MULTIPLEXING METHODS FOR FIBER OPTIC LOCAL COMMUNICATION NETWORKS

by

STUART SCOTT WAGNER

S.B., Massachusetts Institute of Technology (1979)

S.B., Massachusetts Institute of Technology (1980)

S.M., Massachusetts Institute of Technology (1983)

E.E., Massachusetts Institute of Technology (1984)

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1985

Signature of Author____________________________
Department of Electrical Engineering and Computer Science
May 15, 1985

Certified by____________________________
Thesis Supervisor

Accepted by____________________________
Chairman, Electrical Engineering and Computer Science Departmental Committee

MULTIPLEXING METHODS FOR FIBER OPTIC

LOCAL COMMUNICATION NETWORKS

by

STUART SCOTT WAGNER

Submitted to the Department of Electrical Engineering and Computer Science on May 15, 1985 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

ABSTRACT

This Thesis explores various methods for utilizing the bandwidth of fiber optic local networks. Most of the work presumes the use of single mode fibers, implying that channel bandwidths are orders of magnitude higher than those of conventional networks. This enormous bandwidth may be exploited to achieve two purposes: to reduce delay in the network and to simplify network control. We consider both frequency multiplexing and spatial multiplexing as methods of reducing delay. To simplify control, we consider both a fixed frequency assignment scheme and an approach that uses Pure Aloha techniques.

We also examine the effect of power division limitations on network design and capability. The relationship between power limitations and bandwidth usage techniques is described. The power division problem is developed in detail for both the bus and star topologies. We discuss the impact of emerging optics technology, such as heterodyne detection and optical amplifiers, on power division limitations. It is found that optical amplifiers can be extremely valuable in alleviating these limitations for the bus configuration.

Thesis Supervisor: Robert S. Kennedy, Ph.D.

Title: Professor of Electrical Engineering

Acknowledgments

I am indebted to Professor Robert Kennedy, my thesis supervisor, who first stimulated my interest in communications many years ago. Throughout the course of this research, he provided many valuable suggestions and insights that aided in both problem formulation and solution.

Many thanks should also go to the two readers on my Thesis committee, Professor Jeffrey Shapiro and Professor Pierre Humblet. Throughout my graduate study, in both course work and Thesis research, Professor Shapiro has contributed extensively to my knowledge of optics and optical communication. Professor Humblet offered many helpful comments and observations, particularly concerning the idiosyncracies of queueing and network operation.

Table of Contents

Chapter 1

Introduction

Background

The last fifteen years have witnessed enormous effort in the design, development and construction of local communication networks. Such structures, which include local area networks, private branched exchanges and combinations of the two, allow exchange of information among individual users in a community of limited geographic extent, such as a campus or office complex. (The word local typically implies a system diameter of no more than a few kilometers.[1]) To date, a vast number of local networks have been built, or at least have been proposed, with the different designs exhibiting a wide variety[2,3,7] of topologies, data rates, protocols and device technologies.

This Thesis will focus on local networks that employ optical fibers. Over the last decade, the emergence of fiber optics as a viable communications technology has led naturally to widespread effort to incorporate this technology into local networks. Some of the resulting designs have already been built, such as the Xerox Fibernet,[4] Fibernet II,[5] and Chaos of Hewlett Packard.[6] Many others have been proposed.[8,9] Some discussion is therefore in order concerning how this Thesis research will differ from the considerable work that has already been performed in the field of fiber optic local networks.

Goals of the Research

A number of characteristics distinguish this Thesis from previous work in the field. First, most of this research will presume the use of single mode optical fibers. Single mode fiber technology is receiving increased attention,[10-12] particularly for application in long haul, high capacity communication links. However, potential use of these fibers in local networks has yet to be explored fully.

The intrinsic low-dispersion properties of single mode fibers can yield bandwidth-distance products in excess of several hundred GHz·km with a sufficiently narrow linewidth light source.[10,13-15] Given the short propagation distances typical of local networks, it is evident that the channel bandwidth available in a single mode fiber network will be orders of magnitude higher than that available in present networks that employ multimode fiber or coaxial cable. A chief goal of this Thesis will be to determine if and how this tremendous bandwidth can be exploited to yield greater network capabilities and performance.

While pursuing these questions, this research will not attempt to generate a specific network design, a feature that again distinguishes this work from much of the literature on the subject. Rather, we will seek general design principles and techniques. The motivation for this approach is simple. A communication system in which bandwidth is cheap would seem to be fundamentally different from a system in which bandwidth is severely limited. Thus it seems logical to explore networking techniques in the former case that may be fundamentally different from those typical in the latter. If one designed a single mode fiber local network by applying the same philosophies and techniques used in networks of much lower bandwidth, such an attempt would most likely result in most of the single mode fiber bandwidth being wasted, or being used impractically or inefficiently at best. Of course, efficient use of hundreds of Gigahertz may not be necessary in many practical situations. A network that utilizes only a few percent of this total bandwidth may work quite well, depending on the nature of the traffic. However, as data processing needs and capacities increase, we might envision circumstances in which more extensive exploitation of bandwidth could prove highly beneficial.[16]

Yet another characteristic that distinguishes this research from much of the literature is that, in discussing implementation issues, we will not limit our investigations to those devices and techniques that are currently available or practical. Indeed, a major goal of this Thesis will be to examine the potential applicability of emerging optics technology to local networks, and to see how these new developments might allow us to take greater advantage of fiber bandwidth. Heterodyne detection, optical amplifiers, high-Q optical filters and frequency-selective waveguide couplers will be discussed. While such devices cannot yet be considered standard technology, all may be available within the next decade,[17] and it is hoped that their inclusion here will increase the future relevance of this work. Perhaps more important, we may find that a particular device, if available, could lead to substantially greater networking capabilities. Such an observation might provide greater impetus for the development of that device.

Having now described the general motivations and purposes underlying this research, we now discuss the specific issues and problems that this Thesis will examine in the Chapters that follow.

Contents of the Thesis

We seek methods that will enhance network capabilities and performance through the exploitation of fiber bandwidth. This Thesis will investigate two objectives that may be achieved through increased bandwidth usage:

(i) The reduction of network delay.

(ii) The simplification of network control.

Objective (i) will be explored in Chapter 2, while objective (ii) will be the subject of Chapter 5. In addition, Chapters 3 and 4 will investigate various technological issues that underlie our attempts to exploit fiber bandwidth.

The reduction of network delay will involve using increased amounts of fiber bandwidth to achieve network concurrency. Loosely stated, concurrency refers to the ability of more than one packet or conversation to exist on a network simultaneously, without information being lost via collisions (the packets remain completely separable and do not interfere with one another). Concurrency will be the main topic of Chapter 2. In Chapter 2 we will describe and contrast the two principle methods of achieving concurrency in a local network: frequency multiplexing and spatial multiplexing. The chief goal of Chapter 2 is to investigate the potential value of concurrency in local networks, i.e., to determine under what circumstances concurrency can lead to reduction in network delay. We shall find that, in most typical networks, the usefulness of concurrency is confined to two cases. In the first case, the load on the network is so heavy that the network would be unstable unless concurrency were used. In the second case, highly bursty network traffic results in occasionally excessive short term delays. For some types of data (voice traffic, for instance), such delays may be intolerable.

Chapter 3 will examine power division limitations in local networks. At first, this issue may seem to have little connection with the portions of the Thesis that deal with fiber bandwidth usage. We have two motivations for examining the power division problem here. First, power limitations in fiber optic local networks tend to be so severe that any research seeking design principles for such networks should probably take the power division problem into account. Second, we shall see that power limitations have a fundamental bearing on our choice of multiplexing methods. Specifically, we will find that, in a power limited environment, frequency multiplexing has distinct advantages over spatial multiplexing as a means of achieving concurrency.

In Chapter 3 we examine the possible application of emerging optics technology in alleviating power division limitations. The potential impact of both heterodyne detection and optical amplifiers will be investigated. We shall see that the use of heterodyne detection is not particularly effective in dealing with the power division problem. We will find, however, that optical amplifiers have tremendous potential for alleviating power limitations. Furthermore, it will be shown that the gains of these amplifiers do not have to be very high, an encouraging fact, since a lower gain device should be easier to realize.

Chapter 4 analyzes various methods for implementing frequency multiplexing in fiber optic local networks. Both heterodyne detection and direct detection techniques are examined. For direct detection, the use of both optical filters and subcarrier modulation will be considered. The different methods will be compared on the basis of complexity and the maximum number of channels and simultaneous users that are possible. We will find that the subcarrier method is (at least at present) by far the easiest to implement, but may be severely limited in the number of simultaneous usages that it provides. On the other hand, the heterodyne and direct detection/optical filter methods can permit extensive bandwidth usage, but their practicality will depend on future developments in device technology.

Chapter 5 investigates two ways in which increased utilization of fiber bandwidth can lead to simplification (and perhaps elimination) of network control. Specifically, we will examine the potential use of Pure Aloha packet broadcasting techniques, as well as possible use of fixed-assignment frequency multiplexing. The former method involves increased bandwidth usage by individual interfaces, thus allowing for shorter packets and a lower probability of collision in the Pure Aloha scheme. The latter method advocates comparatively low interface bandwidth usage, instead spreading users across the network bandwidth by assigning to each a separate frequency channel. We shall find that both methods are viable possibilities, representing an effective means of simplifying control within the network.

Chapter 2

The Role of Concurrency in Local Networks

Introduction

The purpose of Chapter 2 is to investigate the potential value of concurrency in local networks. This Chapter begins with definitions and illustrations of concurrency and the ways that it may be achieved. We will then describe how concurrency may lead to reductions in network delay.

The chief result of this Chapter is the Concurrency Principle. This Principle states in mathematical terms the circumstances under which the introduction of concurrency into a local network will yield significant reduction in delay. We then interpret the Principle in terms of frequency multiplexing and spatial multiplexing, comparing the two methods.

Definitions

A network that allows concurrency is one in which two or more packets may exist in the network simultaneously without interfering with one another, i.e. without a collision occurring. The two principle methods of achieving concurrency are frequency multiplexing and spatial multiplexing.* Figures 2.1 - 2.4 illustrate various examples. Figure 2.1 shows a snapshot of a bus network that allows no concurrency, a "non-concurrent" network. Only one packet may exist on the network at any given time. Figure 2.2 illustrates concurrency via spatial multiplexing. With this method, the packets may overlap in frequency, but they remain spatially separate (assuming that all packets on the bus propagate in the same direction). The packets arrive at a given receiver at different times and thus remain distinct. Spatial multiplexing involves more extensive use of bandwidth, since to achieve a significant level of multiplexing, the packets must be quite short. This implies the use of high speed (i.e. high bandwidth) interfaces. (The implied relationship between packet length and packet bandwidth assumes that the number of bits per packet is constant.)

Figure 2.3 shows a network allowing concurrency through frequency multiplexing. The packets may now overlap spatially, but remain in disjoint frequency bands and thus are separable by filtering at the various receivers along the bus. Clearly, frequency multiplexing permits more extensive use of bandwidth, with packets being transmitted in previously unused bands.

*There is in fact a third method, involving the use of spread spectrum techniques. It will not be considered here.

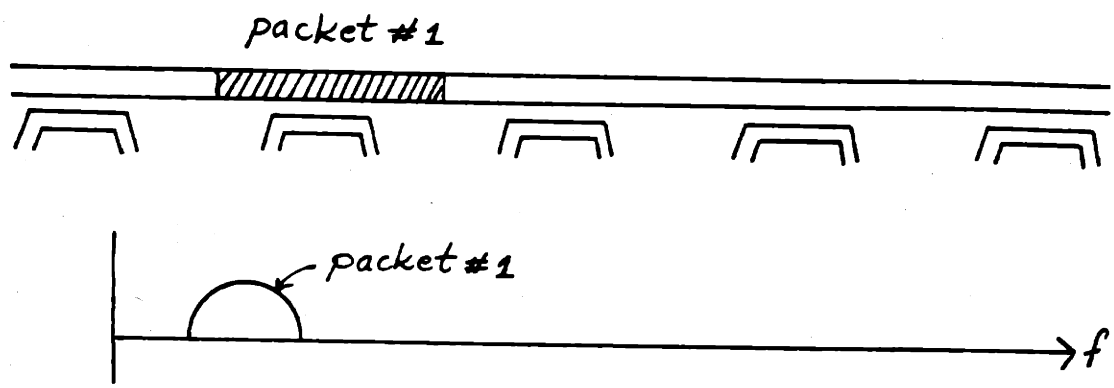


Figure 1

A network employing no concurrency. Only one packet may exist on the network at a time.

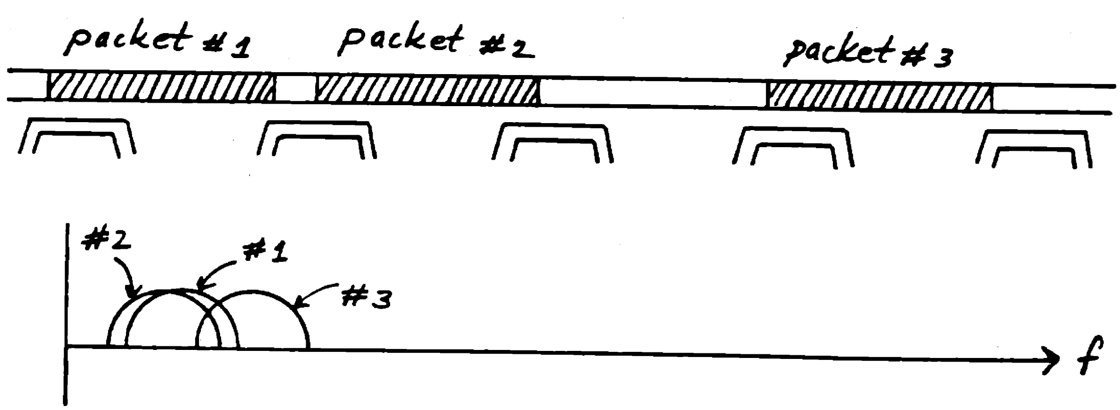


Figure 2

A network employing spatial multiplexing. Packets occupy spatially disjoint portions of the network (unidirectional propagation assumed).

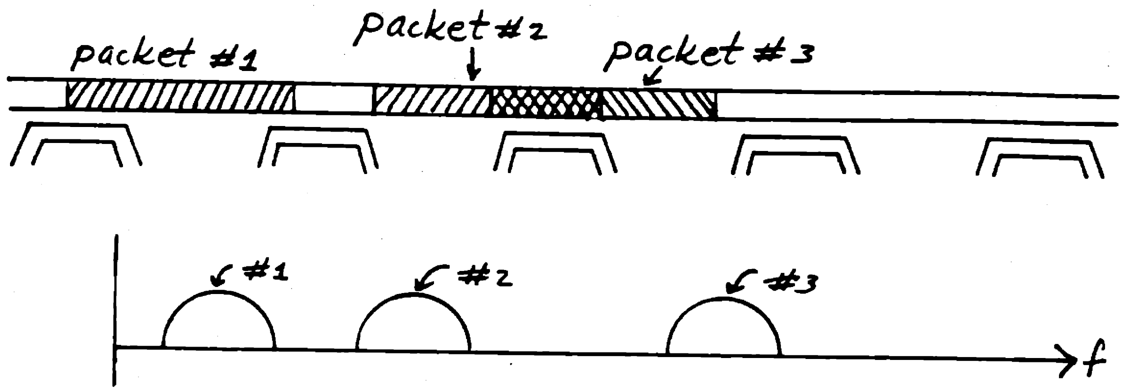


Figure 3

A network employing frequency multi-multiplexing. Packets occupy disjoint frequency bands.

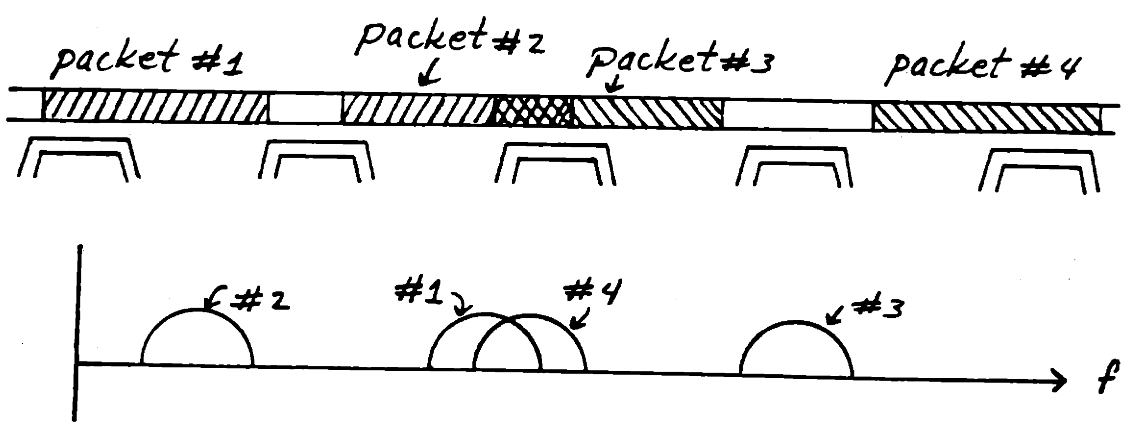


Figure 4

A network employing both spatial and frequency multiplexing. Packets may overlap spatially if they are disjoint in frequency, and vice versa.

Figure 2.4 illustrates the most general case, in which the network employs both spatial and frequency multiplexing simultaneously. In this case, two packets may occupy the same frequency band if they remain separate spatially, and vice versa. It is obvious that the use of both forms of multiplexing together is more optimal than the use of either form alone, more optimal in the sense that it would allow more extensive use of network space and bandwidth. However, such a scheme would also be considerably more complex, both analytically and physically, than the networks of Figures 2.2 and 2.3. For simplicity, this research will confine itself to networks in which concurrency is achieved through either frequency multiplexing or spatial multiplexing, but not both simultaneously. In future Chapters, frequency multiplexing will receive most of the attention, due to a number of technological factors as mentioned in Chapter 1.

Use of Concurrency to Reduce Delay

We regard network space and bandwidth as resources to be allocated to the users in a manner that allows the network to satisfy user requirements. Concurrency involves the allocation of network resources to a number of users simultaneously. To illustrate how such a technique may lower delay in a network, we begin with the following plausibility argument.

Consider a network whose total available bandwidth is so high that no single user interface can utilize more than a minute fraction of the bandwidth at a time. (Such will typically be the case in a single mode fiber network.) Thus, if this network did not allow concurrency, only a small portion of the total bandwidth would be in use at any given time. Such waste may well have no consequences, but consider what happens under heavy traffic conditions, with a high rate of packets arriving to the network for transmission. In the non-concurrent case, we may think of the network as a single server system. Only one packet can be broadcast at a time and the rest must wait, introducing a queueing delay. Now say that we exploited the previously unused bandwidth to achieve concurrency, turning the network into a multiple server system. This could be accomplished with either frequency multiplexing or spatial multiplexing, as described in the previous Section. Clearly, if enough bandwidth could be accessed (i.e. if enough servers could be provided), the queueing delay would be eliminated. If the queueing delay were significant compared to other delays in transferral of information, such as packet preparation time, propagation delay, and packet duration, then the reduction in delay experienced by the users would be significant.

Let us now investigate this reduction in delay quantitatively. Define D to be the overall delay between users, the time that elapses between the instant that the information becomes available and the instant that the packet is received in its entirety at the destination (for the moment, we assume error-free transmission and ignore acknowledgment issues). The delay D has five components:

$$D = P + Q + C + T_p + \tau_p \tag{2.1}$$

where P is the packet preparation time, Q is the queueing delay, C is a control delay, T_p is the packet duration (number of bits per packet divided by bit duration) and τ_p is the propagation delay between source and destination. The nature of the control delay C depends on the network; in a frequency multiplexing scheme, for instance, C would be the time required to establish a virtual channel between source and destination. For typical numbers ($\tau_p <$ 50 μsec, $C \approx \tau_p$, 1 msec $< P <$ 10 msec, bit rate in excess of 10^6/sec and thus $T_p <$ 1 msec for a 1000-bit packet), the packet preparation time P will clearly dominate the overall delay unless Q is significantly large compared to P.

In a network that allows concurrency we will have Q = 0, assuming that the network always has enough space or bandwidth to satisfy user demand (this assumption will be examined more carefully later). In a non-concurrent network we will in general have $Q > 0$ with Q increasing as the load on the network increases. We now define a concurrency improvement factor I to quantify the effectiveness of concurrency in reducing delay:

$$I = \frac{E[D(\text{no concurrency})]}{E[D(\text{concurrency})]} - 1 \qquad (2.2)$$

From the above discussion, D(concurrency) is the D of (2.1) with Q set equal to zero. $I \approx 0$ indicates that little or no reduction in delay would be afforded by concurrency. $I \gg 0$ indicates that concurrency would yield substantial reduction in delay. Based on the previous observations vis a vis Eq. (2.1), we conclude that $I \gg 0$ only when Q for the non-concurrent network is much greater than P. We now determine how high the offered load must be in order to have $Q \gg P$ in the non-concurrent network.

