision detection is particularly difficult on a broadband system because of the *capture effect*. The capture effect occurs in FM and PM receivers if the power of one signal is roughly an order of magnitude greater than any others. Since the upstream modulation is some form of FM or PM, receiver capture can occur. Each transmitter will overpower all other signals at its receiver, so that no collision is detected.

At best, a receiver for the upstream channel can detect only a transmission in progress on the same feeder cable. It gives no information about collisions or upstream transmissions on other feeder cables. The little information gained by such a receiver does not justify its expense. For these reasons, all network state information is obtained from the downstream channel. Contentionless packet transmission requires that the state of the upstream channel be known, but the upstream state can be deduced only from information on the downstream channel. The delayed knowledge of the upstream channel state places severe limitations on the type of access scheme that a CATV network can use and limits the efficiency of contention access schemes.

5.2.4 An Example CATV System

To compare access scheme performance, parameters of the CATV network on which they will operate must be known. For evaluation of access schemes, a typical CATV network might have the following parameters:

1. A data transmission rate, R, of 1 megabit per second
2. A maximum packet length, L_{max}, of 8192 bytes
3. The maximum propagation delay from the hub to a station, P_{max}, of 100 microseconds
4. The average propagation delay from the hub to a station, P_{ave} of 70 microseconds
5. The maximum number of stations on the network, N, of 1000

The chosen data rate meets the requirements, and in addition it is the minimum bit rate recommended by the IEEE 802 committee for LANs and MANs. The packet length is chosen as a compromise between longer, more efficient packets for file transfer and the need for a short access delay for interactive applications. The propagation delay, P_{max}, is that for a system with a 25 kilometer maximum cable length to the farthest station, enough to cover a large community. The value for P_{ave} is calculated by assuming that the network covers a circle of radius r_{max} and that the stations are evenly distributed across the area covered by the network.

$$\pi {r_{max}}^2 = A, \quad \frac{A}{2} = \pi {r_{ave}}^2, \quad r_{ave} = \frac{r_{max}}{\sqrt{2}}$$

Therefore, half of the stations must be enclosed within a circle of radius r_{ave}. Since propagation delay is proportional to the cable length, the average propagation delay to a station must be $P_{max}/\sqrt{2}$, so P_{ave} is 70 microseconds. The maximum number of stations is difficult to estimate, but a high-speed data network would be used by only a small fraction of all cable subscribers. For the purposes of discussion, we assume that on a typical CATV hub, not more than 1000 subscribers would use such a network within a period of time of concern to the access system.

5.2.5 Expected Traffic

The performance of an access scheme depends on the expected data traffic for the network. The traffic for a CATV network is expected to resemble that found on LANs in distributed computing environments.

Measurements have been made on LANs in distributed computing environments at MIT and Xerox PARC. Traffic monitoring shows that packets are either short packets (a packet header plus a few bytes) from interactive sessions or acknowledgments, or long packets (a packet header and several disk blocks) from file transfers. Most packets are short, but most data is moved in the long packets [10, 26]. The CATV network access scheme ideally should provide low delay for short packets from interactive sessions and high throughput for long packets from file transfers.

Traffic measurements on a LAN at MIT suggest that packet arrival is not memoryless.

because a packet is most likely to arrive just after a previous packet arrival [10]. This is not a surprising result because most packets are from either a virtual circuit connection or a file transfer. A virtual circuit provides a perfect channel with no errors and all packets delivered in the correct order. The virtual circuit is implemented on a datagram network by having the receiver acknowledge each received packet, so that if a packet is lost, then the transmitter will retransmit it automatically. This transmit-acknowledge pattern forms a burst of packets that alternate between the source and destination. During a file transfer, packets are transmitted one after the other from a single source until all of the data have been sent. This packet-burst behavior is due to the applications being run on the network, rather than the specific software implementation.

5.2.6 Traffic Modeling

Although analytical descriptions exist for many access schemes, most analyses are based on packet arrival models that do not accurately model measured packet arrivals. A common assumption is that packet arrivals are memoryless and may be modeled by a Poisson process. Traffic measurements show that packet arrival is not memoryless because packets arrive in bursts.

This packet-arrival behavior can be modeled by packet trains [17]. A packet train is a sequence of packets between a source-destination pair on the network that occur during a burst. The inter-car interval is longer than the packet transmission time and the inter-train interval is much longer than the inter-car interval. Network measurements on a LAN at MIT found that trains exhibit a mean inter-car arrival time of 51.1 milliseconds, a mean inter-train time of 23.8 seconds, and an average number of cars per train of 17.4. A packet train model is a generalization of the usual models for packet arrivals used for analytical network analysis.

A single, fixed packet size or exponential distribution of packet sizes is another common assumption; network measurements show a bimodal packet-size distribution. The simplest model that reflects the reported network measurements has two packet sizes - one small and one large. The smallest should be roughly 50 bytes and the longest should be the most disk blocks that can fit into the maximum size packet with a packet header.

Although common modeling assumptions differ from the results of measurements, their benefit is that they greatly simplify the mathematics of delay and throughput analysis. Not enough analysis has been done with packet train modeling for a rigorous treatment of CATV-network access schemes, thus a more intuitive approach will be used with the packet train model rather than a rigorous analysis with an ill-fitting model.

5.2.7 The Comparison Metric

A comparison of systems requires a method of measurement to determine the best system. For computer networks, two important figures of merit are the network access delay and efficiency. The access delay is the amount of time between the queueing and transmission of a packet. Network efficiency is a measure of the amount of network capacity devoted to data transmission. A higher efficiency implies lower network overhead and higher average throughput. Low delay is important for interactive applications, and high throughput is important for file transfer applications. We are also interested in how well the network performs at low and high utilizations.

Two types of systems have been developed to work well in the case of delayed feedback. Satellite schemes are efficient because the system is designed in such a way that everything is planned so far ahead that the state of the upstream channel can be determined by the state of the downstream channel. Random access schemes, such as Aloha, are also promising because they totally ignore the state of the upstream channel.

Access scheme analysis focuses on the two types of access schemes: deterministic systems and contention systems. The term deterministic applies to polling and reservation systems in which all stations are queried for data in some fixed order and packet slots are assigned on the basis of these reservations. Deterministic systems guarantee a minimum network bandwidth to each station, and they also assure a minimum delay time because stations must reserve a packet slot before sending.

Deterministic access schemes can be compared with a single parameter - the walk time of the network. The walk time is the amount of time spent on network overhead per polling cycle. The cycle time of the network is the amount of time that it takes for a complete cycle of polling and associated packet transmissions to occur. The average queueing delay and the network efficiency are dependent on the cycle time of the network.

The access delay is related to the cycle time of the network in such a way that if access delay is to be minimized, then the cycle time should be minimized also. The mean cycle time, C, is

$$C = W + \sum_{k=1}^{N} T_k(C)$$

where W is the mean system walk time and $T_k(C)$ is the time that it takes to service the average traffic that queues at station k over a time period of C.

The cycle time is the sum of the network walk time and the service time of the packets that have accumulated during the last cycle. No matter how packets arrive, the number of packets that queue during any cycle time monotonically increases with the cycle time, thus the cycle time can be reduced by reducing the walk time. Therefore, on deterministic networks with equivalent numbers of stations and packet arrival, access schemes may be directly compared by comparing the walk times.

The efficiency of the network is related to the cycle time and the walk time

$$E = \frac{C-W}{C} = 1 - \frac{W}{C} = 1 - \frac{W}{W + \sum_{k=1}^{N} T_k(C)}$$

The network efficiency is just the fraction of time spent on data transmission exclusive of network overhead. Clearly, the efficiency is increased by reducing the walk time. We shall compare deterministic access schemes by comparing the walk times for various schemes, since reduced walk time results in not only lower access delay but also higher efficiency.

Contention access schemes are more difficult to analyze for a packet train arrival model. Contention systems, unlike deterministic systems, do not assure bandwidth to any station and therefore more than one station may transmit simultaneously. If this occurs, the packets are said to collide; a collision destroys all packets involved and the packets must be retransmitted at some later time. In exchange for unreliable transmission, contention systems have a lower access delay time than deterministic systems for light network traffic. Comparison among contention schemes or between contention and deterministic schemes requires a knowledge of network parameters, not only propagation delay and topology, but also the number of stations, the maximum packet size and the expected network traffic. All access scheme analysis in this thesis is done with first order statistics because this roughly determines how well an access scheme works. An analysis that includes the variance of access delay would be useful, but beyond the scope of this thesis.

The access scheme analysis requires a knowledge of the network round-trip time, which depends on not only the propagation delay of the cable, but also on the delay of the repeater at the headend (D_H). The repeater delay may be as small as a single bit time if the repeater does no data processing; if error-correction coding is used on the upstream channel, then the delay may be significantly longer because an entire block of code must be received before the data can be recovered. In chapter 3, it was suggested that a code of block length 50 might be used to overcome noise on the upstream channel. Since the entire block of code must be received before the block can be decoded, even if decoding is instantaneous the minimum headend delay is 51 bit times. For network evaluation purposes, the headend repeater delay, D_H, is assumed to be 51 microseconds. Although this is realistic for packet transmission, packet reservations contain less data and should use a code of a shorter block length.

5.3 Deterministic Access Schemes

In this section, deterministic systems are compared by examining their walk times. The three systems considered are polling, token passing, and a fixed reservations scheme.

5.3.1 Polling

Polling is a common access scheme for CATV networks, in which a central hub of the system queries each station in turn for traffic. An example of a polling system is the Warner-Amex QUBE system [15]. The hub queries a station and the station returns one of three responses. If the station has traffic, it begins packet transmission immediately; if not, a simple reply stating so is sent. If a station does not answer at all, then the hub eventually gives up and continues polling. As soon as the channel is clear of the result of the previous poll, the next polling packet is sent, thus only a single polling packet is on the network at a time. Polling is particularly inefficient on a network with a long propagation delay because the polling walk time is proportional to propagation delay, as well as the number of stations polled.

The time for a polling packet to reach the average station from the hub is P_{ave} and the response propagation delay is also P_{ave}. M_p is the polling message length in bits, M_r is the poll reply, and R is the network data rate in bits per second. The average time to poll each station is

$$P_{ave} + \frac{M_p}{R} + P_{ave} + \frac{M_r}{R}.$$

For a network with N stations, the total walk time is

$$N\left(2P_{ave} + \frac{M_p}{R} + \frac{M_r}{R}\right).$$

For the assumed system parameters, M_p is at least $\lceil \log_2 1000 \rceil$ or 12 bits since each polling message must contain the address of the station being polled. M_r contains a single bit of data, and even with coding it should not exceed three bits. The network walk time is

$$1000\left(2\times 70\,\mu s + \frac{12}{10^6} + \frac{3}{10^6}\right) = 155\,ms.$$

The network efficiency for a single station sending a maximum length packet is

$$\frac{\frac{L_{max}}{R}}{N\left(2P_{ave} + \frac{M_p}{R} + \frac{M_r}{R}\right) + \frac{L_{max}}{R}} = 0.3$$

These calculations imply a maximum efficiency under best case conditions of 30% of 1 Mbps - not enough to sustain the minimum desired throughput of 500 Kbps. Polling is simply not suited to the network scale and data rate we are trying to achieve, although it might be acceptable if the physical extent, number of stations, and data rate were all substantially smaller.

