# SATELLITE COMMUNICATIONS NETWORKS

by

Terrence P. McGarty
Computer Sciences Corporation
Falls Church, VA

First Draft: July, 1981

CONTENTS

# CHAPTER 1
# INTRODUCTION

Communications systems provide for the interconnection of various forms of information generators and ensure that the transportation of that information is performed in an accurate and timely fashion. As the demands for information have increased the distribution of it both on a local and global scale has called for new and sophisticated technologies. One of the newest of these is that of satellite communications which started as a commercial entity in 1962 with the signing of the Satellite Communications Act.

The initial systems were quite simple with the prime focus on the satellite with its highly sophisticated technology base. The major issues in the system were in turn technologically driven with initially high entry costs and minimal network complexity. As evidenced in the early literature, the focus of activity was to get the message through and this implied that the satellite could be treated merely as a repeater within the terrestrial network. The terrestrial network, however, had significantly different characteristics and capabilities than the satellite channel.

Thus for years the full potential of the satellite was not employed and the systems were not fully integrated within the terrestrial networks. In addition the concept of augmented network based on satellites first and terrestrial plant second did not evolve until the late 1970s with the advent of such concepts as SBS, XTEN, and others. As will be brought to the readers attention throughout the text, one of the more serious impacts on the development of any such set of systems in the regulatory environment that exists.

It is this environment and its lists of limitations and adversary proceedings that will delimit any extensive commercial growth. In the military area, however, such limits are of a different form, based upon interdepartmental requirements and funding.

The system architectures developed in this text are therefore based both upon what exists and what could be in a technical context. Consideration is given to such factors as utilization and cost, with the discussion of policy relegated to a footnote status.

## 1.1 Generic System Structure

The purpose of any communication network is to efficiently and in a cost effective manner interconnect a wide variety of user sources located at various geographical locations and provide highly reliable and expandable service. This purpose is often overlooked by the designers who may focus their attention on some subsections of the system design.

The basic elements that must be considered in a system design are as follows:

i. Sources - these are the basic information generators that produce the signals that are to be transmitted across the network. There are a wide variety of sources including voice, data, video, facsimile, and others. Associated with these sources are interfaces which may be electrical, mechanical, and logical. This implies that there must be a consistency or transparency across the system that permits for such things as protocol interface, E&M lead signalling, timing, and synchronization.

ii. Interconnection - this factor determines what sources are to be connected to each other. The driving forces here are the time factors, protocols, and other compatibilities. There are two major time factors that impact on the design; short term and long term. Short term specifies use of the available transmission capacity on an almost real time basis, e.g., the sharing of the assets. The network must be capable of near term reconfigurability to adapt to near-instantaneous traffic changes. Long term changes relate to growth in the number of users and gross changes in interconnectivity. For example, the international traffic growth is at a 12% to 15% annual rate with many new modes occurring. Thus the system must be able to accommodate these factors.

iii. Cost - for both commercial and government systems, cost plays a dominant role. Large front-end capital investments are required in satellites, earth stations and other network hardware. From the commercial aspect the cash flows associated with revenue and expenses can be signficantly impacted upon by changes in system design and network configuration. Thus, it will be important to develop accurate cost models and optimize designs around cost criteria.

The above items are considered in terms of the generic system model of the satellite communication network. This network is shown in Fig. 1 and is composed of a basic set of building blocks. The inputs and outputs of the networks are sources or other networks. These other networks may be terrestrial networks or even other satellite networks. An example of the latter would be a domestic satellite network with an interconnect with the international network operated by INTELSAT. The sources may be geographically distributed and in fact may be combinations of various types of information such as integrated voice and data.

There are four basic levels to the communications network architecture.

i. Network Interface - these units are distributed interface units which coordinate the information from each of the sources and place it into a format that is useful for satellite transmission. It exercises local control for access purposes and matches electrical, mechanical, and logical interfaces. The network interface units could even communicate between each other at their own level.

each

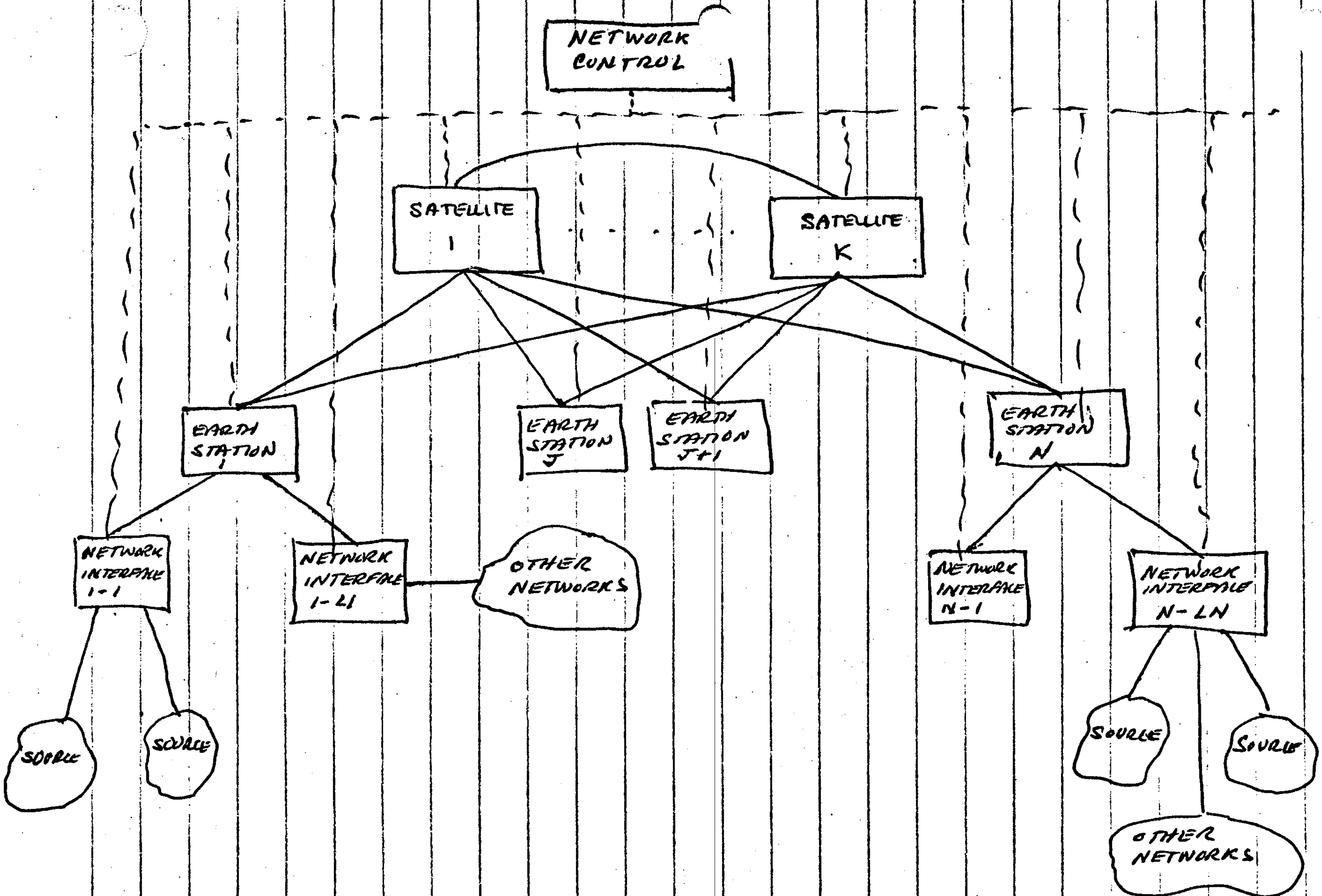



FIG 1 : Generic Satellite Communications Network Architecture

The basic elements of the network interface are depicted in Fig. 2. We have implicity assumed that local protocol control has been taken care of either within the source itself or passed through on a parallel channel. The major elements then are

- Source Coder used to compress the data where possible. This device recognizes statistical regularities and removes redundancy. An example would be a vocoder used on speech.

- Encryption is used to ensure that the data from each source is held secure. Various encryption devices are presently commercially available and these are used on a source by source basis.

- Multiple Access devices provide for the shared use of the transmission links. This multiple access system may allocate real-time on demand or it may be static. It may be centrally controlled or controlled in a distributed fashion. In addition it may work with the processor that is on-board the satellite to fully integrate all sources into the network.

- Channel Coding provides for error protection on the transmission channel. It is generally used for a digital communications link where the encryption is done on a digital bit stream. The coding scheme usually introduces a ~~brown~~ known amount of redundancy into the data stream and upon reception decoding this is an optimum fashion. ? restate sentance

- Modems are used to convert the baseband signals up to a level and into a signalling format that can be used for transmission through the satellite. This may be accomplished with either an analog or digital techniques.

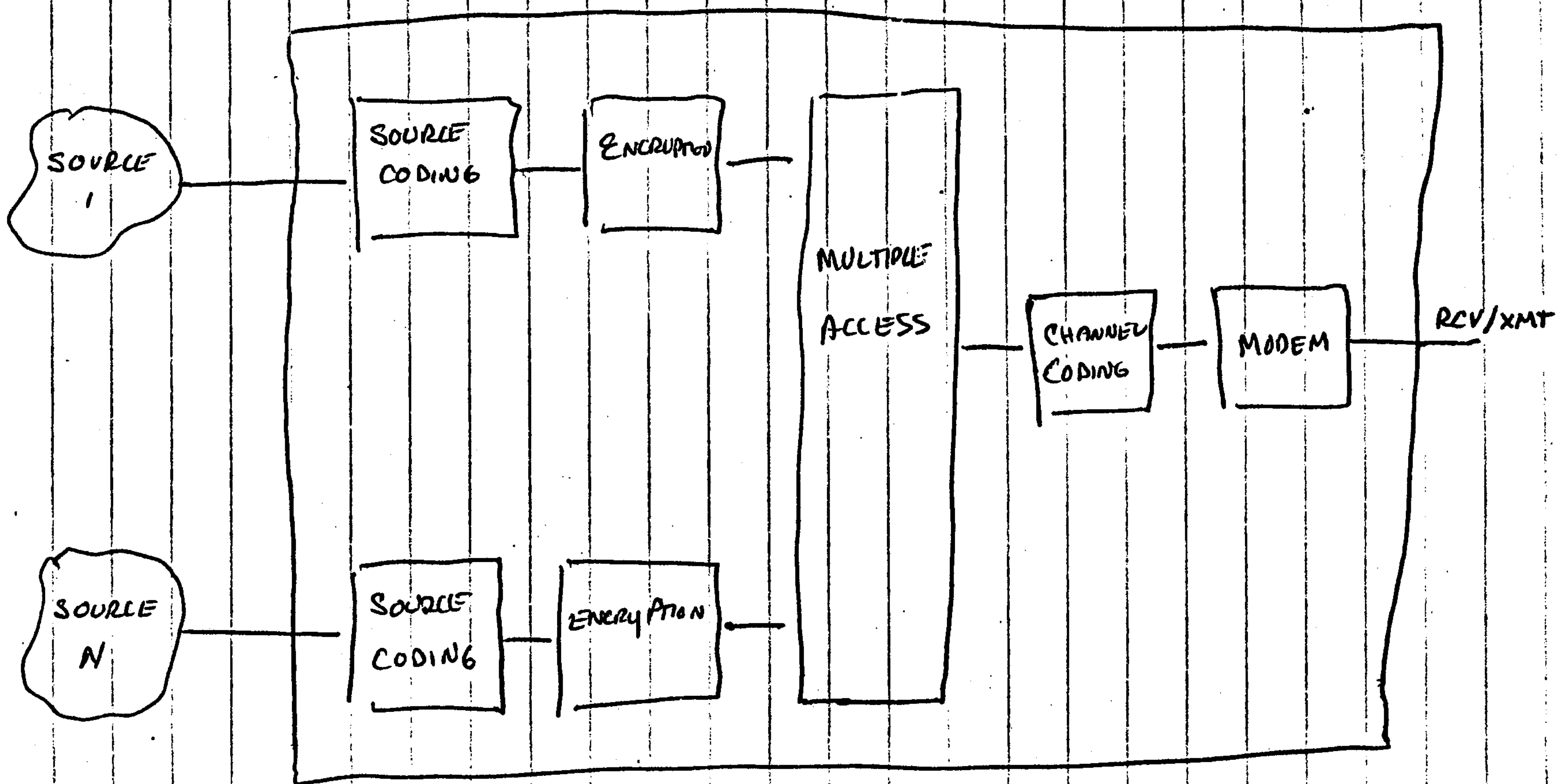



FIG 2 : Structure of Network Interface

ii. <u>Earth Station</u> - this provides for the interface with the transmission link, and it changes signals that are capable of movement on terrestrial facilities alone to those propagated through the atmosphere. Figure 3 depicts the standard structure for an earth station.

There is a transmit and receive leg with up and down converters respectively. These are oscillators that shift the signals center frequencies into the bands that are appropriate. The transmitter and receivers are basically amplifiers at high and low levels respectively. The antenna is used for both transmit and receive.

There are today a wide variety of earth stations ranging from small 1m. dishes used in maritime communications (even smaller at UHF frequencies) up to 35m. INTELSAT Standard A antennas. In terms of data rates they range from the tens of bits per second to well in excess of 100s of Mbps. ✓

iii. <u>Satellite</u> - this represents the key node in the overall network. Historically the satellite acted merely as a repeater, that is ~~it~~ amplified and frequency shifted an up link carrier and redirected the signal down to the earth. The geosynchronous satellites, by rotating at the same rate as the earth, remain relatively stationary and can cover almost one third of the earth's surface. rewrite! ✓

Fig. 4 depicts the communications section of a typical satellite. The antenna, receiver, transmitter, and up/down converters are functionally similar to those in earth stations. In most satellites to date that is all that is necessary. However, with added complexity on on-board processors using a mod/demod and encode/decode capability can be employed. This processor becomes an intelligent switching node. It can route, sort, process, monitor and control system traffic as well as allocate overall system resources. The extent to which this can be done in reality however depends upon two major factors;

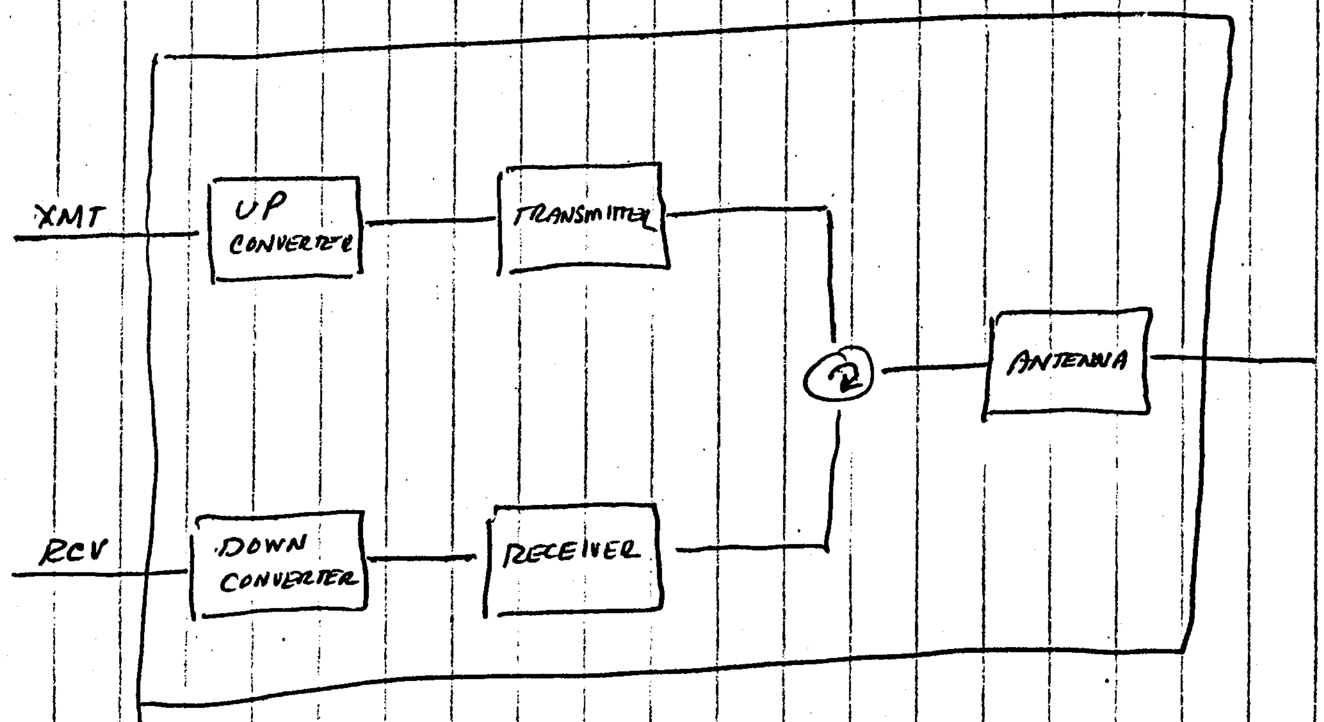



FIG 3 : Structure of Earth Station

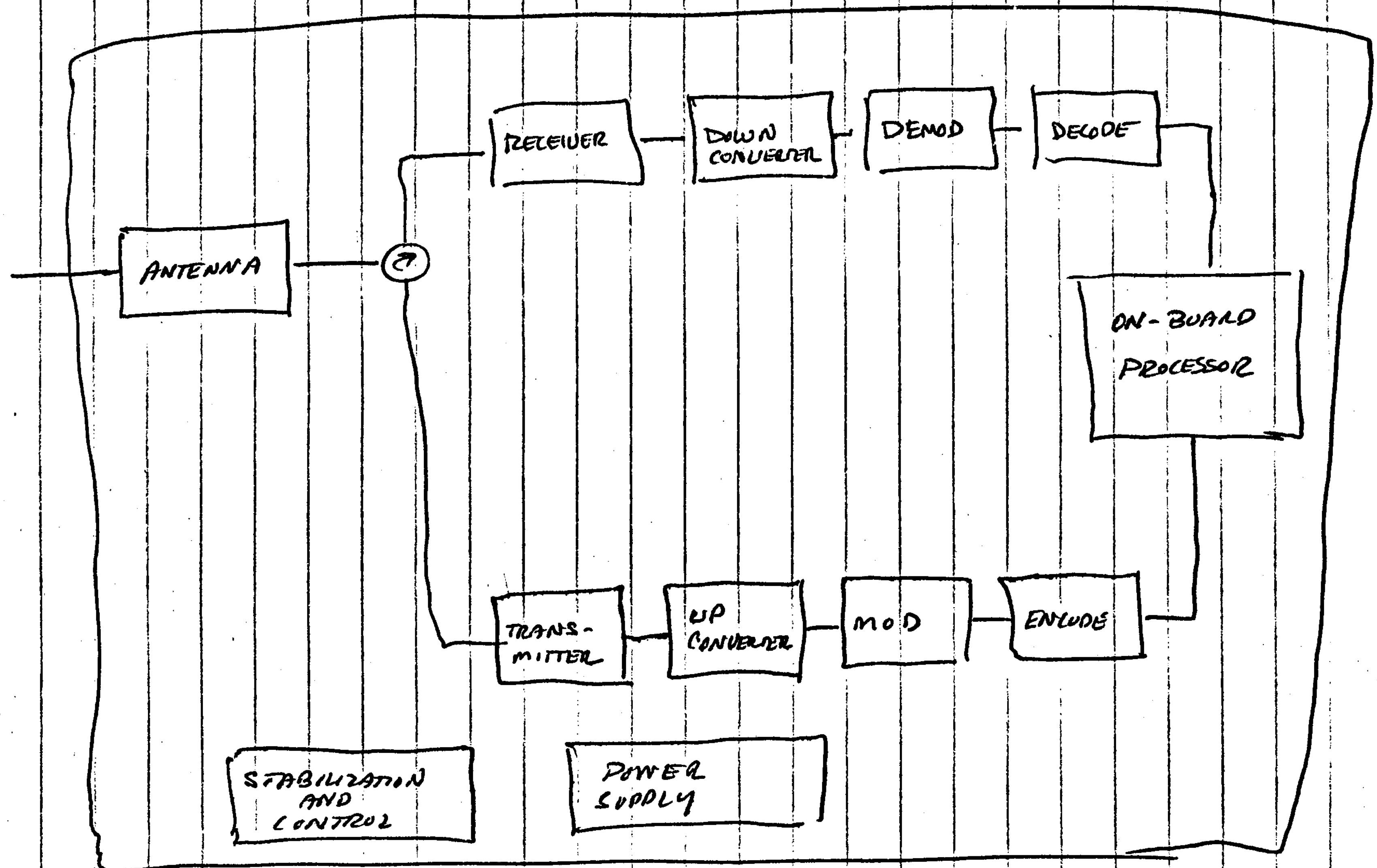



FIG 4 : Structure of Satellite

reliability and cost. To date a limited amount has been accomplished as in the case of the experimental LES 8/9 satellites (Lincoln (MIT) Experimental Satellites).

The other portions of the satellite systems include power sources, stabilization and control, bus structures, TTC and M (Telemetry, tracking Command and Monitoring) systems. Although important they do not interact heavily upon the communications network.

iv. Network Control - supplies overall network monitoring, command and control. This area is rapidly evolving particularly in view of the importance of overall network utilization. In the commercial areas the Intelsat system has a CSM (Communication System Monitor) which is a very simple monitoring network that measures power, frequency, and bandwidth occupied and does so only for analog systems. Since this network is almost totally pre-assigned and over capitalized due to its international funding, it can afford such. In contrast, the proposed SBS (Satellite Business Systems) network has a combined centralized and distributed control using a SMF (System Monitoring Facility) and SCC (Satellite Communications Controller) to monitor and control its demand assigned all digital network. The U. S. military is developing a system called RTACS (Real Time Adaptive Control Systems) to go a step further in assigning in an integrated voice/data, analog/digital multi-user network.

The network control area is evolving rapidly and with the development of low cost distributed processing will most probably evolve into such a fully distributed area.

These four major units comprise a satellite communications network. The specific choice of design elements and overall architecture will depend upon the three major factors of constraints, requirements, and cost. The constraints are these parameters within which the network must operate. The requirements are those factors that it must equal and the cost is that factor that must be minimized. The approach taken in this text is to quantify these wherever possible and develop a methodology that permits an overall network synthesis.