Bux[21] has derived expressions for mean delay between users for various network topologies and contention protocols.* A key parameter in these expressions is the network usage parameter

$$\rho = \lambda E(T_p + \tau_p + C) \qquad (2.3)$$

where λ equals the average rate of arrival of new packets

*Bux's expressions for mean delay do not include packet preparation time.

to the network for transmission. In queueing theory terms,[22] a non-concurrent network is a single server system in which the jobs (packets) arrive at a rate λ and require an average service time $E(T_p+\tau_p+C)$. If $\rho \ll 1$ the network is lightly loaded, and we would expect $Q \ll P$, $I \approx 0$. If $\rho \gg 1$, the offered load exceeds the maximum throughput and the queues will become infinite, as will Q and I. Let us now investigate what happens when ρ falls between these two extremes.

We examine two specific cases of non-concurrent networks: the single token ring and the ordered access bus. Borrowing from the analysis of Bux, we find the following expressions for the improvement factor I for the two cases:

<u>Ring</u>

$$I = f(\rho)\left[\tau + \frac{\rho(E(T_p^2) + 2\,E(T_p) + \tau^2)}{\tau + E(T_p)}\right] \qquad (2.4)$$

<u>Bus</u>

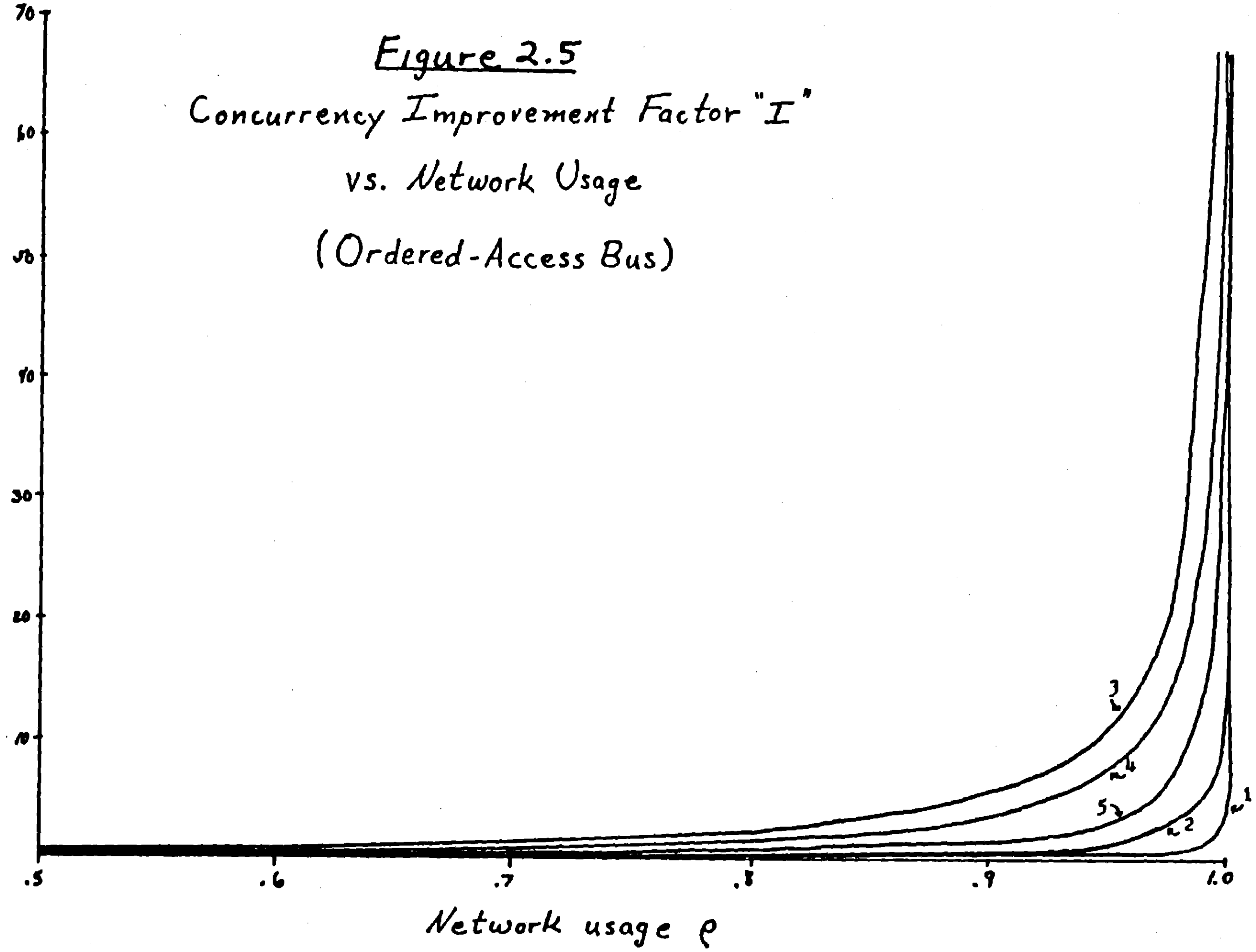
$$I = f(\rho)\left[T_s(3-\rho) - 2\tau(1-\rho) + \frac{\rho(E(T_p^2) + \tau E(T_p) + \tau^2/3)}{E(T_p) + \tau/2}\right] \qquad (2.5)$$

where $\tau \equiv E(\tau_p)$ is the mean propagation delay, T_s is the time needed for scheduling in the ordered access bus, and

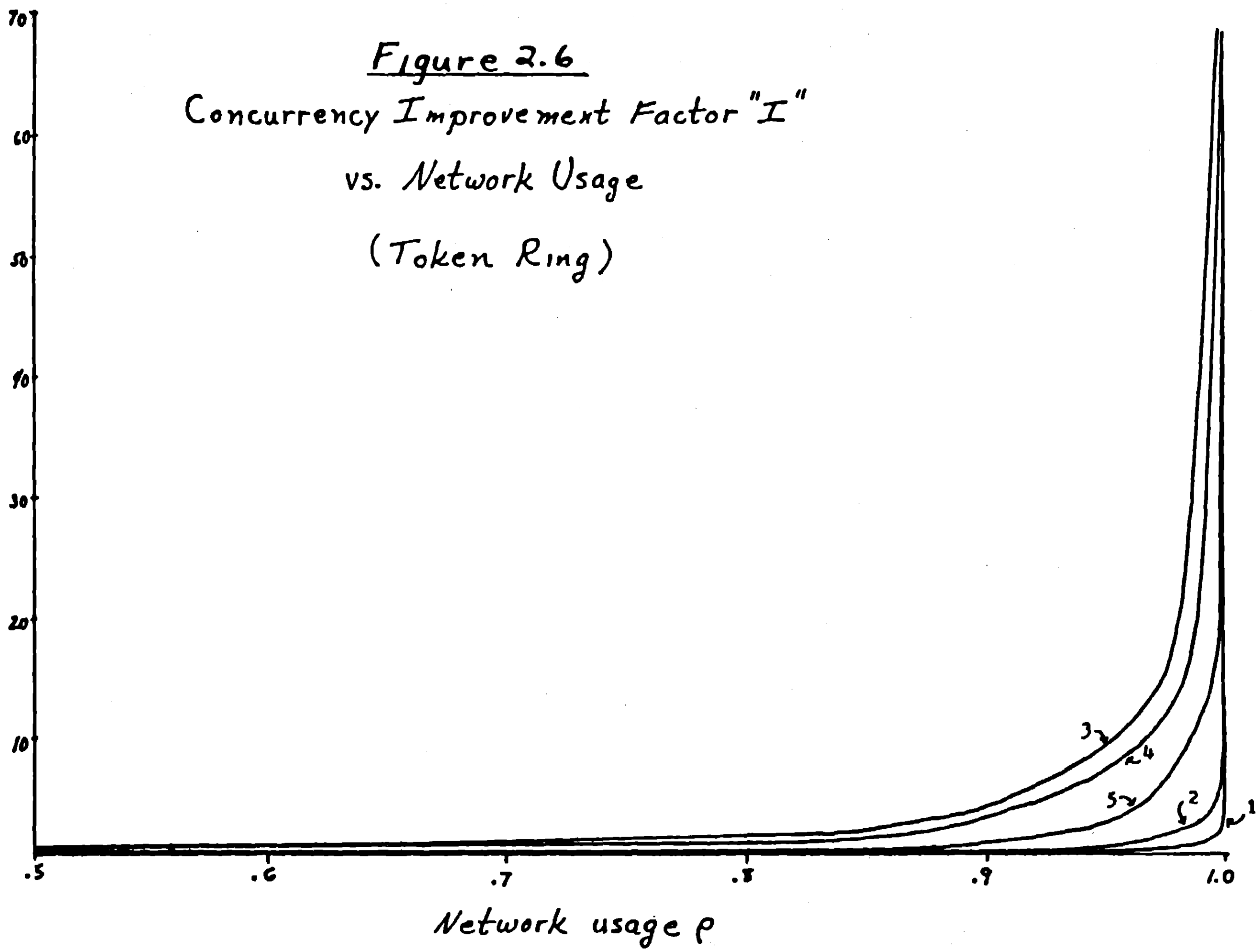
$$f(\rho) = 2(1 - \rho)(P + E(T_p) + 3\tau/2)^{-1} \qquad (2.6)$$

Figures 2.5 and 2.6 show the concurrency improvement factor I plotted as a function of ρ for the two cases. Each figure has five different curves, with each curve having been calculated using a different statistical packet distribution and different τ. The five combinations used are summarized in Table 2.1. The packet length distributions were chosen to represent a wide range of possibilities for the first and second moments of T_p. The exponential distributions are used commonly in local network analyses and simulations. The discrete distributions (distributions 3 - 5) could stem from networks whose interface bandwidths vary widely from user to user.

Figures 2.5 and 2.6 suggest that, over a wide range of packet length distributions, propagation delays and topologies, the reduction in delay afforded by concurrency is negligible unless ρ is near 1 or greater than 1, i.e. unless the non-concurrent network is at the threshold (if not over the threshold) of stability. Based on these observations, we assert the following Concurrency Principle.



Concurrency Improvement Factor vs. Network Usage. The curves 1-5 stem from the different packet length distributions in Table 2.1. Packet preparation time P = 1 msec was assumed throughout.



Concurrency Improvement Factor vs. Network Usage. The curves 1-5 stem from the different packet length distributions in Table 2.1. Packet preparation time P = 1 msec was assumed throughout.

	Distribution $p_{T_p}(T)$ (T in seconds)	$E(T_p)$ (sec)	$E(T_p^2)$ (sec^2)	τ (sec)
1	$10^5 e^{-10^5 T} u_{-1}(T)$	10^{-5}	10^{-10}	$3\text{x}10^{-6}$
2	$10^3 e^{-10^3 T} u_{-1}(T)$	10^{-3}	10^{-6}	$3\text{x}10^{-5}$
3	$.4\delta(T-10^{-6}) + .3\delta(T-10^{-5}) + .2\delta(T-10^{-4}) + .1\delta(T-10^{-3})$	$3\text{x}10^{-5}$	$3\text{x}10^{-8}$	$3\text{x}10^{-6}$
4	$(8/9)\delta(T-5\text{x}10^{-7}) + (1/9)\delta(T-5\text{x}10^{-6})$	10^{-6}	$3\text{x}10^{-12}$	$3\text{x}10^{-6}$
5	$.5\delta(T-10^{-7}) + .25\delta(T-10^{-6}) + .25\delta(T-10^{-5})$	$3\text{x}10^{-6}$	$3\text{x}10^{-11}$	$3\text{x}10^{-5}$

Table 2.1

Packet length distributions and propagation delays used for curves in Figures 2.5 and 2.6.

<u>The Concurrency Principle</u>. Define the following variables:

$T_{S,D}(i,j)$ = duration of a packet generated by i^{th} user and addressed to j^{th} user.

$\lambda_{S,D}(i,j)$ = average rate of production of packets generated by i^{th} user and addressed to j^{th} user.

For most useful networks of interest, it is necessary for at least one of the following conditions to be satisfied in the non-concurrent system, in order for concurrency to provide substantial reduction in delay:

1) $$(\sum_{i,j} \lambda_{S,D}(i,j))(E_{S,D}[T_{S,D}(i,j)] + C + \tau) \gtrsim 1 \qquad (2.7)$$

2) $$\mathrm{Prob}\left[(Q + T_{S,D}(i,j) + C + \Upsilon) > M\right] > \delta \qquad (2.8)$$

where $E_{S,D}()$ is the expectation over all sources and destinations, "$\gtrsim$" means nearly equal to or greater than, and where M and δ depend on the requirements of user i.

<u>Discussion</u>. The quantity on the left hand side of condition (1) is the network usage parameter ρ. Thus condition (1) states that, for most networks, we can expect $I \gg 0$ only if $\rho \gtrsim 1$, a fact suggested by Figures 2.5 and 2.6.

The justification for condition (2) is not as obvious.

The quantity in parentheses is the delay experienced by a packet between its creation and arrival at its destination. The idea behind condition (2) is that users might have a maximum tolerable delay in delivering data, perhaps due to finite buffers or due to the nature of the data itself. This maximum might be exceeded if Q were to become large for even a short time span. Consider, for example, a network carrying voice traffic. Certainly in this case there is a maximum length of time that may elapse between delivery of successive packets, because the data is being handled in real time. Another example would be an emergency network that spends most of its time transmitting little or no data. Because of these long dormant periods, we are bound to have $\rho < 1$ (recall that ρ involves the average packet arrival rate which in this case would be quite low). However, when an emergency occurs, the amount of traffic in the network increases drastically, and this heavy load may lead to long queueing delays over the short term (i.e. during an emergency). Unfortunately, this is the time interval during which delays can be tolerated least. Such bursty traffic is common to local networks, whose peak to mean traffic ratios often exceed 1000.[1] Although the network queues may remain stable over the long term, the above examples illustrate that such long term stability is not always the most relevant measure of network performance.

A number of other comments are in order regarding the Concurrency Principle. This Principle is not a totally general statement that applies to all local networks. One can easily conceive of examples (albeit rather pathological ones) designed specifically to violate the Principle. Further, we do not dismiss the possible existence of a packet length distribution that leads to $I \gg 0$ when, say, $\rho = .4$. However, Figures 2.5 and 2.6 suggest that for typical networks, the ability of concurrency to reduce delay is limited to those cases in which the non-concurrent network is unstable, which is condition (1).

Several other observations can be made. In particular, note that conditions (1) and (2) were stated as being <u>necessary</u> but not <u>sufficient</u>. Indeed, there exist at least three situations in which the use of concurrency may still fail to reduce delay:

(i) We have thus far assumed that network resources available through concurrency will always be sufficient to handle user demand, i.e., that the achievable level of concurrency will always exceed the level required. However, we shall see later that, depending on implementation, this may not always be so. If

$$(\sum_{i,j} \lambda_{S,D}(i,j))(E_{S,D}(T_{S,D}(i,j)) + C + \tau) \gtrsim m \qquad (2.9)$$

where m is the maximum achievable level of concurrency, the network will remain unstable even with concurrency. Eq. (2.9) was written assuming for simplicity that all packets are of equal length and occupy equal bandwidth. This equation is clearly the m-server equivalent of Eq. (2.7).

(ii) Assuming that a user can transmit only one packet at a time, it is possible to have a bottleneck at some transmitter:

$$\left(\sum_j \lambda_{S,D}(i,j)\right)\left(E_D(T_{S,D}(i,j)) + C\right) \gtrsim 1 \qquad (2.10)$$

for one or more i, where $E_D()$ is the expectation taken over all destinations. Condition (2.10) would lead to an unstable queue of packets at station i, and concurrency in the network could not alleviate the problem.

(iii) Assuming that a user can receive only one packet at a time, it is possible to have a bottleneck of packets waiting to be sent to that destination:

$$\left(\sum_i \lambda_{S,D}(i,j)\right)\left(E_S(T_{S,D}(i,j)) + C\right) \gtrsim 1 \qquad (2.11)$$

for one or more j, where $E_S()$ is the expectation taken over all sources.

One final observation should be made. It was noted earlier that the packet preparation time P typically dominates all other contributions to the delay D in Eq. (2.1), sometimes by an order of magnitude or more. Therefore, it is conceivable that we could have network stability ($\rho < 1$) while having

$$(\sum_j \lambda_{S,D}(i,j))\, P \gtrsim 1 \qquad (2.12)$$

since P is usually much larger than $E(T_p) + C + \tau$. Condition (2.12) indicates an unstable queue of information waiting to be put into packet form at user i. If (2.12) holds for some i, then the packet preparation process at user i must be accelerated, either by decreasing P for that user or perhaps by using multiple "packet preparers" in parallel.

The possibility of condition (2.12), while being a viable concern in its own right, does not have significant relevance vis a vis the Concurrency Principle or any of the accompanying discussion, for (2.12) is most appropriately viewed as a <u>user</u> problem rather than a <u>network</u> problem. No change in network design can alleviate the instability indicated in (2.12) if it exists, nor is the rest of the network affected, except of course for those users who desire information that is hung up in the (unstable) queue in question.

Methods for Achieving Concurrency

For most of this Chapter we have regarded concurrency as a general networking technique, without considering how the concurrency would actually be achieved in the network. As mentioned in Chapter 1, frequency multiplexing and spatial multiplexing are the two chief methods of realizing concurrency. Let us now take a closer look at these two methods, examining the application of each in reducing network delay through stabilization of unstable conditions. Along the way we will interpret these methods in light of the Concurrency Principle, as well as compare their relative effectiveness and utility.

Consider a network whose interfaces operate at relatively low bandwidths, such that $E(T_p) \gg \tau$. For example, assuming a 1000-bit packet and $\tau = 10$ µsec, this corresponds to packet bandwidths of 50 MHz or less. Say that this network is unstable ($\rho > 1$), and that we wish to stabilize it through concurrency. We have two choices, spatial multiplexing and frequency multiplexing. If the interface bandwidth is constrained such that $E(T_p) \gg \tau$, then spatial multiplexing is not an appropriate alternative, since the packet length far exceeds the network "size," and typically only one packet could fit spatially onto the network at once. On the other hand, frequency multiplexing seems quite well suited to this particular problem, since with each user occupying only a minute portion of the total available bandwidth, a large number of users could be accomodated simultaneously. In fact, assuming that the total network bandwidth is B, and that each packet requires bandwidth W, we may say via the Concurrency Principle that a necessary condition for stabilization by frequency multiplexing is

$$\lambda(E(T_p) + C + \tau) < B/W \qquad (2.13)$$

where $\lambda \triangleq \sum_{i,j} \lambda_{S,D}(i,j)$ is the total packet generation rate in the network.

A dual argument could be made in the case of a network whose interfaces operate at or near the link bandwidth. Here, frequency multiplexing would lose favor because individual packets require so much bandwidth that a sufficient number of separate channels may not be available. On the other hand, with high enough data rates the packets become much shorter than the network size ($E(T_p) \ll \tau$) and a large number of packets can be multiplexed spatially.

Again, the above arguments exploit a certain duality between packet length and packet bandwidth, a duality that assumes the number of bits per packet to be a constant.

A number of other significant comments can be made about the previous examples. Consider again the case of an unstable network with low bandwidth interfaces. We saw that frequency multiplexing was an effective method of stabilizing this type of network. Alternatively, we might think of stabilizing the network by increasing the interface bandwidth capabilities, if feasible, to the point where a sufficient degree of _spatial multiplexing_ were possible. In fact, if we increased the interface bandwidth capabilities we might even be able to stabilize the network _without going to concurrency_ as long as $\lambda(\tau + C) < 1$. To see this, recall Eq. (2.7). With sufficiently high data rates we will have $E(T_p) \ll \tau + C$ and $\rho \approx \lambda(\tau + C)$. Physically, increasing the data rate in this case has lowered the service time $E(T_p) + \tau + C$ to the point where the network can handle packets at a rate high enough to maintain stability without concurrency. It is important to note, however, that if $\lambda(\tau + C) > 1$ then even infinite interface bandwidth usage cannot stabilize the network.

The above examples have suggested a degree of duality or interchangeability between spatial multiplexing and frequency multiplexing. This duality does exist, at least in principle. In practice, however, there are a number of technological issues inherent to each form of multiplexing, and the various tradeoffs will often favor one method over the other. For example, spatial multiplexing requires the use of high speed interfaces, the cost of which may be prohibitive for many users in the network. On the other hand, frequency multiplexing with optics technology is not a trivial undertaking, either, as we shall see in Chapter 4. A further consideration is the effect of power division limitations to be studied in Chapter 3. We shall find that these limitations favor the use of frequency multiplexing when peak laser power is limited. However, if the peak source power is sufficiently high, the advantages of frequency multiplexing in this situation are not as clear. Overall, it may be said that the multiplexing method of choice will often depend on technological tradeoffs and capabilities, and that in practice the two methods will not be completely interchangeable.

In the next two Chapters we investigate a number of implementation issues inherent to the exploitation of fiber bandwidth. In examining these technological considerations, we draw a number of conclusions that not only are significant in themselves, but should also clarify the tradeoffs between the multiplexing methods that we have discussed.

Chapter 3

Power Division Issues in Local Networks

Introduction

In a local network, a user transmits information by injecting signal power into the communications medium. The power is then distributed in some way among the other users, according to the network topology. Clearly, a given amount of signal power may be divided up among only so many users before the received energy levels drop below the minimum required for acceptable error rates. This limitation on the number of users that a (passive) network can support is often called the power division problem. It tends to be particularly troublesome in networks employing fiber optics, principally due to the cumulative effects of many slightly lossy optical components, as we shall see later in this Chapter.