5.3.2 Token Passing

Token passing is a distributed polling system used on both ring and bus networks. On a physical ring, the token is passed from station to station around the ring. Stations on a bus network form a virtual ring in order of access. Each station on a virtual ring has a list of stations which is used to determine the token-passing sequence. A CATV network is not a physical ring, and therefore a virtual ring must be formed.

Since all signals on a CATV network must pass through the headend, token passing has no walk-time advantage over centralized polling. The average time for the token to reach the next station through the headend is $2P_{ave} + D_T$, where D_T is the delay for the headend to decode the token. If the token length is M_t, then the walk time is

$$N\left(2P_{ave} + D_T + \frac{M_t}{R}\right).$$

Just as the polling message needs to contain the address of the station being polled, the token must contain the address of the token owner, so the data in the token is 12 bits. Since the data length is short, the coded token might be 24 bits long, thus $M_t = 24$ bits.

The headend delay for receiving and decoding the token must be at least 24 μs. The total system walk time is 188 ms, slightly longer than polling.

An advantage that token passing offers over polling is that it is distributed, and thus avoids having a polling controller as a single point of failure. The disadvantage is that if the token is lost, somehow the token must be regenerated, and a new virtual ring formed. A central polling system recovers from a lost polling message more easily because it does not have to communicate the new virtual ring list on a virtual ring that is not working yet. The important point is that if polling performance is inadequate, then token passing performance is also. The only reason to use token passing is for the elimination of a single point of failure and increased reliability.

5.3.3 Fixed Reservation

A modified satellite access scheme should be considered for a CATV network. Since the CATV network is a single, shared channel, only time-division multiplexing access schemes are suitable. Fixed Reservation (FR) is a modification of the time-division satellite access scheme Fixed Priority Oriented Demand Assignment (FPODA) [4].

A satellite network topology is similar to that of a CATV network. On a satellite network, all stations transmit on the *uplink* to the satellite, which rebroadcasts the data to all stations via the *downlink*. Ground stations cannot directly monitor uplink transmissions and only receive data via the satellite on the downlink. Therefore, the state of the uplink channel can be determined only from the state of the downlink channel.

An FPODA polling cycle begins with reservation minislots that are used to reserve fixed-length packet slots for data transmission. Each station with traffic marks its reservation slot on the uplink. The satellite receives these reservations and rebroadcasts them on the downlink for all stations to examine to determine transmission order.

Because all stations know the network data rate, the length of a packet, the propagation delay to the satellite and the transmission order, the state of the uplink can be deduced from the state of the downlink. For example, if a station begins to receive a packet of known length, then the finishing time for the packet at the receiver is the current time (T_c) plus the packet transmission time (L/R). Since the propagation delay to the satellite is known, the finishing time of the packet at the satellite receiver becomes

$$T_c + \frac{L}{R} - P_s - D_S$$

where P_s is the propagation delay to the satellite and D_S is the satellite repeater delay.

To avoid a collision, the beginning of the next station's packet must arrive at the satellite after the end of the previous packet. The next station must begin its transmission at time T such that

$$T > T_c + \frac{L}{R} - 2P_s - D_S.$$

After a propagation time of P_s, the packet reaches the satellite. If $P_s + D_S > L/R$, even immediate transmission leaves a gap between the current packet and the next packet.

A station may begin transmission of its packet before it has finished receiving previous packets and yet avoid a collision. The order of transmission must be known in advance so that each station knows which packet slot to fill. This is the purpose of the reservation minislots at the beginning of a satellite polling cycle. This same scheme could be used on a CATV network for efficient access.

The above scheme requires a knowledge of data rate, packet length, propagation delay to the network hub and transmission order. For the CATV network, the data rate is 1 Mbps, and the propagation delay to the CATV network hub depends on the station location. In the current analysis, it is assumed that the propagation delay to the network hub is known. Stations can determine the delay automatically, and some practical ways of doing this are discussed later.

The packet length must also be known and for FPODA, there is a single, fixed packet length. The expected network traffic, however, would not run efficiently with a fixed packet size. The two expected packet sizes are short packets of about 50 bytes and long packets of several thousand bytes. If the fixed packet size is short, then long packets are broken into smaller pieces and a packet header attached to each piece. This is a very inefficient way to send large amounts of data because of the overhead imposed by the packet headers. If, on the other hand, the fixed packet length is long, then short packets would fill only the first 50 bytes of the packet slot and waste the remaining several thousand bytes of space. Fortunately, the propagation delay of a CATV network is short enough to permit an efficient way to send variable length packets.

A satellite network uses a fixed packet size because the long propagation delay means that many packets may be on the channel simultaneously. The next station to transmit must be able to determine the total length of all packets in order to begin transmission correctly. The shorter propagation delay of a CATV network implies that by the time a packet finishes on the upstream channel, its beginning has already been received by all stations on the downstream channel.

The next station to transmit must have available the length of the current packet in order to correctly begin its transmission. A length field preceding the packet enables the next station to determine the location of the end of the packet on the upstream channel. The position of the tail of the current packet can be calculated because both the packet length and the data rate are known. As soon as the tail of the current packet is closer to the headend than to the next station to transmit, the next packet transmission may begin. The length field is prefixed to the packet by the network hardware, and the new packet length is $L' = L + \lceil \log_2 L_{max} \rceil$, where L is the old packet length. The slight overhead of H_L ($\log_2 L_{max}/R$, about 16 μs for the parameters assumed) saves an average propagation time of $2P_{ave}$ per packet because it is not necessary for a packet to be completely received on the downstream channel before the next transmission can begin.

The minimum packet size is $L/R = 2P_{max} + D_H + H_L$, or the time for a packet from the farthest station to propagate to the headend, through the headend delay, back to the farthest station and the for the length field to be received. This calculation gives a minimum packet size of 34 bytes for a 1 Mbps network with a P_{max} of 100 μs, 51 μs headend delay and a length field of 16 bits. The imposition of a minimum packet size is not unprecedented. An Ethernet LAN, for example, requires a 60 byte minimum packet size. It is also not a burden; in current protocol families, packet headers require 20 to 40 bytes anyway.

If the propagation delay to the first station that reserved a packet slot is P_{first} and if

$$(N-F)\frac{M_r}{R} \geq 2P_{first}$$

then the station with the first transmission will not leave a gap between the end of the reservation slots and the first packet. Otherwise, the station must wait $2P_{first}$ after it makes its reservation before transmission in order to assure that it has reserved the first slot.

The walk time for the FR scheme is

$$F\left(\frac{M_r}{R}\right) + max\left(2P_{first} + D_H, (N-F)\frac{M_r}{R}\right)$$

where F is the reservation slot number of the first station with a packet. If M_r is 3 bits, the worst case walk time is 3.2 milliseconds, about 2% of the polling walk time.

There is no wasted time between the end of the last packet and the first reservation minislot of the next cycle because stations can properly time their reservation replies just as they properly time their packets. This is better than the

$$N\left(\frac{M_r}{R}\right) + 2P_{first} + D_H$$

time required if the first station had to wait until the entire reservation reply is received. Even this saves $(N - 1)(2P_{ave} + D_H)$ over polling or token passing. Of the deterministic schemes considered, so far FR has the smallest walk time and therefore the lowest access delay and the highest throughput.

5.4 Contention Access Schemes

Contention access schemes minimize the access delay under light loads at the expense of network throughput under heavy loads. Contention access schemes are not suited to CATV networks for several reasons. One is that carrier sensing is not possible, and carrier sensing is important to the optimization of several contention access schemes.

Because a CATV system is a frequency-division multiplexed system, collisions could cause frequency splatter, which would interfere with signals at other frequencies. Frequency splatter is caused by two signals being transmitted simultaneously on the same channel and mixing in some non-linear element of the cable system, such as a corroded connector, which causes difference products to be generated. Also, too many transmitters operating simultaneously could overload an upstream amplifier and cause signal distortion by driving the amplifier into non-linear operation.

To combat noise on the upstream channel, some form of frequency modulation is used for the upstream signal. All frequency modulation receivers exhibit the capture effect, where the strongest signal is received and weaker signals are rejected. The capture effect makes access unfair because a strong signal acquires the channel, even if another transmission is in progress. To avoid the capture effect, all signals arriving at the headend must be closely matched in signal level.

5.4.1 Aloha

The simplest contention access scheme is an Aloha system. Any station with traffic simply begins transmission immediately without regard for the network state. The station then listens to the downstream channel to determine whether the packet transmission was successful. Either the transmission is successful or else a collision occurs, that is two or more stations transmit packets simultaneously. If a collision occurs, all packets are destroyed and must be retransmitted.

An improved version of the Aloha system is the slotted Aloha system. On a slotted Aloha channel, all stations are constrained to transmit only at the beginning of a packet slot, rather than at any random time. This reduces the chance of collision because the contention interval is reduced. Although a slotted Aloha scheme normally assumes a fixed slot length, a length field preceding the packet allows variable length packets with no performance penalty. All stations compute the beginning of the next slot time based on the length field of the current packet.

The exception, of course, occurs during a collision in which the length field is corrupted. Assuming that there is no capture effect, then the length field is certain to be altered. If the packet length field is increased, time is wasted while the network waits until the beginning of the next slot. If the packet length field is decreased, then the next station might begin transmission before the collision has finished and become part of the collision.

To avoid drawing additional stations into a collision, the next station must refrain from transmitting until all stations involved in a collision have stopped transmitting. Stations stop transmitting as soon as they realize that a collision has occurred. This system, that might be called abortive Aloha, is the closest to carrier sensing that can be obtained on a CATV network. The worst case occurs for two stations at the maximum distance from the headend. The transmitting station begins to receive its own packet at the time $2P_{max} + D_H$ after transmission begins. The length field takes H_L to receive and if it has been altered, transmission stops immediately. An additional $2P_{max} + D_H$ seconds are required for the end of the packet to reach all stations, for a total of $4P_{max} + 2D_H + H_L$ or 518 microseconds. This also sets the minimum packet size of $L/R = 518$ microseconds or an L of 65 bytes because the next transmitter cannot be sure that a collision is over until this interval passes.

5.4.1.1 Collision Detection

For the maximum network performance, the collision interval should be minimized. To minimize the collision interval, the collision must be detected first. Collision detection on a baseband Ethernet is done by comparing the signal on the Ethernet with the data being transmitted. If the signal level does not represent the data being transmitted, then there must be more than a single transmitter on the network. This approach does not work on a CATV network because stations cannot monitor the transmission on the upstream channel. Transmitting stations on a CATV network compare the received packet on the downstream channel with the transmitted packet and if there is a difference, then a collision must have occurred. Although a bit error may be interpreted as a collision, this is advantageous because it minimizes the transmission time of a packet that has been corrupted.

Once the collision is detected, the station detecting the collision must notify all other stations. On an Ethernet network, any station that detects a collision transmits a *jamming signal* to inform all stations of the collision. The jamming signal serves two purposes. One is to guarantee that all transmitting stations detect the collision and stop transmitting. A jamming signal is necessary because different waveforms are received by different stations on the network during a collision, and there is a possibility that some stations may not recognize a collision.

Although all stations on a broadband network receive identical signals on the downstream channel, if the signals of colliding stations received at the headend are not similar in amplitude, the capture effect could occur and one signal could overpower the others. If this occurs, the transmitter that captures the headend receiver will not detect the collision. The stations that do detect the collision must somehow notify this errant station.

The other reason for the jamming signal is to inform all stations that the current packet slot is finished and to begin a new slot immediately. This assures that as little time as possible is wasted during a collision. A problem is that stations that are not transmitting do not know what the downstream packet should look like, so they cannot detect a collision in the same manner as the transmitting stations.