## 1.2 Services

Before describing the many existing and planned systems, it is useful to put such systems in an overall perspective. Any system results from a logical progression of needs through cost effective implementation. The development process typically considers the following phases.

i. Market - a careful identification of who the services are to be provided to. This includes a detailed definition of their locations, the prices they are willing to pay relative to the utility (usually marginal utility) of the service, the growth in their needs on some time scale and what alternatives they have at hand.

ii. Services - the services are the types of communications needs that are satisfied by the system to the customer. As an example would be switched voice networks, high speed data, etc. The services are usually defined by levels of performance, interconnectivity, reliability and availability, capacity and interoperability.

iii. Systems - these are physical embodiments in architectural sense that provide services to a set of markets. No system is unique in its structure. They are, however, built up on certain canonical units as described in the previous subsection. In each block, there exists alternatives that can be characterized eventually on a cost/performance basis.

iv. Technology - represents the actual physical alternatives available to realize the units in the system architecture. For example antennas can be specified from a system viewpoint by gain, weight (mass), beam isolation and other factors. The technology then would admit realizations with reflector, lens, or array antennas.

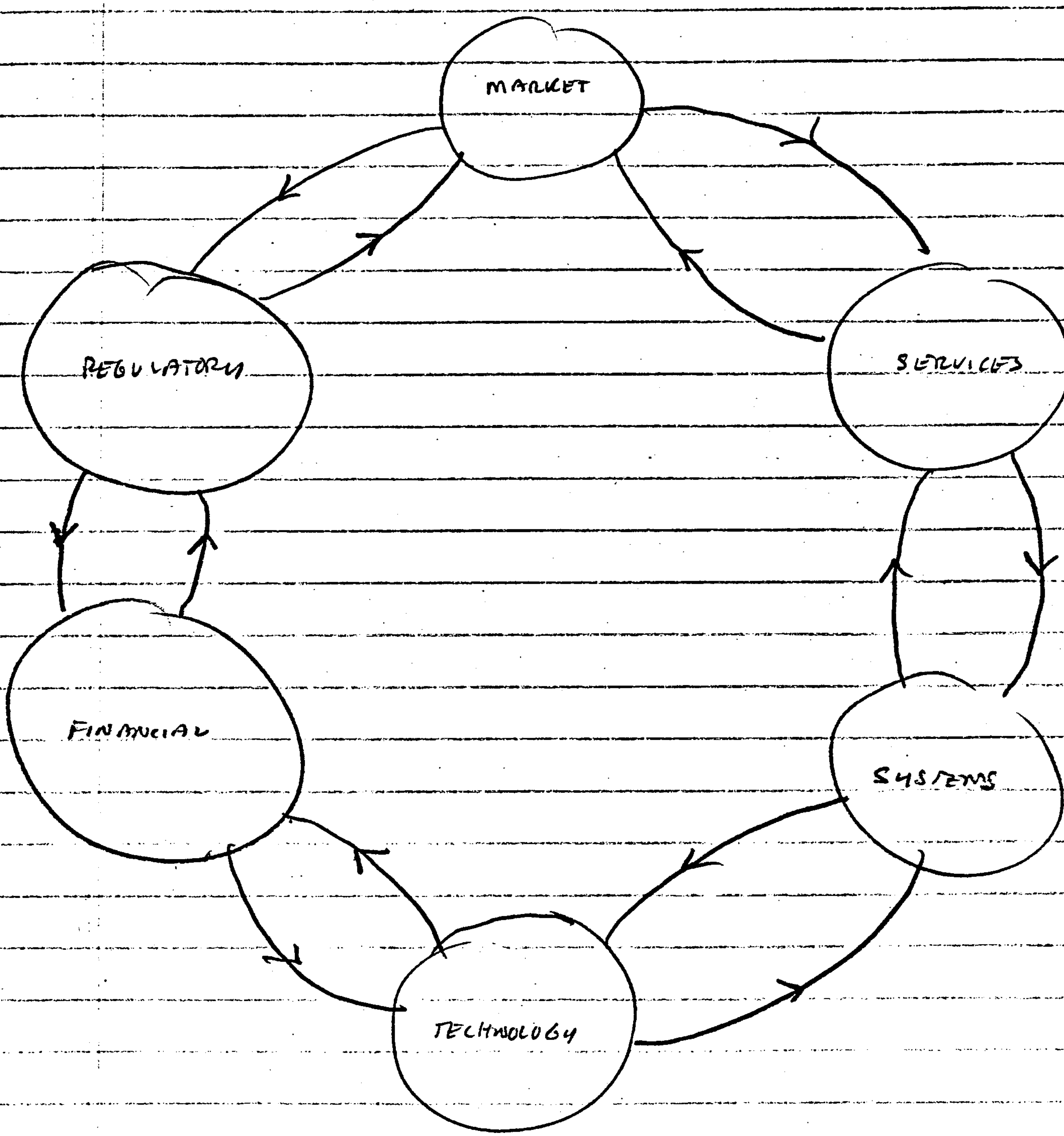



FIG 5: Development Phases and Methodology

v. <u>Cost</u> - the economic factors are most critical. System cost can be determined based on cost/performance measures for each of the system elements based upon a selected technology. The full economic analysis of a system, however, depends upon revenue and expenses to determine such factors as return on investment, internal rates of return and cash flow profits. Thus, a quantified market and system are required.

vi. <u>Policy</u> (REGULATORY) - communications is a regulated business and as any such business logic and quantitative measures are often meaningless in view of the rules and regulations that control both domestic and international areas. A discussion adequate for the reader is well beyond the scope of the text however references abound.

This text directs itself toward the development of systems and the modelling of costs. We relegate the market and plethora of services to those whose task it will be to satisfy the ever growing requirements in this expanding field. However, it is worthwhile to briefly outline some of these services within a categorical context of applications.

1. <u>Telephony</u>

Telephony, defined as the transmission of voice from point to point, has been the basis for most of the initial sets of requirements in satellite communications. The INTELSAT series of satellites, as discussed in the previous section, provide voice service on a global scale from a group concentrator to another group concentrator, i.e., a trunking service. With the advent of the domestic systems, various other modes of connectivity are being used.

2. <u>Video and Radio Broadcast</u>

The video and radio broadcast category of services is concerned with the distribution of video or radio programming from a central source to a number of users. These services are characterized by their one-to-many connectivity nature and simplex transmission requirement. Typical systems involve one or a few transmission sites and many receive-only earth stations.

### 3. Data and Wideband Services

The data and wideband services category includes all digital data services, except mobile, and wideband non-telephony analog channels, which are arbitrarily defined as channels with bandwidths greater than 4 kHz. This category of services is associated with a very heterogeneous set of applications ranging from computer networking to point-to-point video transmission. Some of the services in this category might well be integrated into a telephony-based satellite communications system. This is, indeed, the case today for the INTELSAT system and some services, e.g., voice-band compatible data services and some wideband analog services, for several of the U. S. domestic systems, and is planned for the proposed SBS system. As digital techniques become more prevalent in the transmission of telephony, the distinction between telephony and some forms of data communication, e.g., switched data service, will become increasingly blurred.

### 4. Mobile

Communications to mobile platforms, land, air or sea, is a service that can be uniquely provided by either radio or satellite. Considerable effort has been spent by the U. S. military in this area and several experimental and operational systems have been deployed. Recently, commercially operational maritime services have become available in the Atlantic and Pacific Oceans via the MARISAT (COMSAT General Corporation) satellites.

What characterizes the mobile communications area is the large number of users, the small size of user antennas allowed due to the mobility and agility of the vehicle, the limited terminal power, the required high terminal reliability and the noise and interference environment. In addition, in order to be a viable market, the cost of a user terminal should be kept at a minimum.

In all areas, land, sea and air, there is an existing mode of communication that provides limited service. In most cases, the need for increased and/or improved communication services is recognized as a useful method of increasing operational efficiencies and reducing costs. However the actual introduction of satellite communications into the commercial market is cost and service sensitive.

Define an orderwire!

- DEDICATED PRIVATE LINE (U-U) - *A voice circuit connecting two users on a full time basis. An example would be an orderwire.*

- NON-DEDICATED PRIVATE LINE (U-U) - *This service would provide a direct user-to-user connection for telephony on a demand assigned basis. The network switching would be provided in a distributed fashion throughout the network, i.e., at each earth station. Such a network would be useful in devloping countries or those areas with few and scattered users.*

- THIN LINE (U-LC) OR ACCESS (U-LC, U-GC) - *This service would provide access of remote users to a larger network, normally on a demand assigned basis.*

- PORTABLE ACCESS (U-LC, U-GC) - *This service involves the use of a portable earth station to provide a telephony connection into a major network for areas not customarily provided service or needing service on a temporary basis.*

- SWITCHED NETWORK (LC-LC) - *This service, involving only a two level hierachy, would provide full interconnection between several (>2 ) sets of users grouped on their switches. The transmission capacity associated with the individual links interconnecting the local concentrator could be either dedicated or dynamically allocated on the basis of demand.*

- PRIVATE NETWORK (LC-LC, LC-GC) - *For LC-LC connectivity, this service essentially corresponds to the provision of the switched network service within a closed community, e.g., intra-company communications. If centralized rather than distributed switching is involved then LC-GC connectiv is implied with the centralized switch being the group concentrator.*

- <u>SMALL PBX (LC-LC)</u> - *This service provides the interconnection of a concentration of users into a large network, e.g., the direct distance dialing (DDD) network. It could also provide the interconnection of a private network into a public switched network.*

- <u>TRUNK (GC-GC, LC-GC)</u> - *In this service, long haul transmission capacity is provided on a dedicated basis for high density traffic originating and terminating on large centralized switches.*

- <u>INTERNATIONAL DIRECT DISTANCE DIALING (DDD) (LC-GC)</u> - *Both switched and private networks could request direct access into gateways for international service. The gateways would be earth stations possibly in the INTELSAT system. Such direct connections could minimize interconnect costs.*

- <u>WIDE AREA DIALING (LC-GC)</u> - *As with international DDD, this service would provide interconnect between various switched or private networks, but possibly on a domestic or regional scale.*

Table 1. Telephony Services

- NETWORK DISTRIBUTION - *Broadcast distribution from a network or national center to local stations for re-broadcast over normal facilities. Examples are commercial and national networks.*

- NETWORK INTERCONNECT - *Program origination and broadcast from a local station to other local stations for re-broadcast. The assumption is that the programming does not go to a network center for re-broadcast. Examples are commercial or national networks and specialized or special event networks.*

- DISTRIBUTION TO LIMITED RE-BROADCAST STATION OR DISTRIBUTION NETWORKS - *Distribution from a network center or national source to local communities. This service may involve the use of a community antenna and very localized re-broadcast (isolated communities), or may be to a cable distribution system head-end. Examples include the TELESAT system and the Home Box Office distribution of TV programming to cable head-ends using the WESTAR satellite.*

- DIRECT BROADCAST - *Broadcast signal directly available to user (on small receive-only antenna). As well as entertainment distribution, this service may be used for a variety of educational and training purposes.*

- INTERNATIONAL DISTRIBUTION - *Broadcast distribution to national centers for redistribution. This service includes coverage of events of international interest and regional programming (e.g., Arab Union).*

Table 2: Video and Radio Broadcast Services

- <u>DATA TRUNK</u> - *Dedicated digital link connecting two local concentrators (switches) or group concentrators on a full-time (dedicated) basis.*

- <u>SWITCHED DATA SERVICE</u> - *Link actuated by data call. The link will typically be between two local concentrators, but it could be between two group concentrators. This service corresponds to demand access use of the satellite resource. The use of the INTELSAT SPADE system for data is an example of this type of service.*

- <u>DATA COLLECTION</u> - *Many-to-one connectivity with the primary data sources (users) typically being relatively low data rate and low duty cycle. The data sink is typically a centralized resource such as a large storage or data processing facility. Some topologies will concentrate a collection of data sources at an earth station. There is typically dynamically shared use of the satellite channel.*

- <u>DATA DISTRIBUTION</u> - *One to many broadcast of data. This service might be used to distribute data from a central location to users, perhaps after data collection and processing by a central collection point. Facsimile broadcast for remote publishing is another example application.*

- <u>SATELLITE DATA NETWORK</u> - *Implies dynamically shared use of a common communication resource and use of the broadcast property or satellite switching to obtain connectivity. If network is formed by use of point-to-point links connecting switches or concentrators, such use is considered under the data trunks or switched data service categories. The satellite network organization may range from fully distributed, with each user directly accessing the satellite, through varying levels of hierarchical structure with the satellite access occurring after some terrestrial concentration and/or switching. The satellite data network service may be organized around a centrally located resource, such as a large computer facility, or the network resources may be distributed throughout the network.*

- <u>TELECONFERENCING</u> - *Digital or analog video transmission among n users. May involve, for example, simultaneous availability of video signal from all participating sites at each site (fully connected network), transmissions from each site to a central site (conference director) with broadcast of a single transmission to all sites, or at any time a single site broadcasting to all other sites with the choice of the broadcasting site under the control of a central site (conference director). Other possibilities exist and the choice depends on the requirements of the specific application and many human factor issues, as well as economics.*

- PRIVATE LINE - Dedicated digital link connecting two users.

- DIGITAL ACCESS LINE - Digital link connecting user to a switch or concentrator. It may be a full period link or it may be demand assigned.

- WIDEBAND ANALOG CIRCUITS - Access and other circuits which transmit analog, non-telephony, signals whose bandwidth exceeds 4 KHz. Many of these signals will be imagery based, or from wideband sensors. A specific application is point-to-point distribution of video. The signals might be multiplexed together and sent over a single link.

Table 3. Data and Wideband Services

- <u>TELEPHONY</u> - *This entails the transmission of a voice signal from one node in the topology to another. This will typically represent only a single voice channel in each direction.*

- <u>FACSIMILE</u> - *The transmission of printed matter by means of a slow scan technique. Such things as map reproduction, photographic facial images and detailed graphical information are typical. The rate of transmission depends on the specific user and the sophistication of the encoding equipment.*

- <u>VIDEO</u> - *Video imagery consistent with TV transmission may be required in several markets. For most cases the quality required is not that of network TV and would be used for conferencing and other types of very specialized services.*

- <u>TRUNKING</u> - *This service entails the transmission of more than one voice channel between two locations. The higher voice channel density places added requirements on performance and access but potentially permits the use of DSI techniques to reduce capacity requirements.*

- <u>DATA TRANSFER</u> - *One mobile user may gather data that it needs to transfer to another user or LC or GC. Typically data transfer will be from a user to a local or group concentrator.*

- <u>DATA PROCESSING</u> - *When data is centrally stored and shared or when additional processing capability, above that locally available, is required, there is a requirement to transfer data, usually processed, from a central location to the user (or local concentrator). This service is essentially complementary to the data transfer service; it is treated separately because it may involve different rates and performance criteria.*

- <u>BROADCAST</u> - *This service involves the transmission of a message from a local or group concentrator to all or a subset of the users asspicated with it, by use of the broadcast capability of the satellite system. Such applications as disaster warnings would use this service.*

- <u>MONITORING</u> - *Equipment or environmental conditions can be continuously or aperiodically monitored on various mobile platforms. This service may entail many independent monitors that use a single transmitter to multiplex the signals. Such a service is typically only one way from user to local or group concentrator.*

- TELEX - *This service involves the transmission of character information and is typically of low data rate. It normally provides the user with hard copy which may be a requirement in certain applications.*

- RADIO - *This service corresponds to providing user to user connectivity with a minimally constraining access technique, e.g., random access. Examples of this type of service are amateur radio and citizen band (CB) radio.*

- NAVIGATION - *This service would provide the capability for mobile units to determine their locations autonomously. The service might be provided by a combined communication and navigation system. The navigation signal would typically be provided to the user from a fixed base station via the satellite or from the user himself via the satellite. The navigation service will typically require ranging systems which could require large bandwidths.*

*SURVEILLANCE - This service involves the maintenance of user location information by a central facility. It typically requires that the individual mobile units provide beacon responses to a base station. Surveillance may be initiated by the base station or it may be done in a free running fashion. This may be accomplished by means of an adjunct to the communication system or separate from it.*

Table 4. Mobile Services

## 1.3 Present and Pre-1980 Systems

At the present time, there are a number of communications satellites in use, providing service in the UHF, L, C, X, and Ku bands. The satellites are both spin and three-axis stabilized and are used for international, domestic, mobile military and other purposes. By the late 1980s, there will be an appreciable increase in the number of such systems as systems that are now under development become operational. These systems will provide the boundary conditions from which development in the 1980s will proceed.

### 1.3.1 INTELSAT

The INTELSAT system is presently composed of over one hundred members and users some of which are not members, e.g., U.S.S.R., and it provides global satellite communications services to three ocean regions.

At present, there are a total of almost 200 antennas, typically 30m dishes, that operate in the 4/6 GHz band. There are two operational satellites in the Atlantic plus and operational spare. There is one satellite plus a spare in both the Indian and Pacific Ocean regions. There are over 35,000 half-circuits in use worldwide (with approximately 22,000 in the Atlantic, 9,000 in the Indian, and 4,000 in the Pacific).* This is expected to require three operational satellites in the Atlantic, one or two in the Indian, and one in the Pacific.

---

*Note: A half-circuit or voice channel is defined as a one-way voice circuit. For telephony applications, since there is always a two way path with symmetric interconnectivity, a two way full duplex link, called a circuit, is often used as a measure of traffic. In this text where data, which may not be symmetric, is also being considered, we will use the concept of a channel or half-circuit. These two terms will be used interchangeably.

The INTELSAT IV series, first launched in 1971, had a capacity of 8,000 half circuits (channels) plus one TV transponder and a SPADE transponder and was launched on an Atlas/Centaur. The INTELSAT IV-A provides an increased operational capacity of 12,000 channels plus TV and SPADE transponders (in the Atlantic primary) and was first launched in 1975. The INTELSAT V, an Atlas/Centaur launchable satellite, which is illustrated in Figure 3, was launched in 1980 and in use in the Atlantic Ocean Region and has a capacity of almost 25,000 channels. All of these capacities are relative to use in an FM/FDMA (Frequency Modulation/Frequency Division Multiple Access) mode.

define beam/reuse!

The increased capacity of INTELSAT V is obtained through the use of dual polarization frequency reuse in the 4/6 GHz band plus beam isolation reuse in the same band. In addition, a 11/14 GHz capability has been included, with spot beam antenna coverage over North America and Europe in the Atlantic Ocean Region. The satellites will be channelized into 80 MHz and 40 MHz transponders (and a 240 MHz channel at 14/11 GHz). In addition to FM/FDMA service, the 80 MHz transponder in the INTELSAT V satellite will have the capability of providing Time Division Multiple Access (TDMA) service at 120 Mbps. The introduction of TDMA will, along with the use of digital speech interpolation techniques, provide an increase in the effective satellite capacity over that available in the FM/FDMA mode.

### 1.3.2 Domestic

Domestic satellite systems have been developing at a rapid pace over the past several years. The first such geosynchronous system was initiated by TELESAT of Canada with the ANIK satellite launched in 1972. This system is comprised of more than 80 earth stations and three satellites operating at 4/6 GHz, providing voice, video and data communication services.

deo

The system also incorporates the first operational TDMA link at a transmission rate of 61 Mbps. Both heavy route and thin route services are provided and a mix of earth segment equipment exists. On the thin route, the single channel per carrier (SCPC) multiple access technique is used with small earth stations having a G/T of 20 dB/K and an EIRP of 58 dBW. The future TELSAT satellites will have increased capacity provided by dual polarization, use of the 12/14 GHz bands or both.

define terms!

The U.S.S.R. has had, since the mid-1960s, and operational nonsynchronous system employing the Molniya satellites. These satellites are in elliptical orbit and use tracking earth station antennas.

Sat's use earth station antennas

In late 1975, the U.S.S.R. launched the first of the STATSIONAR series of synchronous satellites to be used for domestic service. A total of ten STATSIONAR orbital slots are planned, seven of which would be used for global coverage. Three of the satellites would be for domestic use only. All of the STATSIONAR satellites are expected to operate in the 4/6 GHz bands overlapping, in part, the bands that are presently used by the INTELSAT satellites.

In the U.S. there are presently four operational systems--Western Union's WESTAR, RCA's SATCOM, COMSAT General Corporation's COMSTAR which is leased to AT&T and will be used by both AT&T and GTE Satellite earth stations and the SBS system. Both WESTAR and COMSTAR are spin-stabilized satellites and all three operate in the 4/6 GHz band. The RCA SATCOM satellite is three-axis stabilized and was the first satellite to be launched on the Thor-Delta 3914 vehicle. WESTAR was launched on a Thor-Delta 2914 and COMSTAR on an Atlas-Centaur. In addition, the RCA and COMSTAR satellites use dual linear polarization for frequency reuse while WESTAR operates on a single polarization. All three of these systems currently operate into a few large ($\geq$10m.) antennas. However, smaller antennas ($\simeq$5m.) are being introduced into the RCA SATCOM System. See Table 1 for a comparison of the space and earth segment capabilities of the several systems.

The fourth system is Satellite Business Systems (SBS), a partnership between IBM, COMSAT General and AETNA Life and Casualty. This system uses 12/14 GHz band and operates into smaller (5 m.) rooftop antennas. Using this higher frequency band eliminates interference with existing terrestrial radio relays and allows on-premise facilities.

In addition to the above systems of owned earth and space segment there are other systems which use leased transponders to provide communication service. An example is the American Satellite Corporation which leases transponders from WESTAR and provides the earth segment.

Table 2 presents a synoptic summary of the present capabilities of the existing domestic systems. In this table operational transponders are those routinely devoted to carrying traffic, while spare transponders are those on-board operational satellites that are held in reserve to replace failed operational transponders. In-orbit spare satellites are used to back-up operational satellites, and do not carry traffic except for service which can be immediately pre-empted. One-way circuit figures are based on single carrier per transponder FM, saturated TWT operation, earth station antennas designated as "standard" by the system operator and the performance standard designated by the system operator. Other modes of operation may result in a considerable reduction in transponder capacity.