The power division problem is well known in the context of fiber optic local networks, and has been discussed to some extent in the literature.[22] There are at least two motivations for its inclusion in this Thesis. First, signal power limitations tend to be so severe that the issue probably warrants consideration in most any treatment of fiber optic network design. Indeed, we will see that power division limitations can have fundamental bearing on our choice of frequency multiplexing versus spatial multiplexing. Second, we would like to expand on previous work in power division by investigating the potential impact of emerging technology, such as heterodyne detection and optical amplifiers.

In this Chapter, we shall examine the power division problem for some typical local network topologies, comparing the topologies on the basis of the maximum number of users that each can accomodate. We then investigate a number of potential methods for alleviating power division limitations.

We shall see that the severity of the power division problem is a strong function of network topology. The most common topologies for local networks are the ring, bus and star. Typically, the ring structure consists of a closed loop of point to point links. Such a configuration is an interesting case, for with point to point links (packet regeneration at each node), the ring suffers from no power division problems. Thus, the ring would seem to be an extremely attractive topology for use in fiber networks, considering the burdensome power limitations typically involved with optics technology. However, the ring is not a universal solution to the local network problem. Recall our initial motivation in Chapter 1; we seek useful methods of exploiting fiber bandwidth. However, in a ring structure with (non-optical) regeneration at each node, the usable bandwidth is necessarily restricted to the postdetection bandwidth of the receivers in the ring. Even with state of the art technology, this bandwidth is at least two orders of magnitude lower than that available from the fiber. Perhaps a more significant disadvantage of the ring topology is that, if a subset of users wishes to operate at very high data rates, all others in the network must be able to operate at these data rates, i.e., a ring with packet regeneration at each node requires identical interface bandwidth capabilities for all users. This requirement may well be unattractive economically for users interested only in interactive keyboard traffic or other low bit rate applications. On the other hand, buses and stars are much better suited to a network whose interfaces utilize widely varied amounts of bandwidth, for these topologies do not require packet regeneration. Thus, we have good reason for studying these structures in depth.

It is worth noting that the above argument against the ring assumes that packet regeneration occurs at each node. It is of course possible to make a ring network by instead using waveguide taps at each node, in which case the bandwidth problems mentioned above do not apply. Such a topology really amounts to a bus with the ends connected, and as such faces essentially the same power division problems as the bus. Thus, for purposes of this Chapter, we will not consider this type of ring as a separate case.

The Bus Topology

We wish to determine the maximum number of users that a bus topology can support. This specific problem has been treated in the literature, at least for direct detection. An analysis will be performed here for both direct and heterodyne detection, and the results will be compared. We recalculate the direct detection case not only for the convenience of the reader, but also to ensure that any comparisons between direct and heterodyne detection are made on the basis of equal assumptions and parameter values. In addition, we will need the specific results of both analyses in the sequel.

Consider the bus topology of Figure 3.1a. Assume that the bus is unidirectional, carrying information from left to right. (Obviously, a true network would need two such buses, one for each direction. The power division issue would be the same for each bus, so we consider only one.) Assume initially that the coupling constants K are fixed and are identical for each coupler. For simplicity, we neglect loss in the fiber between the users.

Figure 3.1b shows an individual waveguide coupler.[26]

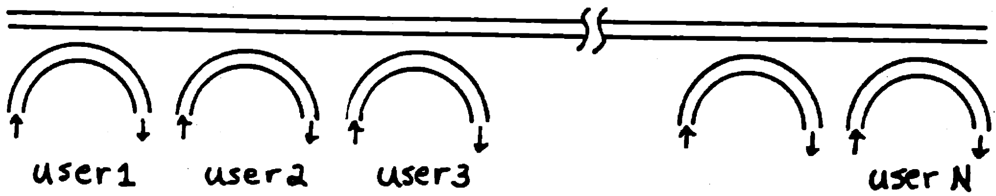


Figure 3.1a

The bus topology. Propagation is assumed to be unidirectional, from left to right.

Figure 3.1b

An individual waveguide coupler. A user injects signal power into port b, and receives power at port b'.

Signal power from upstream users enters the coupler at port a, and a fraction K of that signal power is coupled off the main waveguide, into the user's detector at port b'. To transmit information, the user injects his signal into port b, and a fraction K of that power is coupled into the main channel, out port a', to users downstream. Note that the coefficient for coupling power onto the main channel is the same as for coupling power off, a significant fact in network applications as we shall see.

Any real waveguide coupler has an excess loss associated with it, attributable to factors such as loss at the fiber connections and imperfections in the walls of the coupler channels. Assuming equal loss in both channels of the coupler, we have the following equations relating the signal powers at the various ports:*

$$P_{a'} = KP_b + (1-K)P_a \, 10^{-\alpha/10} \qquad (3.1)$$

$$P_{b'} = (1-K)P_b + KP_a \, 10^{-\alpha/10} \qquad (3.2)$$

where α is the excess loss of the coupler in dB. To determine the maximum number of users N that the bus can support, let us examine the worst case, in which information from user 1 must travel the entire length of the

*These equations assume sufficient orthogonality of the modes entering the coupler at ports a and b.

bus to user N. If P_s is the signal power entering port b of coupler 1 and P_r is the received signal power at port b' of coupler N, it is easy to show that

$$\frac{P_r}{P_s} = K^2(1-K)^{N-2}\, 10^{-N\alpha/10} \qquad (3.3)$$

For K, we choose the value that maximizes the received signal power P_r in this worst case. That value is K = 2/N, and (3.3) becomes

$$\frac{P_r}{P_s} = \frac{4}{N^2}\left[\frac{N-2}{N}\right]^{N-2} 10^{-N\alpha/10} \qquad (3.4a)$$

For large N, (3.4a) can be approximated by

$$\frac{P_r}{P_s} \approx \frac{4}{N^2}\, e^{-2}\, 10^{-N\alpha/10} \qquad (3.4b)$$

We see from (3.4b) that the ratio of received power to transmitted power decays exponentially with N, and will be dominated by that dependence. Such a dependence is not favorable if we are interested in accomodating a large number of users in the network.

To calculate N, we need values for P_s and P_r. In this and all subsequent analyses we assume -10 dBm for the optical source power P_s. This is a rather conservative figure.[17] Interestingly, however, the maximum value of N has a relatively weak dependence on P_s. From Eq. (3.4b), it can be seen that N will be approximately proportional to $\log_{10}(P_s/P_r)$. Two important observations can be made. First, the numerical values that we calculate for N, and the resulting conclusions, are not strongly coupled to the value that we use for P_s. Second, and perhaps more crucial, we will not be able to increase the maximum value of N significantly for the bus by merely increasing the source power.

We take P_r to be the minimum signal power required at the receiver to achieve a 10^{-9} error probability. The numerical value of P_r will depend on the signal set, signal duration T, detection method and various detector parameters. Parameter values used here are $k\theta = 4.1\times10^{-21}$, where k is Boltzmann's constant and θ the absolute temperature, wavelength $\lambda = 1.3\ \mu$, receiver capacitance C = 5 pF, detector load resistor R such that RC = .25T, detector quantum efficiency $\eta = 1$, and APD gain excess noise factor x = .5. For direct detection, the APD gain $\langle G\rangle$ is chosen to be optimum in all calculations. For heterodyne detection, we assume sufficient local oscillator (LO) power to achieve quantum limited operation. Effects such as intersymbol interference, finite laser linewidth and excess LO noise are ignored here. The resulting P_r values are listed in Table 3.1.

The maximum $N \equiv N_{max}$ may now be calculated from (3.4).

Detection Method	Binary Signal Set, Demodulation	P_rT (joules)
Direct/Baseband	ASK Incoherent	1.9×10^{-16}
Direct/Subcarrier	ASK Incoherent	5.6×10^{-16}
Heterodyne	ASK Incoherent	2.0×10^{-17}
	FSK Incoherent	1.0×10^{-17}
	PSK Incoherent	5.0×10^{-18}

Table 3.1 Receiver Sensitivity

Numbers used:

$k\theta = 4.1\times10^{-21}$ $x = .5$

$RC = .25\ T$ $h\nu_o = 2.3\times10^{-19}$ Joules

$C = 5\ pF$ $P_s = -10$ dBm

$\langle G\rangle$ = optimum value for all direct detection calculations

Figure 3.2 shows N_{max} versus signal duration for heterodyne detection and direct detection/baseband, assuming an excess coupler loss $\alpha = 1$ dB.

Several comments can be made about Figure 3.2. The values for N_{max} are in general quite low, considering the number of users that one might typically want to connect with a local network. Clearly, the bus structure represents a severely power-limited environment. In such an environment, it is natural to seek modulation and detection methods that offered high receiver sensitivity. Table 3.1 reflects the well-known fact that heterodyne detection yields over 10 dB greater receiver sensitivity than direct detection, and even more if the relatively exotic FSK and PSK are used.[43] But Figure 3.2 illustrates that, when we translate this increased receiver sensitivity into values for N_{max}, the increase in N_{max}, though significant when compared to direct detection, is still not sufficient to make the bus a particularly attractive network topology. It may also be concluded from Figure 3.2 that, at least for the case of the bus topology, the benefits of fancier signal sets such as FSK or PSK certainly would not outweigh the increase in receiver complexity inherent in their use. (Simplicity and economy in receiver structure are important in a local network, since there are a total of N receivers.)

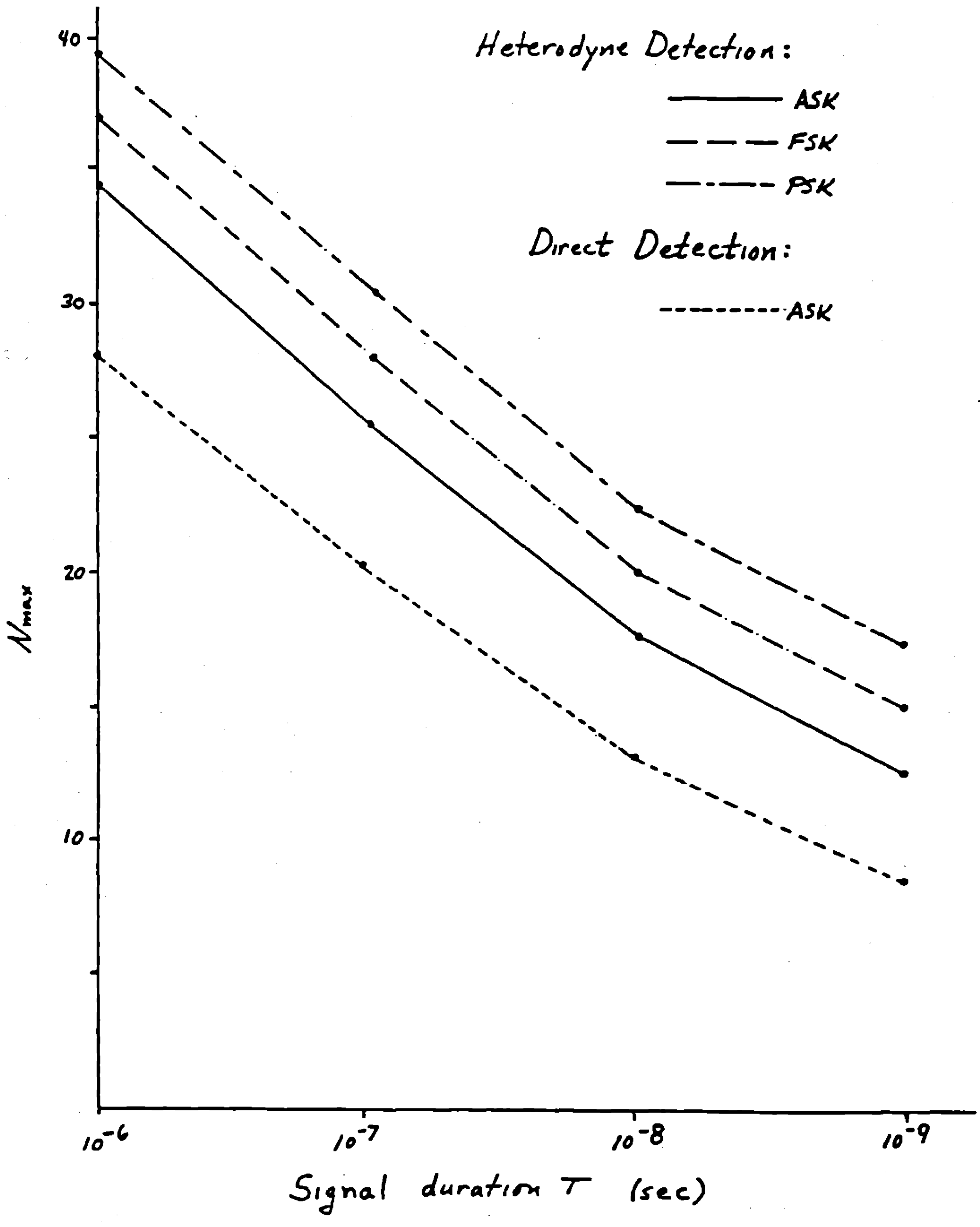


Figure 3.2. Maximum N vs. Signal Duration for Bus Topology, using parameters of Table 3.1. All waveguide couplers were assumed to have identical coupling coefficients and excess loss α = 1 dB.

A number of factors are responsible for the low N_{max} values of Figure 3.2. First, recall from Eq. (3.1) that only a fraction K of the available source power P_s enters the network for transmission to other users. For example, with N = 100 the optimum K is .02, and so 98% of the total source power injected into port b leaves port b' and is wasted, i.e., the signal power available to the rest of the network is only $.02P_s$.

Recall that our bus analysis assumed identical K's for all users. One might ask if N_{max} can be increased by <u>varying</u> the K's along the bus in some manner. This issue is addressed in Appendix A, where we show that varying K along the bus will not yield significant improvement in N_{max}.

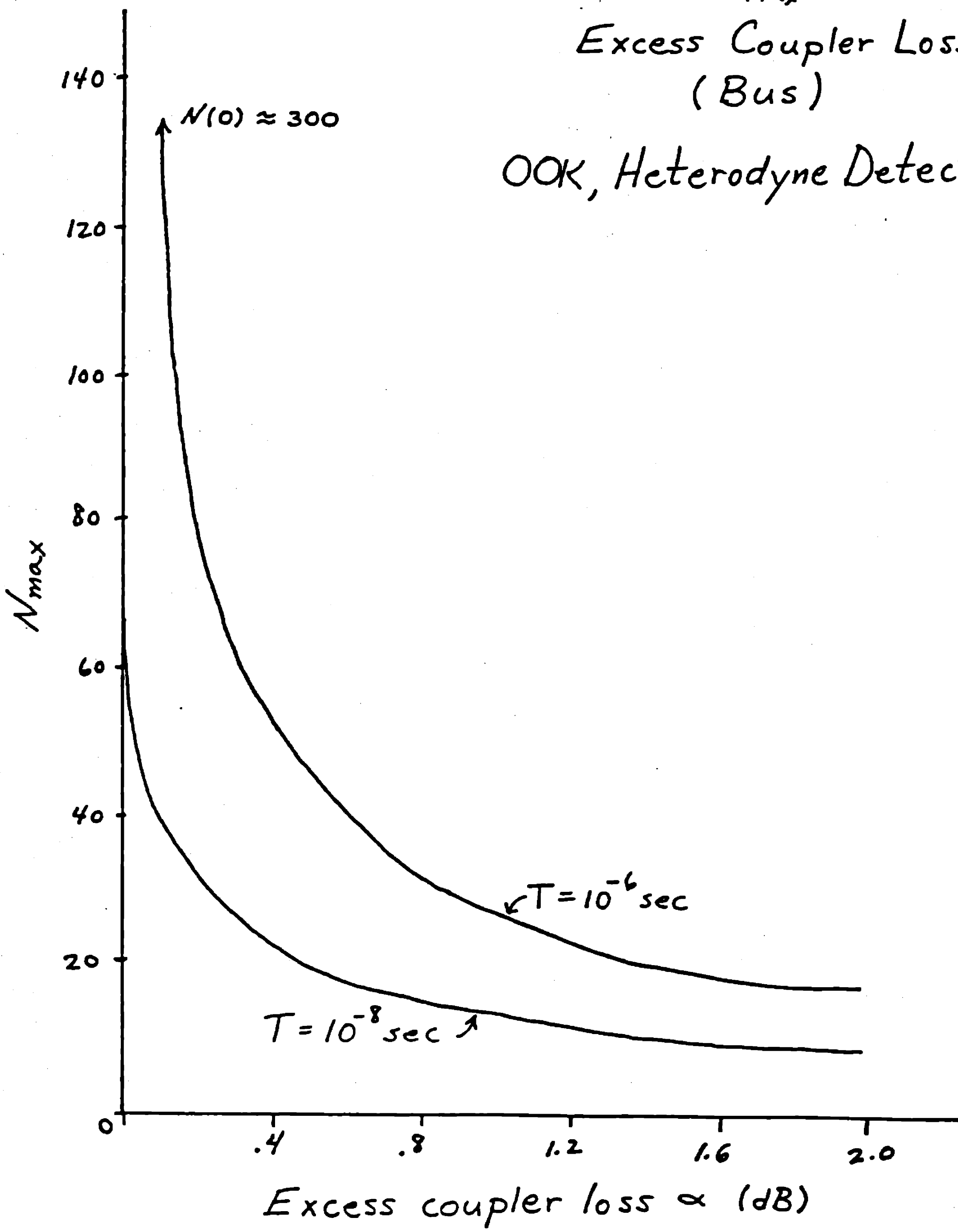
By far the most dominating effect is the end to end excess loss $N\alpha$ dB. We first noticed this dependence in Eq. (3.4b). If α is around 1 dB, a typical number for off the shelf components, it is obvious that the end to end excess loss for N = 100 will be prohibitive by at least 60 dB. Figure 3.3 illustrates the rather dramatic dependence of N_{max} on α. Clearly, we cannot afford to have α greater than about .2 dB if we desire an N over 100.

This temporarily concludes our discussion of power division problems in the bus topology. We will return to this topic later in this Chapter, but first, let us examine the second topology of interest, the star network.

Figure 3.3

N_{max} vs.
Excess Coupler Loss
(Bus)

OOK, Heterodyne Detection



The Star Topology

Consider the star topology in Figure 3.4a. Each user is connected to the central star coupler by a pair of unidirectional links. Signal power transmitted by individual users enters the star coupler on the "inbound" links. The optimum distribution of power among the outgoing links is an even distribution, for this even distribution makes the worst case received power P_r the least severe. (One could imagine an active star coupler that might be capable of reading packet addresses and concentrating the power on the appropriate link. The present analysis will focus on passive couplers only.) Physically, there are a number of coupler designs that (ideally) can accomplish even power distribution.[19] One type of coupler that is popular in larger star networks is shown in Figure 3.4b. In this device, the signal power from any user is divided up by an array of 50/50 waveguide couplers. Smaller arrays can be combined into larger ones, to connect up to 2^m users, where m can be as large as necessary.

At first, the star topology may seem too cumbersome for practical network application. Clearly, connecting a significant number of users in a star configuration will require an enormous amount of fiber, compared to that required in the linear bus topology of Figure 3.1a. However, the star handles signal power in a much more ef-

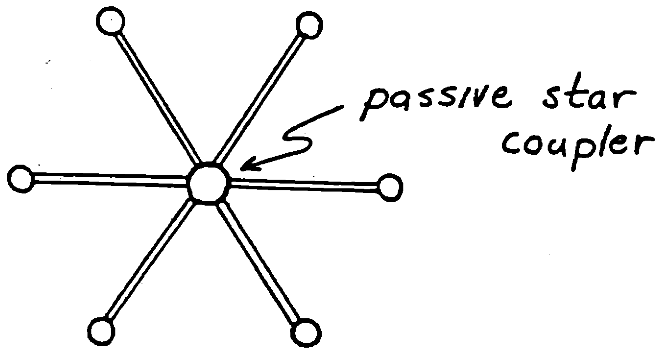


Figure 3.4a. Star Topology

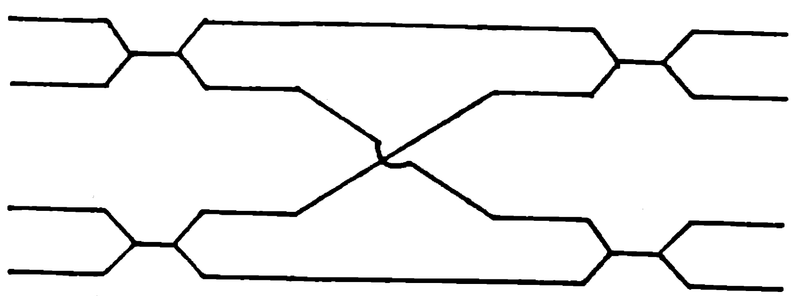


Figure 3.4b. A 4x4 star coupler that uses 4 50/50 waveguide couplers.

ficient manner than does the bus, an efficiency that enables the star to support a far greater number of users. For instance, with a star the source power is injected directly into the fiber rather than through a low-coefficient waveguide coupler at the source, as was the case with the bus. Signal power is then divided naturally in an equal manner among the users. The most important difference, however, is that the star is much less sensitive to excess loss of the waveguide couplers. In the bus, lossy waveguide couplers caused a worst case excess loss of $N\alpha$ dB, and this factor was the major limitation on N_{max}. On the other hand, to connect N users together with a star coupler of the type in Figure 3.4b, the signal need propagate through only $\log_2 N$ couplers. Thus, with an excess loss of α dB per coupler, the excess loss between users is $\alpha \log_2 N$. For large N, the difference between αN and $\alpha \log_2 N$ is quite significant.