A jamming signal, as used on the Ethernet, is not possible because the Ethernet is baseband and the CATV network is broadband. A solution suggested by Digital Equipment Corporation for use in a broadband Ethernet sets aside a special collision enforcement channel [2]. Any station that detects a collision broadcasts a signal on the collision enforcement channel to signal all stations that a collision has occurred. The collision signal is a noise-like signal to avoid destructive cancellation at the headend when two or more stations simultaneously signal a collision. The headend receives the collision signal on the collision enforcement channel, and notifies all station of a collision on the network control channel.

5.4.1.2 P-Persistent CSMA

The most general form of this access scheme is *p-persistent* carrier-sense multiple-access. The channel is slotted, and the packet length is a multiple of the slot length. When the channel is idle, a station with traffic attempts transmission with a probability of p, and defers transmission to a future slot with a probability of (1-p). CSMA/CD is a special case of p-persistent CSMA with p set to 1.

The probability of a successful slot acquisition is the probability that exactly one station transmits in a slot. If n stations have packets queued and all have a probability of p that they will transmit in a given slot, then the probability that a given station will succeed is the probability that it transmits while others refrain or

$$p(1-p)^{n-1}$$

or the total probability that any station successfully transmits is

$$A = np(1-p)^{n-1}$$

If we maximize the probability of successful transmission with respect to p, the optimal value of p is $1/n$. Although it is unknown how many stations have packets queued, the assumption that n is known leads to an upper-bound on network performance.

The probability that the contention interval has j slots is $A(1 - A)^{j-1}$. The mean number of slots per contention is [29]

$$\sum_{j=0}^{\infty} jA(1-A)^{j-1} = \frac{1}{A}$$

Even for two stations, only half of the slots are successful acquisitions. If packets are the size of a slot, then the network efficiency is only 50%. The efficiency decreases with an increasing number of stations to *1/e* as n approaches infinity. FR requires a reservation period in addition to the packet transmission time in order to transmit two packets. If the reservation period for each station is a single bit, the efficiency of FR is

$$\frac{2(518\,\mu s)}{3000\,\mu s + 2(518\,\mu s)} = 51\%$$

The efficiency of FR increases with more packets while the utilization of p-persistent CSMA decreases, thus as expected, p-persistent CSMA does not improve network utilization over FR. The advantage of contention access schemes is lower average access delay for light traffic.

The expected contention interval is two slots. For two stations active, the expected time to transmit two packets is four slots or 2072 μs. The transmission time for two packets in FR is 2036 μs. This implies that the access delay of FR is better than that of p-persistent CSMA even if only two stations are active at any given time. Of course, coding for FR reservations would reduce the efficiency of FR, but the point is that for a network with many stations, FR will probably work better in almost all cases.

5.4.2 Contention Reservation

Contention Reservation (CR) is a modified version of FR that has a shorter average walk time. If the number of stations on the network is large, but the number of stations active simultaneously is small, then network bandwidth is being wasted by assigning each station its own reservation slot. One way to decrease the walk time is to have all stations contend for k reservation slots. Let the number of stations with traffic be s; the hub should select the number of reservation slots k that maximizes expected throughput or minimizes expected delay. Unlike p-persistent CSMA, this access scheme will be compared with FR on the basis of expected walk time. A complete analysis would also take into account the variance in access delay. If the variance of the access delay is very high, then the access scheme performance may be unacceptable.

In any given reservation slot, one of three situations are possible: no reservations, one reservation, more than one reservation. If there are no reservations in a slot, then no packet slot is assigned. If there is a single reservation, then a packet slot is assigned and packet transmission proceeds normally. The problem occurs when there are multiple reservations in a single slot.

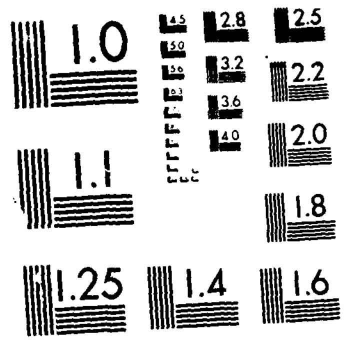
It is assumed that the network controller cannot reliably determine the number of responses in a reservation slot. Even if multiple stations reserve a packet slot, the packet slot is assigned as usual. Unfortunately, if more than one station makes a reservation in the same slot, then these stations will begin transmission simultaneously and their packets will collide. The collision increases the network overhead and increases the average walk time because the wasted time during a collision can be considered network overhead.

There are two collision costs: one is due to the network bandwidth lost to collisions and the other is due to the requeueing of packets that collide, thus increasing the number of stations trying to access the network during a future cycle. The collision time for a CATV network is $4P_{max} + 2D_H + H_L$ or 518 μs for the assumed network parameters.

Although contention for packet slots reduces the average walk time, CR still can be improved. Recall that our traffic model is trains of packets that arrive randomly over long periods, but that packets in the trains arrive over short periods of 50 milliseconds. It makes little sense for each car in a train to have to compete for network access; efficiency is improved if additional requests for packet slots are piggy-backed on previous packets. In this way, only the first car of every train competes. Once a train is started, it progresses without further collisions. In addition, fewer stations are competing for the normal reservation slots, and the number of slots and walk time are decreased. For a file transfer, once the first packet is transmitted, each additional packet slot is requested by the previous packets. A virtual circuit protocol follows a packet from A to B with an

AD-A171 666 A CATV (CABLE TELEVISION)-BASED HIGH-SPEED PACKET-SWITCHING NETWORK DESIG.. (U) MASSACHUSETTS INST OF TECH CAMBRIDGE LAB FOR COMPUTER SCIENCE.. 2/2

UNCLASSIFIED D C FELDMEIER APR 86 MIT/LCS/TR-359 F/G 17/2 NL



CROCOPY RESOLUTION TEST

acknowledgment from B to A. If station A could request a data slot for either itself or station B, virtual circuits could also run more efficiently. If the network is lightly loaded, then the requested packet slot might arrive before the next packet is ready, but if the network is lightly loaded it is easy to reserve slots normally.

Minimizing the expected walk time for any particular cycle is insufficient because it does not take the cost of requeued packets into account. To account for the number of slots, the collision cost and the cost of requeued packets, a way to determine the optimum number of reservations slots is to minimize the expected overhead time per successful reservation

$$\frac{kT_r + T_c E_c(k, s)}{E_s(k, s)}$$

where k is the number of reservation slots, T_r is the time for each reservation slot, T_c is the collision recovery time, E_c is the expected number of collisions as a function of k and s, E_s is the expected number of successful reservations as a function of k, and s is the number of stations attempting access. See Appendix C for more details of the calculation of the optimum number of slots.

Once this function has been minimized with respect to k, the average walk time for the network as a function of the the number of stations attempting access.

$$E_w(k, s) \;=\; kT_r + T_c E_c(k, s) + (s + t)2T_r$$

where s is the number of stations attempting reservations for the first packet in a train, and t is the number of packets reserved with piggy-backed reservations on the previous polling cycle. The $(s + t)$ term accounts for the overhead of piggy-backed reservations.

Assuming a collision cost of 518 microseconds and a reservation slot time of 3 μs, Table 5-1 shows the number of reservation slots needed and the estimated walk time for different values of s. The variable s, the number of stations trying to transmit the first packet of a train during a single polling cycle, is likely to be a small number. Since the expected number of packets per train is 18, the network is carrying about 90 packets per polling cycle if the network traffic is slowly changing and s is 5.

Number of Reservations s	Number of Slots k	Expected Walk Time E_w (μs)
2	24	95.8
3	37	150.4
4	51	199.9
5	66	248.9

Table 5-1: Table of k and E_w versus s

CR is the most efficient access scheme because it has the lowest expected network overhead. It is better than any of the deterministic access schemes, and the best deterministic scheme is better than p-persistent CSMA.

5.5 Propagation Delay Determination

Correct operation of these access schemes requires that the network be synchronized. All stations use a single clock that is broadcast from the headend, so all stations can time packets based on the length field. Another way that stations are synchronized is that the propagation delay to the headend is known. All stations are synchronized to headend clock and adjust their own time by the propagation delay to the headend. Although the delay to the headend could be set manually on each interface at installation time, having each interface automatically determine this delay has several advantages. All interfaces can be identical and any interface can be installed anywhere on the cable system without special setup. This prevents the network from the effects of an incorrectly set card which could cause collisions. Also, portable computers could be moved at any time and still work correctly without any attention from the user. Assuming symmetrical transmission times on the upstream and downstream channels, calculation of a station's delay from the headend is simple - send a packet, note the round-trip time, subtract the headend delay and divide by two.

Making a reservation requires that a station's delay to the headend be known in order to assure that its reservation response arrives at the correct time. Until the first packet is sent, the delay to the headend is unknown and the station is unable to respond in the correct reservation slot. Unfortunately, if no reservation can be made, no first packet can be sent. A way out of this dilemma is to allow for an alternate means of reservation for uncoordinated stations, such as a separate Aloha channel. An uncoordinated station sends its address to the headend on the Aloha channel. The headend makes note of the address and for FR, on the next polling cycle makes a reservation for that station in its reservation slot. This allows a station to transmit its first packet to determine the delay to the headend, and thereafter to make reservations as usual. For CR, it is a little more difficult. At the end of the usual poll response the headend transmits the address of the new station. Then, when all other packets are finished, the system waits for the new station to send its packet. After this packet, reservation continues as usual.

A low bitrate Aloha channel should be adequate because stations only need to use the Aloha channel once each time the propagation delay to the headend is determined, such as when an interface is turned on. The first time that a packet is sent the delay is unknown, so the station waits until the previous packet reaches the headend. This procedure guarantees that the previous packet has already passed when transmission begins. On the average this leaves a gap of P_{ave} seconds between the two packets. If a station needs to recalculate its distance from the headend only when it is turned on, this waste of P_{ave} seconds is spread throughout the day and is negligible.

5.6 Summary

Current metropolitan area networks use polling or CSMA/CD access schemes, neither of which is adequate for high-speed communication on a CATV-based network. Token passing has no efficiency advantage over polling on a network that passes all signals though a central point.

CATV networks have a propagation delay that is three orders of magnitude less than satellite networks, allowing more interaction so that modified satellite access schemes to work efficiently with the expected network traffic. The slots on satellite networks are long which leads to inefficient use of bandwidth under highly-variable loads. Fixed packet lengths are also inefficient for computer traffic because computer traffic packet lengths are bimodally distributed. A length field placed at the beginning of packets allows a CATV network to use efficiently a reservation access scheme with variable length packets.

The best access scheme is a contention reservation system in which stations contend for packet reservation slots. Since packets arrive in trains, each packet contains reservation slots so that once the first packet in a train is successfully transmitted, the following packets have contentionless reservations in the previous packet. This scheme provides a good compromise of delay and throughput for computer traffic.

Access schemes must be synchronized with the headend. The clock frequency must be synchronized, and this is done by broadcasting a clock signal from the headend. All stations must also have available the propagation delay to the headend in order to correctly time packet transmission. Although this time could be set manually at installation time, it is no more difficult to calculate it each time the network interface is turned on. A separate Aloha channel allows stations with unknown delay from the headend to communicate with the headend without interfering with packet transmissions.