Algeria, the Arab States, Brazil, Colombia, India, Indonesia, Iran, Japan, Malaysia, and Nigeria are also operating, developing or seriously planning their own domestic or regional systems. The Algerian, Brazilian, Malaysian, and Nigerian systems have used leased transponders on the INTELSAT satellites. Norway is also using leased INTELSAT capacity to provide service to the North Sea oil rigs. As a typical example, the Algerian system leases a 40 MHz transponder from INTELSAT and has capacity for one TV and 65 single channel per carrier voice channels.

The Indonesian system using Hughes HD 333D satellites, has a capacity of 5000 voice channels or 12 TV channels. Among others both Brazil and Japan plan to use their own satellites in the near future. Such domestic systems can readily provide communications on a cost-effective and rapid basis that would be difficult to meet with extension of the existing terrestrial networks. Table 3 summarizes the currently operational non-U.S. domestic systems.

### 1.3.3 Mobile

Satellites have recently been successfully applied in the area of mobile communications with the launch of three MARISAT satellites in February, June, and October 1976. The MARISAT satellites, operated by COMSAT General, provide services to the U. S. Navy in the Atlantic, Pacific and Indian Ocean regions and commercial voice and teletype communications in the Atlantic and Pacific Ocean regions. The commercial portion of MARISAT is a C-band (4/6 GHz) earth station satellite link crossed to an L-band (1.5/1.6 GHz) satellite/ship link. The ship terminals use 1.3 m. tracking antennas. The military portion of the satellite uses frequencies in the lower UHF band. The satellite is a spin stabilized design using helical antennas for the UHF and L-band portions, as shown in Figure 4.

The European Space Agency (ESA) has developed a maritime satellite called MAROTS, to be located over the Indian Ocean. In contrast to MARISAT, the MAROTS satellite will be three-axis stabilized using the OTS (Orbital Test Satellite) bus. It will use an L-Band ship link, but will use Ku-band for the shore link.

1.3.4

## 1.3.4 Military Systems

Military satellite communication systems entered the 1970s, especially in the United States, using research and development assets as operational systems. Satellite communications has, however, demonstrated the unique capabilities that it can provide in both the strategic and tactical environments.

At the present time, excluding the U.S.S.R., there are three military systems. The U. S. system includes both UHF and X-band capability, while the British and NATO systems operate at X-band (7/8 GHz).

These systems are designed to satisfy the unique and vital requirements of the military as well as ensure the existence of a communication capability whose availability is not subject to political pressures to nearly the same degree as alternative systems. The military X-band systems are primarily used for strategic communications including intelligence, trunking, extension and restoral of terrestrial-switched systems, and, in the United States, support of the command and control requirements of the World Wide Military Command and Control System (WWMCCS). They provide some communication to major Navy ships and potentially provide communication to command aircraft. The UHF systems primarily provide tactical command and control communication, normally in a net environment, for mobile platforms, including ships, aircraft, and ground vehicles.

The military systems must satisfy certain requirements which tend to make them expensive and often technologically advanced. The operating environment is characterized as, both, very hostile and very volatile. Thus, the system should be able to provide secure communication to a large mix of earth terminal types with rapidly changing requirements, even in the advent of electronic warfare; or, in some cases, physical attack. Satisfying these requirements implies a need for such features as a secure and protected command and telemetry system, narrow-beam satellite antennas, satellite antennas with nulling capabilities, and an essentially real time control capability.

At the start of the 1980s, the U.S. strategic system, the Defense Satellite Communications System (DSCS), is composed of DSCS II and DSCS III spacecraft. Their characteristics are shown in Table 4 along with those of their predecessor DSCS I. This table indicates the increasing capability. A notable feature of DSCS III is its multibeam transmit and receive antennas which will provide great flexibility and some jamming protection.[11]

The characteristics of the important U.S. military X-band terminals are shown in Table 5.[12]

The NATO system will be composed of NATO III spacecraft, while the British SKYNET system, whose future evolution is uncertain, will have SKYNET III.[13]

The U.S. UHF system, which currently consists of the leased UHF portion of three MARISAT satellites and UHF transponders on certain host satellites as part of AFSATCOM, should be augmented by FLTSATCOM spacecraft, starting in 1977. These spacecraft will provide service to both the Navy and the Air Force. The evolution of UHF capability is shown in Table 6.

The U.S. military is also trying to develop a satellite system for the 1980s which will give them the required amount of electromagnetic and physical survivability for their most vital command and control traffic. The Lincoln Experimental Satellites, LES 8 and 9, which were recently successfully launched, are concerned with testing various concepts for the potential evolution of this program.[14]

1,3,5

## 1.3.5 Special Systems

A number of satellites have been orbited in the last ten years for the purpose of communication satellite application and technology evaluation. Of particular note has been the NASA Applied Technology Satellite (ATS) series. Our discussion of the category of special systems in the 1970's will be limited to those systems which are pioneering the technology and operational experiments for broadcast satellite applications and to those systems which are major precursors to potential operational satellite-based telecommunication systems. The first class includes ATS-6, the Canadian Technology Satellite (CTS), the Japanese Broadcast Satellite Experiment (BSE), and the Russian Statsionar T system. The second class includes the ESA Orbital Test Satellite (OTS), the French-German Symphonie satellite, the Italian Sirio satellite and the Japenese Experimental Test Satellite (ETS-2), Experimental Communications Satellite (ECS), and medium capacity Communications Satellite (CS). Both classes of systems are not only important technology experiments, particularly in the areas of higher satellite EIRP and/or the use of higher frequency bands but also involve significant applications oriented experiments and demonstrations.

Table 7 lists the major parameters of the broadcast satellite systems, including receiving earth terminals. Papers by Redisch, Chapman, Donoughe, Down, et al., Whalen, Henry, Ippolito, King and Hyde, and Ishida, et al. provide greater detail in both technology and applications experiments.[15, 16, 17] The ATS-6 direct broadcast TV experiments in India (using 3 m. wire mesh antennas tied to television receivers through a front-end (UHF-VHF, FM-AM) converter) and the educational TV experiments in the Rocky Mountain States and Appalachia are particularly significant from the viewpoint of evaluating the sociological impact and value of these types of broadcast applications.[15,18] additional ATS-6 experiments planned for 1976-77 and the CTS experiments will permit a realistic evaluation of the potential utility of broadcast services in a number of additional applications, such as telemedicine. The broadcast experience of TELESAT in Canada with the Anik system provides a very important example of the utility of TV broadcast services in sparsely settled and remote areas, and further provides important operational benchmarks with respect to the provision of such service with conventional satellites in the 4/6 GHz regime.[19] From a technology point of view, the systems of Table 7 should provide much of the basic information needed to evaluate the performance and economies of providing broadcast services at various frequencies.

While not expected to be the direct precursor to an operational system, the 3-axis stabilized Symphonie satellite is proving useful for various technical and operational type experiments.[20] The satellite will be used, along with 4 receive only (9 meter) and 2 transmit/receive earth stations to provide educational TV in Africa. It is also being used for teleconferencing experiments and to demonstrate the potential of satellite communication in natural disaster relief and peace-keeping, through experimentation with a 3 meter air-transportable earth terminal. Symphonie also played an important role in the development and demonstration of European know-how and technology in the space field.

The orbital Test Satellite (OTS) being developed by the European Space Agency (ESA), is important as a test bed for technology verification and for technical and operational experiments in preparation for the planned operational European Communications Satellite (ECS) system.[21,22] It is also important as an example of a modular satellite design; the same bus will be used for the MAROTS satellite.

The ECS System is expected to provide intra-European international telephony service, exchange of TV programs among the members of the European Broadcasting Union, and potentially such new services as high-speed data, T.V. broadcasting, teleconferencing, communications to North Sea oil rigs, and computer communications, etc. The planned characteristics and uses of the ECS are discussed by Bartholomé.[23]

OTS is a 3-axis stabilized satellite which will operate in the 11/14 GHz band and use polarization frequency reuse. It has two "Eurobeam" antennas and one spot beam (2.5°). (ECS is expected to have "Eurobeam" coverage and three spot beams.) Important technology experiments include depolarization measurements for both linear and circular polarization, TDMA tests at 60 and 120 Mb/s, and digital QPSK transmission tests of 60 and 180 Mb/s. A number of tests of an operational or experimental service nature are also planned.

Italy launched the SIRO satellite. This satellite operates in the 11/14 GHz band, will be used for propagation and communication experiments.

The Japanese ETS-2 satellite, which contains an S-band transponder, has the primary mission of enabling Japan's National Space Development Agency (NASDA) to develop and test its ability to launch and control a spacecraft in synchronous orbit. It also contains a propagation experiment transmitter which can provide S-band (1.7 GHz), X-band (11 GHz), and Ka-band (34 GHz) coherent signals.

The ECS satellite, which will also be N-rocket launched, is planned for 1979. It will contain C-band and K-band (34.8 GHz uplink and 31.6 GHz downlink) transponders for digital data, wide-band FM color television transmission (20-40 MHz), and K-band propagation experiments.

The CS satellite is important because it is the first commercial satellite operating in the 30/20 GHz K-band region. The satellite, which will contain six 200 MHz transponders in the 30/20 GHz band and two 200 MHz transponders in the 6/4 GHz band, will be launched by a Delta 2914 in November 1977. It has been described by Ishida, et. al. and Hyams.[25,26] used with a variety of earth stations ranging in size from 3 to 13 meters.[25,27] Experiments and measurements are planned in the areas of spacecraft technology, multiple access, propagation, communications system operation, and spacecraft operation and control.

was it launched?

Appendix A contains a list of all satellites in orbit or planned as of early 1980.

e

| SYSTEM | Satellite EIRP (dBW) | Satellite Receive G/T (dB/K) | Typical Earth Station Antenna Size (m) | Earth Station EIRP/36 MHz Channel (dBW) | Earth Station G/T (dB/K) | Ratio (Single Carrier per Transponder)*** (dBHz) |
|---|---|---|---|---|---|---|
| COMSTAR* | 33 (CONUS, Alaska only, & Hawaii/ Puerto Rico)<br>31 (Combined Alaska-Conus) | -9 | | | | |
| AT&T Earth Station | | | 30 | 92.3 | 41.4 | 106.3 (CONUS) |
| GTE Earth Station | | | 30 | 92 | 43 | 107.9 (CONUS) |
| SATCOM* | 32.5 (U.S./Alaska)<br>25.9 (Hawaii) | -4.8 (U.S./Alaska)<br>-9.8 (Hawaii) | 10** | 88 | 32.4 | 97.8 (U.S./Alaska) |
| WESTAR | 33 (CONUS)<br>26 (Alaska/Hawaii) | -7.4 (CONUS)<br>-14.4 (Alaska/ Hawaii) | 17 | 83 | 37.4 | 101.4 (CONUS) |
| AMSAT | Lease Satellite Capacity | | 11 | 85 | 33 | |
| TELESAT | 33 | -7 | 8-30 | 53-85 | 19-37 | 101.7 (Large Earth Stations) |

* Dual Linear Polarization Used To Provide Frequency Reuse

** Smaller 5 m antennas are also being added.

*** Calculated Based on Table Entries

TABLE 1. Nominal Characteristics of U.S. and Canadian Domestic Satellite Communications Systems

[Reference: *Communications Satellite Systems: An Overview of the Technology*, ed. R. G. Gould and Y. F. Lum[8]]

| SATELLITE | YEAR OF FIRST LAUNCH | ORBIT | BANDWIDTH | E.I.R.P. | SECURE COMMAND | A/J COMMAND | PROTECTED BEACON | X-BAND TT&C | MULTIPLE ACCESS | SPECIAL ANTENNA FEATURES |
|---|---|---|---|---|---|---|---|---|---|---|
| DSCS I | 6/66 | 33,000 Km equatorial ($30^{o}$ drift per day) | 26 MHz | 7 dBW | No | No | No | No | FDMA<br>CDMA<br>(spread spectrum) | None |
| DSCS II | 11/71 | Synchronous equatorial | 4 Channels *<br>EC-EC 125 MHz<br>EC-NB 50 MHz<br>NB-NB 185 MHz<br>NB-EC 50 MHz | (EC) 28 dBW *<br>(NB) 40-43 dBW<br>(20 W TWTAs) | Yes | No | No | No | FDMA<br>CDMA<br>(spread spectrum) | 2 moveable $2.5^{o}$ antennas |
| DSCS III<br>[illegible] | Est. 1979 | Synchronous equatorial | 6 Channels<br>4 - 60 MHz<br>1 - 70 MHz<br>1 - 50 MHz | (EC) 25 dBW<br>Multibeam transmit antenna - 40 dBW (max.)<br>(2-40 W TWTAs)<br>(4-10 W TWTAs) | Yes | Yes | Yes | Yes | FDMA<br>CDMA<br>(spread spectrum)<br>TDMA | <u>Planned</u><br>Multibeam receive antenna for interference nulling<br>2 Multibeam transmit antennas |

* EC - Earth Coverage, NB - Narrow Beam

TABLE 4: U.S. Military X-Band Communication Satellite Evolution

| PROGRAM | NOMENCLATURE | ANTENNA SIZE (Ft.) | TRANSMIT BANDWIDTH (MHz) | MAXIMUM AVERAGE TRANSMIT POWER (Watts) | E.I.R.P. (dBW) | G/T (dB/°K) | REMARKS |
|---|---|---|---|---|---|---|---|
| DSCS | AN/MSC - 46 | 40 | 40 | 10 kW | 98 | 34 | Modified for reliability, flexibility, capacity |
| | AN/TSC - 54 | 18 | 10 | 5 kW | 87.9 | 25.3 | Modified |
| | AN/FSC - 78 | 60 | 500 | 8 kW | 97 | 39 | New |
| | AN/MSC - 61 | 35 | 500 | 5 kW | 91 | 34 | New; Same electronics as FSC-78 |
| | AN/TSC - 86 | 18-20 | 40 | 1 kW | 80 | 26 | New; Can be transported by 2½ ton truck. Values shown are nominal. |
| Ground Mobile Forces (GMF) (Tactical SHF Communi-cations) | AN/TSC - 85 (V)(1) | 8 | | 500 (2,000 W peak) | 76.9* | 18 | New; Mount in shelter; Carried by 5/4-ton vehicle; Point-to-point use. |
| | AN/TSC - 85 (v)(2) (Multi-point) | 8 | | 500 (2,000 W peak) | 76.9* | 18 | New; Mount in shelter; Carried by 5/4-ton vehicle; Multi-point use. |
| | AN/MSC -.59 | 8 | | 100 (2,000 W peak) | 76.9* | 18 | New; Jeep trailer mounted |
| NAVY | AN/WSC - 2 | | | | | | |
| | ADM | 4 | | | 68 | 13 | |
| | STM | 8 | | | 76.5 | 18 | |

* Peak power E.I.R.P.

TABLE 5: Principal U.S. Military SHF (X-Band) Earth Terminals.

| [illegible] | | LES 6 | TACSAT* | MARISAT** (GAPSAT) | FLTSATCOM*** |
|---|---|---|---|---|---|
| ...unch Date | | 1968 | 1969 | 1976 | 1977 (Est.) |
| ...atus | | Turned off | Failed 12/72 | Operational in both oceans | |
| ...ght | | 350 lbs. | 1,593 lbs. | 720 lbs. | 1,850 lbs. (approx.) |
| [illegible] | | Spinner | Spinner | Spinner | 3-axis stabilized |
| [illegible] | Uplink<br>Downlink | 300 MHz<br>250 MHz | 300 MHz<br>250 MHz | 300 MHz<br>250 MHz | 300 MHz<br>250 MHz |
| [illegible] | | 500 KHz | 425 KHz<br>100 KHz<br>50 KHz | 500 KHz<br>2 @ 25 KHz | 1 @ 500 KHz<br>10 @ 25 KHz<br>12 @ 5 KHz |
| ...ain | | 10 dB | 15 dB<br>(17.1 dB peak;<br>14.7 dB min.) | 10 dB | 15 dB (4.9 m) |
| ...Transmitter Power | | 100 W | 240 W (max.) | 1 @ 65 W<br>2 @ 20 W | 8 @ 26.4 W<br>1 @ 35.5 W<br>3 @ 42.5 W |
| E.I.R.P. | | 27 dBW | 40.7 | 28 dBW — Channel 500 KHz<br>23 dBW — 2 @ 25 KHz | 27 dBW — Channel 500 KHz<br>26 dBW — 8 @ 25 KHz<br>28 dBW — 2 @ 25 KHz<br>28 dBW — Combined 5 KHz Channels. |

* ...erated 2 - 20 W X-Band TWTAs; Various X-Band/UHF cross straps possible

** ...tains commercial L/C band service.

*** ...of capability shared with Air Force as part of AFSAT program; X-band receive capability ... fleet broadcast mode. Data Source: TRW Pamplet, "FLTSATCOM Spacecraft," ..., 1976 and FLTSATCOM Security Guide, OP NAV/INST, C5510.129.

TABLE 6: Evolution of U.S. Military UHF Tactical Satellite Capability.

| SATELLITE | E.I.R.P. | DOWNLINK FREQUENCY | ANTENNA TYPES |
|---|---|---|---|
| ATS - 6<br>(Applications Technology Satellite) | 52.5 dBW | 2.67 GHz | • HET (Health Education Telecommunications Experiment)<br>3 m Both Receive Only and Receive/Transmit |
| | 51 dBW | 860 MHz | • SITE (Satellite Instructional TV Experiment)<br>10 m Wire Mesh |
| CTS<br>(Communications Technology Satellite) | 59 dBW | 12 GHz | • Canadian<br>3m Transmit and Receive TV<br>2m Receive Only<br>• U.S.<br>2 - 3 m |
| BSE<br>(Broadcast Satellite Experiment [Japan]) | ≃ 60 dBW | 12 GHz | 1 - 1.6 m Receive Only<br>2.5 - 4.5 m Receive Only<br>2.5 - 4.5 m Transmit and Receive |

## 1.4 Outline

The objective of this book is to provide the reader with an overall understanding of the structure, design, and analysis of communications satellite networks. As such it requires that the interplay between technology, system requirements, and financial considerations be developed. It therefore requires a quantitative approach that provides measures for performance and comparison. The book, therefore, addresses first each of the elements of the system, providing a detailed analysis of performance. Then we draw all of the elements together into an overall system design that focuses on cost/performance tradeoffs.

Chapter Two focuses on the system environment. Here we address many of the regulatory issues that tend to dominate this field. We first begin with a discussion of the organizational issues ranging from the FCC, INTELSAT, COMSAT, IRCs, RCCs, and SCCs. We discuss their structures, roles, and missions and potential impacts on any network designer.

We then discuss the environmental factors as regards to frequency choice and orbit slot allocation. As the need for space segment increases, these factors will push systems into higher frequencies and more complex space segment leg satellite design. Since this is so rapidly changing a field, it behoves the network designer to address these issues, since his network will last 20 years, which spans four satellite generations.

Finally in this chapter, we briefly address the issues of network interconnect. The satellite network does not stand alone. In almost all cases it communicates with other networks. Thus, it must be compatible in terms of both logical and physical interfaces, and in terms of performance factors.

Chapter Three discusses the network interface element. Since we are devoting our attention to all digital systems, we assume that all interfaces are digitized. We avoid introducing voice or video elements in the design. Thus the four basic interface elements are the modulator schemes, channel encoding, encryption /authentuation , and data link control interface.

The development of modulation and demodulation follows standard lines but it focuses on the more standard modulation schemes of FSK and PSK. We stress the systems performance issues such as error performance, bandwidth, and demodulator design and processing loss.

The next element of the interface is that of channel coding. This is used to tradeoff between power and bandwidth and as we show later in the text, it provides a cost effective result. Our emphasis in on convolutional codes with Viterbi decoding since most of the existing systems function in this manner. These codes are most suitable for satellite networks since the noise on the links are Gaussian and white.

We then discuss authentication and encryption concepts. These are critical issues in today's world where this data could be open to the public. If the user wants to retain his privacy then he must avail himself of some measure of data protection. The schemes we discuss are those commercially available.

Finally we develop the concept of data link control. Since our emphasis in on data networks, we expect computer communications to be key. Such communication is highly interactive and demands a high level of performance. Computer communication is a highly layered architecture with protocols used tc communicate. The data link control is one of those layers. We herein discuss one such protocol and analyze the effects of the channel on performance. We particularly evaluate the automatic repeat request schemes (ARQ).

Chapter Four discusses the earth stations. We first introduce the basic elements and then discuss each in detail. The emphasis is on system consideration and not on discussing the detailed design.

The first element is the earth station antennas. Its elements and performance are analyzed. The next element is the high power amplifier. Here we discuss several of the options that are available and detail their performance factors. We finally introduce a sample specification sheet. This latter element is used throughout the chapter for we feel the reader should understand how to specify an earth station design. The reader should develop a recognition that vendors supply catalogs contain considerable information.

The third element is the low noise amplifier (LNA) which is also developed as the HPA. The final element is the up/down converter. Here we develop the concept of phase noise and relate its effects to the demodulation problem. A good converter design requires a stable oscillator, accurate multiplier and efficient mixers. These are all discussed in detail.

We end this chapter with an analysis of a typical earth station design. We carefully go through the detail of signal levels, operations, and bandwidth.

Chapter Five presents a systems overview of the satellite. This overview stresses the elements of the satellite that impact on system performance. We first introduce the overall architecture stressing the channelization of several typical satellites. The rational for channelization is important. We then focus on the antenna. This is a key element. These antennas have shaped beams, dual polarization and multiple frequencies. As satellites advance, a great deal of performance gain will be obtained with this element.

The second set of elements are the transmitters and receivers. These, to a great degree, expand upon the discussion in Chapter Four. The third element is the on-board processor. This is probably the key element of most advanced designs. It allows the satellite to be more than just an RF repeater. It permits adaptive routing, storage, data reconfiguration, and switching.

Finally, we discuss the area of inter-satellite links (ISL). The ISL holds the key to linking satellites together into spatial networks. Some preliminary designs are discussed and the key system design issues developed.

Chapter Six presents the key systems consideration. Having developed the issues and elements in previous chapters, here we tie it all together. We first start with the link budget. This is developed to show how all the elements are linked. We then present a discussion of link anomalies. These are the factors that cause losses on the link above and beyond the simple elements usually considered. They are both random and time varying.