If the available source power is P_s and the power received by each user is P_r, it follows from the previous discussion that

$$\frac{P_r}{P_s} = \frac{1}{N} 10^{-(\alpha/10)\log_2 N} \qquad (3.8)$$

Solving for N,

$$N = \left[\frac{P_s}{P_r}\right]^{\frac{1}{1 + (\alpha/10)\log_2 10}} \qquad (3.9)$$

A comparison of Equations (3.4b) and (3.8) support the observations of the previous paragraph. For large N, the attenuation in the star will be significantly less than for the bus. Both have an exponential dependence, but the bus attenuation varies exponentially with N, whereas that for the star varies only as N to a power.

Values for N_{max} were calculated, using P_s = -10 dBm, α = 1 dB and the P_r values of Table 3.1. The results are displayed in Figure 3.5. It is clear that, even with direct detection, power division limitations for the star are negligible compared to those for the bus. Indeed, before we run out of signal power in the star, we will most likely run into other limitations. For instance, this analysis assumed perfect 50/50 splitting of power by each element in the star coupler. In practice, we would expect small deviations from 50/50 at various stages within the coupler array. This effect would tend to lower N_{max} slightly. Nevertheless, it is evident that, for typical network populations, the star is an effective way to avoid power division limitations.

Power Division and Concurrency

Thus far in this Chapter, we have taken an extensive look at the power division problem for two of the most popular network topologies. It is evident that this issue

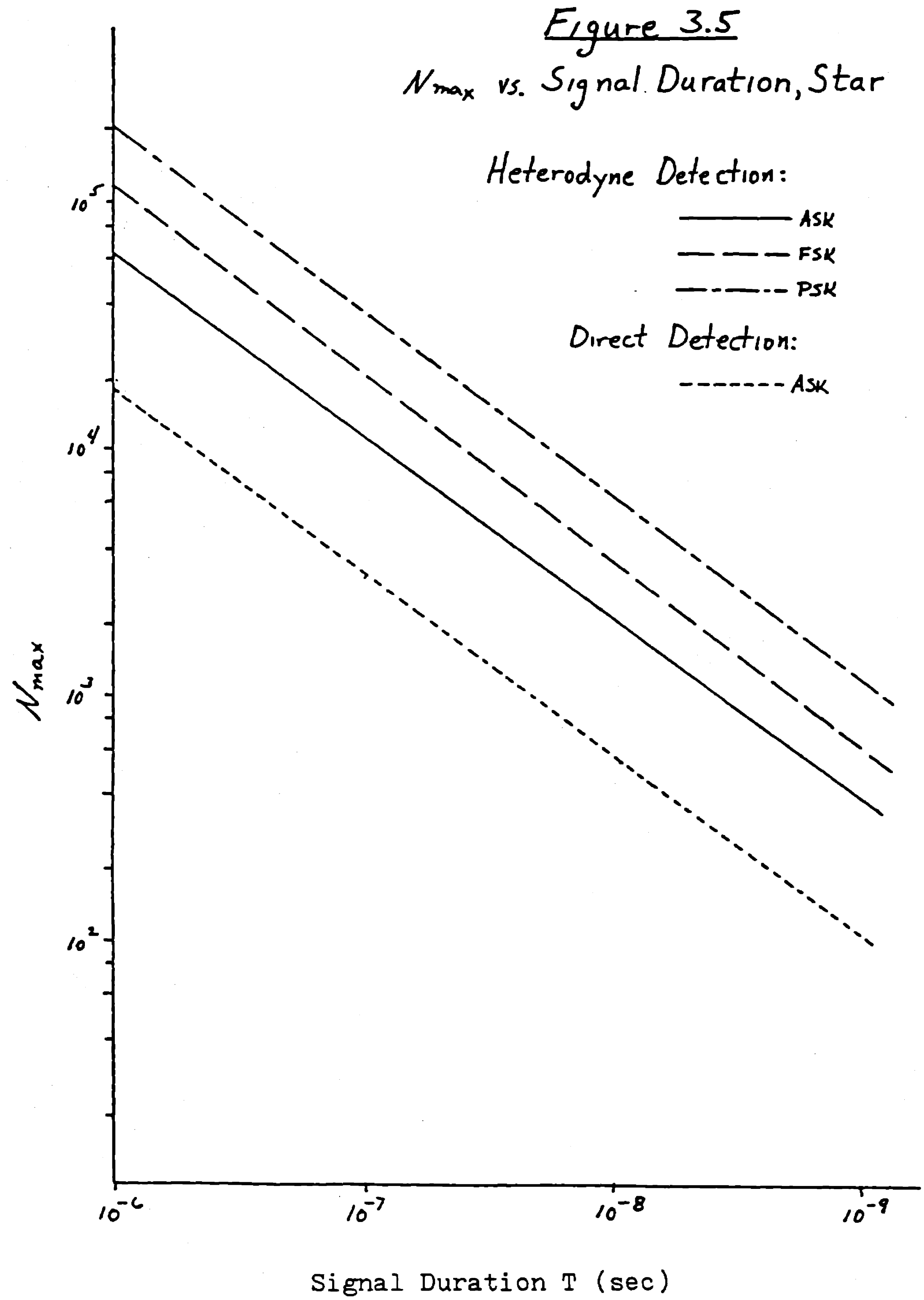


Figure 3.5
Nmax vs. Signal Duration, Star
Heterodyne Detection:
ASK
FSK
PSK
Direct Detection:
ASK
10^5
10^4
10^3
10^2
Nmax
10^-6
10^-7
10^-8
10^-9
Signal Duration T (sec)

by itself is an important component of fiber network design. It is perhaps not as clear how power division considerations relate to concurrency and multiplexing, the main topics of this Thesis.

The relationship between power division and concurrency is simple but important. Recall once again Figures 3.2 and 3.5, where N_{max} has been plotted versus the signal duration T for the bus and star, respectively. Note that the power division limitation on N_{max} becomes decidedly more pronounced as the signal duration decreases. This dependence results from the well-known fact that the error rate is a (monotonic increasing) function of the energy per bit $P_r T$, and not of the signal power P_r alone. The severity of power division limitations at smaller T has an influence on our choice of frequency multiplexing versus spatial multiplexing as a means of achieving con-concurrency, at least in the case where the peak source power is constrained. To obtain a significant level of concurrency via spatial multiplexing requires signal durations small enough such that $T_p \ll \Upsilon$. With $\Upsilon = 10$ μsec, for instance, a signal duration $T \ll 10^{-8}$ sec is necessary, assuming a 1000-bit packet. Based on Figures 3.2 and 3.5, it is easy to conclude that spatial multiplexing may lose favor in a severely power-limited environment.

On the other hand, frequency multiplexing can actually assist in alleviating power division problems. A fundamental fact implicit in Eq. (2.1) is that, since the packet preparation time is so large ($P \approx 10$ msec), the bit rate, and thus the packet duration T_p, may be varied over orders of magnitude without affecting the overall delay D significantly. In other words, we may increase the signal duration T (and thus ease power division constraints on N_{max} somewhat) without increasing the overall delay, as long as T_p remains less than P. Of course, when we increase T_p we also increase the network usage parameter $\rho \approx \lambda T_p$. If the packets become so long that $\rho \geq 1$, and if the network does not allow concurrency, then the network will become unstable with Q (and D) going to infinity. However, if we allow concurrency through frequency multiplexing, then we could maintain stability within the network at the higher T_p values. Furthermore, since with larger signal durations the bandwidth requirements of each user are comparatively low, we should have little problem in achieving an adequate level of concurrency through frequency multiplexing, i.e., the network should be able to handle even heavy loads.

In fairness, however, two comments should be made concerning these observations. First, the above argument against spatial multiplexing applies only when the peak source power is constrained. If we decrease T, we could (at least for heterodyne detection[19]) recover the original N_{max} by increasing P_s accordingly. Thus we could go to very small T if sufficiently large peak source power were available.

Second, the above argument in favor of frequency multiplexing exploits the fact that increasing T has the same effect as increasing P_s. But we have noted previously that increasing P_s will not reduce significantly the effects of lossy waveguide couplers, particularly for the bus. Thus we cannot claim that frequency multiplexing alone is an adequate solution to the power division problem, though it is true that sufficient increase in T can mitigate power limitations somewhat.

A Second Look at the Bus Topology

Figures 3.2 and 3.5 indicate that the star is vastly preferable to the bus in local network applications, at least as far as power division limitations (i.e. limitations on N_{max}) are concerned. The disadvantages of the bus are unfortunate, since its linear structure seems inherently better suited for network use, being more adaptable and requiring much less fiber than the star. There are also a number of applications for which the linear topology is better suited. (Most current spatial multiplexing schemes, such as D-Net,[9] use a bus structure.) If there were some device or method that could yield substantial increases in N_{max} for the bus, it would clearly have great significance, allowing use of the bus without paying the price of power division limitations. We have already seen that increased receiver sensitivity obtained via heterodyne detection, FSK and PSK did not improve N_{max} appreciably for the bus, but a number of other approaches are possible. Thus let us return to the bus topology in a further attempt to increase N_{max}.

Use of Coding. Severe signal power limitations at the receiver are rather common in communications. Attempts to improve error rate for a given signal level often involve use of error correcting codes.[30] The basic idea is to allow an increased uncoded bit error rate, thus relaxing signal power requirements at the receiver, then correct the errors through use of additional check bits transmitted along with the data (i.e. a parity check code). In the case of local networks, the relaxed signal power requirements should result in higher values for N_{max}. Figure 3.6 illustrates the dependence of N_{max} on uncoded bit error rate for the bus of Figure 3.1a, for two different values of excess coupler loss α. The resulting improvement in N_{max} is negligible for $\alpha = 1$ dB, even when a low rate code is used. Even at $\alpha = .1$ dB, the increase in N_{max} is probably not worth the additional complexity and

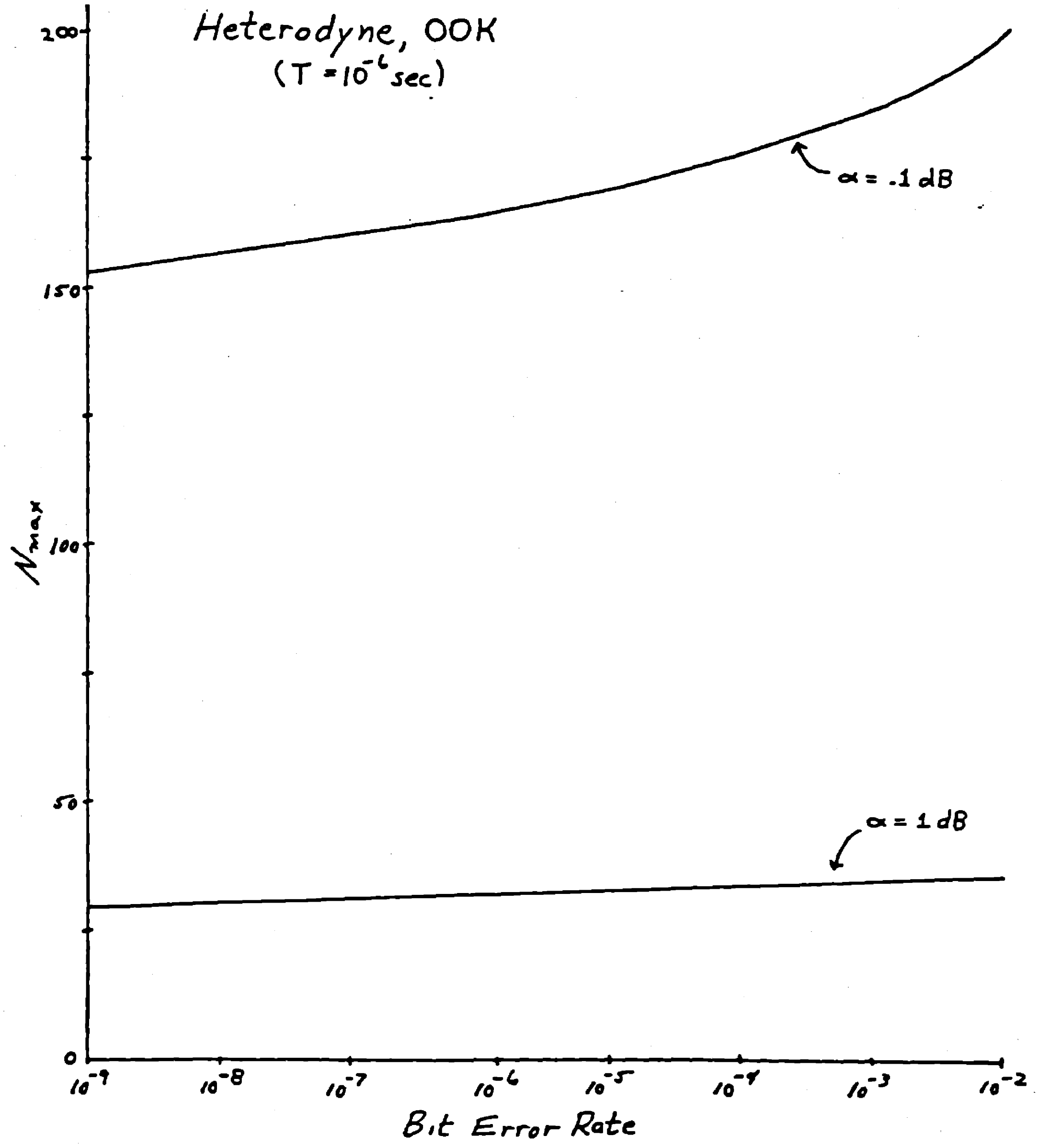
Figure 3.6

Effect of Coding - Bus

N_{max} vs. Bit Error Rate

Heterodyne, OOK

($T = 10^{-6}$ sec)



delay inherent to the use of large numbers of parity bits.

Use of Squeezed States. Another potential method for increasing N_{max} involves improving receiver sensitivity through use of so-called squeezed states or two-photon coherent states (TCS).[31] Yuen and Shapiro[32-34] have shown that the novel noise reduction properties of TCS radiation can be exploited via heterodyne/homodyne detection to achieve improved SNR, and thus lower signal power requirements for a given error probability. However, it has also been shown that TCS radiation rapidly loses its quadrature noise asymmetry when propagating through a lossy channel. Thus, in our bus application, it is doubtful that using TCS to transmit signal energy will improve N_{max} to a significant degree. Shapiro has shown[25] that a waveguide coupler requiring infinitesimal insertion loss ($K \approx 0$) is possible by injecting TCS into port b of the receiving tap. This approach has perhaps greater promise for increasing N_{max}, but recall that the dominant limitation on N_{max} is the excess loss of the taps, not the insertion loss (at least for α greater than a few tenths of a dB). Moreover, it should be noted that TCS radiation has yet to be observed experimentally, and its use in communications is still quite speculative.

Use of Optical Amplifiers. We have investigated a number of potential methods for increasing N_{max} for the bus. None of these methods have provided substantial benefit. The reason for the failure should by now be clear. All the techniques discussed to this point (heterodyne detection, use of FSK and PSK, coding and TCS) attempt to increase N_{max} by improving receiver sensitivity. The real problem with the bus, however, is the excess loss incurred in propagating through the channel. It is impossible to build an optical receiver with sufficient sensitivity to overcome the attenuation on the bus when N is significantly large (50 or more). In order to find an effective solution to the power division problem for the bus, we must deal more directly with the channel excess loss.

One method that accomplishes this task involves the use of optical amplifiers. The basic approach is illustrated in Figure 3.7a. We position amplifiers at regular intervals along the bus, with each amplifier having sufficient gain to offset the attenuation (insertion loss + excess loss + fiber loss) suffered by the signal since the previous amplifier. Thus, in Figure 3.7b, $G_a = L^{-1}$. By periodically restoring the signal power, we hope to be able to extend the bus to additional users, thus increasing N_{max}. Unfortunately, besides boosting the signal power, each amplifier also adds its own noise (and amplifies noise

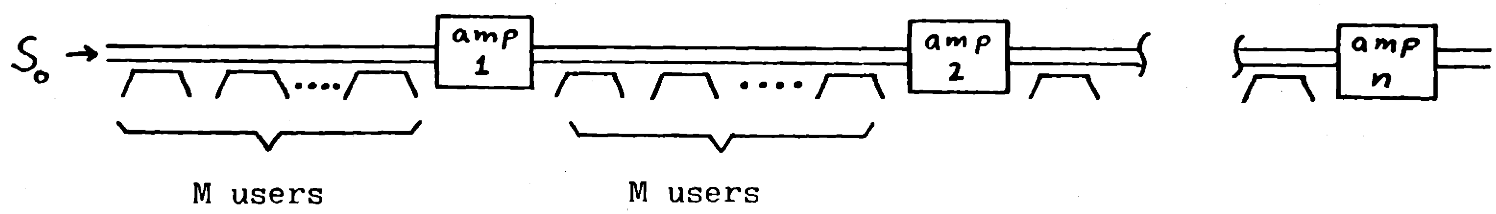


Figure 3.7a

Bus employing optical amplifiers.

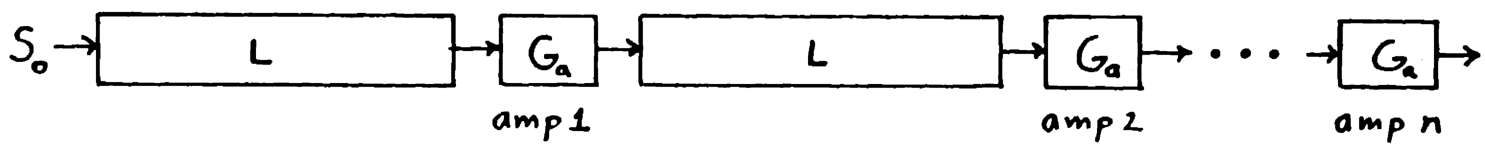


Figure 3.7b

Equivalent model showing gains
and losses along bus.

from previous amplifiers), so we expect that there will be some maximum number of amplifiers that can be used.

At this point one might wonder why we do not simply use regenerative repeaters, for with such devices the bus could be extended indefinitely. In this treatment we disregard (non-optical) repeaters for the same reason that we avoided the ring topology at the beginning of this Chapter. If the data must be regenerated, the usable bandwidth is limited to the bandwidth of the post-detection electronics, which is orders of magnitude less than the available channel bandwidth of single mode fibers.

The optical amplifier[35] consists of an atomic medium whose population is inverted. The signal beam makes a single pass through the medium. The amplifier model that we shall use is illustrated in Figure 3.8. Here $\underline{p}(t)$ represents the signal (and possibly noise) that physically enters the medium for amplification. The noise from the amplifier itself is modeled at the input, as shown, by a noise process $\underline{w}(t)$ that we assume to be zero mean Gaussian and (for the moment) white over the frequency bands of interest. The spectral height of $\underline{w}(t)$ is assumed to be $h\nu_o$ watts/Hz, which is valid for[36] $G_a \gg 1$.

As illustrated in Figure 3.7, the overall structure consists of a number of sub-buses, each accomodating M users. A signal propagating the entire length of a sub-

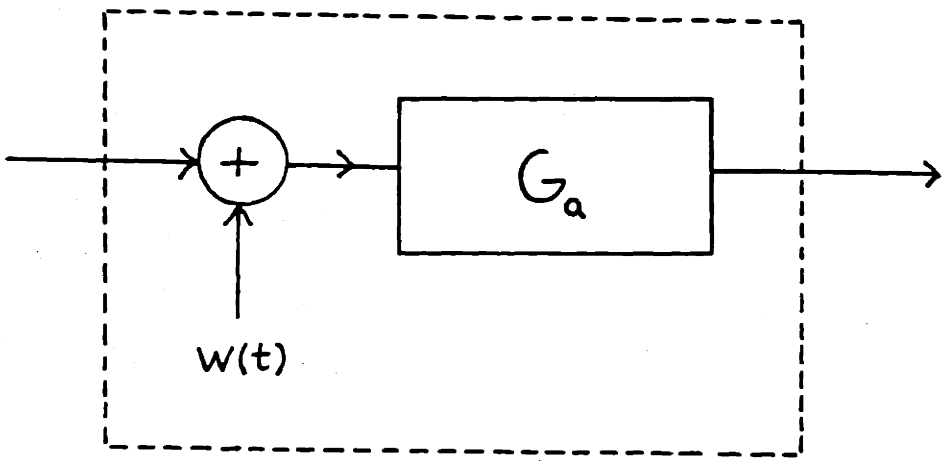


Figure 3.8

Model for Optical Amplifier

Power gain G_a; w(t) white Gaussian noise process of spectral height $h\nu_0$ (assuming $G_a >> 1$).

bus, from one amplifier to the next, experiences a loss

$$L = (1 - K)^M \, 10^{-\alpha M/10}$$

$$= (1 - \tfrac{2}{M})^M \, 10^{-\alpha M/10} \qquad (3.10)$$

assuming identical couplers. Clearly, M must be small enough that the error probability at the end of each sub-bus remains sufficiently low. The maximum value of M is thus given by the analysis performed earlier in this Chapter (Eq. (3.4) or Figure 3.2 assuming $\alpha = 1$ dB).