Chapter Six

System Design

This chapter provides an overview of the entire network, including upstream transmission, downstream transmission, access scheme, network controller design and network interface design. Many parts of the network use complex electronics and algorithms to overcome the inherent problems of using a CATV system for a high-speed data network. Although the solutions are complex, the cost of the interface is relatively independent of the complexity of the system if the network interface is implemented as a VLSI chip. To keep cost low, the VLSI chips should be general, and parameters of the specific CATV system can be given to the chips allowing network performance is adjusted appropriately.

The first sections of this chapter select the best subsystems for the network based on the findings in Chapters 3, 4, and 5. The latter sections of the chapter discuss some aspects of the network control center and the network interface design.

6.1 Downstream Transmission

The downstream modulation technique is vestigial sideband (VSB) modulation in order to insure compatibility with the CATV system, video transmissions, and television receivers. The Sony corporation has demonstrated the feasibility of VSB downstream transmission with a high speed (7.4 Mbps), low bit error rate (10^{-7} errors/bit) system. Although the circuitry is complex, mass-produced television receiver chips should keep the cost low.

The downstream channel has greater communication capacity than the upstream channel because the downstream channel has less noise. Some of this extra bandwidth can be used for error-correction coding to assure reliable transmission. In addition to echoing data from the upstream channel, the downstream channel can be used for network control. Various systems on the network need feedback from the headend to operate properly, particularly the upstream modems.

Another use for the network control channel is to broadcast the parameters of the specific CATV system in use. These parameters include propagation delay, P_{max}, and the number of stations. When an interface is connected to the network, it can automatically optimize itself for operation on the specific CATV system. The network interfaces can be mass-produced regardless of system specifics, which lowers the interface cost by increasing the production run.

The network interface needs a clock to determine the frequency for the transmitter and the data rate. The advantage that a CATV-based communication system has over many conventional communication systems is a central clock. A shared clock for the transmitter and receiver allow coherent detection to be used on the upstream channel for higher noise immunity. A shared clock also assures that all stations operate at the same bit rate and make identical time measurements, both of which are important for the access scheme and simplicity everywhere. Such a clock can be received on the downstream channel with low phase jitter.

6.2 Upstream Transmission

Upstream transmission is complex because of the noise and interference on the upstream channel. Several noise reduction methods must be used simultaneously to assure robust transmission. The upstream transmitters use Binary Phase Shift Keying (BPSK) modulation, because coherent detection can be used and BPSK has a performance 3 dB better than other transmission systems. BPSK requires the receiver to have a phase reference for the transmitted signal. Transmission of the phase reference with the data requires a higher S/N ratio on the channel than for pure BPSK. If the transmitter and receiver share a master clock that is synchronized on the downstream channel, then the relative phase of the clock and the data signal at the receiver is constant. The receiver need only adjust its clock phase at the beginning of a signal and thereafter it will remain in phase.

To combat stationary noise sources, such as ingress noise and common mode distortion, the transmitters selectively use those parts of the upstream channel with the lowest noise. Each network interface has five transmitters, and each of these transmitters operates in a 300 kHz clear spot on the channel at a 250 Kbps rate, for a total rate of 1.25 Mbps. The five best frequencies are chosen at the headend and sent to the modems via the network control channel.

To combat impulse noise, the modems use error-correction coding to correct error bursts caused by impulses. Because the desired bit rate of the channel is 1 Mbps, the minimum acceptable coding rate is 0.8 data bits per code bit. The proposed system uses a Reed-Solomon coder that operates the five binary transmitters as a single 2^5-ary channel to combat impulse noise most effectively.

Despite the above constraints, the error rate of the channel still may be too high. The desired packet error rate of the system is one packet in a thousand. If the average packet length is 10^4, then for independent errors, the error rate of the channel must below 10^{-7}.

Since only five channels of 300 kHz each are used in a 6 MHz channel, the carrier power can be increased so that the power for a 6 MHz channel is spread across the five subchannels. Since a single return carrier is 20 dBmV, the carrier power for each subchannel can be

$$13\, dBmV = 20\, dBmV - 10 \log 5$$

A carrier power of 13 dBmV raises the S/N ratio of the example system to between 15 and 22 dB, which is more than enough for a 10^{-7} bit error rate. If a contention access scheme is used, this power level may be too high because two stations transmitting simultaneously would overload the upstream amplifiers. Since most collisions will involve only two stations, the amplifier will not overload unless three or more stations collide if the power for each subchannel is reduced to 10 dBmV.

The subchannel carrier power must be maintained at the proper level for the lowest bit error rate without overloading the upstream amplifiers. Variations in drop cable and tap attenuation mean that the transmitter output level must be adjusted for each interface. This can be done automatically if the headend measures the power of the incoming carrier and relays the received power measurement to the upstream modem via the network control channel. To reduce the chance of having the signal attenuated so much that the signal is never heard, the first transmission should use only one of the five transmitters and should boost its power by $10\log 5 = 7$ dB. The total power on the channel remains the same, but all of it is in a single carrier for ease of detection. Once the received power of this single carrier has been measured, the upstream modem should be able to set the power carrier levels so that only minor adjustments during regular packet transmission are needed.

Coding can also be used to lower the error rate on a channel. For the example CATV system in Chapter 3, the channel S/N ratio is 3 dB below that needed for the desired error rate. Therefore, coding is used to increase the effective S/N ratio. Since the amount of error correction could change with time to compensate for changing amounts of noise, the code parameters are transmitted from the headend on the network control channel.

If the S/N ratio on the upstream channel is still too low, the best alternative may be to reduce the noise rather than increase the signal power. A way to reduce the upstream noise with no modification to the CATV system is to divide the network into segments and to allow only one segment to transmit to the headend at a time. The number of segments should be as small as possible, consistent with the necessary noise reduction. If each of n segments has the same amount of noise, then the S/N ratio of the channel is increased by $10\log(n)$ dB. A disadvantage of this approach is that a more complex network controller is needed at the headend to correctly switch the bridger amplifiers at the right time. Another problem is that bridger switching could interfere with the network access scheme or other services on the upstream channel.

6.3 Proposed Access Scheme

The best access scheme for a CATV system is the Contention Reservation (CR) system. CR provides reservation slots as a function of network load, and CR has a low average walk time. Although network overhead is lowered because fewer reservation slots are needed, contention also allows collisions, which raises the network overhead. The system selects the number of reservation slots that minimizes the expected walk time per successful packet reservation.

CR is improved by allowing packets to piggy-back reservations for additional packets. If the current packet is from station A to station B, then a packet slot for the next access cycle may be reserved for either a packet from A to B or B to A.

Reservations are more efficient if done in parallel rather than serially. Instead of each reservation response using all five transmitters simultaneously, the stations should divide such that each upstream frequency is used by a fifth of the stations. Since the reservation is only a single bit, reservation is inefficient if five transmitters are used simultaneously to send a single bit. Reservation is more efficient if each station uses a single transmitter and five stations reserve simultaneously. The response may have to be more than a single bit in length because the receiver needs time to synchronize to the signal phase.

For each station, which of the five reservation channels to use must be known. For Fixed Reservation and station numbers that are evenly distributed, the first transmitter is for stations whose addresses are 1 modulo 5, the second transmitter for stations whose addresses are 2 modulo 5, and so on. For CR, assume that the number of reservation slots is a multiple of 5. Each station simply chooses one of the transmitters at random with a probability of 1/5. Once a transmitter has been selected at random, then one of n timeslots on that transmitter is chosen at random, for total of 5n slots. If the number of reservation slots is not a multiple of 5, then the probability of choosing each transmitter must be adjusted to reflect the number of reservation slots that each transmitter handles in order to retain equi-probable selection of reservation slots.

The data link layer format of a packet includes a length field at the beginning of the packet that is important for the efficient operation of the network. Stations determine the state of the network by observing the downstream channel, and must deduce the state of the upstream channel from this information. If a station knows its propagation delay from the headend, the length of the packet being received, and the data rate of the network, it is a simple matter to compute where the end of the packet is on the upstream channel. If the location of the end of a packet is known, transmission of a station's packet begins in such a way that it does not collide with the previous packet and yet leaves a small inter-packet gap. All stations must have clocks that are closely matched in frequency for this scheme to work properly, and the easiest way to assure this is to transmit a master clock signal to all stations on the downstream channel.

The access scheme described above requires that the propagation delay from the headend be known for all stations. Ideally, stations should be self-adjusting and automatically determine this delay each time they are turned on. A simple way to determine delay from the headend is to send a packet and measure its round-trip delay. Unfortunately, the only way to send a packet is to reserve a slot first, but to reserve correctly requires a knowledge of the delay from the headend. A way out of this dilemma is to provide an alternate reservation scheme for uncoordinated stations. A station of unknown delay to the headend sends its network address to the network controller via a separate, low-speed Aloha channel. The network controller appends the address of the new station to the end of the next reservation reply. After all other stations have transmitted their packets, the new station is allowed to transmit to determine its delay from the headend.

Although the headend needs an additional receiver for the Aloha channel, the network nodes can use one of the transmitters that is usually used for packet transmissions. If a station is actively participating in the network, then its delay to the headend must be known and there is no need for an Aloha channel. If an Aloha channel is needed, then it must be unable to transmit packets. Since only one of these two functions is needed at a time, the Aloha channel can be accessed with one of the five transmitters already on the card, using the Aloha channel frequency.

6.4 The Network Center

The network center is the network facility at the headend that controls the network. The most important function of the network center is the upstream-to-downstream repeater. The headend has five BPSK receivers that feed into a Reed-Solomon decoder. The data channel bits from the output of the decoder are interleaved with the bits of the network control channel, and this data is transmitted on the downstream channel with a VSB modulator. Five upstream receivers are needed for the packet channel, and the network controller needs a sixth receiver for the Aloha reservation channel.

A spectrum analyzer at the headend allows the network controller to monitor the upstream channel for ingress and common mode distortion interference. The network controller uses this information to determine the optimal frequencies for the upstream transmitters. This data is transmitted to the upstream modems via the network control stream on the downstream channel. This network control channel is also used for transmitting changes in coding rate if upstream noise is present.

A network monitor at the network center records the system and network performance for operation, maintenance and billing. The network center is the ideal place for these operations because it is centralized and all network traffic must pass through the network center.

On a public system, access control is desirable in order to deny service to a subscriber who interferes with the system in a malicious manner. A way to remove disruptive stations from the network is to isolate the upstream channel of the feeder cable of the offending stations at the bridger switch. The station can be located with a binary search and isolated to a neighborhood so few stations are removed from the network and the offender can be found.

6.5 The Network Interface

The network interface should be simple to enhance reliability, but the inherent shortcomings of a residential CATV network cause the interface to be complex. The price should be similar to that of a local area network interface or high-speed modem, or $700 in 1985 dollars and dropping. Although this is expensive by CATV standards, high-speed data communication to the home is likely to be a specialized service for those willing to purchase local area networks or high-speed modems, and thus are willing to pay a similar price for a CATV-network interface.

The network interface connects a computer to a network, but unless an interface is well designed, it becomes the bottleneck in the system. Experience with local area network interfaces has influenced the design of the interface for a CATV-based network described in this section.

The interface is full duplex. Because of the long propagation delay of a CATV network, it is possible to be transmitting a packet while receiving a packet from another station. It is undesirable to stop receiving a packet when sending and equally bad to miss an opportunity to transmit while receiving a packet.

The interface is intelligent enough to receive several commands simultaneously from the processor and execute them in the correct order. For example, if a card is instructed to clear the output packet buffer, transfer a packet from computer memory into the output buffer and to transmit a packet, the sequence of events should occur in this order or else the packet is not correctly transmitted.