We then develop the traffic matrix concept. This is the key to system design. This can be fixed, time varying or random. It defines who talks to who. We then consider ways to satisfy this traffic using satellites, frequencies, beams, polarization, and finally multiple access schemes. It is the latter issue we discuss last in the chapter. We present both the classic schemes (FDMA, TDMA, etc.) plus some more recent concepts such as the Capetenakis algorithm.

Chapter Seven is the final chapter. It develops the basic cost concepts for the system. We first construct models for earth station and satellite costs. Then we provide examples of these cost models to earth station design, system design, and investment analysis.

The purpose of this chapter is to allow the reader to see the cost performance tradeoffs placed in a quantitiative fashion.

# CHAPTER 3

## TERRESTRIAL INTERFACE

The terrestrial interface portions of the satellite communications network is that portion that takes the baseband signals that have been digitized and converts them to waveforms that can be used for transmission. In this chapter we shall consider four elements of this interface; the modem, channel encoder, encryption , and data link control (Fig. 1 ).

The development of the modem area concentrates on the more typical modulation formats and analyzes the impact on performance by demodulator designs. Particular emphasis is paid to the processing loss that results in modems. Channel encoding or decoding is necessary to increase performance. We focus on convolutional codes and soft decision Viterbi decoders which are most frequently used for satellite links.

The encryption schemes presently in use are discussed in detail. It is becoming more important to use such schemes to protect the data from unlawful intervention and interception.

We complete this chapter with a discussion of data link control procedures. These are used on most data links to pace the data flow and ensure integrity. In particular we analyze several ARQ techniques and study their impact when used in satellite links.

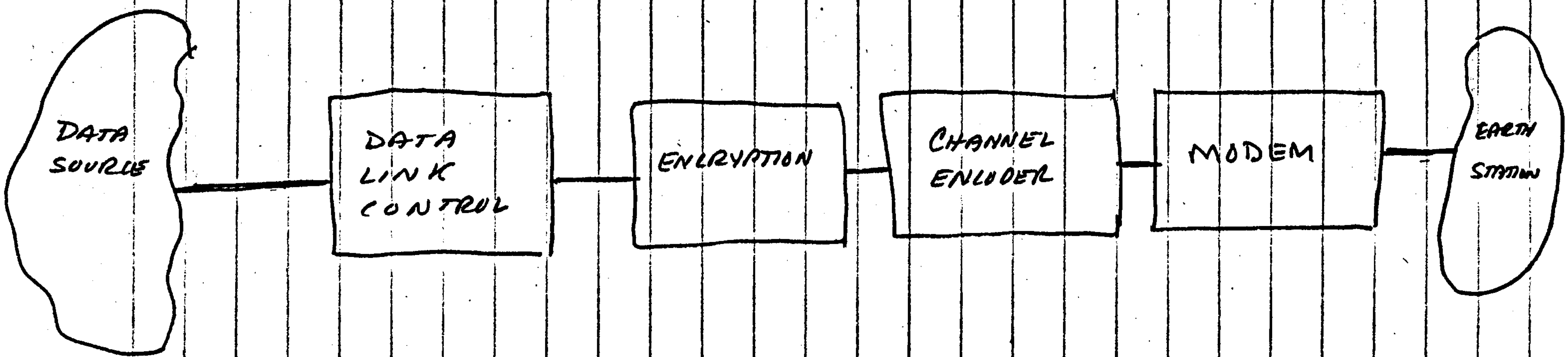
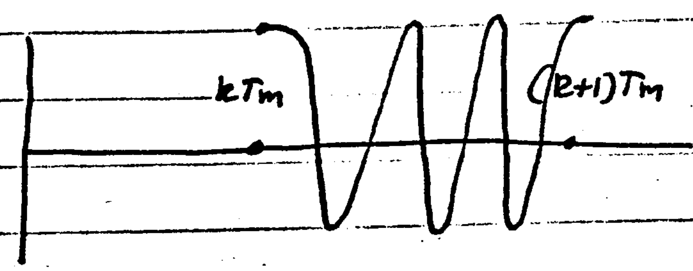
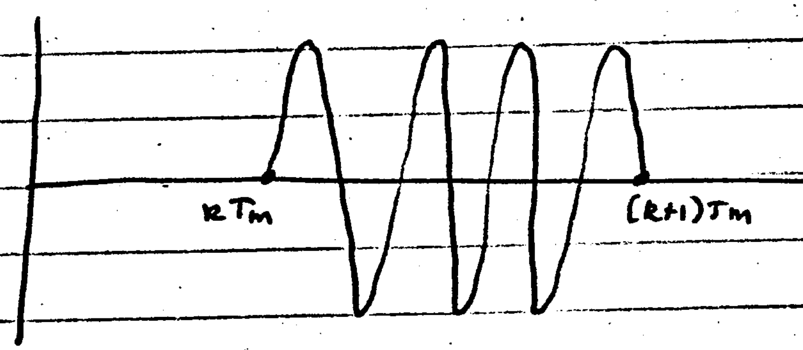



FIG 1 : Terrestrial Interface Elements




(a)

~~2φ~~ B-PSK ; Cosine wave carrier




(b)

~~2φ~~ B-PSK ; sinewave carrier

Fig 2 : 2φ-PSK waveforms

### 3.1 Modulation

The modulation process converts the binary data stream into a waveform that is suitable for transmission across the channel. In satellite communications, constant envelope waveforms are most widely used due to the nonlinearities that may occur in the amplifier throughout the system. We will therefore, in this section, restrict our attention to those types of modulation schemes.

The output of the encoder is at a rate of $R_S$ symbols per second or one symbol per $T_S$ second ($R_S = 1/T_S$). The modulator will also generate modulation symbols at a rate of one modulation symbol every $T_m$ sec. Thus it is possible to have a modulation scheme that maps several data symbols into a set of modulation symbols. For simplicity we shall assume that $T_m$ is a multiple of $T_s$ and we shall call the output of the encoder bits/sec. $T_S \leq T_M$

A constant envelope signal can be represented by

$$s(t) = \cos(f(t))$$

where f(t) is some arbitrary signal ~~waveform~~ phase function. The choice of cos is arbitrary since f(t) would allow sin or combinations. We restrict f(t) as follows. First f(t) is defined on an interval $(kT_m, (k+1)T_m)$ where $T_m$ is the modulation period. Secondly we allow f(t) to be from a finite set of possible waveforms.

Consider the first class of such waveforms that are used to generate ~~2Ø-PSK~~ BPSK (~~two phase~~ Binary, phase shift keyed) modulation.

Then

$$f_o(t) = 2\pi f_s t$$

or

$$f_1(t) = 2\pi f_s t+\pi$$

with duration $T_m$ and $f_s$ is some multiple of $1/T_m$. We call this cosine PSK. If we add a $-\pi/2$ phase shift we obtain a sine PSK waveform. The information in the data bits are mapped into the $f_i(t)$ function. Thus 0 maps into $f_o(t)$ and 1 into $f_1(t)$. It is important to note that the critical factor is phase because that is what controls the mapping. Fig. 2 depicts those sample waveforms.

In ~~2∅~~ B-PSK, one bit of data is encoded into one modulation waveform. We can expand this by increasing the $f_i(t)$ set on $T_m$. Let us assume that $T_m$, as with 2∅-PSK, equals kT, T the data duration. Then k bits can be encoded into the modulation waveform. For this to happen there must be $2^k$, f(t) waveforms. Consider the case of k=2. We can write;

$$f_o(t) = 2\pi f_s t$$

$$f_1(t) = 2\pi f_s t + \frac{\pi}{2}$$

$$f_2(t) = 2\pi f_s t + \pi$$

$$f_3(t) = 2\pi f_s t + \frac{3\pi}{2}$$

This is called ~~4∅-PSK~~ Quadriphase since the constant envelope is phase modulated with four constant phases in $T_m$. Note also that we have 2 bits per waveform. We will return to this when we determine the modulation schemes bandwidth efficiency. In a similar fashion we can write a ~~$2^k$∅-PSK~~ as MPSK $M=2^k$ is the usual notation!

$$f_i(t) = 2\pi f_s t + (i-1)\frac{2\pi}{k}$$

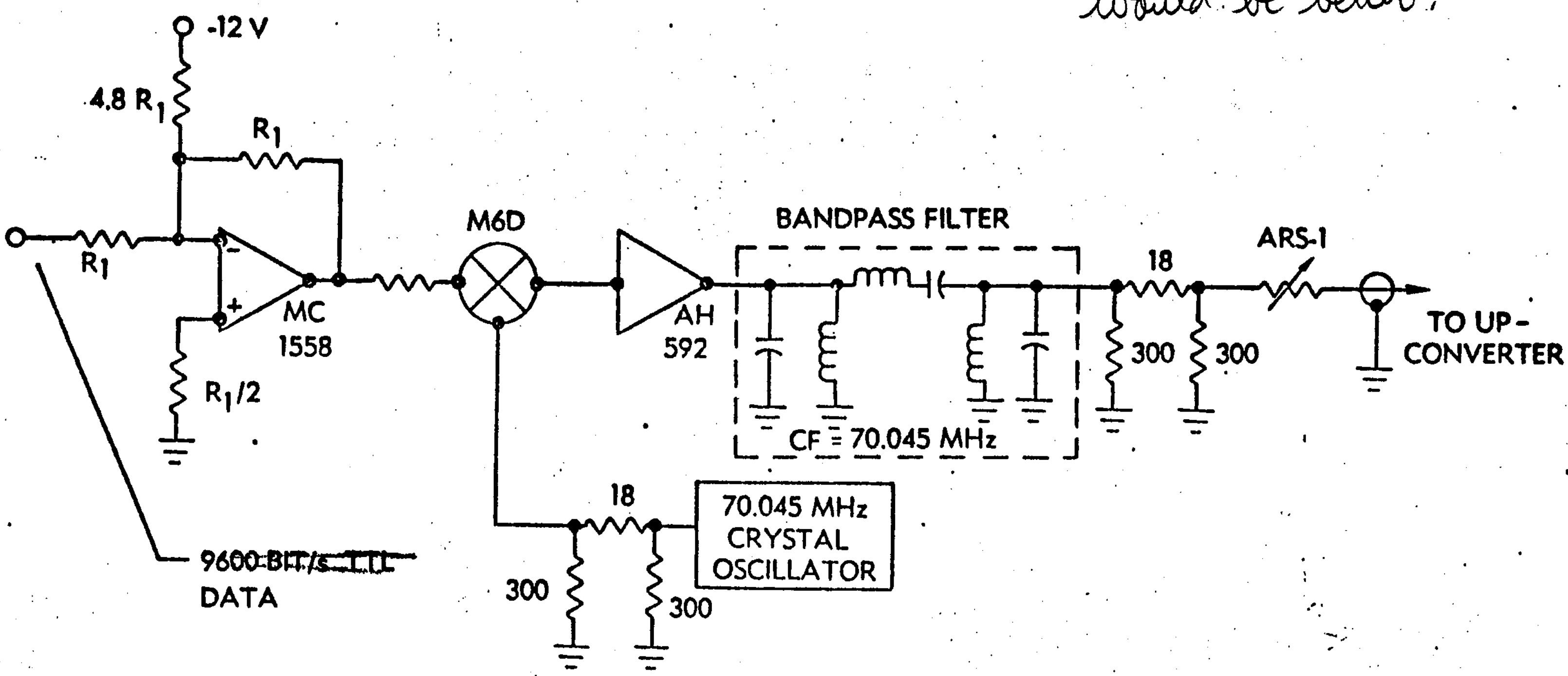
In the higher order PSK systems, we have more bits per waveform. However we shall see that when we consider performance, some degradation occurs.

A typical 2∅-PSK modulator is shown in Fig. 3. The input is a TTL (Transistor-Transistor Logic) to an operational amplifier. A 70 MHz oscillator is used as the IF section source and corresponds to $f_s$. The digital waveform is coupled to an operational amplifier which goes into a multiplier along with the 70 MHz signal. This is then isolated and filtered and then

3

Figure ~~3.5.15~~: ~~REMOTE-STATION-DATA-LINK~~ B- PSK MODULATOR

I think a function diagram would be better!



sent to the uplink chain. This is a fairly inexpensive modulation scheme except that the oscillator may have an uncertain offset. This would be unacceptable in the 2Ø-PSK that we have just presented. The way we avoid having to phase reference the oscillator and keep cost low is to use a differential PSK (DPSK) system. For this system

$$f_o(t) = 2\pi f_s t - \frac{\pi}{2} + \theta$$

$$f_1(t) = 2\pi f_s t + \frac{\pi}{2} + \theta$$

where $\theta$ is the unknown but constant offset. To avoid the uncertainty of the phase, we encode the input data streams. The encoder works as follows. If a 1 is sent, then the phase remains the same. If a zero is sent, then the phase changes. Thus we have, for example, the data streams

$$(1,1,0,0,1,0,1,1)$$

maps into;

$$(f_1,f_1,f_o,f_1,f_1,f_o,f_o,f_o)$$

Thus the raw data stream must first be encoded according to the above rule and then demodulation is done on a comparison basis. We shall discuss that later in this section.

A second general type of modulation is frequency shift keyed (FSK) where now

$$f_i(t) = 2\pi f_s t + 2\pi f_i t$$

Here $f_i$ is a deviation about the standard IF frequency $f_s$. For example, in the binary case, could use better clearer notation!

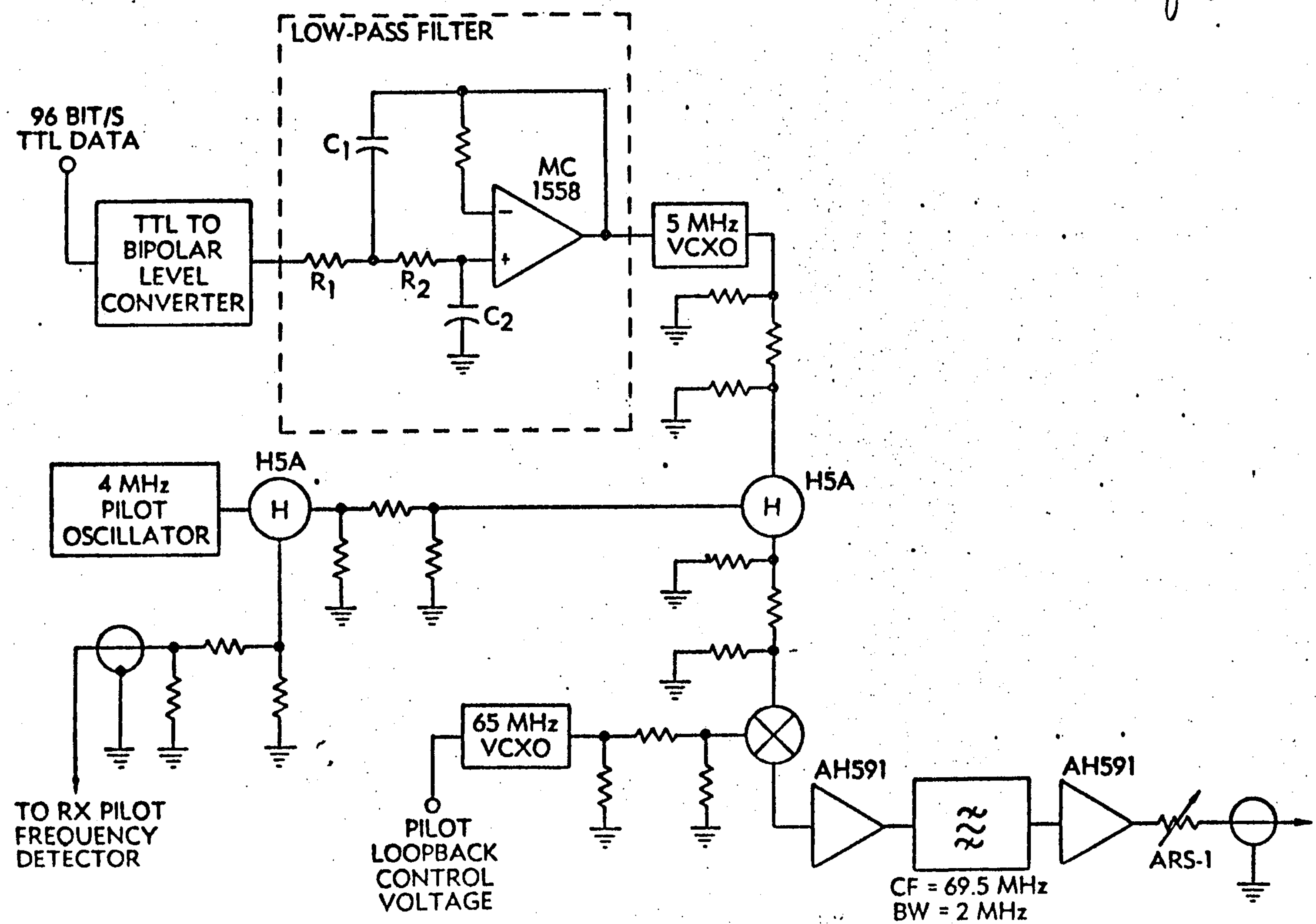
$$f_o(t) = 2\pi f_s t - 2\pi f_o t$$

$$f_1(t) = 2\pi f_s T + 2\pi f_o t$$

The offset frequency between these two is $2f_o$ Hz. As with PSK we may include an unknown phase to make the transmission incoherent and the result is an incoherent (noncoherent) FSK system. A sample modulator is shown in Fig. 4.

Figure 4~~-4-5~~: ~~COMMAND-LINK~~ FSK MODULATOR

functional diagram!



One of the most important characteristics of these modulation schemes is the bandwidth that they occupy. We shall now analyze ~~the case of~~ spect~~rum~~ral occupancy for each of the previous cases. Consider first the coherent 2Ø-PSK system. We can let s(t) be the signal and this can be represented by

$$s(t) = a(t) \cos 2\pi f_s t$$

Here a(t) is a random $\pm 1$ waveform of period T. We want to determine the spectrum of s(t), $S_s(f)$. Our approach is to note that

$$S_s(f) = \tfrac{1}{2} S_a(f-f_s) + \tfrac{1}{2} S_a(f+f_s)$$

due to the properties of cos $2\pi f_s t$. Only if the transitions are not coherent with $\cos 2\pi f_s t$ ! Thus all that is necessary is to obtain $S_a(f)$. Now this can be done by determining $R_a(\tau)$, the correlation function and obtaining its Fourier Transform. Now by definition ;

$$R_a(\tau) = E\left[a(t)a(t+\tau)\right]$$

This is not a stationary process unless you randomize the epoch time!

but we know if $\tau<T, \tau>0$

$$a(t)a(t+\tau) = \begin{cases} +1 & \text{if it is same interval} \\ +1 & \text{if different interval but same sign} \\ -1 & \text{if different interval but different sign} \end{cases}$$

Call the three events above $A_1$, $A_2$, $A_3$. Then it is easily shown;

$$P(A_1) = \frac{T-\tau}{T}$$

$$P(A_2) = \tfrac{1}{2} \frac{\tau}{T}$$

$$P(A_3) = \tfrac{1}{2} \frac{\tau}{T}$$

Thus we obtain

$$E\left[a(t)a(t+\tau)\right] = \frac{T-\tau}{T} + \frac{1}{2}\frac{\tau}{T} - \frac{1}{2}\frac{\tau}{T}$$

Performing the same analysis for $\tau<0$ we obtain;

$$R_a(\tau) = \frac{T-|\tau|}{T}$$

and is shown in Fig. 5. To determine $R_a(\tau)$'s Fourier Transform we note the following. Let g(t) be given by

$$g(\tau) = \begin{cases} 1/\sqrt{T} & |\tau| < T/2 \\ 0 \end{cases}$$

You better define convolution since you are using it!

Then clearly

$$R_a(\tau) = g(\tau) * g(\tau)$$

where * indicates convolution. However,

$$S_g(f) = \int_{-T/2}^{T/2} \frac{1}{\sqrt{T}} e^{-j2\pi f\tau} d\tau$$

is the spectrum of g $(\tau)$. This equals

$$S_g(f) = \frac{\sin\left(2\pi f \, T/2\right)}{2\pi f \sqrt{T} \, \frac{}{2}}$$

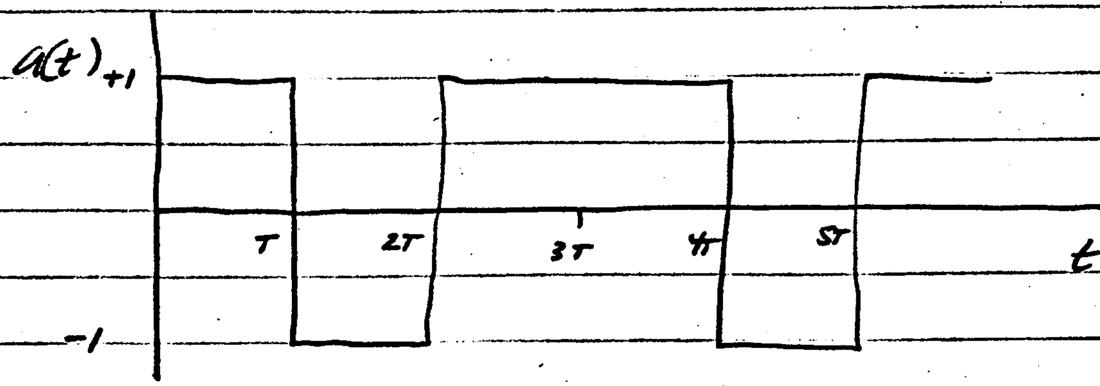
Thus by the properties of the convolution in the frequency domain

$$S_a(f) = S_g^2(f) = T\left(\frac{\sin 2\pi f \, T/2}{2\pi f \, T/2}\right)^2$$

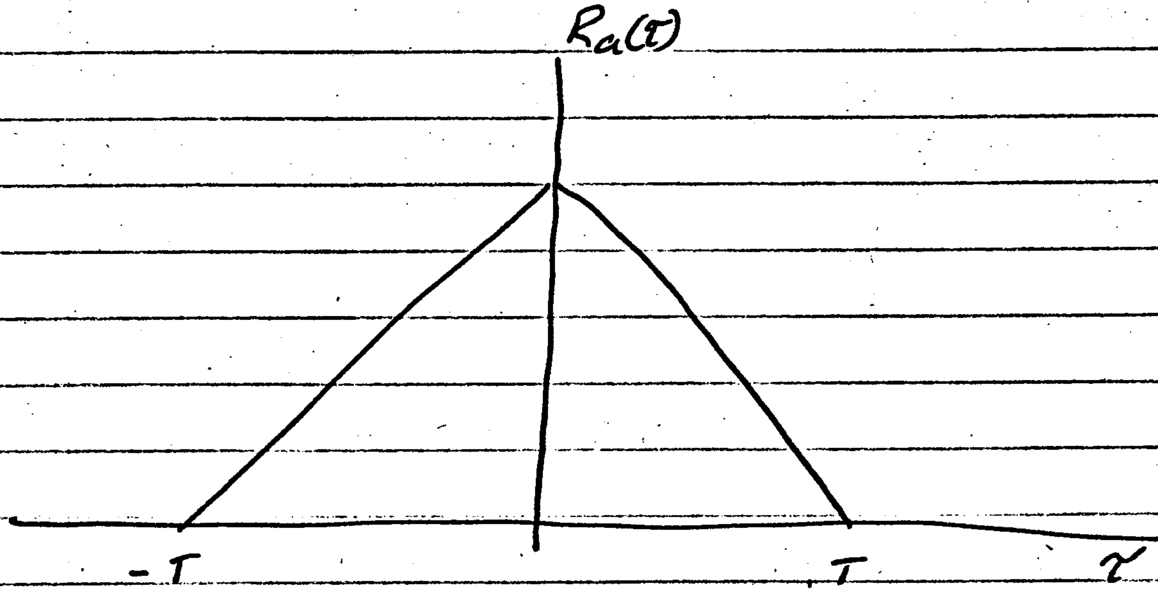
This spectrum is shown in Fig. 6. Note the first zeros occur at $2\pi f \frac{T}{2} = \pi$ or $f = \pm 1/T$. The definition of the bandwidth of this signal is loosely defined. If we consider the width to the zeros, it is 2/T. However, this is generally not the definition. What is usually called the bandwidth is the width to the 10 dB points. This gives an effective bandwidth of 1/T.