Consider Figure 3.7b, an equivalent model for the overall bus structure. If a signal power S_o (and no noise) is launched from the left end of the bus, then just to the left of the n^{th} amplifier we will have on the bus a signal power LS_o and a noise process of spectral height $(n-1)h\nu_o$. Thus as more sub-buses are joined together, the noise level grows on the downstream portions of the network. We would like to determine n_{max}, the maximum possible value of n (and thus the maximum value of $N \approx Mn$) subject to constraints on error probability and signal power similar to those used in generating Figure 3.2. We shall see that n_{max}, and thus N_{max} for the overall structure, will depend on the number of users M between each amplifier. Of course, n_{max} and N_{max} will also be a function of receiver sensitivity. We consider both heterodyne detection and direct detection, assuming on-off keying (OOK) in both cases.

Optical Amplifiers/Heterodyne Detection. Figure 3.9a illustrates a model for quantum limited heterodyne detection.[28,43] The noise process n(t) is zero-mean white Gaussian of spectral height $h\nu_o/4\eta$. In our problem we have

$$s_m(t) = \begin{cases} P_r^{\frac{1}{2}} \cos \omega_{IF}t + v(t) & m = 1 \\ v(t) & m = 0 \end{cases}$$

over the signal duration $0 \leq t \leq T$, where v(t) is the noise due to the amplifiers on the bus. Clearly the worst case SNR's will be at the right end of each sub-bus, so we focus on the users at those positions. Just to the left of amplifier n:

$$P_r = K^2(1 - K)^M \, 10^{-\alpha M/10} \, P_s \tag{3.11}$$

$$S_v(f) = \frac{1}{2} K \, h\nu_o(n - 1) = \frac{1}{M}(n - 1)h\nu_o \tag{3.12}$$

where P_s is the laser source power. Figure 3.9b shows an equivalent heterodyne model that takes into account both n(t) and v(t).

Figure 3.9c illustrates the optimum receiver. The decision rule consists of comparing y to an optimum threshold value:

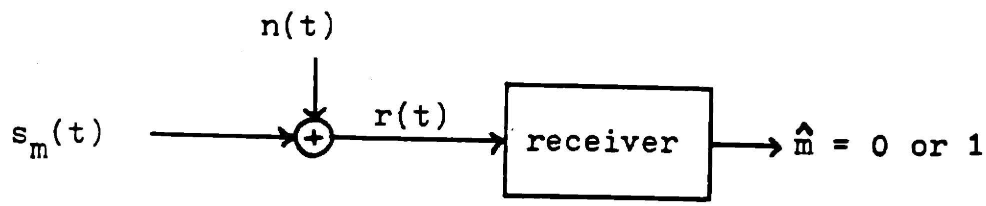


Figure 3.9a

Model for quantum-limited heterodyne detection. The noise process n(t) is white with spectral height $h\nu_o/4\eta$, while $s_m(t)$ is an IF signal with additive noise.

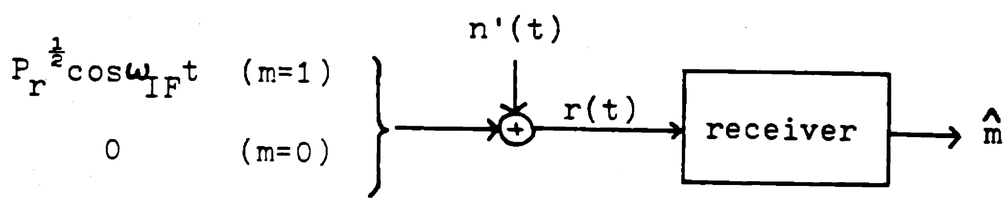


Figure 3.9b

Equivalent model in which all noise has been combined into n'(t), which is white with spectral height $(h\nu_o/4\eta)(1+4\eta(n-1)/M)$.

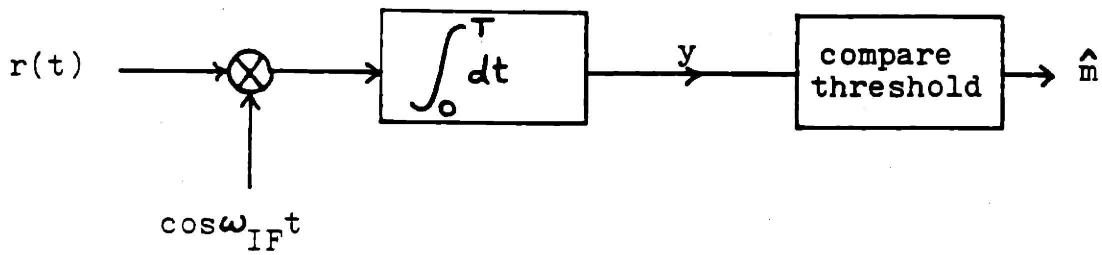


Figure 3.9c

Optimum receiver used in analysis for heterodyne detection.

$$y \underset{\hat{m}=0}{\overset{\hat{m}=1}{\gtrless}} \frac{1}{4} T \cdot P_r^{\frac{1}{2}} \qquad (3.13)$$

assuming equally likely 0 and 1.

The resulting probability of error is[37]

$$\Pr(\text{error}) = Q((\eta P_r T/2h\nu_o')^{\frac{1}{2}}) \qquad (3.14)$$

where $Q(x) = \int_x^{\infty} (2\pi)^{-\frac{1}{2}} \exp(-y^2/2)\, dy$, and $\nu_o' = \nu_o(1+\frac{4\eta}{M}(n-1))$. For $\Pr(\text{error}) = 10^{-9}$ we set the argument of the Q function equal to 6. Solving for n,

$$n = \frac{1}{M}(1 - \frac{2}{M})^M \, 10^{-\alpha M/10} \frac{P_s T}{72h\nu_o} - \frac{M}{4\eta} + 1 \qquad (3.15)$$

$$= n_{max}$$

In Figure 3.10, $N_{max} = M \cdot n_{max}$ is plotted versus M for various signal durations, assuming coupler excess loss α = 1 dB and source power P_s = -10 dBm. It is evident that the use of optical amplifiers can increase N_{max} by one or two orders of magnitude, at least in the case of heterodyne detection.

<u>Optical Amplifiers/Direct Detection</u>. Let us now see if direct detection can yield similar impressive results. Binary OOK signaling will again be assumed.

Figure 3.11 illustrates our model for the direct detection process and receiver. The optical signal arriving at the photodetector is

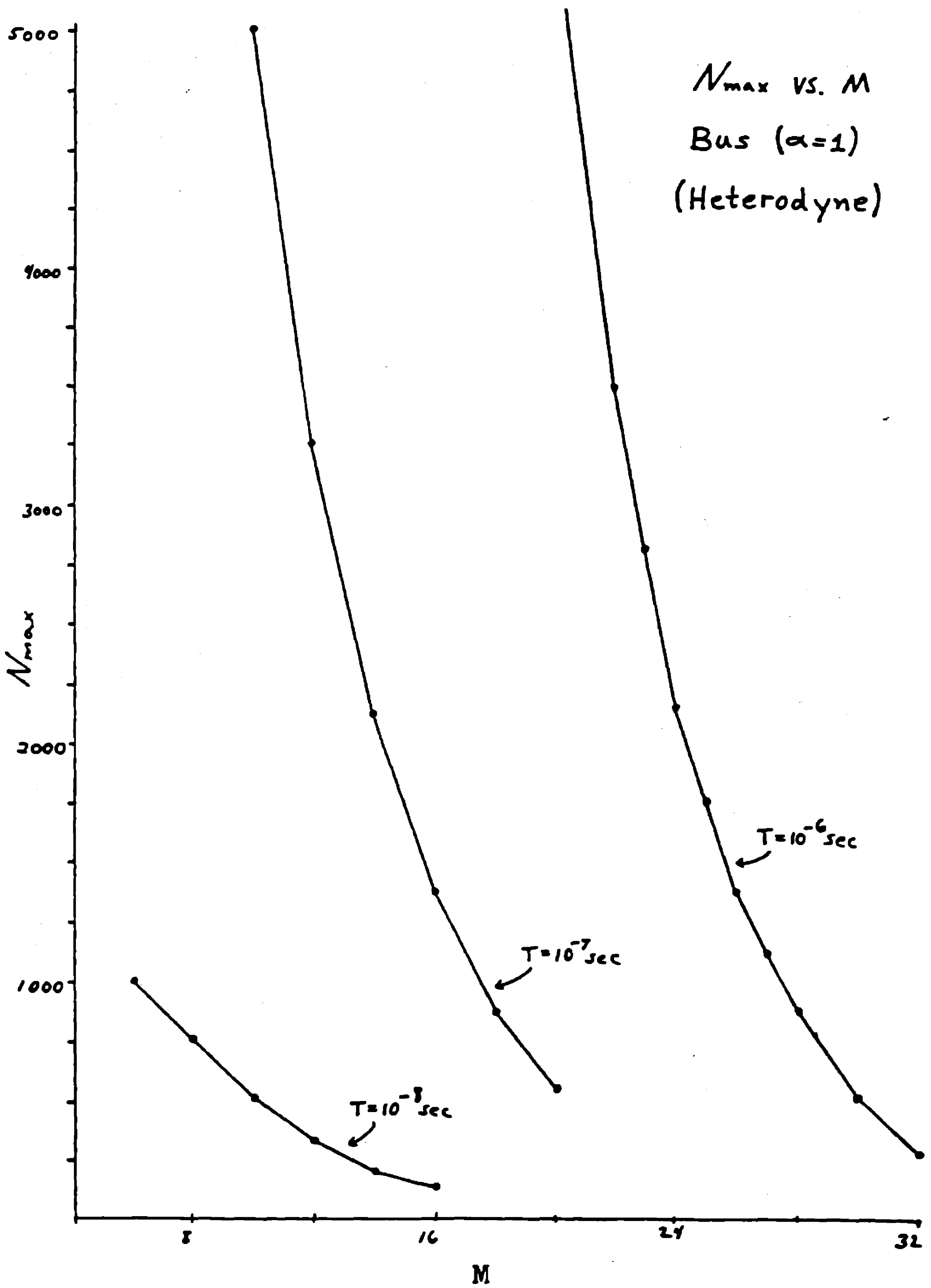


Figure 3.10. Improvement in maximum N afforded by optical amplifiers when heterodyne detection is used. Binary ASK signalling, coupler excess loss α = 1 dB and parameter values of Table 3.1 are assumed.

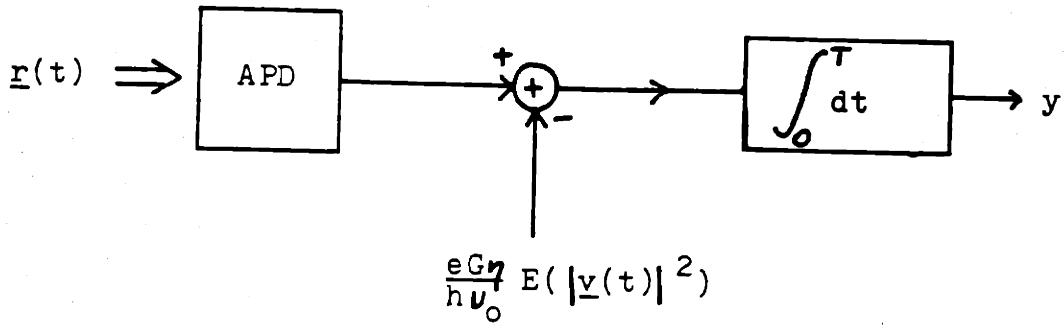


Figure 3.11

Model for direct detection and post-detection processing.

$$\underline{r}(t) = \underline{s}_m(t) + \underline{v}(t)$$

$$\underline{s}_m(t) = P_r^{\frac{1}{2}} \delta_{m1} \qquad 0 \leq t \leq T$$

where δ_{ij} is the Kronecker delta. We view the power $P(t) = |\underline{r}(t)|^2$ incident on the photodetector as having a known component (given m) plus a random component (due to the amplifier noise) that we include via a covariance $K_{pp}(t,u)$. Thus

$$E(P(t)|m) = |\underline{s}_m(t)|^2 + E(|\underline{v}(t)|^2) \qquad (3.16)$$

$$K_{pp}(t,u) = 4s_m(t)s_m(u)\, K_{vv}(t,u) + K_{vv}^2(t,u) \qquad (3.17)$$

where $v(t) = \mathrm{Re}(\underline{v}(t))$. Clearly v(t) has the same physical nature as it did in the case of heterodyne detection, but in the present direct detection problem we cannot model v(t) as white noise, since such noise has infinite mean square value, resulting in infinite average power incident on the photodetector. We thus model v(t) as having a flat spectrum over the range $|f| < W$:

$$S_v(f) = \begin{cases} \frac{1}{2}K(n-1)h\nu_o & |f| \leq W \\ 0 & |f| \geq W \end{cases} \qquad (3.18)$$

$$K_{vv}(\tau) = \frac{\sin 2\pi W\tau}{2\pi W\tau}\, K\, W\, h\, \nu_o(n-1) \qquad (3.19)$$

For the photocurrent we have, conditioned on knowledge of m,

$$E(i(t)|m) = \frac{eG\eta}{h\nu_o}\left(|\underline{s}_m(t)|^2 + E(|\underline{v}(t)|^2)\right) \quad (3.20)$$

$$K_{ii}(\tau) = \left[\frac{e^2G^{2+x}\eta}{h\nu_o}\left[|\underline{s}_m(t)|^2 + E(|\underline{v}(t)|^2)\right] + \frac{2k\theta}{R}\right]\delta(t) + \left(\frac{eG\eta}{h\nu_o}\right)^2 K_{pp}(\tau) \quad (3.21)$$

After the bias subtraction

$$E(i'(t)|m) = \frac{eG\eta}{h\nu_o} E(|\underline{s}_m(t)|^2) \quad (3.22)$$

$$K_{i'i'}(\tau) = K_{ii}(\tau) \quad (3.23)$$

The optimum receiver, assuming conditional Gaussian statistics, is the integrator shown in Figure 3.11. It follows that

$$E(y\,|\,m=1) = \frac{eG\eta}{h\nu_o} P_r T \equiv Y$$

$$E(y\,|\,m=0) = 0$$

$$\mathrm{Var}(y\,|\,m=1) \equiv \sigma_1^2 = \left(\frac{e^2G^{2+x}\eta}{h\nu_o}\left[P_r + E(|\underline{v}(t)|^2)\right] + \frac{2k\theta}{R}\right)T + \left(\frac{eG\eta}{h\nu_o}\right)^2\left[4P_r\int_0^T dt\int_0^T du K_{vv}(t,u) + \int_0^T dt\int_0^T du K_{vv}^2(t,u)\right]$$

$$\mathrm{Var}(y \mid m=0) \equiv \sigma_0^2 = \left(\frac{e^2G^{2+x}\eta}{h\nu_o} E(|\underline{v}(t)|^2) + \frac{2k\theta}{R}\right)T + \left(\frac{eG\eta}{h\nu_o}\right)^2 \int_0^T dt \int_0^T du\, K_{vv}^2(t,u)$$

with P_r as in Eq. (3.11).

For W large compared to T^{-1}, the integrals can be evaluated easily, yielding

$$\sigma_1^2 = \frac{Te^2G^2\eta}{h\nu_o}\left[P_r\left(G^x + \frac{4\eta(n-1)}{M}\right) + \frac{4h\nu_o W(n-1)}{M}\left(G^x + \frac{2\eta(n-1)}{M}\right)\right] + \frac{2k\theta T}{R} \qquad (3.24)$$

$$\sigma_0^2 = 4Te^2G^2\eta W(n-1)M^{-1}\left(G^x + \frac{2\eta(n-1)}{M}\right) + \frac{2k\theta T}{R} \qquad (3.25)$$

Optimum (i.e. minimum Pr(error)) processing of y is given by the Likelihood Ratio Test (LRT)[38]

$$\frac{p_{y|m}(y \mid m=1)}{p_{y|m}(y \mid m=0)} \underset{\hat{m}=0}{\overset{\hat{m}=1}{\gtrless}} 1 \qquad (3.26)$$

for equally likely 0 and 1. We make the standard Gaussian assumption for direct detection, i.e. y is Gaussian under both hypotheses.[28,43] After some manipulation, (3.26) reduces to the decision rule

$$\gamma_- < y < \gamma_+ \qquad \text{choose } \hat{m} = 0$$

$$y < \gamma_- \text{ or } y > \gamma_+ \qquad \text{choose } \hat{m} = 1$$

with

$$\gamma_+ = C^{\frac{1}{2}} - \frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2} Y \tag{3.27}$$

$$\gamma_- = -C^{\frac{1}{2}} - \frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2} Y \tag{3.28}$$

$$C = \left(\frac{Y\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\right)^2 + \frac{\sigma_0^2\sigma_1^2 \ln(\sigma_1^2/\sigma_0^2) + \sigma_0^2 Y^2}{\sigma_1^2 - \sigma_0^2} \tag{3.29}$$

The expressions for the thresholds are of course complicated by the fact that both the means and variances of y differ under the two hypotheses.

The probability of error is given by

$$2\Pr(\text{error}) = Q(\gamma_+/\sigma_0) + Q(-\gamma_-/\sigma_0) + Q((Y-\gamma_+)/\sigma_1) - Q((Y-\gamma_-)/\sigma_1) \tag{3.30}$$

The arguments of the Q functions contain many parameters. The parameter of greatest interest to us is n, the number of amplifiers. We wish to determine n_{max}, the maximum number of amplifiers for which $\Pr(\text{error}) \leq 10^{-9}$ can be achieved in the worst case. Assuming the same values for load resistor R, kθ, etc., as used in Table 3.1, and approximating the Q function by

$$Q(x) \approx (2\pi)^{-\frac{1}{2}} x^{-1} e^{-\frac{1}{2}x^2}$$

n_{max} was calculated for various values of M, T and noise bandwidth W. Figure 3.12 illustrates the resulting $N_{max} = Mn_{max}$ versus M.

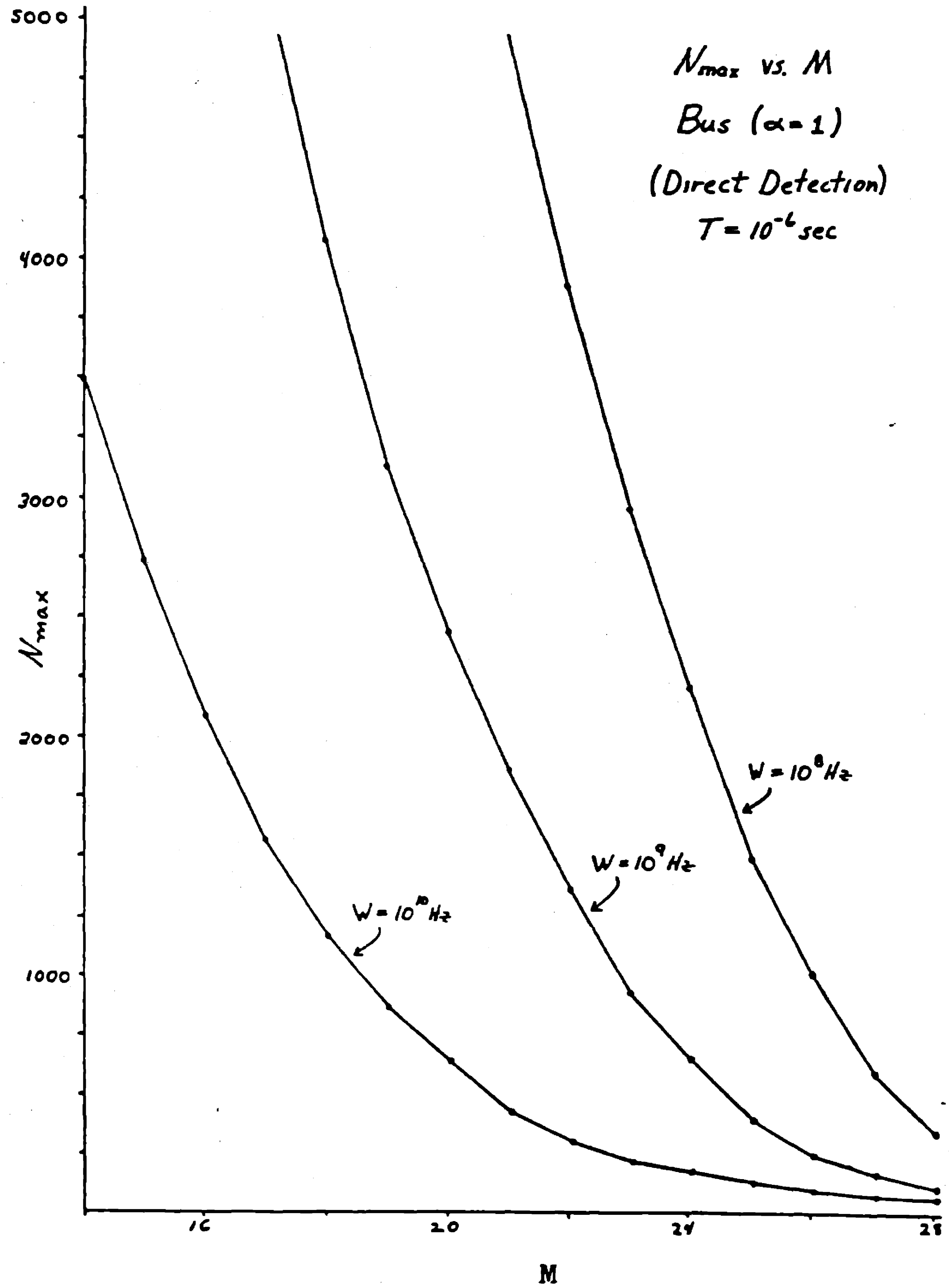


Figure 3.12a. Maximum N vs. M for bus using optical amplifiers, with direct detection and signal duration T = 1 μsec. W is the bandwidth of the received noise. Parameter values of Table 4.1 were assumed, along with an excess coupler loss α = 1 dB.