To speed the transmission of a large block of data, the interface has multiple transmit buffers. The simplest way is to have two buffers: one receives data from the computer and the other transmits on the network. Then once the packet is sent, roles are reversed. A packet buffer is inaccessible until the packet that was just transmitted is received without error on the downstream channel. This assures that if a packet is damaged during transmission that it can be repeated without reloading the transmit buffer.

Sometimes packets arrive at the network interface faster than the associated computer can process them. To avoid missing these packets and forcing a retransmission, all interfaces should have multiple receive buffers, so that multiple packets are received from the network without processor intervention. Each network interface has a circular receive buffer that is written to by the interface and read from by the host processor. A single circular buffer is more efficient than multiple fixed-sized buffers because packets can be stored back-to-back rather than one per buffer, and more packets can be stored in a given amount of memory.

6.6 Summary

Downstream transmission uses vestigial sideband modulation to minimize interference with downstream video transmissions. Mass-produced television integrated circuits at the receiver keeps the receiver cost low. The downstream channel consists of two streams: a data stream and a network control stream.

The upstream channel uses five transmitters that transmit in the clearest parts of the upstream channel to avoid ingress noise. Each channel uses BPSK modulation because of its excellent performance at low S/N ratios. All the channels are used together to form a 2^5-ary channel at the output of the coder. Coding is used on the system to overcome burst errors on the channel due to impulse noise. A Reed-Solomon code of rate 0.8 is used for error correction because it is an efficient code with a short block length. Coding can also be used to increase the S/N ratio on the upstream channel if necessary.

A Contention Reservation system is used because it has the highest network efficiency. The number of reservation slots assigned varies with the network traffic to minimize the network overhead per successful packet reservation. The slotted reservation system is also modified to cause each of the five upstream transmitters to handle the reservations for a fifth of the stations.

The network controller repeats data from upstream to downstream, makes reservations, handles the Aloha reservation channel, determines the optimum frequencies for the upstream transmitters, operates bridger switches, provides access control, and monitors the network for maintenance and billing. Network control information is multiplexed with the packet data on the downstream channel.

The network interface should have a price similar to that of local area network interfaces. The interface design takes advantage of lessons learned from local area networks. The interfaces are full-duplex and can execute multiple commands from the processor in the correct order. They also have at least two transmit packet buffers for fast transmission of a train of packets and a circular receive buffer to hold packets that are received faster than they are processed.

Chapter Seven

Conclusion

A CATV-based network brings high-speed computer communications to the home. Although a CATV system can be the basis of a high-speed data communication system to the home, there are many issues yet to be resolved before such a system can be built.

7.1 Major Points

High-speed data communication can be brought to the home on a network built on a residential CATV network. Communication on a CATV system has an advantage over other communication systems because a CATV system can provide a central clock. A central clock allows transmitters and receivers to be synchronized and efficient coherent detection schemes to be used, and packet transmission can be correctly synchronized for efficient network utilization. The network design can be divided into three pieces: upstream transmission, downstream transmission, and access scheme.

The upstream channel is noisy, but the noise can be overcome through the use of several noise reduction techniques. A modulation technique with good performance in noise is needed. Since coherent detection is available, BPSK is the best choice because it outperforms other schemes by 3 dB. Although BPSK performs well in noise, it cannot overcome some of the spikes of ingress noise on the upstream channel. The best approach is to avoid ingress noise. The bandwidth of any transmitted signal should be restricted to allow the signal to slip in between noise spikes. Since this bandwidth is smaller than that required for the desired upstream data rate, multiple transmitters must be used.

Another problem is impulse noise that causes burst errors on the channel and cannot be avoided because it occurs randomly in time. The way to solve this problem is to use forward error correction. *Even though burst errors will still occur,* the data can be recovered in the decoding process. If the noise level is still too high for the desired error rate, the carrier power or the amount of coding can be increased, or the network can be segmented into smaller pieces to reduce the ingress noise received by the headend.

Because the upstream noise is so high, feedback to the modems via the downstream channel is important to their performance. The two important feedback considerations are selection of the best transmission frequencies and the received carrier power level. To assure the best error rate on the upstream channel, the upstream transmitters operate in the quietest part of the spectrum. Power measurements at the headend assure that the transmitters are operating at the maximum allowable power level.

The high S/N ratio on the downstream channel makes downstream transmission simple. Vestigial sideband modulation should be used for compatibility with existing CATV system and television receivers. The downstream channel has low error rate and greater transmission capacity than the upstream channel, and this excess capacity is used to provide feedback to the upstream modems. Another use is to carry network parameters to the stations.

Satellite access schemes can be modified for efficient performance on a CATV network. Variable length packets are essential for network performance. Trains of packets can be sent efficiently because additional packet slot reservations can be piggy-backed on the previous packets.

7.2 The CATV System as a Metropolitan Area Network

It has been argued that an appropriate design of a network can bring high-speed packet-switched communications to the home. The recommendations for upstream transmission, downstream transmission and access scheme are relatively independent.

The premise of these arguments is that complex electronics and algorithms can overcome the inherent difficulties of using the CATV system as for a high-speed data network to the home. If the network interfaces are implemented in VLSI, then the complexity has a low cost. A general VLSI chip can be constructed, with the appropriate inputs for network-specific parameters. Interfaces should receive all commands via the network control stream, thus the network controller is the only thing that needs to be manually adjusted. A generalized network controller would also decrease cost if enough CATV-based data networks are built. If such a network can be constructed, it will have to be along the lines suggested in Chapter 6. If cost is not a factor, then a better communication system undoubtedly could be designed by modifying the cable system.

CATV networks could be interconnected at community boundaries by installing a gateway at the border with cables from both systems. Alternatively, the network hubs could be linked together by fiber optic cable, microwave and satellite because some hubs already have facilities for microwave and satellite reception.

7.3 Future Research

The most important unknown factor for data communication on a CATV network is the noise on the upstream channel. More measurements need to be made on the upstream channels of CATV systems. The next step would be to build the upstream transmission system. This subsystem is the most difficult to design. The effect of impulse noise on the transmission system also needs to be measured.

The most difficult aspect of building the CATV network is the upstream transmission system. The upstream system must overcome high noise on the upstream channel, and the transmitters depend heavily on the network center for feedback about the channel noise. The upstream transmission system, including the controller at the network center, should be constructed and tested on a variety of residential systems.

The master clock system is necessary for the upstream transmitters and the proper operation of the access scheme. For good performance of the upstream transmitters, the phase jitter must be low on the upstream and downstream channels. This jitter needs to be measured to assure that a master clock broadcast on the downstream channel has little enough jitter to be a phase reference for the upstream receiver.

The access schemes in this paper are compared with first-order statistics. An in-depth analysis of the Contention Reservation system should be made before a system is built to assure that the variation in network utilization and access delay are acceptable. Because the access scheme is probabilistic, its stability must be determined for a large number of users. A computer simulation of the network with the expected traffic patterns would be valuable.

The network center was only mentioned briefly in this paper, but it is important to proper network operation and maintenance. Not only must it make network measurements for correct operation of the upstream channel, but this information needs to be multiplexed onto the downstream channel with the packet data. The systems for channel measurement and data multiplexing must be determined.

Some important points about the network interface have been raised, but to produce a sufficiently inexpensive interface for such a complex network is difficult, especially the analog systems. The network interface design should be examined in more depth.

7.4 Conclusion

A CATV-based network to the home can provide high-speed communications soon and at low cost. The system, however, is marginal for the needs of distributed computation and the need for high-speed communications will increase in the future. The CATV-based network is superior to other types of communication to the home, but soon networks specifically dedicated to data communications will be needed, especially as powerful engineering workstations drop in price and begin to appear in the home.

In ten to twenty years, the telephone company will probably find it economically attractive to replace the copper twisted-pair telephone line to each home with a fiber optic cable. Probably at about at the same time, it will become economical to replace the aging CATV system, probably with fiber optic cable to provide high-bandwidth digital television to the home, and this will provide another high-bandwidth link to the home. When this occurs, then there will be enough bandwidth into the home for high-speed communication.

Although the bandwidth into the home will be tremendously large, the type of services that such systems support is determined by the current and future applications in the home. It is important that we gain experience now with high-speed data communication to the home, so that future network design can adequately handle future data communication needs. The time to begin building CATV-based data networks to the home is now. The experience gained by providing communication to the home now could influence future network design.

References

[1] Allora-Abbondi, Georgio.
Interactive Communications Networks Design Concepts and Operational Experiences.
In *Proceedings of 14th International TV Symposium*, pages 574-598. June, 1985.

[2] Abraham, Menachem.
Proposal for an AUI-Compatible 10 MB/S CSMA/CD Broadband Medium Access Unit.
Presented to IEEE 802.6 on July 14, 1983.

[3] Beyer, William H. (editor).
CRC Standard Mathematical Tables, 25th Edition.
CRC Press, Inc., 1978.

[4] Binder, Richard.
Packet Protocols for Broadcast Satellites.
Protocols and Techniques for Data Communication Networks.
Prentice-Hall, Inc., 1981, Chapter 5.

[5] Cafarella, John.
Personal Communication.

[6] Citta, Richard and Mutzabaugh, Dennis.
Two-way Cable Plant Characteristics.
In *Technical Papers, 1984 NCTA Convention*, pages 270-277. NCTA, June, 1985.

[7] Couch, Leon W. II.
Digital and Analog Communication Systems.
Macmillan Publishing Company, 1983.

[8] Ennis, Gregory.
Design Considerations for Broadband Metropolitan Networks.
General Instruments working paper.

[9] Estrin, D. L.
Data Communications Via Cable Television Networks: Technical and Policy Considerations.
Technical Report TR-273, MIT Laboratory for Computer Science, May, 1982.

[10] Feldmeier, D. C.
Empirical Analysis of a Token Ring Network.
Technical Memorandum TM-254, MIT Laboratory for Computer Science, January, 1984.

[11] Feller, William.
An Introduction to Probability Theory and Its Applications. vol. 1.
John Wiley & Sons, Inc., 1957.

[12] Forney, G. David Jr.
Concatenated Codes.
The M.I.T. Press, 1966.
Research Monograph No. 37.

[13] Gifford, David K.; Lucassen, John M.; and Berlin, Stephen T.
The Application of Digital Broadcast Communication to Large Scale Information Systems.
IEEE Journal on Selected Areas in Communications SAC-3(3):457-467, (May, 1985).

[14] Gregg, W. David.
Analog and Digital Communication.
John Wiley & Sons, Inc., 1977.

[15] Gur, Mete M.
An Analysis of the Networking Requirements for a Residential Two-way CATV System: MetroNet.
General Instruments working paper.

[16] Hatamian, M. and Bowen, E.G.
Homenet: A Broadband Voice/Data/Video Network on CATV Systems.
The AT&T Technical Journal 64(2):347-67, (February, 1985).

[17] Jain, Raj and Routhier, Shawn.
Packet Trains: Measurements and a New Model for Computer Network Traffic.
Technical Memorandum TM-292, MIT Laboratory for Computer Science, November, 1985.

[18] Kojima, Yuichi; Fujita, Etsumi; Toyoshima, Masakatsu; and Hideshima, Yasuhiro.
A New Digital Audio and Data Transmission System Using the CATV Network.
IEEE Transactions on Consumer Electronics CE-30(3):462-7, August, 1984.