We shall see later what effect this assumption has on performance. Thus for a 10 Kbps signal the bandwidth is 10 KHz. For 2Ø PSK this yields an efficiency of 1 bit per second per Hz of bandwidth. The spectrum for 2Ø-PSK is shown in Fig. 7. Note that if $f_s$ is chosen too small the tails can overlap. Thus $f_s$ is typically chosen as $>>10(1/T_s)$. The actual choice of $f_s$ depends on available IF oscillators.

For 4Ø-PSK we can readily determine the spectrum by performing a slight trick. If we draw two orthogonal axes, then 4Ø-PSK can be represented as one of four points. The axis are basis functions, cos $2\pi f_s t$ and sin $2\pi f_s t$ which are orthogonal on (O,T) or multiples thereof. That is, we select a single period QPSK waveform to be one of




(a)
Random Modulation Wave form




(b)
Correlation Function

Fig 5 : Random Waveform

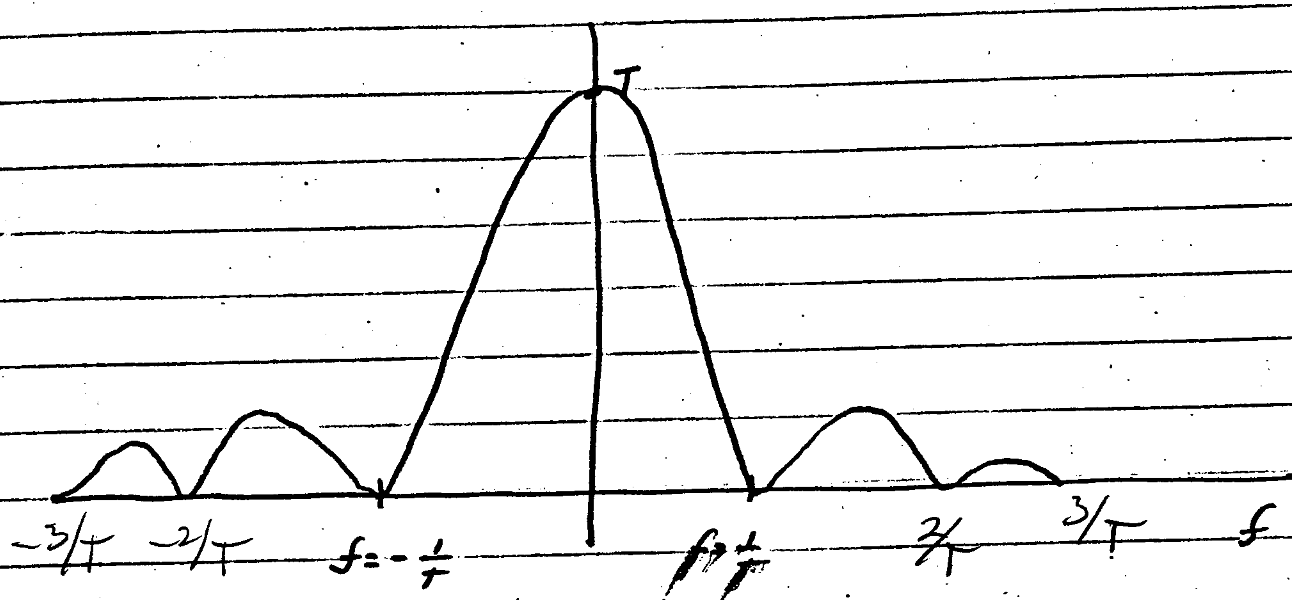



Fig 6: Spectrum of Random Sequence

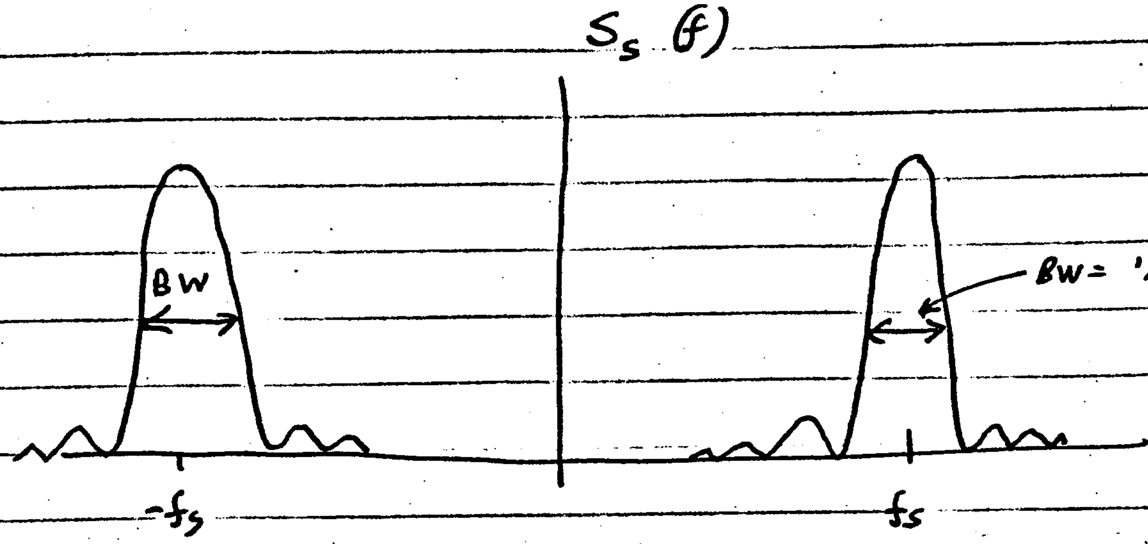



Fig 7 : 2φ-PSK Spectrum

$\pm\sqrt{E}\,\phi_1(t),\ \pm\sqrt{E}\,\phi_2(t)$

You probably should define basis functions since many readers of this book might not know what they are.

where

$$\phi_1(t) = \sqrt{\frac{2}{T}}\cos 2\pi f_s t$$

$$\phi_2(t) = \sqrt{\frac{2}{T}}\sin 2\pi f_s t$$

Note that

$$\int_0^T \phi_i^2(t)\,dt = 1$$

$$\int_0^T \phi_i(t)\phi_j(t)\,dt = 0 \qquad i\neq j$$

These are called orthonormal basis functions. The four wave forms are shown in Fig. 8 a. Now if we rotate these by $45^o$ we can obtain the same signalling set. Yet it is easy to show that we can represent this by

$$S(t) = a(t)\left(\sqrt{\frac{E}{T}}\cos 2\pi f_s t\right) + b(t)\left(\sqrt{\frac{E}{T}}\sin 2\pi f_s t\right)$$

where a(t) and b(t) are independent random sequences as defined before. Thus the spectrum for QPSK (4Ø-PSK) is

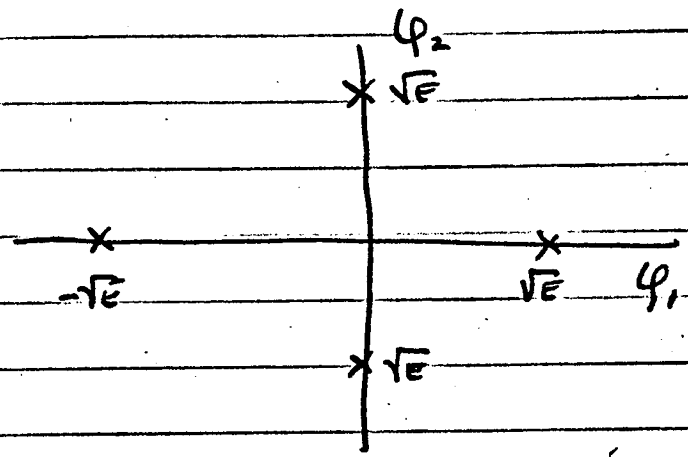
$$S_s(f) = \frac{E}{T}\left(\frac{S_a(f-f_s)}{2} + \frac{S_a(f+f_s)}{2}\right)$$

$$+\frac{E}{T}\left(\frac{S_a(f-f_s)}{2} + \frac{S_a(f+f_s)}{2}\right)$$

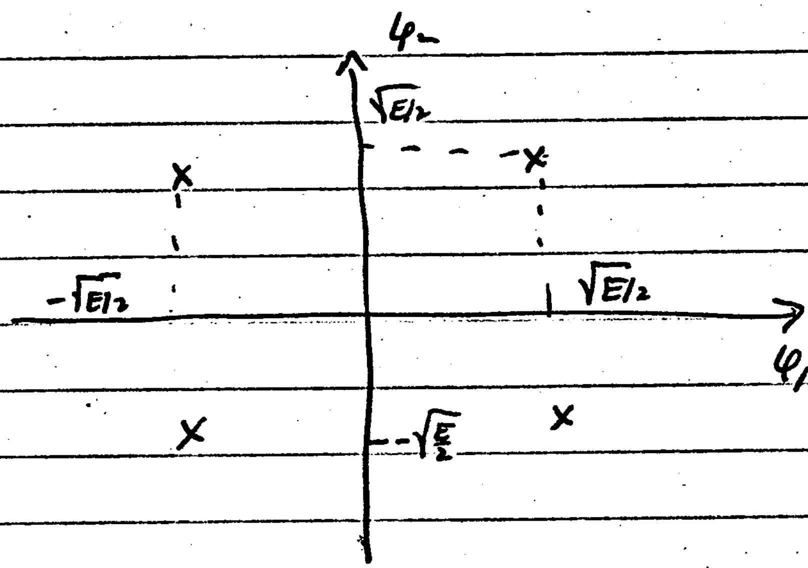
since a(t) and b(t) are assumed to be statistically independent!

which is identical to 2Ø-PSK. Thus they have identical spectra and use the same bandwidth. However we now have 2 bits per waveform, so the efficiency is 2 bits/sec./Hz.

We can readily extend this to 8Ø-PSK. Fig. 9 shows the signal constellation using the same decomposition as before. 8Ø-PSK has signal vector lying on a circle. It then appears as 2, 4Ø-PSK signals. Thus it is immediately obvious that it too has the same spectrum as 2Ø-PSK but now we have 3 bps/Hz as the efficiency.




(a)
QPSK




(b)
Rotation

Fig 8: Analysis of QPSK

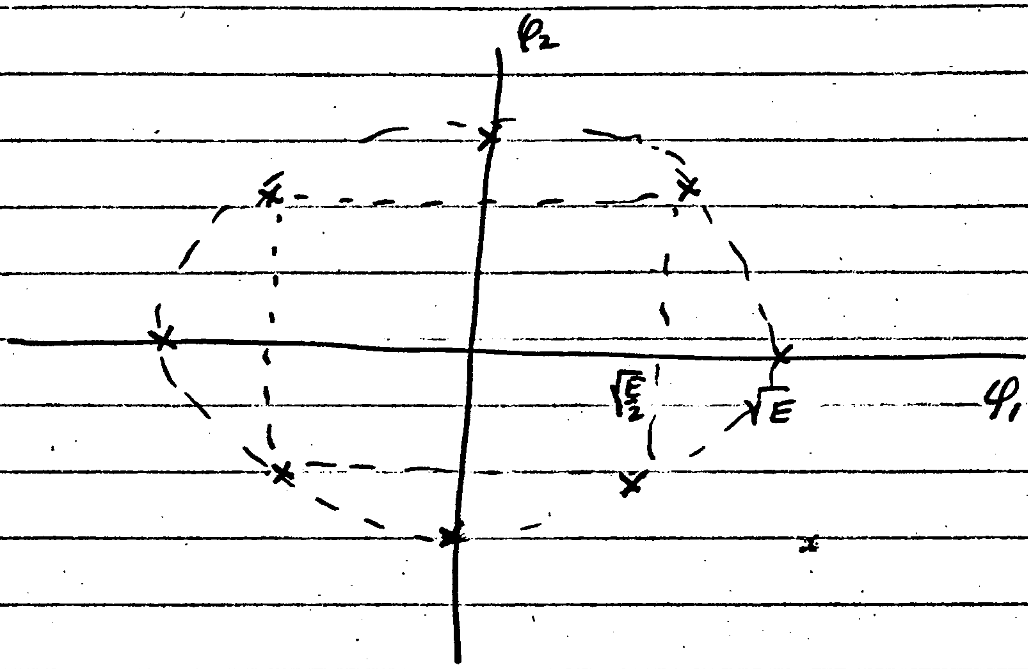



Fig 9 : 8φ-PSK Signal Set

This process can be continued for all higher order PSK systems with identical results.

Let us now turn our attention to FSK. Let us assume a coherent system. Now we can write $s(t)$ as;

$$s(t) = \frac{1}{2}(1-a(t))\sqrt{\frac{2E}{T}} \cos(2\pi f_s t - 2\pi f_o t)$$

$$+ \frac{1}{2}(1+a(t))\sqrt{\frac{2E}{T}} \cos(2\pi f_s t + 2\pi f_o t)$$

as the data waveform, where $a(t)$ is the random data sequence. Note that we have again normalized the signals to energy E. The duration of $a(t)$ is in periods of T seconds. The ~~spectrum~~ power spectral density then is

$$S_s(f) = \frac{E}{4T}\,\delta(f-f_s+f_o) + \frac{E}{4T}\,\delta(f-f_s-f_o)$$

$$+ \frac{E}{4T}\,\delta(f+f_s+f_o) + \frac{E}{4T}\,\delta(f+f_s-f_o)$$

$$+ \frac{E}{2T}\Big[(S_a(f-f_s-f_o) + S_a(f-f_s+f_o)$$

$$+S_a(f+f_s-f_o) + S_a(f+f_s+f_o))\Big]$$

What is immediately evident is that the spectrum contains a carrier component (Fig. 10). There are methods of surpressing (surpressing) this carrier. These are used because the carrier robs power from the system and it carries no information.

We can now consider the spectrum of the final modulation format DPSK. Now DPSK is given by

$$s(t) = a(t) \cos(2\pi f_s t + \theta)$$

However $a(t)$ has memory. Let us consider the two states that $a(t)$ can be in (1 or 0). Now using the transition definition, we can write the expression for the probability of being in state j at time k. This yields

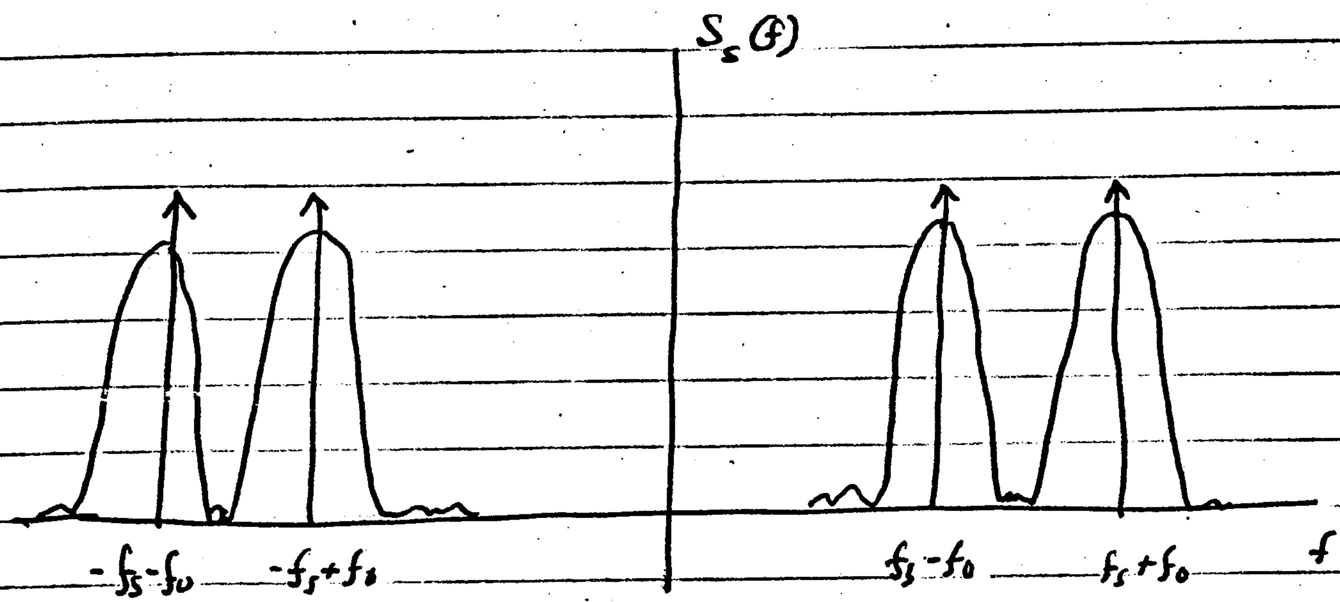



Fig 10 : FSK Spectrum

$$P_1(k+1) = P_1(k)\ P(1 \text{ is input}) + P_o(k)\ P(o \text{ is input})$$

not lower case o but rather should be zero!

$$P_o(k+1) = P_1(k)\ P(o \text{ is input}) + P_o(k)\ P(1 \text{ is input})$$

But both inputs are equally likely so it is clear that $P_o(k+1)$ and $P_1(k+1)$ are both 1/2 and independent of the previous state. Therefore a(t) is the same process as for 2Ø-PSK and therefore the ~~spectrum~~ spectral density is the same. This then provides us with an adequate basis for bandwidth characterization for a wide class of modulation formats.

We now want to concern ourself with the channel characterization. Fig. 11 depicts a typical channel. The modulator feeds the up converter which is then amplified, transmitted over the channel, received, and then down converted. The resulting signal, r(t), contains a replica of the signal plus noise. The noise comes from the amplifier front ends, space or sky noise, antenna noise and other sources. The noise, n(t), has received extensive attention in the satellite case and has been readily characterized as white Gaussian noise. That is n(t) is a zero mean process;

$$E[n(t)]=0$$

and has a correlation function

$$R_n(\tau) = \frac{N_o}{2}\ \delta(\tau)$$

where $N_o/2$ is the spectral density height. The noise power spectrum is

$$S_n(f) = \frac{N_o}{2} \quad \text{(Two sided)}$$

which accounts for it being called white; e.g., constant at all frequencies.

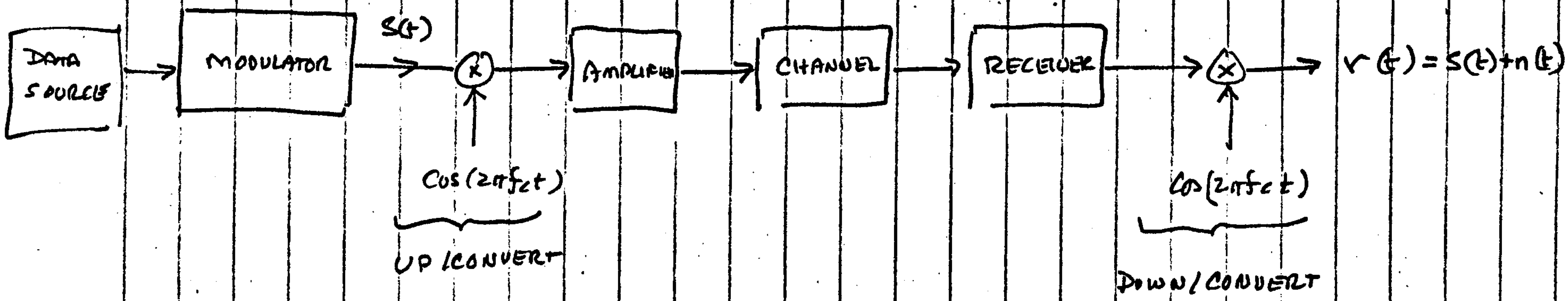



FIG 11 : COMMUNICATION CHANNEL

The spectral density height is also related to the noise temperature by

$$\frac{N_o}{2} = kT$$

where T is the noise temperature in $^oK$ and k is Boltzmanns constant.