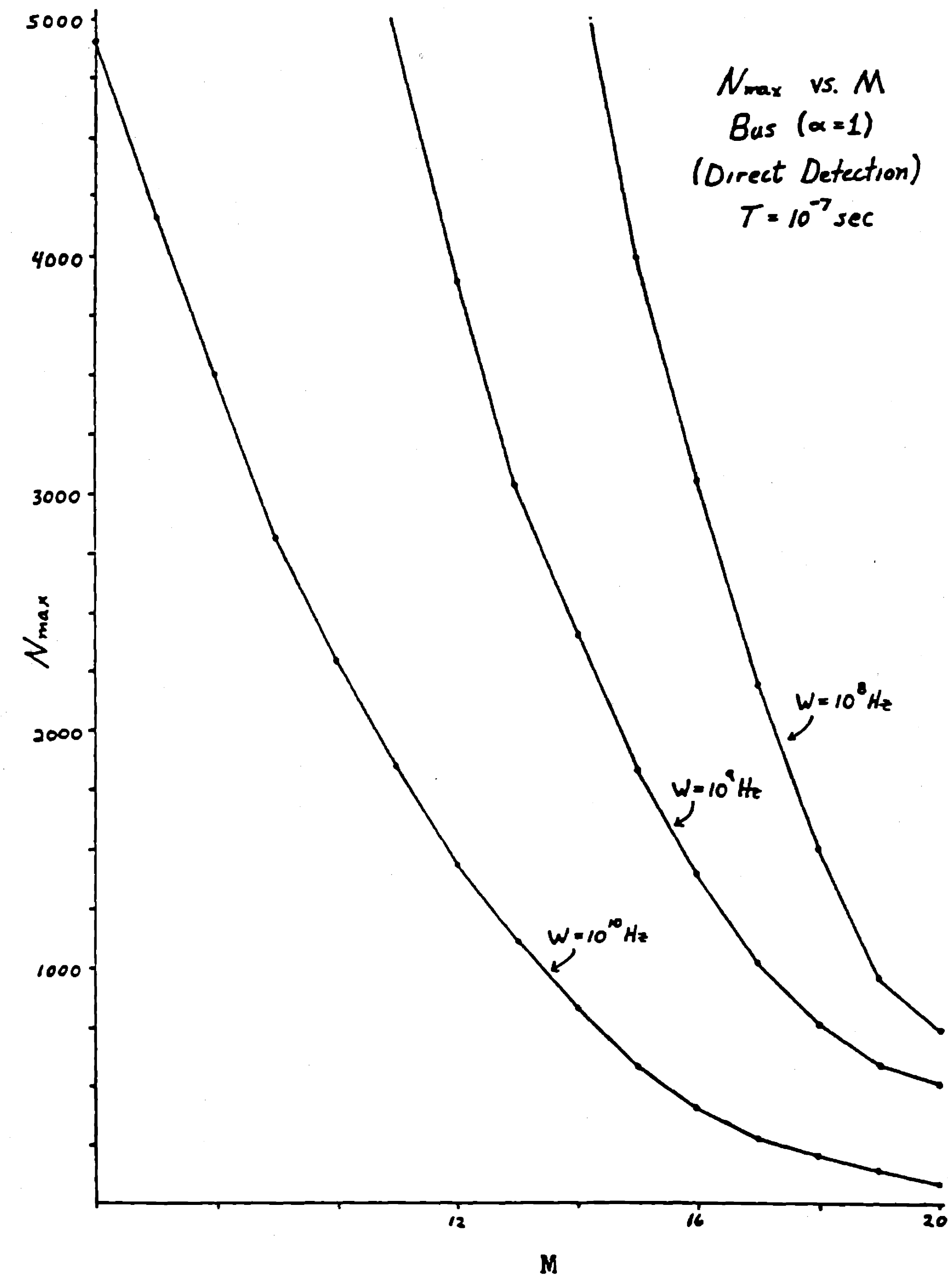


Figure 3.12b. Maximum N vs. M for bus using optical amplifiers, with direct detection and signal duration T = .1 μsec.

The preceding analysis was performed on the premise that the noise process v(t) was bandlimited. This is certainly true of any real noise process, but thus far we have avoided the question of what determines W in Eq.(3.18). W could be influenced by any of a number of factors. The original noise w(t) in Figure 3.7 will have some finite bandwidth. The amplifier gain curve will also have some finite linewidth. The photodetector has associated with it a predetection bandwidth. We could also conceivably place a narrowband optical filter in front of the photodetector, or use a frequency selective waveguide coupler, thereby limiting the bandwidth of the incident radiation. Thus the original amplifier noise w(t) is subject to many bandlimiting mechanisms before it reaches the post-detection electronics. The effective value of W will be essentially the smallest of the bandwidths mentioned above. Since it is difficult to assign a definite value to W, Figure 3.12 shows a family of curves, with W ranging over two orders of magnitude.

Figures 3.10 and 3.12 illustrate clearly that the use of optical amplifiers in the bus topology yields impressive increases in N_{max}, regardless of whether direct or heterodyne detection is used. In the case of direct detection, we see that the improvement in N_{max} is quite substantial even at $W = 10^{10}$ Hz, thus our conclusions are not strongly coupled to assumptions about the noise bandwidth.

A number of other observations can be made about these graphs. The curves indicate that N_{max} is increased by decreasing the number of users M between amplifiers. This result is perhaps surprising, since a decrease in M implies an increase in the number of amplifiers n, and as n increases so does the amplifier noise level at the end of the bus (see Eq.(3.12)). Thus, decreasing M would seem to degrade the SNR. However, a more dominant effect is that as M decreases, attenuation of the signal power over the length of each sub-bus also decreases (fewer lossy taps between amplifiers). Thus the ratio of signal power to noise power actually increases as M decreases, a fact illustrated in Figures 3.10 and 3.12.

Another result of great practical significance is that the gain G_a of the original amplifiers need not be high, if we are willing to use enough of them. The total number of users on the bus determines the overall loss that a signal would suffer propagating end to end if no amplifiers were used. This total loss in turn determines the amount of overall gain that must be supplied end to end. It is clear that we have the option of supplying this gain with either a relatively small number of high gain amplifiers or a larger number of low gain amplifiers. We have noted that the latter case yields more favorable SNR performance. Perhaps more important, however, is that the use of low gain amplifiers should be more attractive in terms of the device technology required to realize them. (A high gain device might also be prone to nonlinearities and other non-ideal effects.) In fact, we might envision the limiting case where M = 1, i.e. each waveguide coupler comes packaged with a low gain amplifier, the amplifier gain being just high enough to offset the loss that the signal would experience in propagating through that coupler. It should be noted, however, that our analysis assumed that the noise of the amplifiers (referred to the input) was $h\nu_o$, an assumption valid only when the gain of the amplifier is much greater than 1. If we go to a low value of M, the gain will be no more than a few dB, and the noise can no longer be modeled as having a spectral height $h\nu_o$. The case of low gain amplifiers is discussed in Appendix B, where we show that the SNR dergadation due to amplifier noise may increase as the gain of the individual amplifiers drops.

Conclusions

In this Chapter we have studied power division limitations in depth. For the bus topology, the excess loss of the waveguide couplers posed by far the most dominant limitation on N_{max}. We found that, because the attenuation on the bus (end to end) increases exponentially with N

(Eq. 3.4b), the effects of excess coupler loss cannot be canceled by increasing receiver sensitivity or by increasing source power, at least for N significantly large (100 or more).

We saw that the star topology is much less sensitive to excess coupler loss. More significantly, however, it was found that the use of optical amplifiers in the bus configuration was capable of increasing N_{max} for that topology by orders of magnitude. Further, we noted that the gain of the amplifiers need not be high to prove beneficial in increasing N_{max}, a fact that should be of great practical significance in future applications of this approach.

Chapter 4

Implementation Issues in Frequency Multiplexing

Introduction

In previous Chapters, we have seen the potential benefits of concurrency in local networks, particularly when it is achieved through frequency multiplexing. It remains to be seen how this multiplexing would actually be performed in a network employing optics technology. In this Chapter, we examine a number of possible methods for frequency multiplexing of signals in local networks. We begin by investigating the applicability of heterodyne detection. Direct detection methods are then discussed, including subcarrier modulation as well as baseband techniques that employ optical filters. We would like to compare the various methods in terms of complexity, efficiency of bandwidth usage and the number of simultaneous users that can be accomodated. As in earlier portions of this Thesis, the analysis is not confined to technology that is currently practical and economical. In taking this approach, we continue our examination of emerging optics technology, in order to judge its potential relevance to local networks.

Before frequency multiplexing can actually be achieved in a local network, one must address various control issues, such as coordination between users and bandwidth allocation. Control aspects of concurrency and multiplexing will be examined in Chapter 5; they will be ignored in the present discussion.

Heterodyne Detection Methods

Optical heterodyne detection has received widespread attention over the past several years, particularly for use in fiber optic transmission systems.[39,40] The various device considerations germane to heterodyne detection are by now well known. In this Chapter, our interest in heterodyning stems from the fact that heterodyne detection receivers are field detectors, as opposed to direct detection receivers which are intensity detectors. Thus heterodyne detectors are capable of separating signals located in disjoint optical frequency bands. This suggests a method for frequency multiplexing, in which each user tunes his source laser to a slightly different frequency, as shown in Figure 4.1. The maximum number of channels that the system can support (which in this case equals the maximum number of simultaneous users) obviously depends on the band interval $\Delta\nu$. We now determine how large $\Delta\nu$ must be for successful operation.

Let $\underline{s}_i(t)$ be the baseband information signal of user i, extending from -W to W in the frequency domain, and normalized such that

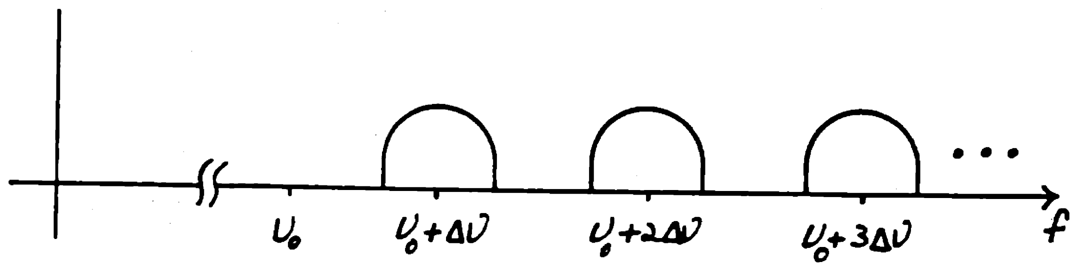


Figure 4.1. Placement of signals in the frequency domain, with individual source lasers being tuned to slightly different frequencies.

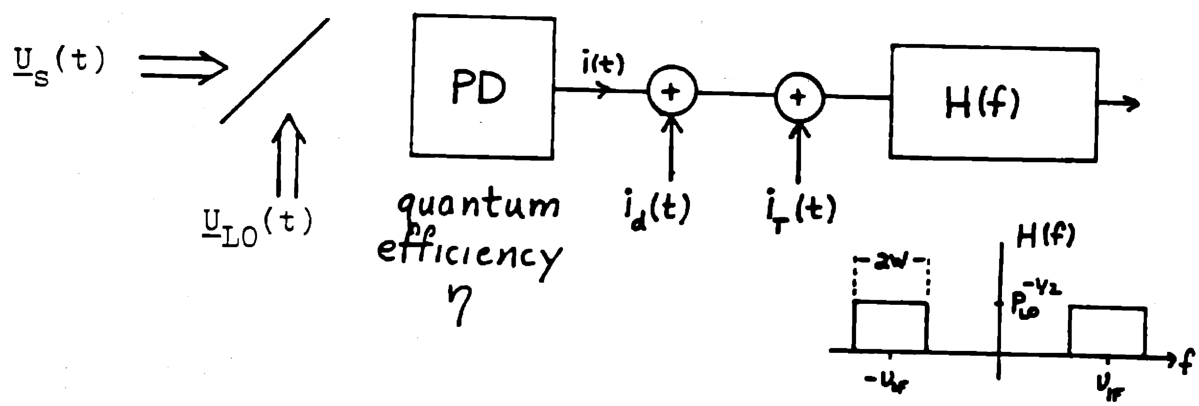


Figure 4.2. Heterodyne detector/receiver configuration. With sufficiently large P_{LO} we obtain quantum-limited operation, in which the effects of the dark current $i_d(t)$ and thermal noise $i_T(t)$ can be neglected in the SNR calculations.

$$\frac{1}{T}\int_0^T |\underline{s}_i(t)|^2 \, dt = 1 \tag{4.1}$$

The complex envelope of the radiation emitted by user i is

$$\underline{U}_{s_i}(t) = (4P_s/\pi d^2)^{\frac{1}{2}} \, \underline{s}_i(t) e^{j(kz - 2\pi i \Delta\nu t)} \tag{4.2}$$

where d is the radius of the detector surface. The field transmitted by user i thus has nominal optical frequency $\nu_0 + i\Delta\nu$. In this analysis we assume for simplicity that all users occupy the same amount of bandwidth (2W at baseband) and utilize the same time-average power P_s.

Figure 4.2 illustrates the general heterodyne detector and receiver structure.[43] In general, the incident signal field at a given detector is a sum of contributions from all users who are transmitting:

$$\underline{U}_s(t) = \sum_i \underline{U}_{s_i}(t) \tag{4.3}$$

Strictly speaking, Eq. (4.3) is incorrect for two reasons. The fields incident on a given detector will show various time delays, depending on the positions of the sources relative to the detector. Also, because of power division within the network, the $\underline{U}_{s_i}(t)$ may not all have the same average power. However, in the present analysis we are interested only in the positioning of the $s_i(t)$ in the frequency domain. Neither the time delays nor the power have an effect on this issue, thus we ignore the inaccuracies in (4.3).

Let the local oscillator be at frequency $\nu_0 - \nu$:

$$\underline{U}_{LO}(t) = (4P_{LO}/\pi d^2)^{\frac{1}{2}} e^{j(kz + 2\pi\nu t)} \qquad (4.4)$$

The rate parameter for the total field incident on the photodetector will be proportional to $|\underline{U}_s(t) + U_{LO}(t)|^2$ integrated across the surface of the photodetector. The only components in this expression that can make it through the passbands of H(f) will be those resulting from the $\underline{U}_s^*(t)\underline{U}_{LO}(t)$ and $\underline{U}_s(t)U_{LO}^*(t)$ terms. The first of those terms is proportional to

$$\sum_i \mathrm{Re}\left[\underline{s}_i^*(t)e^{j2\pi(i\Delta\nu + \nu)t}\right] \qquad (4.5)$$

and the second proportional to

$$\sum_i \mathrm{Re}\left[\underline{s}_i(t)e^{-j2\pi(i\Delta\nu + \nu)t}\right] \qquad (4.6)$$

We would like to find the relationship between ν, ν_{IF}, $\Delta\nu$ and W such that the desired frequency bands (whichever ones they may be), and only those bands, are aligned with the passbands of H(f). In order to receive information in band i, it is easy to see from (4.5) and (4.6) that we must have

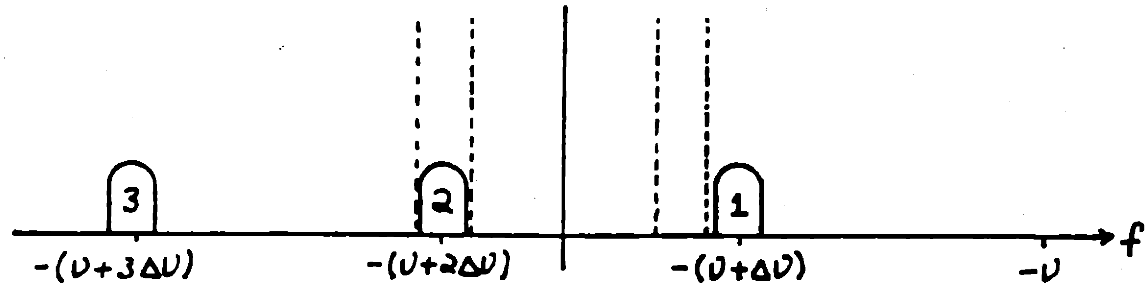
$$\nu = \pm(\nu_{IF} - i\Delta\nu) \qquad (4.7)$$

Thus to receive information in band i, the local oscillator (of frequency $\nu_0 - \nu$) should be adjusted according to (4.7).

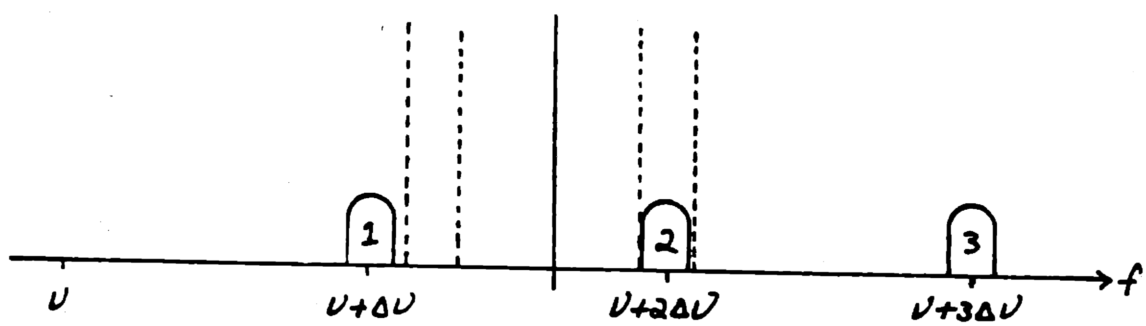
In this approach we hold the passbands of the IF filter fixed while tuning the local oscillator. In principle we could do the opposite. In practice, however, tuning the LO should be considerably easier than constructing an IF filter that is tunable over perhaps hundreds of GHz.

This use of heterodyne detection is in many ways analogous to the superheterodyne receiver of standard radio. We know from the superheterodyne case that one must prevent unwanted signal components in the image band from beating through the IF filter along with the desired signal, thus corrupting reception. In the superheterodyne problem, this difficulty is dealt with simply by passing the collection of received signals (i.e. the signal off the antenna) through a relatively low-Q bandpass filter to reject the image band before any mixing takes place. In our heterodyne detector the mixing process occurs optically and such pre-mixing image band rejection is not possible, at least with electronics. We rely on intelligent placement of the individual frequency bands to avoid image band interference.

Figure 4.3 illustrates the basic idea. For purposes of example, i = 2 has been chosen as the "desired" channel.



<u>Figure 4.3a</u>. Frequency distribution of $\underline{U}_s^*\underline{U}_{Lo}$, with ν set to receive $s_2(t)$. The passbands of H(f) are indicated by the dashed lines.



<u>Figure 4.3b</u>. Frequency distribution of $\underline{U}_s U_{Lo}^*$, with ν set to receive $s_1(t)$.

The only way to assure that all signals with $i \neq 2$ fall in the stopbands of H(f) is to have

$$\Delta\nu > 2\nu_{IF} + 2W \qquad (4.8)$$

Clearly, if $\nu_{IF} \gg W$ the spacing between signals in the frequency domain will be large compared to the bandwidth actually being used, an inefficient scheme.

The preceding discussion suggests selection of a relatively small ν_{IF}. Indeed, an IF frequency as low as W will work, at least in principle. This would result in $\Delta\nu$ values as small as 4W. With W 10^6Hz to 10^7Hz, the system could support a very large number of channels. However, there are other considerations that may dictate larger spacings, and thus less efficient use of bandwidth. Specifically, consider the effect of laser frequency drift. Frequency drifts of the source laser and LO laser can be tracked out at the receiver, to assure that the desired signal falls within the IF filter passbands. However, as the center frequencies of the other source lasers change, the relative positions of the frequency bands in Figure 4.1 also change, and we must be sure that the minimum spacing given by (4.8) is preserved. Thus the value of $\Delta\nu$ in (4.8) can be regarded as a lower bound, attained with lasers perfectly stable in frequency.

Let us now return to the original question of how many channels this multiplexing scheme can provide. The total available bandwidth is either the tunable range of the laser or the predetection bandwidth of the photodetector, whichever is smaller (in practice, it is almost always the laser bandwidth). Either way, the bandwidth is quite large. For instance, laser sources tunable over 300 GHz have been reported.[41] By way of example, let us choose W = 10 MHz, ν_{IF} = 50 MHz, and $\Delta\nu$ = 200 MHz. These numbers permit significant leeway for both image band rejection and laser frequency drift. Assuming a total bandwidth of 200 GHz, 10^3 frequency bands can be accomodated, an enormous amount considering the total number of users typical of many local networks. Clearly, heterodyne detection provides a natural and effective (albeit technologically nontrivial) way to perform frequency multiplexing in local networks.