[19] Marcum, J. I.
Table of Q Functions.
Research Memorandum RM-339, The RAND Corporation, January, 1950.

[20] McGarty, Terrence P. and Clancy, Gerald J. Jr.
Cable-Based Metro Area Networks.
IEEE Journal on Selected Areas in Communications SAC-1(5):816-31, (November, 1983).

[21] Oetting, John D.
A Comparison of Modulation Techniques for Digital Radio.
IEEE Transactions on Communications COM-27(12):1752-7, (December, 1979).

[22] Peterson, W. Wesley.
Error-Correcting Codes.
The M.I.T. Press, 1961.

[23] Rosenbaum, Arnold S.
PSK Error Performance with Gaussian Noise and Interference.
The Bell System Technical Journal 48(2):413-442, (February, 1969).

[24] Saadawi, Tarek N. and Schwartz, Mischa.
Distributed Switching for Data Transmission over Two-Way CATV.
IEEE Journal on Selected Areas in Communications SAC-3(2):323-9, (March, 1985).

[25] Shanmugam, K. Sam.
Digital and Analog Communication Systems.
John Wiley & Sons, Inc., 1979.

[26] Shoch, John F. and Hupp, Jon A.
Measured Performance of an Ethernet Local Network.
Communications of the ACM 23(12):711-721, (December, 1980).

[27] Stiffler, J. J.
Theory of Synchronous Communications.
Prentice-Hall, Inc., 1971.

[28] Sunde, E. D.
Pulse Transmission by AM, FM, and PM in the Presence of Phase Distortion.
The Bell System Technical Journal 40(3):353-422, (March, 1961).

[29] Tanenbaum, Andrew S.
Computer Networks.
Prentice-Hall, Inc., 1981.

[30] Taylor, Archer S.
Characterization of Cable TV Networks as the Transmission Media for Data.
IEEE Journal on Selected Areas in Communications SAC-3(2):255-65, (March, 1985).

[31] Tjhung, T. T. and Wittke, P. H.
Carrier Transmission of Binary Data in Restricted Band.
IEEE Transactions on Communications COM-18(4):295-304, (August, 1970).

[32] Wiggert, Djimitri.
Error-Control Coding and Applications.
Artech House, Inc., 1981.

Appendix A

Filter Power Loss

This appendix contains the derivation of the signal-power loss for the pre-modulation filter, the transmitter bandpass filter and the receiver VSB filter; examples of the transmitter bandpass filter loss and the receiver VSB filter loss for the Sony data transmission system are calculated. This section is included for those who may wish to calculate the S/N ratio of the downstream channel for a VSB transmission system other than Sony's, without rederiving the mathematics.

7.1 Pre-Modulation Filter Loss

The power loss in the pre-modulation baseband filter is the ratio of the output power to the input power. The input power is

$$\int_{-\infty}^{\infty} P_i(f)\, df$$

where $P_i(f)$ is the Power Spectral Density (PSD) of the baseband modulating signal. Assuming an ideal bandpass filter with a cutoff frequency of f_N, the output power is

$$\int_{-f_N}^{f_N} P_i(f)\, df$$

The signal-power loss depends on the (PSD) of the baseband modulating signal. For a random binary signal with equally probable 1s and 0s, the PSD is [7]

$$P(f) = T_b\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2$$

The input power is

$$\int_{-\infty}^{\infty} T_b \left(\frac{\sin \pi f T_b}{\pi f T_b} \right)^2 df$$

Although this cannot be integrated directly, this definite integral can be found in mathematical reference books [3].

$$\int_{-\infty}^{\infty} \frac{\sin^2(px)}{x^2} dx = \frac{\pi p}{2}$$

Therefore, the input power is simply T_b.

For the output power, we must integrate

$$\int_{-f_N}^{f_N} T_b \left(\frac{\sin \pi f T_b}{\pi f T_b} \right)^2 df$$

The more general integral

$$\int_{x}^{y} T_b \left(\frac{\sin \pi f T_b}{\pi f T_b} \right)^2 df$$

cannot be computed directly, but it can be separated into simpler integrals by using the identity

$$\sin^2 x = \frac{(1 - \cos 2x)}{2}$$

The equation becomes (leaving the limits off for simplicity)

$$\frac{1}{2} \int \frac{(1 - \cos 2\pi f T_b)}{(\pi f T_b)^2} df = \frac{1}{2(\pi T_b)^2} \int \frac{df}{f^2} - \frac{1}{2(\pi T_b)^2} \int \frac{\cos 2\pi f T_b}{f^2} df$$

The second part of this equation can be expanded by using the identity [3]

$$\int \frac{\cos ax}{x^m}\,dx = -\frac{\cos ax}{(m-1)x^{m-1}} - \frac{a}{m-1}\int \frac{\sin ax}{x^{m-1}}\,dx$$

or in this case

$$-\frac{1}{2(\pi T_b)^2}\int \frac{\cos 2\pi f T_b}{f^2}\,df = \frac{1}{2(\pi T_b)^2}\left(\frac{\cos 2\pi f T_b}{f} + 2\pi T_b \int \frac{\sin 2\pi T_b f}{f}\,df\right)$$

Although the $\frac{\sin x}{x}$ term cannot be integrated, tables of values can be found in mathematical reference books. [3]

The total solution is

$$\int_x^y \left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 df =$$

$$\frac{1}{\pi T_b}\int_x^y \frac{\sin 2\pi T_b f}{f}\,df - \frac{1}{2(\pi T_b)^2}\left(\frac{1-\cos 2\pi T_b y}{y} - \frac{1-\cos 2\pi T_b x}{x}\right)$$

Equation A-1

Since the input power is T_b, the ratio of output power to input power is

$$\int_{-f_N}^{f_N} \left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 df =$$

$$\frac{2}{\pi}\int_0^{2\pi T_b f_N} \frac{\sin 2\pi T_b f}{f}\,df - \frac{1}{\pi^2 T_b}\left(\frac{1-\cos 2\pi T_b f_N}{f_N}\right)$$

If the pre-modulation lowpass filter meets the Nyquist criterion for elimination of intersymbol interference, then the cutoff frequency f_N is $1/2T_b$ and the filter loss is

$$\frac{2}{\pi}\int_0^{\pi} \frac{\sin f}{f}\,df - \frac{4}{\pi^2} = -1.1\ dB.$$

2 Transmitter Bandpass Filter Loss

The filter loss is the ratio of signal power at the filter output to signal power at the filter input. Because only the ratio of output-to-input power is of interest, the mathematics of the problem can be simplified. Since the transmitted signal is real, the positive and negative frequency components of the AM modulated signal are symmetrical about the zero-frequency axis. To compute the power ratio, it is sufficient to compute the ratio for positive frequencies only. Another simplification is to set the positive carrier frequency to zero, simplifying that some of the mathematics by symmetry arguments.

The transmitter bandpass filter loss depends on the PSD of the input to the filter and for binary data, the PSD depends on the bit rate. If the modulation is binary data, and the probabilities for 1 and 0 are equal, then the PSD of the sidebands of the AM signal is:

$$P_i(f) = \frac{T_b}{4}\left(\frac{\sin \pi(f-f_c)T_b}{\pi(f-f_c)T_b}\right)^2 + \frac{T_b}{4}\left(\frac{\sin \pi(f+f_c)T_b}{\pi(f+f_c)T_b}\right)^2$$

$$f_c - \frac{1}{2T_b} < |f| < f_c + \frac{1}{2T_b}$$

The frequency limits on the PSD assume that the pre-modulation lowpass filter meets the Nyquist criterion for eliminating intersymbol interference.

The ratio of output power to input power is

$$\frac{\int_{-1.25\times 10^6}^{1/2T_b}\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2}{\int_{-1/2T_b}^{1/2T_b}\left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2}$$

The input power can be calculated using equation A-1

$$\int_{-1/2T_b}^{1/2T_b} \left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 df =$$

$$\frac{2}{\pi}\int_0^{\pi} \frac{\sin 2\pi T_b f}{f} df - \frac{1}{\pi^2 T_b}\left(\frac{1-\cos \pi}{1/2T_b}\right)$$

The output power can be evaluated also with equation A-1

$$\int_{2\pi T_b(-1.25\times 10^6)}^{\pi} \left(\frac{\sin \pi f T_b}{\pi f T_b}\right)^2 df =$$

$$\frac{1}{\pi T_b}\int_{2\pi T_b(-1.25\times 10^6)}^{\pi} \frac{\sin 2\pi T_b f}{f} df$$

$$- \frac{1}{2(\pi T_b)^2}\left(\frac{1-\cos \pi}{1/2T_b} - \frac{1-\cos 2\pi T_b(-1.25\times 10^6)}{-1.25\times 10^6}\right)$$

3 Determination of VSB Filter Loss

The power loss in a VSB filter is determined by the bandpass characteristics of the VSB filter. The filter removes most of the lower sideband and part of the upper sideband of the AM signal at the filter input. The filter loss is the ratio of signal power at the filter output to signal power at the filter input.

The signal power at the filter input is

$$\int_{-1.25\times 10^6}^{4.2\times 10^6} P_i(f)\, df$$

where P_i(f) is the PSD of the positive frequencies of the AM input signal.

$$\int_{-1.25\times 10^6}^{4.2\times 10^6} P_o(f)\, df$$

where $P_o(f)$ is the PSD at the filter output and $P_o(f) = P_i(f)|H(f)|^2$, where $|H(f)|$ is the magnitude of the frequency response of the VSB filter shown in figure 1.

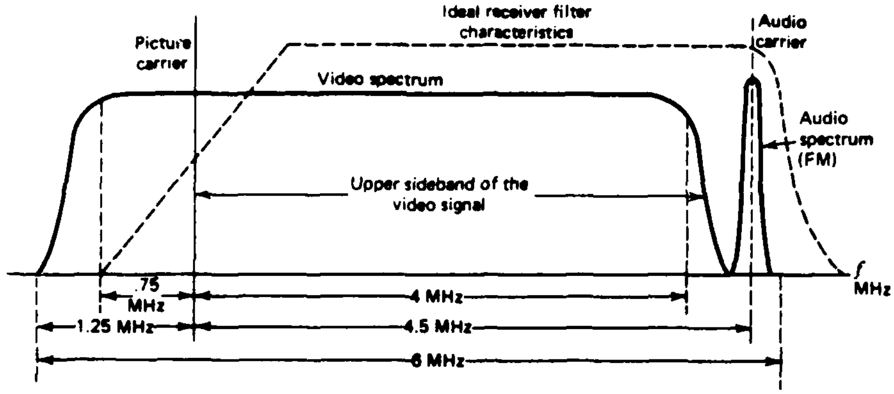


Figure 1: Vestigial Sideband Filter Frequency Response (from [25])

We represent the filter response $|H(f)|$ piecewise

$$|H(f)| = 0 \qquad -1.25\times10^6 < f < -0.75\times10^6$$

$$|H(f)| = \frac{0.75\times10^6 + f}{1.5\times10^6} \qquad -0.75\times10^6 < f < 0.75\times10^6$$

$$|H(f)| = 1 \qquad 0.75\times10^6 < f < 4.2\times10^6$$

The total output power can be calculated piecewise

$$\int_{-1.25\times10^6}^{4.2\times10^6} P_o(f)\,df =$$

$$\int_{-0.75\times10^6}^{0.75\times10^6} \left(\frac{0.75\times10^6 + f}{1.5\times10^6}\right)^2 P_i(f)\,df + \int_{0.75\times10^6}^{4.2\times10^6} P_i(f)\,df$$

The PSD of the sidebands of an AM signal for a random binary modulating signal with 1 and 0 equally probable is

$$P(f) = \frac{T_b}{4}\left(\frac{\sin\ \pi(f-f_c)T_b}{\pi(f-f_c)T_b}\right)^2 + \frac{T_b}{4}\left(\frac{\sin\ \pi(f+f_c)T_b}{\pi(f+f_c)T_b}\right)^2$$

$$f_c - 1.25\times10^6 < |f| < f_c + \frac{1}{2T_b}$$

The lower frequency limit on the PSD is determined by the transmitter bandpass filter and the upper frequency limit assumes that the pre-modulation lowpass filter meets the Nyquist criterion for eliminating intersymbol interference.