There is one further issue with this process. Let $\varphi_i(t)$ be an orthonormal function on (O,T) as we described before. Then

$$\int_o^T \varphi_i^2(t)\, dt = 1$$

and
$$\int_o^T \varphi_i(t)\, \varphi_j(t)\, dt = 0 \quad i \neq j$$

Now consider the random variable $n_i$ defined as;

$$n_i = \int_o^T \varphi_i(t) n(t) dt$$

Clearly $n_i$ is also zero mean and

$$E[n_i^2] = E\left[\int_o^T\int_o^T \varphi_i(t)\varphi_i(s)\; n(t)n(s)dtds\right]$$

$$= \int_o^T\int_o^T \varphi_i(t)\varphi_i(s)\; E[n(t)n(s)]\; dtds$$

$$= \frac{N_o}{2}$$

Thus the probability density for $n_i$ can be written as;

$$p(n_i) = \frac{1}{\sqrt{2\pi \frac{N_o}{2}}} \exp\left(-\frac{1}{2}\, \frac{n_i^2}{N_o/2}\right)$$

Note also that the original process n(t) has an infinite variance!

The problem of designing the demodulator centers around a performance measure. We typically choose the probability of error $P[E]$ as such a performance measure. We typically look at the demodulation process on a bit at a time basis. The coding/decoding process looks at the entire data streams. Thus the demodulation process looks at deciding between alternatives T seconds at a time.

To develop the demodulator structure let us return to the signal structure. Recall that we introduced the concept of a set of orthonormal functions $\varphi_i(t)$. For 2Ø-PSK we have a single function, $\varphi_i(t)$, given by

$$\varphi_i(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_s t)$$

Clearly any set of multiples of $f_s$ would yield increasing codes of $\varphi i(t)$. Now we can let s(t) be;

$$s(t) = a_i \varphi_i(t)$$

where $a_i = \pm \sqrt{E}$, $\sqrt{E}$ if 1 and $-\sqrt{E}$ if 0. Similarly for 4Ø-PSK we have

$$s(t) = a_i \varphi_1(t) + b_i \varphi_2(t)$$

where

$$\varphi_1(t) = \sqrt{\frac{2}{T}} \cos (2\pi f_s t)$$

$$\varphi_2(t) = \sqrt{\frac{2}{T}} \sin (2\pi f_s t)$$

and we have the mapping

$$00 \rightarrow a_i = -\sqrt{E}, b_i = 0 \qquad -0\ ?$$

$$01 \rightarrow a_i = -0, b_i = \sqrt{E}$$

$$10 \rightarrow a_i = 0, b_i = \sqrt{E}$$

$$11 \rightarrow a_i = \sqrt{E}, b_i = 0$$

Now the noise $n(t)$ can also be shown to be expandable; Karhunen Loeve expansion!

$$n(t) = \sum_{i=1}^{\infty} n_i \varphi_i(t)$$

and since $\varphi_i$ are orthonormal the $n_i, n_j$ terms are uncorrelated and thus independent.

The reason for using this characterization is that instead of using the waveforms for the decision process we can use the ~~constants~~ coefficients of the expansion. Thus $r(t)$ also can be expanded;

$$r(t) = \sum_{i=1}^{\infty} r_i \varphi_i(t)$$

where clearly;

$$r_i = \int_0^T r(t) \varphi_i(t)\, dt$$

Now, in general, $s(t)$ is given in terms of a finite number of terms. Thus

$$s(t) = \sum_{i=1}^{N} s_i \varphi_i(t)$$

where $\varphi_i(t)$, $i=1,..N$ are the key decision functions. Note that for all the modulation schemes we discussed so far, there are at most 2. Thus at the receiver, the decision is made on $r(t)$ or equivalently $r_1, \ldots r_N$. Fig. 12 shows how this is obtained.

Let us now consider one of two messages that can be sent, $s_o(t)$ and $s_1(t)$. Recall,

$$s_o(t) = \sum_{i=1}^{N} s_{oi} \varphi_i(t)$$

$$s_1(t) = \sum_{i=1}^{N} s_{1i} \varphi_i(t)$$

Now we want to decide on which of these two were transmitted given $\underline{r}$, where $\underline{r}$ is the NX1 vector of outputs. But we want to do this optimally. The optimization criteria is to have a minimum probability of error $P\{E\}$. Let $R$ be the entire space that $\underline{r}$ can fall into. Let

$$R = R_1 \oplus R_o$$

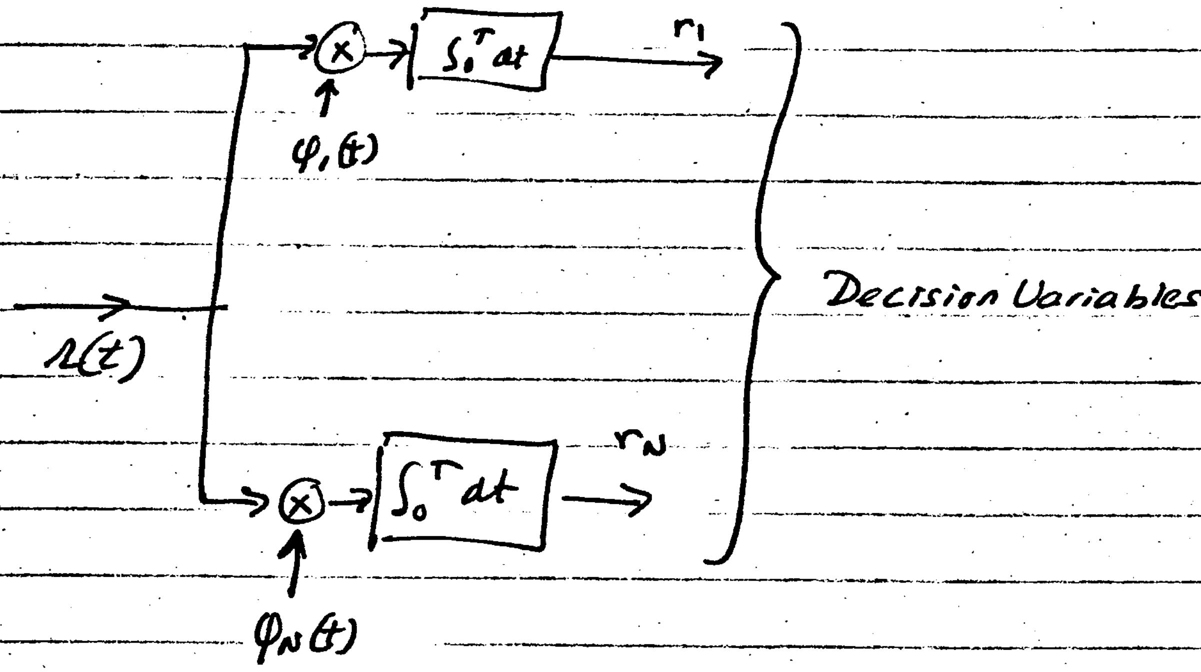



Fig 12 : Demodulator Front End

where $\mathbb{R}_i$ is that ~~part~~ partition of $\mathbb{R}$ where we ~~say~~ decide that $S_i$ occurred. Then errors will occur if

a. If $s_o$ ~~did happen~~ occurs and $r\in\mathbb{R}_1$

b. If $s_1$ ~~did happen~~ occurs and $r\in\mathbb{R}_o$

This can be expressed as follows

$$P[E] = P[r\in\mathbb{R}_1 \text{ and } s_o \text{ occurred}] + P[r\in\mathbb{R}_o \text{ and } s_1 \text{ occurred}]$$

This is easily written as;

$$P[E] = \int_{\mathbb{R}_1} p(r/s_o)\,dr P(S_o) + \int_{\mathbb{R}_o} p(r/s_1)\,dr P(S_1)$$

Now using the fact that $\mathbb{R}_1 = \mathbb{R} - \mathbb{R}_o$ we have

$$P(E) = \int_{\mathbb{R}} p(\underline{r}/s_o)\,d\underline{r}\; P(s_o) + \int_{\mathbb{R}_o} [p(\underline{r}/s_1)P(S_1) - p(\underline{r}/s_o)P(S_o)]\,d\underline{r}$$

We can use this expression to determine $\mathbb{R}_o$. Clearly $\mathbb{R}_o$ should be all those r's that make the second integral negative, thus minimizing $P[E]$. Thus we have;

Choose $s_o$ if $\underline{r}$ satisfies;

$$\frac{p(\underline{r}/s_o)P(s_o)}{p(\underline{r}/s_1)P(s_1)} \geq 1$$

otherwise choose $s_1$.

Now recall that $r_i$ under $s_o$ equals

$$r_i = s_{oi} + n_i$$

and under $s_1$

$$r_i = s_{1i} + n_i$$

Then by defining vectors $\underline{s}_1$ and $\underline{s}_o$ with the components $s_{1i}$ and $s_{oi}$, we can show that the optimum region expression for equally likely $s_i$, $(P[s_o] = P[s_1])$ yields; raise up

$$p(\underline{r}/\underline{s}_o) = C \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \left(\frac{r_i - s_{oi}}{N_o/2}\right)^2\right)$$

We can simplify the result by noting that the ratio for the region can be equivalent to one using ln. That is;

Choose $\underline{s}_o$ iff $\underline{r}$ satisfies

$$\ln\left[\frac{p(\underline{r}/\underline{s}_o)}{p(\underline{r}/\underline{s}_1)}\right] \geq 0$$

otherwise choose $s_1$. This simplifies to;

Choose $\underline{s}_o$ iff $\underline{r}$ satifies

$$\sum_{i=1}^{N} (r_i - s_{oi})^2 \leq \sum_{i=1}^{N} (r_i - s_{1i})^2$$

Using the vector notation this yields;

Choose $\underline{s}_o$ iff $\underline{r}$ satifies

$$\underline{r} \cdot \underline{s}_o \geq \underline{r} \cdot \underline{s}_1 - |\underline{s}_1|^2 + |\underline{s}_o|^2$$

where $\underline{r} \cdot \underline{s}_i$ is the dot product of vectors and $|s_i|$ is the magnitude, $|s_i|^2$ is the energy of $s_i(t)$. Thus for equal energy we have

$$\underline{r} \cdot \underline{s}_o \geq \underline{r} \cdot \underline{s}_1$$

We now note that

$$\underline{r} \cdot \underline{s}_i = \sum_{j=1}^{N} r_j s_{ij}$$

But also note that

$$\int_o^T r(t) s_i(t)\, dt$$

$$= \int_o^T \left(\sum_{j=1}^{N} r_j \varphi_j(t)\right) \left(\sum_{k=1}^{N} s_{ik} \varphi_k(t)\right) dt$$

$$= \sum_{j=1}^{N} \sum_{k-1}^{N} r_j\, s_{ik} \int_0^T \varphi_j(t)\, \varphi_k(t)\, dt$$

$$= \sum_{k=1}^{N} r_k s_{ik}$$

The last step is a result of the properties of orthonormal functions. Thus the optimal decision statistic is

$$\int_0^T r(t)\, s_i(t)\, dt$$

which is compared. This is the optimum demodulator and it is called the integrator-correlator. (See Fig. 13)

<u>Example</u>:

Consider the 2∅-PSK modulator scheme. The waveform chosen is

$$s(t) = \pm \sqrt{\frac{2E}{T}} \sin(2\pi f_s t)$$

The decision statistic is

$$r = \int_0^T r(t) \sqrt{\frac{2}{T}} \sin(2\pi f_s t)\, dt$$

This operator is the demodulator. It assumes the following.

a. It knows when the bit arrives, e.g, $t=0$

b. It knows the duration, e.g., T

c. It knows the IF frequency $f_s$.

awkward

d. It knows that no additional phase in on the signal, e.g., $\theta=0$.

The demodulator is then a synchronized oscillator followed by an integrator. If $r>0$ we say $\sqrt{E}$ or 1 and if $r<o$ we say $-\sqrt{E}$ or 0.

Let us ~~new~~ evaluate the error probability. Now the regions $R_o$ and $R_1$ are

$$R_o = \{r:\ r<o\}$$
$$R_1 = \{r:\ r>o\}$$

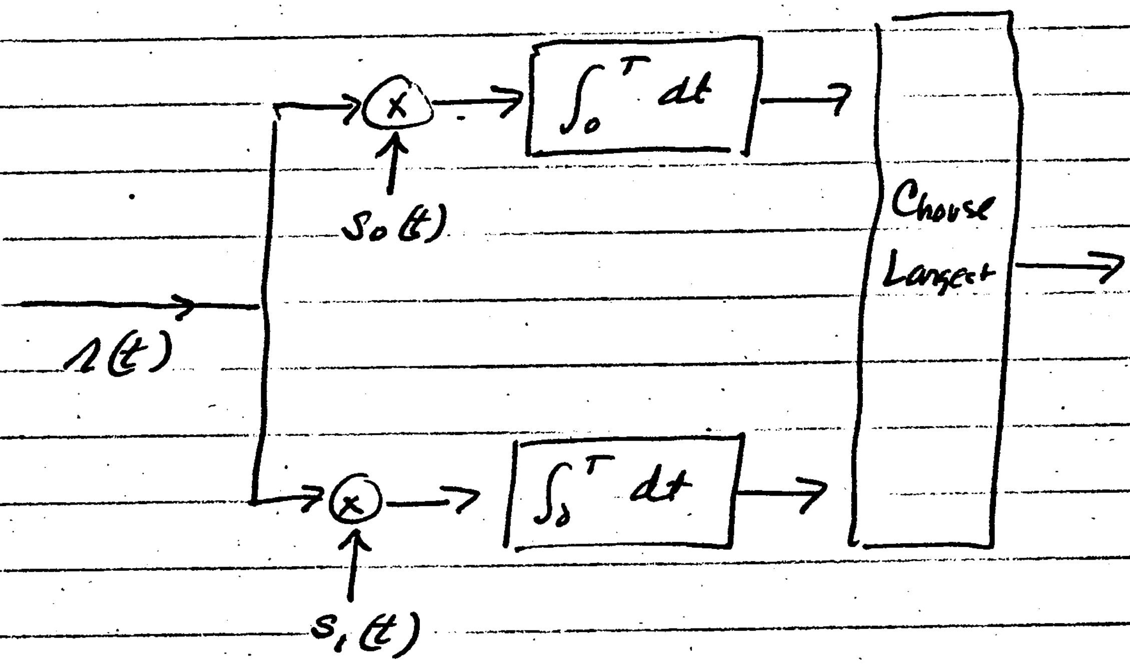



Fig 13: Optimum Demodulator.

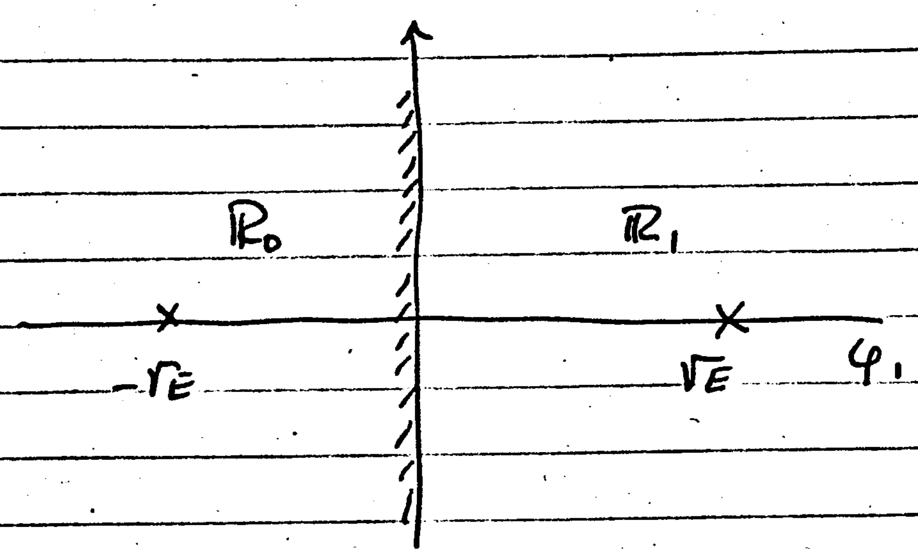



Fig 14 : 2φ- PSK Decision Region

This is shown in Fig. 14. Now we can write

$$\bar{p}(r/so) = \frac{1}{\left[2\pi \frac{N_o}{2}\right]^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{r+\sqrt{E})}{N_{o/2}}\right)\right]^2$$

$N_o/2$

since r under $s_o$ has mean $-\sqrt{E}$ and variance $N_{o/2}$. Similarly

$$p(r/s_1) = \frac{1}{\left[2\pi \frac{N_o}{2}\right]^{1/2}} \exp\left\{-\frac{1}{2}\frac{(r-\sqrt{E})}{N_{o/2}}\right\}^2$$

etc.

Thus;

$$P[E] = \frac{1}{2}\int_{-\infty}^{o} p(r/s_1)\,dr + \frac{1}{2}\int_{o}^{\infty} p(r/s_o)\,dr$$

$$= \int_{-\infty}^{o} \frac{1}{\sqrt{2\pi N_{o/2}}} \exp\left(-\frac{1}{2}\frac{(r-\sqrt{E})}{N_{o/2}}\right)^2 dr$$

$$\stackrel{\Delta}{=} Q\left(\frac{\sqrt{2E}}{N_o}\right)$$

where $Q(\alpha)$ is defined as

$$Q(\alpha) \stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\infty} \exp\left(-\frac{r^2}{2}\right) dr$$

do you mean a coherent detector?

Example: Consider now the FSK case. Let us assume that a coherent signal is used. Then clearly we need

$$r \cdot s_o = \int_{o}^{T} r(t)\sqrt{\frac{2}{T}}\sin(2\pi f_s t - 2\pi f_o t)\,dt$$

$$r \cdot s_1 = \int_{o}^{T} r(t)\sqrt{\frac{2}{T}}\sin(2\pi f_s t + 2\pi f_o t)\,dt$$

Let us further assume that the $f_o$ and $f_s$ are such that the two sine waveforms are orthogonal over (O,T). They therefore are the basis functions. The $\varphi$ plane plot for FSK is shown in Fig. 15. Clearly from this plot we can write

$$P[E] = Q\left(\sqrt{\frac{E}{N_o}}\right)$$

This is clear to me but is it clear to the typical reader? add a few details!

since this is equivalent to PSK but the signals are closer. Thus less noise is necessary to cause an error.

You used QPSK in figure 16 in the title - be consistent!

Example: For ~~4$\emptyset$-PSK~~ QPSK we have a demodulator that generates

$$r_1 = \int_0^T r(t)\sqrt{\frac{2}{T}}\ \cos\ (2\pi f_s t)\,dt$$

$$= \int_0^T r(t)\,\varphi_1(t)\,dt$$

and

$$r_2 = \int_0^T r(t)\sqrt{\frac{2}{T}}\ \sin\ (2\pi f_s t)\,dt$$

$$= \int_0^T r(t)\,\varphi_2(t)\,dt$$

The decision regions are shown in Fig. 16. Note that there are four regions. Since all decision regions are symmetric and since we are assuming each signal is equally likely, we can write;

$$P[E] = P(r_1, r_2 \notin R_1)$$

which equals

$$P[E] = 1-P\ (r_1, r_2 \in R_1)$$

This integral is difficult to evaluate. It becomes simpler if we remember that we can rotate this $45^o$ to generate the region of Fig. 17. Now we note that for $R_1$ the probability of being correct $(P[C/s_1])$ is