Looking back on Eq. (4.8), we see that the maximum percentage of bandwidth used in this method is 50%. A 50% efficiency may be more than adequate if W is small, for in that case a huge number of channels can be accomodated. However, if W is 1 GHz instead of 10 MHz, far fewer channels will be available, and one may wish to improve somehow on the 50% efficiency in order to accomodate more users. Recall that the bound (4.8) stemmed from the need to reject the image band. But say that we placed a frequency-selective device between the main channel and detector, an optical bandpass device (possibly tunable) that is capable of blocking the image band. Then the bandwidth usage could in principle approach 100%. In fact, suppose that we employed a frequency-selective element whose Q was high enough that it could reject <u>all</u> frequency bands other than the one desired. Then all of the necessary frequency selectivity resides with that device, and we would no longer need to use heterodyne detection to perform the multiplexing. This brings us to our next topic.

Direct Detection Methods

There are two principle ways to achieve frequency multiplexing with direct detection. In one method, users tune their source lasers to different frequency bands. At a given receiver, a desired band is separated from the rest by means of a frequency-selective device placed between the main channel and the detector. The second method utilizes subcarrier modulation. Different users wishing to transmit simultaneously use different subcarrier frequencies. The various bands can then be separated by a filtering process in the post-detection electronics. We investigate both methods, with our chief aim being to determine how many simultaneous users can be accomodated.

<u>Use of Optical Filters</u>. The frequency-selective device mentioned previously can take the form of an optical wavelength filter placed in front of the detector, or a frequency-selective waveguide coupler. We will compare the two devices at the end of this section. For now, we refer to them both generically as optical filters.

The total usable bandwidth in this scheme is the same as for heterodyne detection, essentially the tunable range of the laser. The number of channels that the system can support (which again in this case is the same as the maximum number of simultaneous users) depends on the total bandwidth and the width of the passbands of the optical filters. The width of the filter passbands is influenced by three factors: the bandwidth W of the information signal, the amount of laser center frequency drift, and the maximum available filter $Q \equiv Q_{max}$. With so many technology-related variables, it is difficult to say exactly how many channels are possible. We instead look at two important limiting cases.

Let the maximum laser center frequency excursion (from the nominal center frequency) be Δf, and call the total bandwidth B. The required width of the filter passbands is then $2(W + \Delta f)$. In the case where $2(W + \Delta f) > \nu_o/Q_{max}$, we are not limited by the maximum filter Q and the number of channels is approximately $B/2(W+\Delta f)$, assuming that all packets occupy the same amount of bandwidth W (or at least are allocated that much). On the other hand, if $2(W+\Delta f)$ is less than ν_o/Q_{max}, as would be the case with large signal durations T and highly stable lasers, then the filter Q is the primary limitation, and the number of channels is approximately BQ_{max}/ν_o. This gives an upper bound on the number of channels possible with this method, under the constraint $2(W+\Delta f) < \nu_o/Q_{max}$.

It was mentioned earlier that the optical filter may be realized with either a frequency-selective waveguide coupler or a wavelength filter placed in front of the detector. From the point of view of the multiplexing problem, the two types of devices are interchangeable. However, numerous other considerations might dictate the use of one over the other. For instance, the coupler might seem to be the more appropriate device for the bus topology, since it removes from the main bus only that power that the particular user is interested in detecting. Such an advantage is slight, though, since we saw in Chapter 3 that the main contribution to power loss along the bus is not the insertion loss, but the excess coupler loss. In the long run, which type of device is used will probably depend on issues such as cost, device excess loss, maximum obtainable Q and tunability if it is required.

Use of Direct Detection/Subcarrier. In our study of various frequency multiplexing methods, the use of subcarrier modulation techniques is by far the easiest, at least in terms of requisite device capability. This simplicity does carry a price, however. The maximum number of simultaneous users may be severely limited by two factors. First, the usable bandwidth is limited to the postdetection bandwidth of the receiver electronics. This is at most a few Gigahertz, orders of magnitude less than for the methods discussed previously. Thus the maximum number of channels will be much lower than before, with an upper bound of approximately $B_d/2W$, where B_d is the post-detection bandwidth. The second potential limitation is that the shot noise entering the detection process is that contributed by all users who are transmitting, not just the one in the desired band. Because of this effect, the network might be able to tolerate only a limited number of simultaneous transmissions without significant degradation of SNR. In other words, the maximum number of simultaneous users might actually be much lower than the maximum number of channels $B_d/2W$.

The remainder of this section investigates analytically the extent of the shot noise limitation. Before we begin that analysis, however, we discuss one fundamental implementation issue.

Subcarrier techniques may be employed in conjunction with either lasers and single mode fibers (as was the case with all methods discussed earlier) or LED's and multimode fibers. (It is conceivable that LED's could be used together with single mode fibers, but this combination would not be recommended for use in local networks, since the power coupled into the fiber would be so low.) In this problem, the rate parameter for the total field incident on the photodetector is related to $|\underline{U}_s(t)|^2$, where $\underline{U}_s(t)$ is the sum of complex envelopes from all users transmitting. Thus the photocurrent will contain numerous cross terms at various positions in the frequency domain. If LED's are used, the complex envelopes of different sources will be statistically independent. Thus the cross terms will average to zero and will not affect the mean photocurrent. With lasers, on the other hand, the cross terms will in general not average to zero. It is possible to perform subcarrier multiplexing with lasers, if the line centers of the different lasers were sufficiently separated that the beat frequencies fell outside the photodetector bandwidth. Though this is a viable approach, the remainder of this section will focus on the LED/multimode fiber method.

Let the complex envelope of the radiation incident on a given photodetector be $\underline{U}_s(\bar{r},t)$, where

$$\underline{U}_s(\bar{r},t) = \sum_i \underline{U}_{s_i}(\bar{r},t)$$

and the $\underline{U}_{s_i}(r,t)$ are such that

$$E(\underline{U}_{s_i}(\bar{r},t)) = 0 \tag{4.9}$$

$$E(\underline{U}_{s_i}(\bar{r},t)U_{s_j}(\bar{r},t)) = 0 \tag{4.10a}$$

$$E(\underline{U}_{s_i}(\bar{r},t)\underline{U}^*_{s_j}(\bar{r},t)) = 0 \tag{4.10b}$$

$$\int_S d\bar{r}\,|\underline{U}_{s_i}(\bar{r},t)|^2 = P_i(1 + s_i(t)\cos\omega_i t) \tag{4.11}$$

where $s_i(t)$ is the baseband information signal of user i, and S is the photodetector surface.

From (4.10) and (4.11) it follows that the total average power incident on the photodetector is

$$P(t) = \sum_i P_i(1 + s_i(t)\cos\omega_i t) \tag{4.12}$$

We assume the receiver structure of Figure 4.4. On-off signaling (OOK) is assumed. Coherent subcarrier demodulation is employed for analytical simplicity.

This analysis follows closely that of Shapiro.[28] For the photocurrent i(t) we have, assuming the low photon coherence limit,

$$E(i(t)) = \frac{eG\eta}{h\nu_o} P(t) \tag{4.13}$$

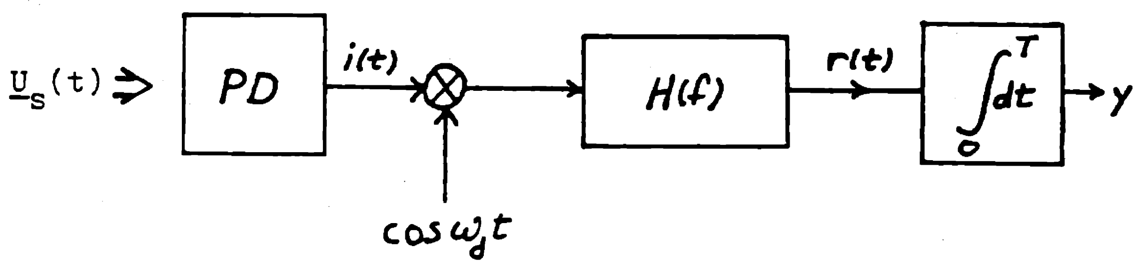


Figure 4.4

Detector and receiver structure for direct detection/subcarrier problem. H(f) is a lowpass filter. The photodetector has quantum efficiency η. The local oscillator has been set to receive $s_j(t)$.

$$K_{ii}(t,u) = \left(\frac{e^2G^{2+x}\eta}{h\nu_o}P(t) + \frac{2k\theta}{R}\right)\delta(t-u) \qquad (4.14)$$

It follows that

$$K_{rr}(t,u) = \int d\tau \cos^2\omega_j\tau\left[\frac{e^2G^{2+x}\eta}{h\nu_o}P(\tau) + \frac{2k\theta}{R}\right]h(t-\tau)h(u-\tau)$$

$$= \frac{1}{2}\int d\tau \cos 2\omega_j\tau \; z(\tau)h(t-\tau)h(u-\tau)$$

$$+ \frac{1}{2}\left[\frac{e^2G^{2+x}\eta}{h\nu_o}\sum_i P_i + \frac{2k\theta}{R}\right]\int df \left|H(f)\right|^2 e^{j2\pi f(t-u)}$$

where

$$z(t) = \frac{e^2G^{2+x}\eta}{h\nu_o}\left[\sum_i P_i(1 + s_i(t)\cos\omega_i t) + \frac{2k\theta}{R}\right]$$

The first integral in the $K_{rr}(t,u)$ expression has magnitude less than

$$\frac{e^2G^{2+x}\eta}{2h\nu_o}\int\left|\tilde{Z}(f)\right|K(f)\,df$$

where $\tilde{Z}(f) = \frac{1}{2}(Z(f-2f_j) + Z(f+2f_j))$ and

$$K(f) = \int df' \left|H(f'+f)\right|\left|H(f')\right|$$

At most one of the $s_i(t)$ can fall within the nonzero portions of $K(f)$. Thus, as an approximation, the noise contribution made by the first integral will be neglected compared to the second. It follows that

$$K_{rr}(\tau) = \frac{1}{2}\left[\frac{e^2G^{2+x}\eta}{h\nu_o}\sum_i P_i + \frac{2k\theta}{R}\right]\int df \left|H(f)\right|^2 e^{j2\pi f\tau} \qquad (4.15)$$

We assume that H(f) is a single pole lowpass filter, with

$$H(f) = \frac{A}{1 + j2\pi f\alpha} \tag{4.16}$$

$$A = 2h\nu_o / eG\eta \tag{4.17}$$

Thus, using (4.13) and (4.16),

$$E(r(t)) = P_j \, s_j(t) \tag{4.18}$$

$$K_{rr}(\tau) = \left[\frac{e^2 G^{2+x}}{h\nu_o}\eta \sum_i P_i + \frac{2k\theta}{R}\right]\frac{A^2}{2\alpha} e^{-|\tau|/\alpha} \tag{4.19}$$

The covariance (4.19) is a function of the powers P_i received from the various users. For simplicity, we assume that all the P_i are the same. Such would be the case in the star network of Figure 3.4, if all users had the same source power.

Again assuming Gaussian statistics for the noise, the decision process consists of comparing y to an optimum threshold. We have

$$\begin{aligned} \mathrm{Var}(r \mid 1 \text{ sent}) &= \mathrm{Var}(r \mid 0 \text{ sent}) \\ &= \left[\frac{e^2 G^{2+x}}{h\nu_o}\eta \, nP_i + \frac{2k\theta}{R}\right]\frac{A^2}{2\alpha} \end{aligned} \tag{4.20}$$

$$E(y \mid 1 \text{ sent}) = P_i T \tag{4.21a}$$

$$E(y \mid 0 \text{ sent}) = 0 \tag{4.21b}$$

$$\mathrm{Var}(y \mid 1 \text{ sent}) = \mathrm{Var}(y \mid 0 \text{ sent}) \tag{4.22}$$

$$= A^2T\left(1 - \frac{\alpha}{T}\left(1 - e^{-T/\alpha}\right)\right)\left[\frac{e^2G^{2+x}\eta}{h\nu_o} nP_i + \frac{2k\theta}{R}\right]$$

The cutoff frequency of the lowpass filter is chosen such that $\alpha = T/4$. Treating the noise as Gaussian, we know that the optimum threshold for y is $P_iT/2$ and the probability of error is

$$\Pr(\text{error}) = Q\left[\frac{P_iT/2}{\sqrt{\mathrm{Var}(y)}}\right] \tag{4.23}$$

For $\Pr(\text{error}) = 10^{-9} = Q(6)$, we solve for n:

$$n = \frac{P_iT}{530h\nu_oG^x} - \frac{2k\theta h\nu_o}{e^2G^{2+x}RP_i} \tag{4.24}$$

Using the parameter values in Table 3.1, this reduces to

$$n = 6.5\text{x}10^{15}P_iTG^{-\frac{1}{2}} - 1.9\text{x}10^{-12}(G^{2.5}P_iT)^{-1}$$

Minimizing over G gives

$$G_{opt} = 3.8\text{x}10^{-14}(P_iT)^{-1} \tag{4.25}$$

$$n = 2.7\text{x}10^{22}(P_iT)^{3/2} \tag{4.26}$$

Table 4.1 shows the maximum number of simultaneous users for the specific case of the star network. Values for source power and coupler loss are the same as those

	Number of Users N in Star		
	128	512	1024
$T=10^{-6}$	128	105	26
$T=10^{-7}$	54	3	0

Table 4.1

The maximum number of simultaneous users n, as determined by the shot noise limitation (4.26), for the star network. Source power P_s = -10 dBm has been assumed.

used in Chapter 3. The optimum value of G for the various cases in Table 4.1 ran between 10 and 100, quite reasonable for APD gain. The Table shows clearly that there are circumstances under which the maximum number of simultaneous users is limited much more by the shot noise than by bandwidth availability.

Conclusions

This Chapter has examined implementation issues germane to frequency multiplexing in fiber optic local networks. Three methods were discussed: heterodyne detection, use of optical filters, and subcarrier modulation. In each case, bounds on the maximum number of simultaneous users were determined.

When one considers the limitation on total bandwidth and the $T^{3/2}$ dependence of the shot noise limitation of Eq. (4.26), it is evident that subcarrier techniques will lose favor in high bit rate applications. However, Table 4.1 shows that with signal durations $T = 10^{-6}$ sec or more, subcarrier multiplexing may in fact be capable of supporting a large number of simultaneous users. Thus this technique should not be disregarded, at least for low bandwidth applications.

The two other methods that were discussed, those employing heterodyne detection or optical filters, are quite similar in nature. Both are really WDM schemes, multiplexing users at optical frequencies. We saw that heterodyning may be the method of choice if a large number of low bandwidth channels are needed, for in this case the optical filters would be required to have an extremely high Q. On the other hand, we observed that optical filters may provide a somewhat more efficient use of system bandwidth when the individual channel bandwidths are very large. In general, two significant statements can be made. First, both of these methods allow users to be multiplexed across tens, perhaps hundreds of Gigahertz, providing substantial levels of concurrency even in high bit rate applications. Second, the feasibility of either method will depend strongly on future developments in device technology. The course of these developments could dictate which method will carry the most relevance for local network application.

So far we have viewed frequency multiplexing as a way to achieve concurrency in local networks, and thus as a potential means of reducing delay between users, vis a vis the Concurrency Principle of Chapter 2. If reduction in delay is our sole objective, one may ask how many simultaneous channels are really necessary. Reflecting on the discussions in Chapter 2, the answer in most cases may well be,"Not very many." One might wonder, then, why we have maintained such great interest in multiplexing techniques that could utilize hundreds of Gigahertz, and provide perhaps a thousand or more channels. The answer lies in the potential simplification of network control that would result if each user could be assigned his own frequency band, i.e., a fixed assignment scheme. This is one of the topics to be explored in the following Chapter.

Chapter 5

Control Issues and Bandwidth Usage

Introduction

In Chapter 1 it was claimed that, in the local network problem, the enormous bandwidth of single mode fibers can be exploited to achieve two purposes: to reduce delay between users and to simplify network control. The last few Chapters have discussed how concurrency can lead to reduction in delay, and how the concurrency can be implemented. Thus far, however, very little has been said about control. In this Chapter we investigate a number of control issues germane to the exploitation of fiber bandwidth in local networks.

This Chapter has three sections. The first two examine methods in which increased bandwidth usage leads to an elimination of explicit control within the network. The two approaches to be discussed are in a sense extreme cases, and utilize the bandwidth in fundamentally different ways. One method, which amounts to an application of Pure Aloha concepts to local networks, involves increased bandwidth usage by individual interfaces throughout the network. The other method consists of allocating to each user a separate, fixed frequency band, i.e. fixed assignment frequency multiplexing.

The third section in this Chapter discusses control aspects of a dynamic assignment scheme. This work will not address issues concerning the actual bandwidth allocation algorithm. Rather, it will focus on the operation of a dedicated control channel, a so-called order wire, to provide coordination among users in the network.

Use of Pure Aloha Techniques

In most local networks, control consists of some sort of access algorithm,[1,22] whose purpose is to eliminate, or at least to reduce drastically, the probability of packet collisions. In these networks, bit rates tend to be such that individual packets occupy the network for periods of time much longer than the propagation delay τ. The probability of a packet becoming available for transmission while another packet already occupies the network increases monotonically with the packet duration. Thus contention algorithms may prove quite beneficial at low bit rates, where the probability of collision would be relatively high if no control or arbitration were exercised. Consider what happens, however, if we increase interface bandwidth utilization in order to decrease T_p (assuming that the number of bits per packet is constant) to the point where the probability of collision is acceptably low without using any contention algorithm. Indeed, in the limit of infinite bit rates the packets would become infinitesimally short, and the probability of two packets overlapping spatially (a collision) would go to zero, assuming finite packet generation rates. Packets could then be transmitted whenever they become available, without any form of network arbitration. We must determine how high the bit rates must be in order for this method to result in acceptable performance.

The transmit when ready scheme is precisely the well-known Pure Aloha packet broadcasting strategy,[23,24] and in this analysis we will draw on some of the established results for Pure Aloha networks. For simplicity, it will be assumed that all packets have identical duration T_p, and that they are generated via a Poisson process with overall rate λ.

It is easy to show that the probability of a packet successfully reaching its destination without collision has a lower bound of $\exp(-2\lambda T_p)$ for the ring and the bus, and also for the star if the "spines" of the star have equal lengths. This Prob(no collision) is agreement with the Pure Aloha results in the literature. The chief question in our local network problem is how high Prob(no collision) must be (i.e. how low T_p must be) in order to achieve adequate performance. For any finite bit rate, $\exp(-2\lambda T_p)$ will be less than one, so collisions are bound to occur occasionally. When a collision occurs, the packets involved must be retransmitted, resulting in an additional delay in transferral of the information between source and destination. To find a bound for T_p, our criterion will be that collisions cannot lead to a significant increase in the average delay which, in the absence of collisions, is given by Eq. (2.1) of Chapter 2.

The rate λ in $\exp(-2\lambda T_p)$ is the overall packet generation rate in the network:

$$\lambda = \lambda_n + \lambda_r \qquad (5.1)$$

where λ_n is the arrival rate of new packets to the network and λ_r is the rate of packets retransmitted because of collisions. Regarding packet transmissions as a Bernoulli process with success probability $\exp(-2\lambda T_p)$, the expected number of transmissions per packet is then $\exp(2\lambda T_p)$. Thus

$$\lambda_r = \lambda_n \exp(2(\lambda_n + \lambda_r)T_p) - 1 \qquad (5.2)$$

$$\lambda_n = \lambda \exp(-2\lambda T_p) \qquad (5.3)$$

The throughput of a Pure Aloha network is $\lambda\exp(-2\lambda T_p)$, which has a maximum value of $(2eT_p)^{-1}$. Thus if λ_n exceeds $(2eT_p)^{-1}$, we are guaranteed network instability, since the offered load exceeds the maximum possible throughput.

Actually, as described above the Pure Aloha scheme is inherently unstable at <u>any</u> value of λ_n. This is because there is always a probability, however small, that enough packets will be transmitted over a sufficiently short period of time to drive the network into unstable operation, from which it cannot recover. What is done typically in practice is to run the network far enough below threshold that the probability of instability can be neglected.

This criterion immediately suggests a condition for T_p. We must have $\lambda_n \ll (2eT_p)^{-1}$, or

$$T_p = \frac{a}{2e\lambda_n} \tag{5.4}$$

where $a \ll 1$. If the network operates at, say, ten percent of capacity ($a = .1$), (5.4) becomes $T_p = .018/\lambda_n$. With $\lambda_n = 10^3$/sec, a liberal estimate for typical networks, then assuming a 1000-bit packet we must use a bit rate of approximately 50 Mbit/sec.

The condition (5.4) resulted from network stability considerations. Note that this effective stability threshold is considerably lower (in fact, about 1/50 the size, for $a = .1$) than the threshold $\rho \approx \lambda_n T_p$ of the Concurrency Principle. This is the price that we pay for removing control from the network.

Let us now determine whether a network that satisfies (5.4) with $a \ll 1$ also satisfies the delay requirement stated earlier in this Section, namely that collisions do not cause a significant increase in the overall delay in transferral of information between users. Let x be the number of retransmissions required for a packet. Then using the geometric probability distribution[27] we obtain

$$E(x) = e^{a} - 1 \tag{5.5}$$

$$Var(x) = e^{2a} - e^{a} \tag{5.6}$$

using $\lambda T_p = a/2$. Assuming a = .1 it follows that

$$E(x) = 0.1 \tag{5.7a}$$

$$Var(x) = 0.1 \tag{5.7b}$$

$$Prob(x>5) = 4.5x10^{-4} \tag{5.7c}$$

From Equations (5.7) it is evident that as long as (5.4) is satisfied with a << 1, collisions will not cause a significant increase in delay, and Aloha techniques should provide a viable method of simplifying local network control. Clearly, however, the practicality of this approach depends on the bit rates required to achieve a << 1. The previous numerical example with $\lambda_n = 10^3$/sec dictated bit rates around 10^8/sec, not an unreasonable speed for fiber optics, power division issues aside (see Chapter 3). If for some reason λ_n were much larger, the requisite speeds might be prohibitive. Nevertheless we may conclude that, for many typical networks, extremely high bit rates are not required in order to take advantage of the simplification of network control that Aloha techniques provide.