If this expression is simplified by removing the negative frequency components, setting the carrier frequency to zero and dropping the constants of proportionality

$$P(f) = \left(\frac{\sin\ \pi f T_b}{\pi f T_b}\right)^2$$

$$-1.25\times10^6 < f < \frac{1}{2T_b}$$

where T_b is the bit time (the reciprocal of bit rate). The frequency limitations are imposed by the transmitter baseband Nyquist filter.

Equation A-1 can be used to calculate the input power to the filter or to calculate the output power for the flat parts of the VSB filter. For the sloping part of the filter response

$$\int\left(\frac{0.75\times10^6 + f}{1.5\times10^6}\right)^2 P_i\, df =$$

$$\frac{1}{4}\int P_i(f)\, df + \left(\frac{2\times0.75\times10^6}{(1.5\times10^6)^2}\right)\int f\, P_i(f)\, df + \frac{1}{(1.5\times10^6)^2}\int f^2\, P_i(f)\, df$$

The first term can be integrated using an equation similar to A-1. The middle of the three terms integrates to zero because it is anti-symmetric about the origin. The third term may be expanded

$$\frac{1}{(1.5\times10^6)^2}\int f^2\left(\frac{\sin(\pi T_b f)}{\pi T_b f}\right)^2 df = \frac{1}{\pi^2 T_b^{\,2}(1.5\times10^6)^2}\int \sin^2(\pi T_b f)\,df =$$

$$\frac{1}{\pi^2 T_b^{\,2}(1.5\times10^6)^2}\left(\frac{b-a}{2}+\frac{1}{4\pi T_b}(\sin 2\pi T_b a - \sin 2\pi T_b b)\right)$$

So the total output power is

$$\frac{1}{8(\pi T_b)^2}\left(\frac{1-\cos 2\pi T_b(-0.75\times10^6)}{-0.75\times10^6} - \frac{1-\cos 2\pi T_b(0.75\times10^6)}{0.75\times10^6}\right)$$

$$+\frac{1}{4\pi T_b}\int_{-0.75\times10^6}^{0.75\times10^6}\frac{\sin 2\pi T_b f}{f}\,df$$

$$+\frac{1}{\pi^2 T_b^{\,2}(1.5\times10^6)^2}\left(\frac{(0.75\times10^6)-(-0.75\times10^6)}{2}\right.$$

$$\left.+\frac{1}{4\pi T_b}(\sin 2\pi T_b(-0.75\times10^6) - \sin 2\pi T_b(0.75\times10^6))\right)$$

$$+\frac{1}{2(\pi T_b)^2}\left(\frac{1-\cos 2\pi T_b(0.75\times10^6)}{0.75\times10^6} - \frac{1-\cos\pi}{1/2T_b}\right)$$

$$+\frac{1}{\pi T_b}\int_{0.75\times10^6}^{1/2T_b}\frac{\sin 2\pi T_b f}{f}\,df$$

Simplifying, the output power becomes

$$\frac{1}{4(\pi T_b)^2}\left(\frac{1-\cos 2\pi T_b(0.75\times10^6)}{0.75\times10^6}\right) - \frac{1}{2(\pi T_b)^2}\left(\frac{1-\cos\pi}{1/2T_b}\right)$$

$$+\frac{1}{\pi^2 T_b^{\,2}(1.5\times10^6)^2}\left(0.75\times10^6 - \frac{1}{2\pi T_b}\sin 2\pi T_b(0.75\times10^6)\right)$$

$$\frac{1}{\pi T_b}\int_0^{\pi}\frac{\sin f}{f}\,df - \frac{1}{2\pi T_b}\int_0^{2\pi T_b(0.75\times10^6)}\frac{\sin f}{f}\,df$$

The input power can be calculated using equation A-1

$$\frac{2}{\pi T_b}\int_0^{\pi}\frac{\sin f}{f}\,df-\frac{1}{(\pi T_b)^2}\left(\frac{1-\cos\pi}{1/2T_b}\right)$$

The signal-power loss is the ratio of output power to input power.

4 Calculation

The bit duration T_b for the Sony system is $\frac{1}{7.4\times 10^6}$ and the pre-modulation bandpass filter cutoff is 3.7 MHz [18]. Since the Sony pre-modulation bandpass filter meets the Nyquist criterion for eliminating intersymbol interference, the equations derived above are directly applicable.

4.1 Transmitter Bandpass Filter

The input power is

$$\frac{2(7.4\times 10^6)}{\pi}\int_0^{\pi}\frac{\sin f}{f}\,df-\frac{(7.4\times 10^6)^2}{\pi^2}\left(\frac{1-\cos\pi}{3.7\times 10^6}\right) = 5.7\times 10^6$$

The output power is

$$\frac{(7.4\times 10^6)}{\pi}\int_{-1.06}^{\pi}\frac{\sin f}{f}\,df-\frac{(7.4\times 10^6)^2}{2\pi^2}\left(\frac{1-\cos\pi}{3.7\times 10^6}-\frac{1-cos(-1.06)}{-1.25\times 10^6}\right)$$
$$= 4.0\times 10^6$$

The transmitter bandpass filter loss in dB is

$$filter\ loss\,(dB) = 10\log\frac{4.0\times 10^6}{5.7\times 10^6} = -1.5\,dB.$$

4.2 VSB Calculations

$$\frac{(7.4\times10^6)^2}{4\pi^2}\left(\frac{1-\cos 2\pi(0.75/7.4)}{0.75\times10^6}\right)-\frac{(7.4\times10^6)^2}{2\pi^2}\left(\frac{1-\cos\pi}{3.7\times10^6}\right)$$
$$+\frac{(7.4\times10^6)^2}{\pi^2(1.5\times10^6)^2}\left(0.75\times10^6-\frac{7.4\times10^6}{2\pi}\sin 2\pi(0.75/7.4)\right)$$
$$\frac{7.4\times10^6}{\pi}\int_0^{\pi}\frac{\sin f}{f}\,df-\frac{7.4\times10^6}{2\pi}\int_0^{2\pi(0.75/7.4)}\frac{\sin f}{f}\,df$$

$$= 2.6\times10^6$$

The input power is the same as the output power above or $4.0\text{x}10^6$

The VSB filter loss in dB for the data signal is

$$filter\ loss\,(dB) = 10\log\frac{2.6\times10^6}{4.0\times10^6} = -1.9\ dB.$$

Appendix B

Bit Error Rate of an Envelope Detector

Although results have been calculated for the error rate of envelope detection of a 100% modulated AM signal, two generalizations are needed to calculate the bit error rate for envelope detection of a VSB signal. Unlike an AM signal, the sidebands of a VSB signal are asymmetrical and as a consequence some of the signal power is in quadrature with the carrier. This is known as quadrature distortion since the signal envelope is distorted from that of a normal AM signal. Noise in quadrature with the carrier also affects the envelope and this is known as quadrature noise.

A method of reducing quadrature distortion leads to the second generalization. As the carrier power increases relative to the data power, the effect of quadrature distortion is reduced. As a consequence, we would like to be able to modulate at less than 100% to *reduce the quadrature distortion to an acceptable level.*

In general, the signal and noise power of the quadrature phase are shaped by the VSB filter response, and this makes calculation of the signals magnitude difficult. To simplify the argument, we shall assume that the VSB system totally eliminates the entire lower sideband of the signal, producing a single sideband system with a carrier. This assumption simplifies the treatment of the signal and noise on the quadrature channel, and it is also the worst case situation because the signal and noise power on the quadrature channel are maximized. The argument presented is a modified version of that presented by Gregg for envelope detection of AM signals [14].

After passing through the VSB filter, the signal envelope $e(t)$ at the detector can be expressed as

$$e(t) = \Big(C + s_c(t) + n_c(t)\Big)\cos\omega_o t - \Big(s_q(t) + n_q(t)\Big)\sin\omega_o t$$

where C is the carrier amplitude, s_c(t) is the modulating signal, n_c(t) is the noise in phase with the carrier, s_q(t) is the signal in quadrature with the carrier caused by VSB filtering, n_q(t) is the noise in quadrature with the carrier, and ω_o is the carrier frequency.

Assuming bipolar square-pulse modulation, $s_c(t) = +/- A/2$. Because the VSB filter sharply cuts off at the carrier frequency, s_q(t) is the Hilbert transform of s_c(t). Therefore, $n_q(t) = +/- A/2$. Since $s_c(t) = s_q(t)$, from now on, s_c(t) and s_q(t) will be referred to as s(t).

The envelope detector detects the magnitude of the received signal, so it is convenient to transform coordinates to magnitude/angle form.

$$r(t) = \left(\Big(C + s(t) + n_c(t)\Big)^2 + \Big(s(t) + n_q(t)\Big)^2\right)^{1/2}$$

$$\phi(t) = Tan^{-1}\left(\frac{s(t) + n_q(t)}{C + s(t) + n_c(t)}\right)$$

where r(t) is the magnitude of the received signal and ϕ(t) is the phase of the signal.

At detection times nT_b, where n is an integer and T_b is the bit duration, the time variables can be replaced by their sampled values. To simplify the mathematics,

$$r = \sqrt{x^2 + y^2}$$

$$\phi = Tan^{-1}\left(\frac{y}{x}\right)$$

$$x = r\cos\phi = C + s + n_c$$

$$y = r\sin\phi = s + n_q$$

Since n_c and n_q are gaussian, zero mean with variance N, and uncorrelated, x and y are gaussian and independent.

$$f(x, y) = f(x)f(y) = \frac{1}{2\pi N} exp\left(\frac{-(x-s-C)^2 - (y-s)^2}{2N}\right)$$

Using the transform identities of $x = r \cos \phi$ and $y = r \sin \phi$, the probability density can be transformed to magnitude/angle form to obtain

$$f(r, \phi) = \frac{r}{2\pi N} exp\left(\frac{-(r^2 + 2s^2 + C^2 + 2sC)}{2N}\right) exp\left(\frac{r\sqrt{2s^2 + C^2 + 2sC}}{N} \cos(\phi - \phi_k)\right)$$

$$\phi_k = Tan^{-1}\left(\frac{s}{s + C}\right)$$

The probability density function of the envelope magnitude is found by integrating over ϕ.

$$f(r) = \frac{\partial}{\partial r} F(r, \infty) = \int_{-\infty}^{\infty} (r, \phi)\, d\phi$$

Since ϕ ranges between 0 and 2π

$$f(r) =$$

$$\frac{r}{N} exp\left(\frac{-(r^2 + 2s^2 + C^2 + 2sC)}{N}\right)\left(\frac{1}{2\pi}\int_0^{2\pi} exp\left(\frac{(r\sqrt{2s^2 + C^2 + 2sC} \cos\phi)}{N}\, d\phi\right)\right)$$

where the phase factor ϕ_k has been dropped since the integral covers 0 to 2π. The above integral does not exist in closed form; the result is a modified Bessel function of the first kind, zero order.