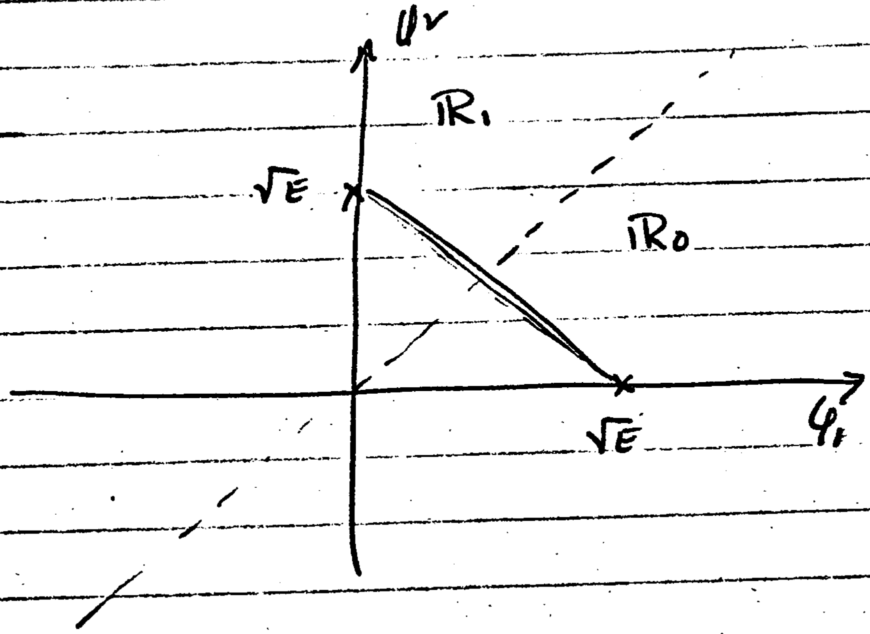



Fig 15 : FSK Decision Region

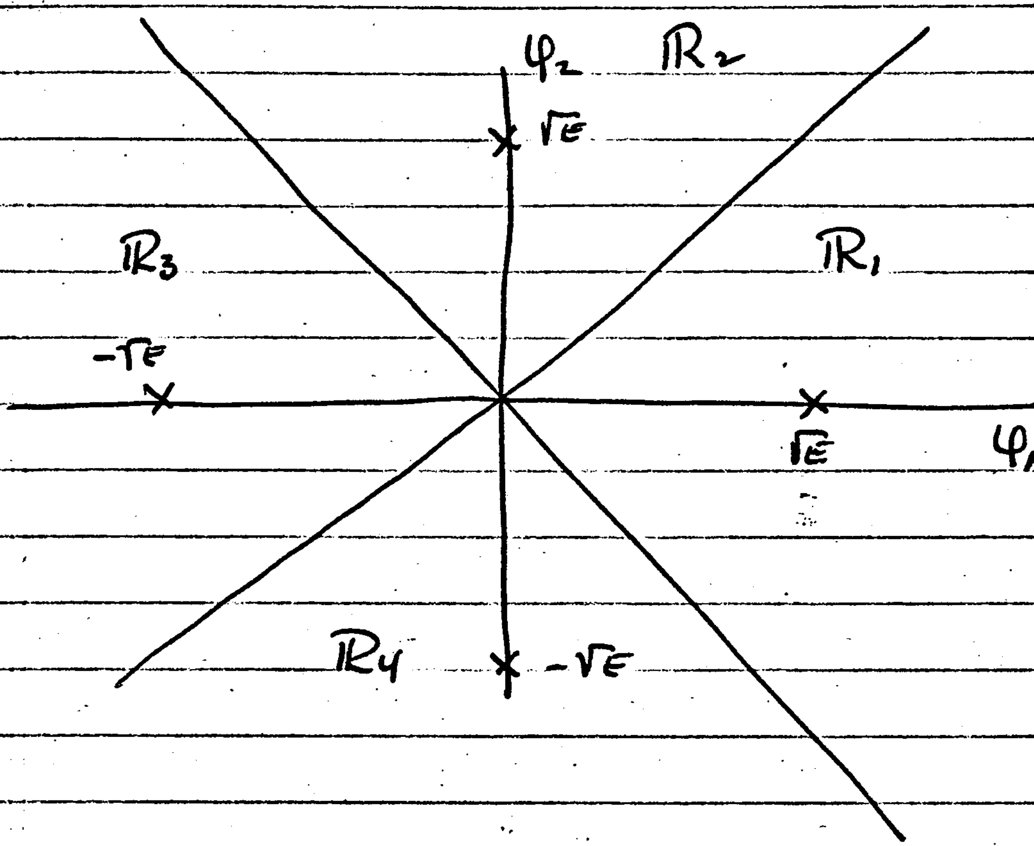



Fig 16.1 QPSK Decision Region

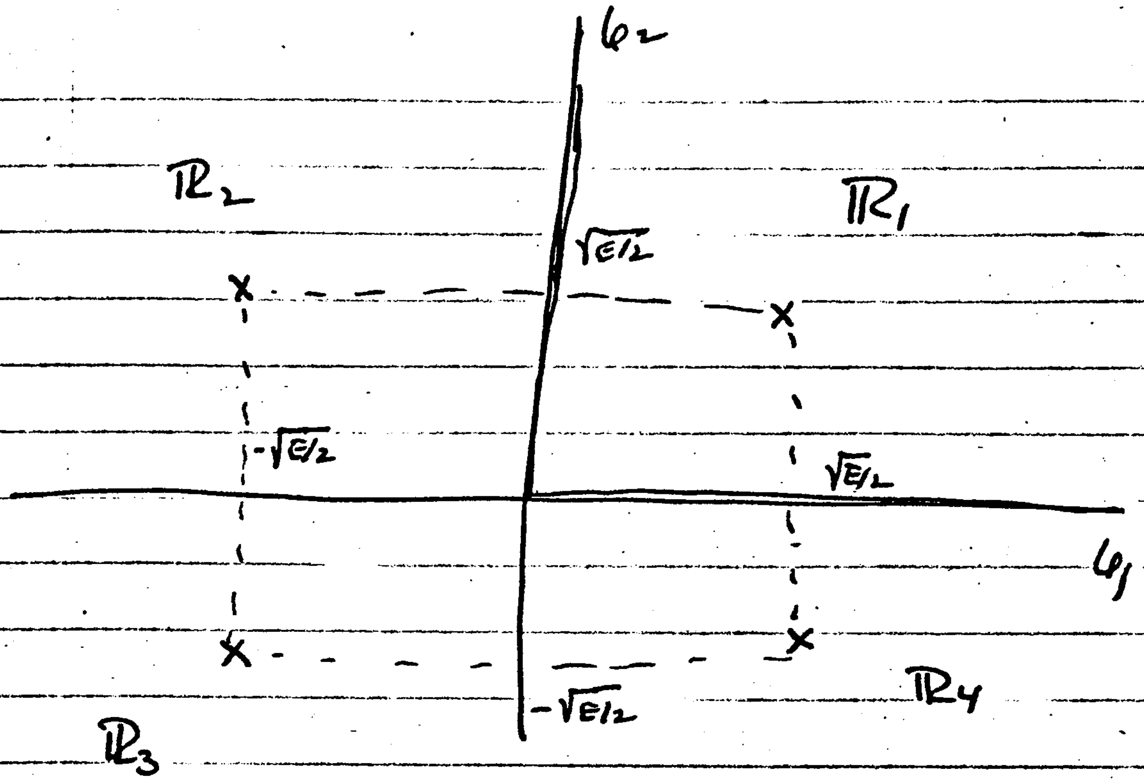



Fig 17 : Rotated QPSK Decision Regions

$$P[C/s_1] = P[r_1 > o \text{ and } r_2 > o]$$

But

$$r_1 = \sqrt{\frac{E}{2}} + n_1$$

$$r_2 = \sqrt{\frac{E}{2}} + n_2$$

where $n_1$ and $n_2$ are zero mean, independent, Gaussian, of variance $N_o/2$. Thus

$$P[C/s_1] = \left[\int_o^\infty \frac{1}{\sqrt{2\pi \frac{N_o}{2}}} \exp\left(-\frac{1}{2} \left(r - \sqrt{\frac{E}{2}}\right)^2\right) dr\right]^2$$

$$= \left[1 - Q\left(\sqrt{\frac{E}{N_o}}\right)\right]^2$$

Thus;

$$P[E] = 1 - \left[1 - Q\left(\sqrt{\frac{E}{N_o}}\right)\right]^2$$

It is important to note that this is not the bit error probability but the probability that the transmitted bit pair is interpreted as some other bit pair. Thus in comparing modulation schemes, the P[E] must be carefully interpreted. Fig. 18 provides a summary of these P[E] vs. $E/N_o$.

This is the word error probability! So state it!

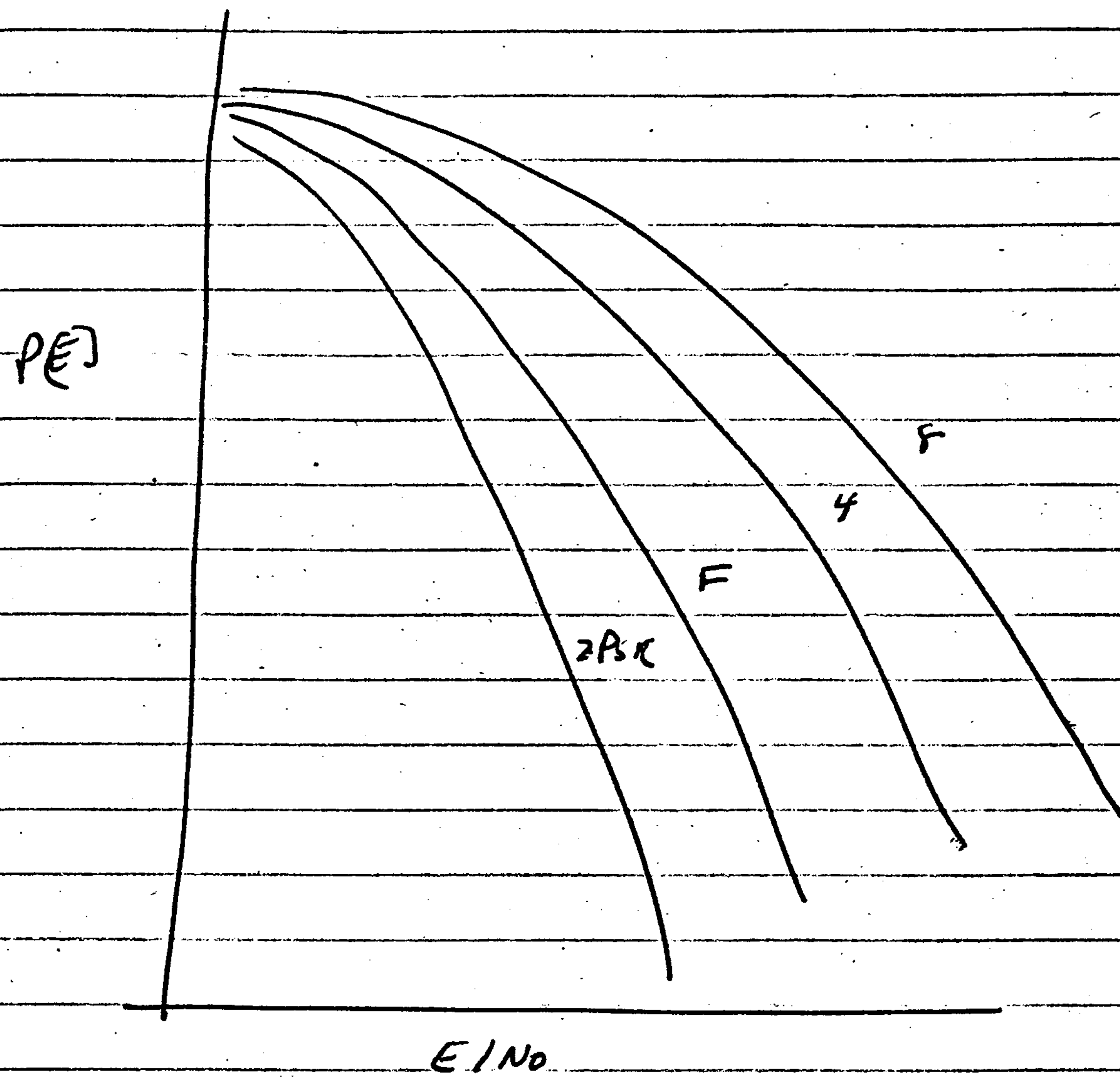



Fig 18 : Error Performance of Modulation Schemes

In the previous analysis we made several assumptions concerning information available to the demodulator. We specifically assumed the following:

1. That the carrier frequency was known and constant
2. That no additonal phases were added to the signal
3. That the time of bit occurrence and its duration were known.

In general we do not know these and they must be estimated in order to perform the optimum demodulation. In this section we will briefly discuss how this is done but it must be pointed out that this material is well beyond the scope of the present text and it is very adequately covered by Gagliardi, Lindsey, of Simon or Holmes.

Consider now the demodulator shown in Fig. 19. Here we assume that the signal is

$$r(t) = A \sin (\omega_c t + \theta_T(t) + \psi(t)) + n(t)$$

where $\omega_c$ is the carrier frequency, $\theta_T(t)$ the data waveform of bit duration T and $\psi(t)$ time varying phases added to the signal in transmission. $n(t)$ is additive white Gaussian noise. In Fig. 20 we show first the carrier recovery circuit that tracks $\omega_c$ and puts out a signal at IF. Second is a timing circuit that uses the data to recover timing. Third is the phase recovery to eliminate the effects of $\psi(t)$. Finally, we have the optimal demodulator.

Phase and carrier are normally obtained with a PLL - one device not two as you have shown. Or are you using first a FLL then a PLL? clarify!

It is important at the outset to note that since noise is present there will be errors due to frequency, phase and timing misalignments. In addition the errors in these misalignments are random. Thus the overall performance will be degraded. Let $P[E/\tilde{\omega}, \tilde{\varphi}, \tilde{\tau}]$ be the probability of error give frequency error $\tilde{\omega}$, phase error $\tilde{\varphi}$ and timing error $\tilde{\tau}$. Let $p(\tilde{\omega}, \tilde{\varphi}, \tilde{\tau})$ be the density of these errors. Then

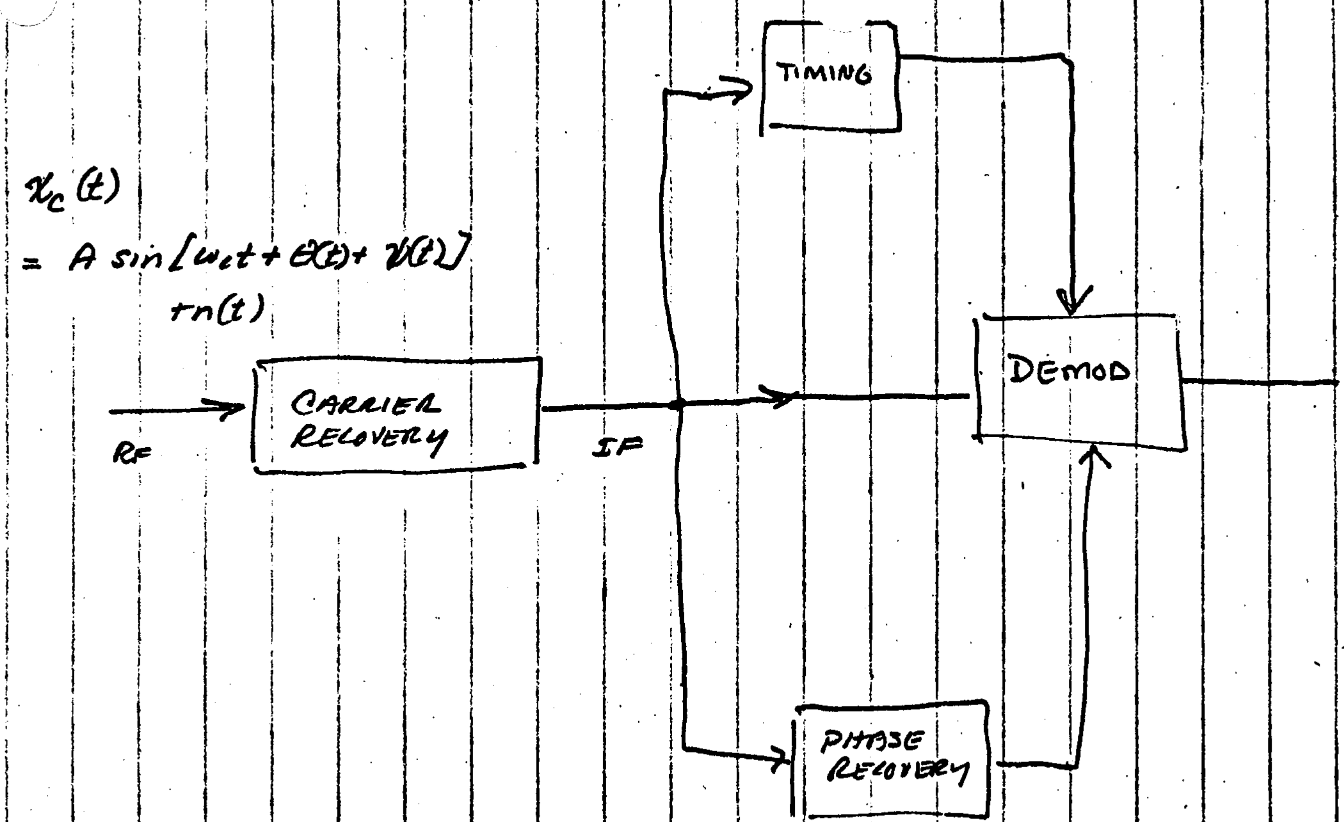



FIG 19 : Demodulator Functions

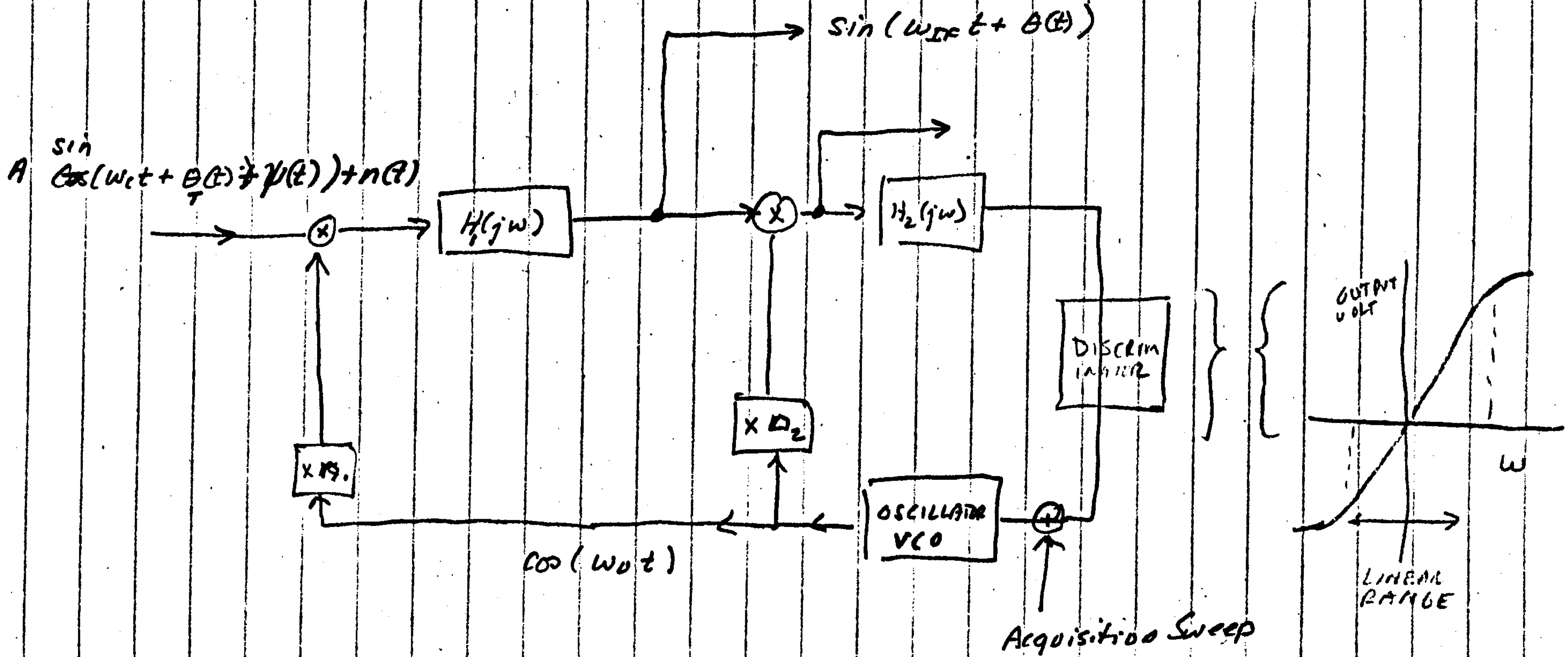



Fig 20 : Carrier Recovery Loop (FLL) ?

$$P[E] = \int P[E/\tilde{\omega},\tilde{\psi},\tilde{\tau}]\,p(\tilde{\omega},\tilde{\varphi},\tilde{\tau})\,d\tilde{\omega}\,d\tilde{\varphi}\,d\tilde{\tau}$$ ← not applicable to QPSK or PSK! Since $P(E|\tilde{\omega}\neq 0,\tilde{\varphi},\tilde{\tau}) = 1/2$

We shall then be able to plot P[E] versus $E/N_o$ before but we will note that the curve differs from theory as shown in Fig. 21. At a fixed P[E], the difference between the $E/N_o$ required in practice and that specified by theory is called the processing loss, or degradation $L_p$, and is typically in dB. In the above mentioned figure, for three cases, it is seen that the processing loss can be small or very considerable. Thus, in selecting a modem, one must be careful in assessing the true processing loss. Recent designs have reduced? these losses to about — awkward, 0.5 dB. Typical losses are 2-5 dB.

We now want to briefly go through the three functions. The first is carrier recovery. Fig. 20 depicts a carrier recovery loop. It operates in a closed loop fashion as follows. A VCO, voltage controlled oscillator, is used which has an output which equals

$$\cos(\omega_o t+\Delta\omega t)$$

where $\Delta\omega$ is proportional to the input voltage. Let us initially assume that $\Delta\omega = o$. Now two frequency multipliers are used, $\omega_1$ and $\omega_2$. At the output of filter $H_1(j\omega)$, a bandpass filter, we have a signal

$$A_1 \sin((\omega_c-n_1\omega_o)t + \theta_T(t) + \psi(t)) + n_1(t)$$ $\omega_{IF}$

We let $\omega_c-n_1\omega_o$ be the $\omega_{IF}$ or IF frequency. A second multiplication occurs and the signal is further filtered to yield at the output of the $H_2(j\omega)$, a baseband filter,

$$A_2\sin((\omega_c-n_1\omega_o-n_2\omega_o)t + \theta_T(t) + \psi(t)) + n_2(t)$$

Now a discriminator circuit is used whose output is proportional to the frequency at the input. Thus the output of the discriminator is;

$$K_d(\omega_c-n_1\omega_o-n_2\omega_o) + n_d(t)$$

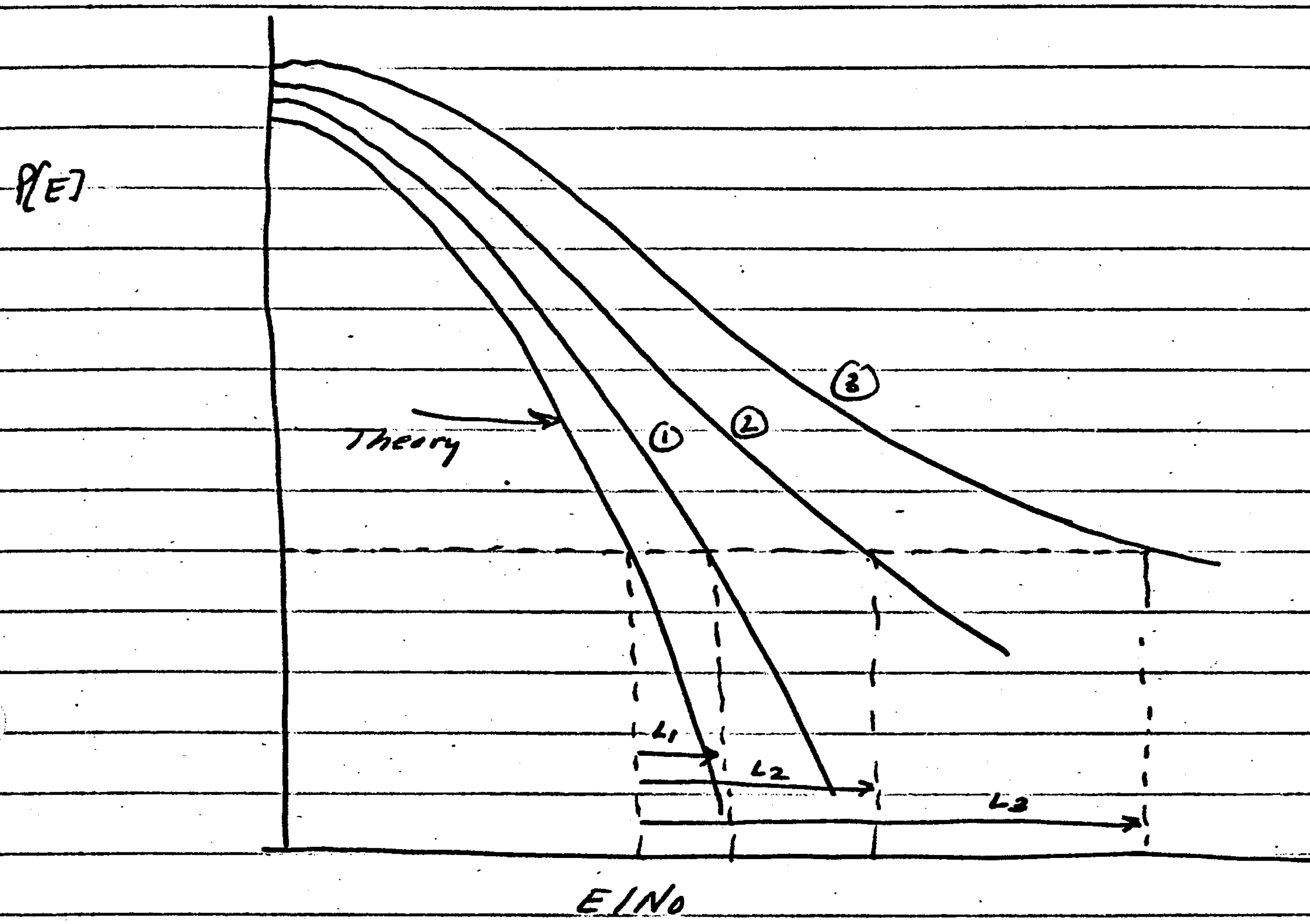



Fig 21 : Processing Loss    Noncoherent FSK?

where $n_d(t)$ is the total noise as reflected through the discriminator. Now if this signal is positive it increases $\omega_o$ which makes $(n_1+n_2)\ \omega_o$ approach $\omega_c$. If negative, it decreases. It thus tracks the carrier. However, there is some variation in $\omega_{IF}$. This will be eliminated in a subsequent loop. This is a FLL so why not call it that!