Use of Fixed Frequency Assignment

We now examine a second approach to simplifying network control. This method is based on the results of Chapter 4, where we found that frequency multiplexing, when achieved through either heterodyne detection or narrowband optical filters, is capable of accomodating a very large number of simultaneous users (at least for low to moderate bandwidth interfaces). If a sufficient number of frequency bands were available, we could again eliminate the need for formal network control by allocating to each user a separate, fixed frequency channel. Then any user wishing to communicate with another would merely tune his source laser to the appropriate frequency band, i.e. to the band assigned to the user for whom the packet is intended.

The fixed assignment method sounds simple enough, but there are at least two issues relevant to this scheme that deserve discussion. First, it is clear that fixed assignment strategies will not be able to use bandwidth as efficiently as a more general dynamic assignment method. Indeed, if the total available bandwidth were not sufficient to provide each user with a separate receiving channel, while at the same time satisfying the bandwidth requirements of individual users, then frequency multiplexing would require dynamic assignment. However, this does not diminish the potential importance of fixed assignment. When using heterodyne detection or optical filters, the total bandwidth is so huge that the number of channels available will suffice for many applications of interest. (Recall that we found in Chapter 4 that about 1000 channels, each of 10 MHz bandwidth, could be accomodated using heterodyne detection.) In fact, if all users in the network can be assigned dedicated frequency bands of sufficient width to satisfy their bandwidth requirements, then there will be no advantage in going to a more general dynamic assignment scheme.

A second issue relevant to fixed assignment concerns the possibility of packet collisions within a given frequency band. Our aim with the fixed assignment method is to eliminate the need for explicit network control. Clearly, if users are free to transmit at any time, it is possible for two users to send packets to the same destination simultaneously, resulting in a possible collision on that band. Packet collisions will cause increased delay in transferral of information, and in general only a certain percentage increase in delay can be tolerated.

To evaluate the effect of these collisions on delay, we note that the fixed assignment approach divides the overall network into a large number of virtual channels, one per user, each of which is run in a Pure Aloha manner. Thus to draw conclusions we need only borrow from the Aloha analysis performed earlier in this Chapter.

Specifically, let λ_i be the rate of generation of packets throughout the network that are addressed to user i. For successful operation we know from (5.4) that

$$T_p = \frac{a}{2e\lambda_i} \qquad (a \ll 1) \tag{5.8}$$

is required for all i. Since λ_i is approximately λ_n/N, where N is the total number of users in the network, (5.8) should be even easier to satisfy than (5.4), i.e., the required bit rate is only (1/N) as large as that necessary for (5.4). This is particularly important in light of the fact that we are multiplexing users in frequency. If high bit rates (and thus high interface bandwidths) were necessary to run each of the channels effectively, then (depending on N) we might not have enough total bandwidth to assign each user a separate band.

On balance, then, we conclude that, for typical network traffic, collisions within individual frequency bands should not degrade significantly the performance of a fixed-assignment frequency multiplexing scheme.

Control of Dynamic Frequency Assignment

We have pointed out that some situations may demand dynamic assignment of bandwidth among users. Such might be the case, for instance, in a network that achieved frequency multiplexing through subcarrier modulation, for with subcarrier techniques the total available bandwidth might not be large enough to supply each user with a dedicated channel.

Clearly, a dynamic assignment method requires some form of coordination among users. Perhaps the most straightforward method of affecting this coordination would be to use a dedicated channel, a so-called order wire, over which only control information would pass. A user wishing to send a packet would first contact a central controlling mechanism over the order wire. The controller could then send information (again over the order wire) to the packet source and destination points concerning which frequency band was to be used.

The order wire is really a separate subnetwork itself. Thus to operate it successfully we must address the same issues, such as contention and delay, as with any network. It would of course be desirable to keep the order wire operation as simple as possible, and one way to do this would again be to apply Pure Aloha techniques. We could do this as long as the probability of a collision between control packets was acceptably low. Borrowing once again from the results earlier in this Chapter, we know that the required condition is

$$T_o = \frac{a}{2e\lambda_o} \qquad (a \ll 1) \qquad (5.9)$$

where λ_o is the rate of packet generation on the order wire and T_o is the duration of the control packets. For every packet transmitted on the main network there may be three packets on the order wire (one from packet source to controller, one from controller to packet source, and one from controller to packet destination). This factor of 3 would tend to make the right hand side of (5.9) smaller, making it more difficult to satisfy. However, the amount of information to be sent in the control packets is small, containing only details such as the center frequency of the band to be used. Therefore, T_o will be small compared to the duration T_p of packets on the main network, and thus (5.9) may actually be easier to satisfy than (5.4). For instance, assuming a packet generation rate on the main network of 1000/sec, then with a = .1, (5.9) yields $T_o = 10^{-5}$ sec. If the control packets are 200 bits long, this translates to a bit rate of slightly over 30 Mbit/sec. Even if the packet generation rate on the main network were several times higher than we just assumed, the bit rate on the order wire could be increased accordingly without burdening the electronics excessively.

An important conclusion to be drawn here is that the overhead* required to coordinate users in a dynamic bandwidth assignment scheme need not be excessive. The operation of an order wire via Pure Aloha techniques should provide an adequate medium over which coordination among users can be maintained.

Chapter 6

Conclusions and Summary

This Thesis began with the assertion that the enormous bandwidth of single mode fibers can be exploited to achieve two purposes: to reduce delay in transferral of information and to simplify network control. Delay reduction was explored in Chapter 2, where the concept of concurrency played a fundamental role. We stated via the Concurrency Principle that the potential utility of concurrency, as a means of reducing delay, was confined to two cases, one involving the stabilization of unstable networks and the other involving the elimination of short term delays. These conclusions are important, though perhaps in a negative sense. Since most current networks operate well within the range of stability, network usage patterns will have to change markedly before concurrency gains widespread applicability.

The use of bandwidth to simplify control was examined in Chapter 5. We investigated two methods, one method using Pure Aloha techniques and the other employing fixed-assignment frequency multiplexing. We concluded that both methods had significant potential, though fixed-assignment frequency multiplexing will most likely require use of either heterodyne detection or narrowband optical filters, technology that is still in a developmental stage.

Chapter 3 investigated power division limitations in detail, and it was here that we encountered some of the most fundamental roadblocks to network capability. For the bus topology, we saw that excess coupler loss was by far the most dominant limitation for typical network parameters. The effect of excess coupler loss could be circumvented only by switching to a star configuration, or by employing optical amplifiers on the bus. (Electronic regenerative repeaters could be used, but such devices would severely limit the bandwidth capability of the network.)

One of the underlying goals of this Thesis was to evaluate the potential applicability of emerging optics technology to the local network problem. For instance, we found that heterodyne detection was not particularly effective as a means of alleviating power division limitations, though heterodyning would be very useful in implementing a frequency multiplexing scheme. The same statement could be made concerning the use of narrowband optical filters or frequency-selective waveguide couplers.

Of the different devices and techniques that we considered, clearly the most useful one (in network applications) is the optical amplifier. This device is capable of increasing N_{max} for the bus by orders of magnitude, regardless of whether direct or heterodyne detection is used. An observation of great practical importance was that the gain of the individual amplifiers did not need to be extremely high, if we were willing to use enough of them.

As this Thesis comes to a close, there are at least two topics that deserve further investigation. One topic concerns the dynamic assignment of bandwidth among users. In Chapter 5, we demonstrated that an order wire run in a Pure Aloha manner was a viable means of providing coordination in the network. However, we did not discuss the actual algorithm that would be used to assign bandwidth, nor did we investigate alternatives to the Aloha order wire approach to control.

A second topic that we avoided was the use of spread spectrum techniques as a multiple access method in local networks. As with frequency multiplexing and spatial multiplexing, spread spectrum could conceivably be used either to stabilize unstable networks or to simplify control. Spread spectrum will of course have its own set of implementation issues, and a comparison between it and the multiple access methods discussed in this Thesis should be worthwhile.

References

1. Andrew Tannanbaum, Computer Networks. Englewood Cliffs: Prentice Hall, 1981.

2. K. Kummerle and M. Reiser, "Local Area Communication Networks - An Overview," Journal of Telecommunication Networks, Vol. 1, #4, 1982, pp. 349-370.

3. D. Clark, K. Pogran and D. Reed, "An Introduction to Local Area Networks," Proceedings IEEE, Vol. 66, #11, 11/78, pp. 1497-1517.

4. E.G. Rawson, R.M. Metcalfe, R.E. Norton, A.B. Nafarrate and D. Cronshaw, "Fibernet: A Fiber Optic Computer Network Experiment," Proceedings Fourth European Conference on Optical Communication, Genova, Italy, September 1978.

5. R.V. Schmidt et al, "Fibernet II: A Fiber Optic Etnernet," IEEE Journal on Selected Areas in Communications, Vol SAC-1, #5, 11/83, pp. 702-710.

6. R. Neff and D. Senzig, "A Local Network Design Using Fiber Optics," Digest of Papers of Spring Compcon 1981, pp. 64-69.

7. I.T. Frisch, "The Evolution of Local Area Networks," Journal of Telecommunication Networks, Vol. 2, #1, 1983, pp. 7-23.

8. C. Yeh and M. Gerla, "High Speed Fiber Optic Local Networks," Proceedings of the Tenth Anniversary Meeting of the National Science Foundation Grantee-User Group in Optical Communications Systems, June 1982, pp. 69-80.

9. C.W. Tseng and B.U. Chen, "D-Net - A New Scheme for High Data Rate Optical Local Area Networks," IEEE Journal on Selected Areas in Communication, Vol. SAC-1, #3, 4/83, pp. 493-499.

10. K. Ogawa, "Considerations for Single Mode Fiber Systems," Bell System Technical Journal, Vol. 61, #8, 10/82, pp. 1919-1931.

11. J. Yamada and T. Kimura, "Single Mode Optical Fiber Transmission Experiments at 1.3μ Wavelength," Review Elect. Comm. Lab., #27, July-August 1979, pp. 611-629.

12. T. Kimura, "Single Mode Systems and Components for Longer Wavelengths," IEEE Transactions on Circuits and Systems, Vol. CAS-26, #12, 12/79, pp. 987-1010.

13. D.E. Payne and W.A. Gambling, "Zero Dispersion in Optical Fibers," Electronics Letters 11, #8, 4/75, pp. 176-178.

14. K. Ogawa, "Analysis of Mode Partition Noise for Laser Diode Systems," IEEE Journal of Quantum Electronics, Vol. 17, #5, 5/82, pp. 849-855.

15. K. Okamoto et al, "Dispersion Minimization in Single Mode Fibers Over a Wide Spectral Range," Electronics Letters 15, #22, 10/79, pp. 729-731.

16. L.M. Branscomb, "Networks for the Nineties," IEEE Communications Magazine, 10/83, pp. 38-43.

17. S.D. Personick, private communication.

18. M. Stahlman, "Inside Wang's Local Net Architecture," Data Communications, January 1982, pp. 85-90.

19. S.D. Personick, Fiber Optics - Technology and Applications, to be published.

20. M. Barnoski, Fundamentals of Optical Fiber Communications, Second Edition. New York: Academic Press, 1981.

21. W. Bux, "Local Area Subnetworks - A Performance Comparison," IEEE Trans. Comm., Vol. COM-29, #10, 10/81, pp. 1465-1473.

22. R.M. Metcalfe and D.R. Boggs, "Ethernet: Distributed Packet Switching for Local Computer Networks," Comm. ACM, Vol. 19, 7/76, pp. 395-404.

23. N. Abramson, "The ALOHA System--Another Alternative for Computer Communications," Proc. FJCC, 1970, pp. 281-285.

24. N. Abramson, "The ALOHA System," Computer-Communication Networks, N. Abramson and F Kuo editors, Englewood Cliffs: Prentice Hall, 1973.

25. J.H. Shapiro, "TCS Waveguide Tap with Infinitesimal Insertion Loss," Optics Letters, Vol.5, #8, 8/75.

26. H.A. Haus, Waves and Fields in Optoelectronics, Englewood Cliffs: Prentice Hall, 1984.

27. A.W. Drake, Fundamentals of Applied Probability Theory, New York: McGraw Hill, 1967.

28. J.H. Shapiro, "Lecture Notes for 6.453 - Optical Detection and Communication," unpublished.

29. Y. Yamamoto, "Receiver Performance Evaluation of Various Digital Optical Modulation-Demodulation Systems in the 0.5 - 10 μm Wavelength Region," IEEE Journal of Quantum Electronics, Vol QE-16, #11, 11/80, pp. 1251-1259.

30. R.G. Gallager, Information Theory and Reliable Communication, New York: Wiley & Sons, 1968.

31. H.P. Yuen, "Two Photon Coherent States of the Radiation Field," Phys. Rev. A, Vol. 13, pp. 2226-2243 (1976).

32. H.P. Yuen and J.H. Shapiro, "Optical Communication with Two-Photon Coherent States--Part I: Quantum state propagation and quantum noise reduction," IEEE Trans. Inform. Theory, Vol. IT-24, 11/78, pp. 657-668.

33. J.H. Shapiro, H.P. Yuen and J.A. Machado Mata, "Optical Communication with Two-Photon Coherent States--Part II: Photoemissive detection and structured receiver performance," IEEE Trans. Inform. Theory, Vol. IT-25, 3/79, pp. 179-192.

34. H.P. Yuen and J.H. Shapiro, "Optical Communication with Two-Photon Coherent States--Part III: Quantum measurements realizable with photoemissive detectors," IEEE Trans. Inform. Theory, Vol. IT-26, 1/80, pp. 78-91.

35. M. Sargent III, M. Scully, and W. Lamb, Laser Physics, Reading MA: Addison Wesley, 1974.

36. A. Yariv, Quantum Electronics, Senond Edition, New York: Wiley, 1975.

37. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering, New York: Wiley, 1965,

38. H.L. Van Trees, Detection, Estimation and Modulation Theory, Part I, New Yory: Wiley, 1968.

39. Y. Yamamoto and T. Kimura, "Coherent Optical Fiber Transmission Systems," IEEE Journal of Quantum Electronics, Vol. QE-17, #6, 6/81, pp. 919-934.

40. S. Saito et al, "S/N and Error Rate Evaluation for an Optical FSK-Heterodyne Detection System Using Semiconductor Lasers," IEEE Journal of Quantum Electronics, Vol. QE-19, 2/83, pp. 180-193.

41. S. Saito et al, "Oscillation Center Frequency Tuning, Quantum FM Noise, and Direct Frequency Modulation Characteristics in External Grating Loaded Semiconductor Lasers," IEEE Journal of Quantum Electronics, Vol. QE-18, #6, 6/82, pp. 961-970.

42. R. Alferness, "Guided Wave Devices for Optical Communication," IEEE Journal of Quantum Electronics, Vol. QE-17, #6, 6/81, pp. 946-959.

43. R. Gagliardi and S. Karp, Optical Communications, New York: Wiley, 1976.

Appendix A

Variation of Coupling Coefficients Along Bus

Equation (3.4) was derived assuming that all waveguide taps had identical coupling coefficients K. In this Appendix we explore the possible improvement in N_{max} to be gained from allowing the coupling coefficients to vary along the bus. The optimum distribution of power among users is an <u>even</u> distribution, in the sense that an even distribution makes the worst case received power as favorable as possible. If all of the coefficients along the bus were identical, as they were assumed to be in Chapter 3, the distribution of power will be far from even. We now derive an expression for K values that yield an even power distribution among users. We then investigate the increase in N_{max} that would result if such coefficients were used.

With the couplers (and users) numbered left to right, let S be the signal power entering port a of coupler i-1. Let P_{i-1} and P_i be the power received at port b' of users i-1 and i, respectively. Then

$$P_{i-1} = SK_{i-1}\gamma \tag{A.1}$$

$$P_i = S(1 - K_{i-1})K_i\gamma^2 \tag{A.2}$$

where K_{i-1} and K_i are the respective coupling coefficients, and $\gamma = 10^{-\alpha/10}$. For equal division of power, we set $P_{i-1} = P_i$, yielding

$$K_{i-1} = \frac{K_i \gamma}{1 + K_i \gamma} \qquad (A.3)$$

a recursion formula for the coupling coefficients, with initial condition $K_{N'} = 1$, where N' is the total number of taps on the bus. The solution to (A.3) is

$$K_i = \gamma^{N'-i} \frac{1 - \gamma}{1 - \gamma^{N'-i+1}} \qquad (A.4)$$

Say that a certain amount of signal power P is available on the main bus, to the left of user 1. The power extracted off the bus by user N' is

$$P_r = P\gamma^{N'} \prod_{i=1}^{N'-1} (1 - K_i)$$

$$= P\gamma^{N'} \left[\frac{1 - \gamma}{1 - \gamma^{N'}}\right] \qquad (A.5)$$

Going back to Eq.(3.3), we see that the corresponding equation in the case where all K's are identical is

$$P_r = 2PN^{-1}(1 - 2/N)^{N-1}\gamma^N$$

$$\approx 2Pe^{-2}N^{-1}\gamma^N \qquad (A.6)$$

Note that N' refers to the number of users on the bus whose coefficients vary, and N refers to the number of users on the bus whose coefficients are identical. Let us compare N' and N on the basis of equal P_r/P, i.e. equal attenuation. We suspect that N' will be higher than N, because the distribution of power is more optimum on the bus whose coefficients vary according to (A.4).

Setting Eq.'s (A.5) and (A.6) equal to each other gives, assuming $\gamma^{N'} \ll (1-\gamma)$,

$$N' \approx N + (|\log\gamma|)^{-1} \log N \qquad (A.7)$$

From (A.7), we see that the increase in N_{max} is proportional to the log (base 10) of the old value. Because of this logarithmic dependence, the improvement to be gained from even power distribution will be quite limited. In other words, the low N_{max} values cannot be blamed on suboptimum distribution of power among users. The dominant limitation is the coupler excess loss, at least for α greater than a few tenths of a dB.

Appendix B

Use of Low Gain Amplifiers

Consider once again the bus of Figure 3.7a. In Chapter 3 we saw that the total amplifier noise at the right end of the bus (the worst case) had a spectral height $nh\nu_o$, where n was the number of amplifiers. In obtaining this result we assumed that the noise w(t) of Figure 3.8 had a spectral height $h\nu_o$. This assumption is valid only when the amplifier gain G_a is much greater than one. We now examine the case in which the gain is low.

In general, the (power) gain of the amplifier is given by

$$G_a = \exp(\gamma z) \tag{B.1}$$

where z is the length of the medium in the direction of propagation and[36]

$$\gamma = (N_2 - (g_2/g_1)N_1)K(\nu) \tag{B.2}$$

where N_2 and N_1 are the populations of the upper and lower levels, respectively, g_2 and g_1 are the degeneracies of the two levels, and $K(\nu)$ is a function (independent of N_1 and N_2) that depends on the atomic lineshape and other factors. It has also been shown[36] that the amplifier noise has spectral height (referred to the amplifier input)

$$q = h\nu_o \frac{N_2}{N_2 - gN_1}(1 - e^{-\gamma z}) \qquad (B.3)$$

where $g = g_2/g_1$. In the high gain case ($N_2 >> N_1$), Eq. (B.3) reduces to the expected $q = h\nu_o$.

We now compare the noise levels at the right end of the bus in the low gain amplifier and high gain amplifier cases. We have that

$$\text{noise(high gain case)} = nh\nu_o = N_h \qquad (B.4a)$$

$$\text{noise(low gain case)} = n'ah\nu_o = N_\ell \qquad (B.4b)$$

where

$$a = \frac{N_2'}{N_2' - gN_1'}(1 - (G_a')^{-1}) \qquad (B.5)$$

Note that we use primed notation to indicate quantities associated with the low gain amplifier bus. Since the total loss end to end on the bus (in the absence of the amplifiers) is the same in both cases, it follows that

$$(G_a')^{n'} = G_a^n$$

$$n' = n\frac{\ln(G_a)}{\ln(G_a')} \qquad (B.6)$$

The ratio of the noises is then

$$\frac{N_\ell}{N_h} = a \frac{\ln(G_a)}{\ln(G_a')}$$

$$\approx \frac{\ln(G_a)}{G_a'} (1 - g(N_1'/N_2'))^{-1} \qquad (B.7)$$

From (B.7) we see that the noise is increased when we go to low gain amplifiers, i.e., the SNR at the end of the bus will be lower than that found in Chapter 3. To find the exact noise level, we need knowledge of g and $K(\nu)$ for the specific medium to be used. With this information, (B.1) - (B.3) can be evaluated.

In general, then, we may conclude that the graphs in Chapter 3 are accurate for $M \gg 1$ ($M > 8$ should be sufficient), but because of the increased noise in the low gain case, N_{max} for low M will be smaller than that predicted by the analysis of Chapter 3.