$$I_o(x) = \frac{1}{2\pi}\int_0^{2\pi} exp(x \cos\phi)\, d\phi$$

Now we compute the envelope density for one by replacing s with A/2.

$$f_1(r) = \frac{r}{N} exp\left(\frac{-(r^2 + A^2/2 + C^2 + AC)}{2N}\right) I_o\left(\frac{r\sqrt{A^2/2 + C^2 + AC}}{N}\right)$$

For the envelope density function for a zero, replace s with -A/2.

$$f_0(r) = \frac{r}{N} exp\left(\frac{-(r^2 + A^2/2 + C^2 - AC)}{2N}\right) I_o\left(\frac{r\sqrt{A^2/2 + C^2 - AC}}{N}\right)$$

Given the probability density functions for the envelopes of a transmitted one and zero, the receiver must set a decision threshold to decide which symbol has been received.

$$r(nT_b) \geq \alpha \qquad \textit{Decide a 1-symbol was sent}$$

$$r(nT_b) < \alpha \qquad \textit{Decide a 0-symbol was sent}$$

where α is the decision threshold. The probability of error for a zero is the probability that a zero is sent and a one is detected. If α is the detection threshold, then

$$P_0(\alpha) = \int_\alpha^\infty f_0(r)\, dr$$

Likewise, the probability of error for a one is

$$P_1(\alpha) = \int_0^\alpha f_1(r)\, dr = 1 - \int_\alpha^\infty f_1(r)\, dr$$

] These integrals cannot be solved in closed form, but the values were calculated numerically and tabulated by Marcum [19]. The integral is called the Q function

$$\int_b^\infty t\, e^{-(t^2 + a^2)/2}\, I_o(at)\, dt \equiv Q(a, b)$$

By employing a change of variables, the error probabilities can be expressed with the Q function.

$$t = \frac{r}{\sqrt{N}}$$

$$a = \frac{\sqrt{A^2/2 + C^2 - AC}}{\sqrt{N}} \qquad \textit{for } P_0$$

$$a = \frac{\sqrt{A^2/2 + C^2 + AC}}{\sqrt{N}} \qquad \textit{for } P_1$$

The error probabilities are

$$P_0 = Q\left(\frac{\sqrt{A^2/2 + C^2 - AC}}{\sqrt{N}}, \alpha'\right)$$

$$P_1 = Q\left(\frac{\sqrt{A^2/2 + C^2 + AC}}{\sqrt{N}}, \alpha'\right)$$

where α' is $\alpha\sqrt{\mathrm{N}}$, the normalize detection threshold. We now need to calculate α.

If p_0 is the probability of transmitting a zero and p_1 is the probability of transmitting a one, then the error probability for the channel is

$$P_e = p_0 P_0(\alpha) + p_1 P_1(\alpha)$$

The total error is minimized by taking the derivative of P_e with respect to α.

$$\frac{dP_e}{d\alpha} = p_0 \frac{d}{d\alpha}\int_\alpha^\infty f_0(r)\,dr + p_1 \frac{d}{d\alpha}\int_0^\alpha f_1(r)\,dr = 0$$

$$\frac{f_1(\alpha)}{f_0(\alpha)} = \frac{p_0}{p_1}$$

Substituting in the expressions for $f_1(\alpha)$ and $f_0(\alpha)$

$$\frac{p_0}{p_1} = \frac{I_o\left(\frac{\alpha\sqrt{A^2/2 + C^2 + AC}}{N}\right)}{I_o\left(\frac{\alpha\sqrt{A^2/2 + C^2 - AC}}{N}\right)} e^{-AC/2N}$$

Unfortunately, it is difficult to solve this equation explicitly for α.

We must now calculate the error rate in terms of the S/N and C/N ratios at the detector input. For equally probable 1's and 0's, the average signal power on the channel is

$$\frac{1}{2}\left(p_1\left(\frac{A}{2}\right)^2 + p_0\left(-\frac{A}{2}\right)^2\right) = \frac{A^2}{8}$$

The average carrier power is

$$\frac{1}{2}C^2$$

The noise on the channel is

$$N = n_o B$$

where n_o is the spectral density of the noise power and B is the receiver bandwidth.

Therefore

$$\frac{A^2}{8N} = \frac{S}{N} \textit{ ratio at the detector input}$$

$$\frac{C^2}{2N} = \frac{C}{N} \textit{ ratio at the detector input}$$

Making these substitutions, the error probability can be expressed as a function of the S/N and C/N ratios at the detector input

$$P_0 = Q\left(\sqrt{4(S/N) + 2(C/N) - 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

$$P_1 = 1 - Q\left(\sqrt{+4(S/N) + 2(C/N) + 4\sqrt{(S/N)(C/N)}},\ \alpha'\right)$$

The normalized detection threshold α' must satisfy

$$\frac{p_0}{p_1} = \frac{I_o\left(\alpha' \sqrt{4(S/N) + 2(S/N) + 4\sqrt{(S/N)(C/N)}}\right)}{I_o\left(\alpha' \sqrt{4(S/N) + 2(C/N) - 4\sqrt{(S/N)(C/N)}}\right)} e^{-2\sqrt{(S/N)(C/N)}}$$

Appendix C

Contention Reservation Slot Calculation

As suggested in Chapter 5, we would like to minimize the expected network overhead time per successful reservation. The network overhead is the expected walk time

$$kT_r + T_c E_c(k, s) + (s + t)2T_r$$

where k is the number of reservation slots, T_r is the time for each reservation slot, T_c is the collision recovery time, s is the number of stations attempting access, t is the number of stations that piggy-backed reservations during the last polling cycle, and E_c is the expected number of collisions as a function of k and s. The expected overhead per successful reservation is

$$\frac{kT_r + T_c E_c(k, s) + (s + t)2T_r}{E_s(k, s)}$$

where E_s is the expected number of successful reservations as a function of k.

$E_c(k, s)$ is the expected number of collisions as a function of the number of reservation slots and the number of stations attempting access. The number of collisions is the number of slots that have two or more reservations. We shall calculate this in two steps; the first is to determine the number of reservation slots with any reservations. Assuming that each of the k reservation slots is chosen with a probability of $1/k$, then the number of slots filled can be derived from the solution to a standard combinatorial problem.

Let $P_m(r, n)$ be the probability that each of r balls placed with a probability of $1/n$ into one of n cells leaves exactly m cells empty. The general solution to this problem is [11]

$$P_m(r, n) = \binom{n}{m} \sum_{i=0}^{n-m} (-1)^i \binom{n-m}{i} \left(1 - \frac{m+i}{n}\right)^r$$

where r is the number of balls, n is the number of cells, and m is the number of cells left empty. If $P(x, y, z)$ is the probability that y balls placed into x cells will fill exactly z of the cells, then

$$P(x, y, z) = P_{x-z}(y, x)$$

because the number of empty cells, m, must be equal to the total number of cells, x minus the number of filled cells z. Making the substitution yields

$$P(x, y, z) = \binom{x}{x-z} \sum_{i=0}^{z} (-1)^i \binom{z}{i} \left(\frac{z-i}{x}\right)^y$$

where x is the number of cells, y is the number balls and z is the number of cells containing at least one ball. In this case, the number of slots (k) replaces x, the number of reservations (s) replaces y, and the number of slots with at least one reservation (f) replaces z.

Although we know the number of slots filled, the distribution of reservations in these slots is unknown. All f slots have at least one reservation, but we want to know the number of slots with two or more reservations. This can be found by applying the above equation recursively. Let the new number of slots be f, the new number of reservations to consider (s - f) and the number of slots filled (c). What has been done is to subtract a reservation for each of the f slots, because each must have at least one reservation in it. Now any of these f slots that still has a reservations must be one of the original slots that had more than one reservation. We recursively apply $P(f, s\text{-}f, c)$; since the probabilities are independent, the probability of c collisions is

$$\sum_{f=1}^{min(k-1, s-1)} P(k, s, f)\, P(f, s-f, c)$$

and the expected value is

$$E_c(k, s) = \sum_{f=1}^{min(k-1,\, s-1)} \sum_{c=1}^{min(f,\, s-f)} c\, P(k, s, f)\, P(f, s-f, c) =$$

$$\sum_f \left(\binom{k}{k-f} \sum_{i=0}^{f} (-1)^i \binom{f}{i} \left(\frac{f-i}{k}\right)^s \right) \left(\sum_c c \left(\binom{f}{f-c} \sum_{j=0}^{c} (-1)^j \binom{c}{j} \left(\frac{c-j}{f}\right)^{s-f} \right) \right)$$

$E_s(k,\ s)\}$ is the expected number of successful reservations. The number of successful reservations is the number of slots with at least one reservation (f) minus the slots with two or more reservations (c) , f - c. So the expected number of successful reservations is

$$E_s(k, s) = \sum_{f=1}^{min(k-1,\, s-1)} \sum_{c=1}^{min(f,\, s-f)} (f-c)\, P(k, s, f)\, P(f, s-f, c)$$

$$= \sum_{f=1}^{min(k-1,\, s-1)} \sum_{c=1}^{min(f,\, s-f)} f\, P(k, s, f)\, P(f, s-f, c) - \sum_{f=1}^{min(k-1,\, s-1)} \sum_{c=1}^{min(f,\, s-f)} c\, P(k, s, f)\, P(f, s-f, c)$$

$$= \left(\sum_{f=1}^{min(k-1,\, s-1)} f\, P(k, s, f) \left(\sum_{c=1}^{min(f,\, s-f)} P(f, s-f, c) \right) \right) - E_c(k, s)$$

$$= \left(\sum_{f=1}^{min(k-1,\, s-1)} f\, P(k, s, f) \right) - E_c(k, s)$$

$$= E_f(k, s) - E_c(k, s)$$

So the expected number of requeued packet is the number of attempts, minus the expected number of slots filled (E_f), plus the expected number of collisions (E_c), where

$E_f(k, s)$ is

$$E_f(k, s) = \sum_{f=1}^{min(k, s)} f\,P(k, s, f)$$

$$= \sum_{f=1}^{min(k, s)} f\left(\binom{k}{k-f}\sum_{i=0}^{f}(-1)^i\binom{f}{i}\left(\frac{f-i}{k}\right)^s\right)$$

Since the minimization is done with respect to k, terms independent of k can be excluded.

$$\frac{kT_r + T_c E_c(k, s)}{E_f(k, s) - E_c(k, s)}$$

The k for which this equation is minimized was determined by evaluating the equation for $k \geq s$ for $s = 2, 3, 4, 5$.

OFFICIAL DISTRIBUTION LIST

1985

Director Information Processing Techniques Office Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209	2 Copies
Office of Naval Research 800 North Quincy Street Arlington, VA 22217 Attn: Dr. R. Grafton, Code 433	2 Copies
Director, Code 2627 Naval Research Laboratory Washington, DC 20375	6 Copies
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12 Copies
National Science Foundation Office of Computing Activities 1800 G. Street, N.W. Washington, DC 20550 Attn: Program Director	2 Copies
Dr. E.B. Royce, Code 38 Head, Research Department Naval Weapons Center China Lake, CA 93555	1 Copy
Dr. G. Hopper, USNR NAVDAC-OOH Department of the Navy Washington, DC 20374	1 Copy

END
DTIC
10-86