Timing recovery can be done in several ways. There are however two modes of operation; burst data and continuous. As we shall see later burst systems are typical of TDMA and continuous of FDMA/SCPC. In the burst mode a preamble is used and a sequence correlation is performed. In the continuous case actual data transitions are monitored. Typically timing errors are a small part of the demodulator error budget.

The more serious concern is signal recovery, and in effect phase recovery. Consider the demodulator shown in Fig. 22. Here the input is

$$A \sin(\omega_{IF}t+\theta_T(t)+\psi)+n(t)$$

$$= A \sin(\theta_T(t)) \cos(\omega_{IF}t+\psi)$$

$$+ A \cos(\theta_T(t)) \sin(\omega_{IF}t+\psi)+n(t)$$

Let us assume a 2$\phi$-PSK system with

$$\theta_T(t) = \begin{cases} \frac{-\pi}{2} & ;\ 0 \\ \frac{\pi}{2} & ;\ 1 \end{cases} \quad \text{every Tsec}$$

Then we can neglect the sin $(\omega_{IF}t+\psi)$ term. However, for the optimal demodulator we have to know $\cos(\omega_{IF}t+\psi)$ and the start and stop times for the integrator. The latter is the bit timing problem. The former combines the carrier and phase recovery problem. Note in the figure we show an estimate $\hat{\omega}_{IF}$ and $\hat{\psi}$. These must be obtained from the data signal.

For the PSK signal we do the above mentioned recovery with some form of a phase locked loop (PLL). Fig. 23 depicts a form called the Costas Loop. This loop operates as follows.

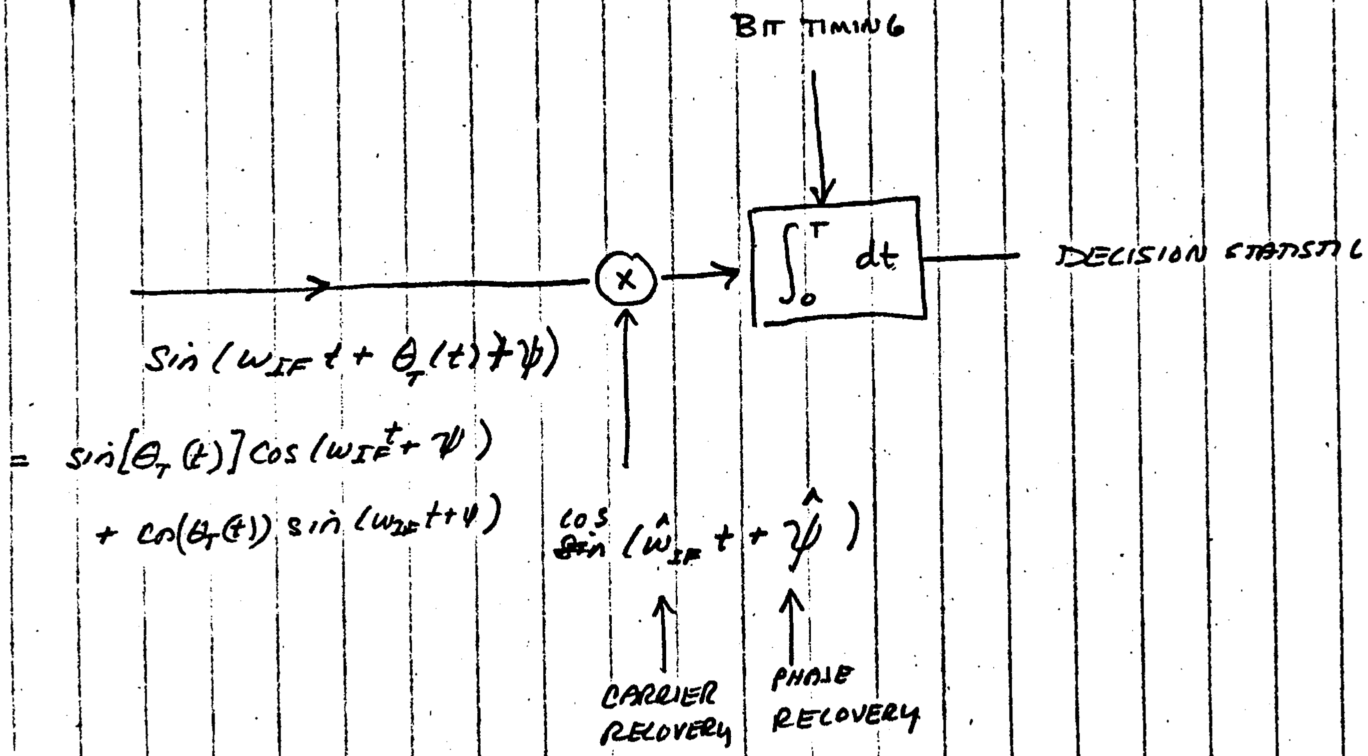



Fig 22: Demodulator Structure

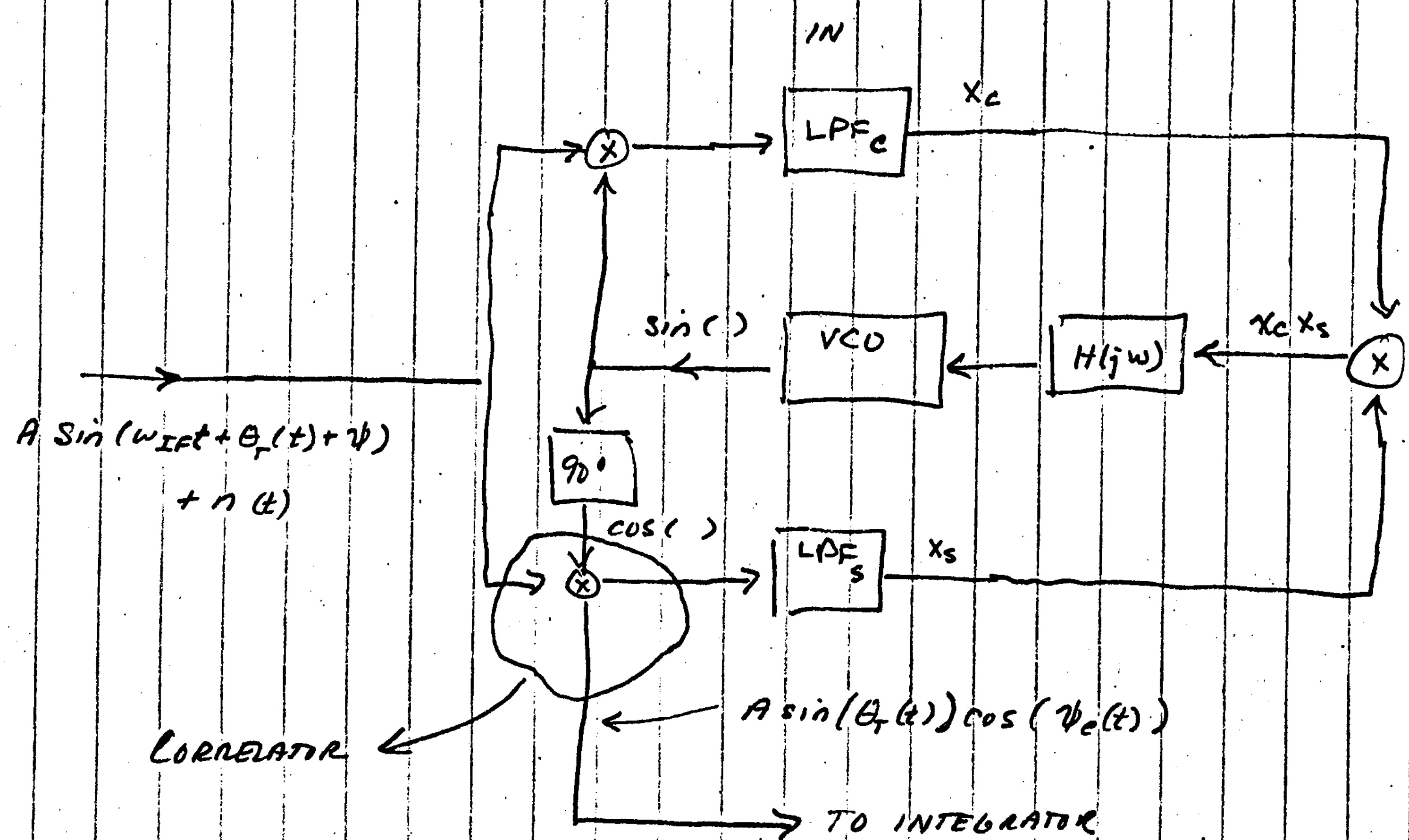



FIG 23: COSTAS LOOP PHASE RECOVERY AND CORRELATOR

At its heart is a VCO as before. It is assumed that the VCO is centered at $\omega_{IF}$, if not we could add another loop. Then the output of the VCO is

$$\sin (\omega_{IF}t+\phi(t))$$

where

$$\frac{d\phi}{dt}(t) = K_v \, e_{in}(t)$$

and $e_{in}$ is the input voltage. Then the true signal is divided as before. Again we separate the signal assuming PSK to be

$$r(t) = A \sin (\theta_T(t)) \cos (\omega_{IF}t \pm \phi(t)) + n(t)$$

with $\theta_T(t)$ as defined before ($\pm\pi/2$). Then $r(t)$ is multiplied by sin and cos of $\omega_{IF}t+\phi(t)$ and low pass filtered. The output of $LPF_c$, called the inphase channel, is;

$$X_c(t) = A_c \sin (\theta_T(t)) \cos (\phi(t)-\psi(t)+n_c(t)$$

and the output of $LPF_s$, the quadrature channel is

$$X_s(t) = A_s \sin (\theta_T(t)) \sin (\phi(t) - \psi(t))+n_s(t)$$

These are then multiplied to generate;

$$\Big[X_c(t) X_s(t)\Big]_{LP} = A_3 \Big[\sin^2(\theta_T(t))\Big]_{LP} \sin (\phi(t)-\psi(t)) + n_3(t)$$

which is the input to $H(j\omega)$. If we use PSK then the square of $\sin (\theta_T(t))$ eliminates the data dependence, and if $\phi(t)$ is close to $\psi(t)$ then $\sin (\phi-\psi) \simeq \phi-\psi$. Thus the loop forces tracking and eliminates data dependence.

The key factor to note is that the multiplier on the quadrature channel is nothing more than the correlator needed for the demodulator. Thus all we need now do is integrate. But we must note that if

$$\psi_e = \phi - \psi$$

is the loop error, the signal amplitude is now

$$A \cos \psi_e.$$

and is thus reduced. A good loop design minimizes $\psi_e$. Fig. 24 to 25 depict P [E] degradated for several of the demodulater errors. Here we note the processing losses.

A similar design holds for FSK signals. Fig. 26 depicts the circuit and Fig. 27 the PLL that is used.

You haven't defined $CNR_L$ and you haven't shown where these curves came from

i.e. $P[E] = \int P(E|\phi)\, p(\phi)\, d\phi$

etc

add more details!

Discuss figure 26!

Draw accurately! ✓

Reference?

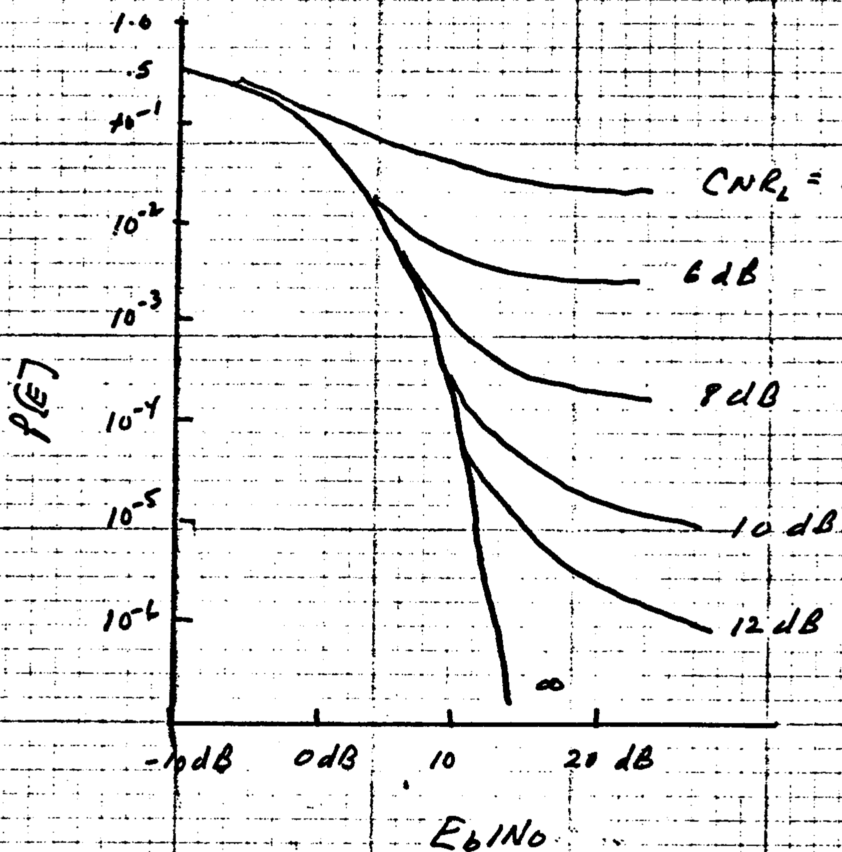



FIG 24: 2ØPSK BER As a Function of the PLL Carrier to Noise Ratio ($CNR_L$).

Reference?

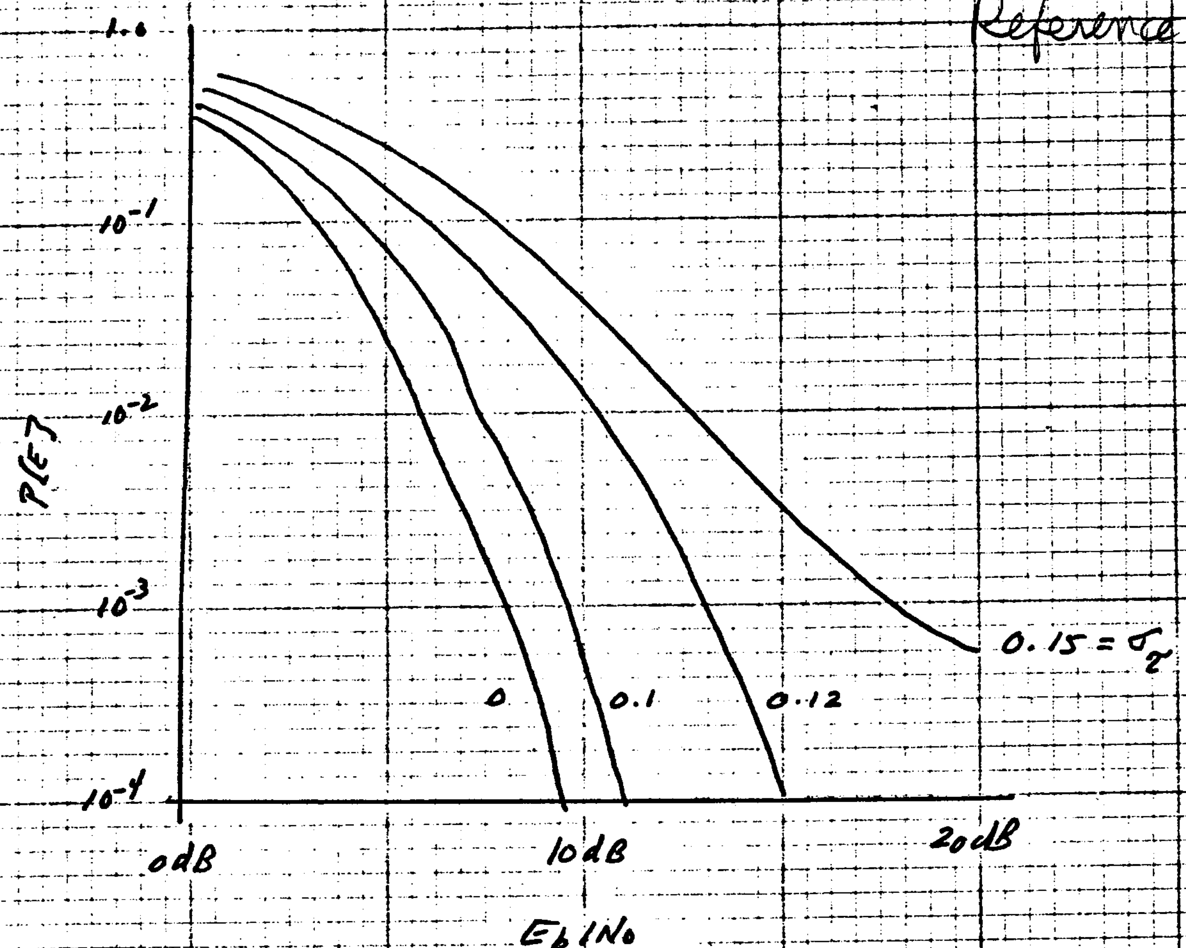
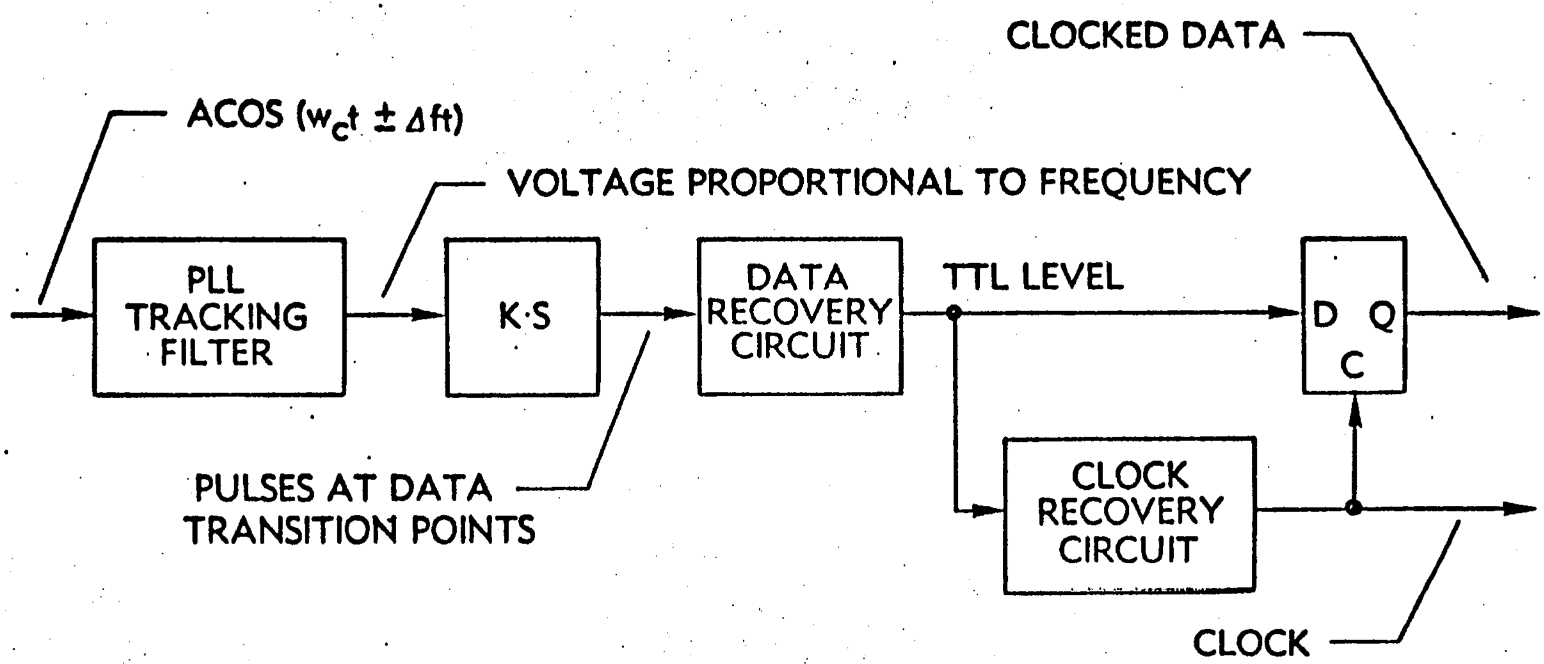
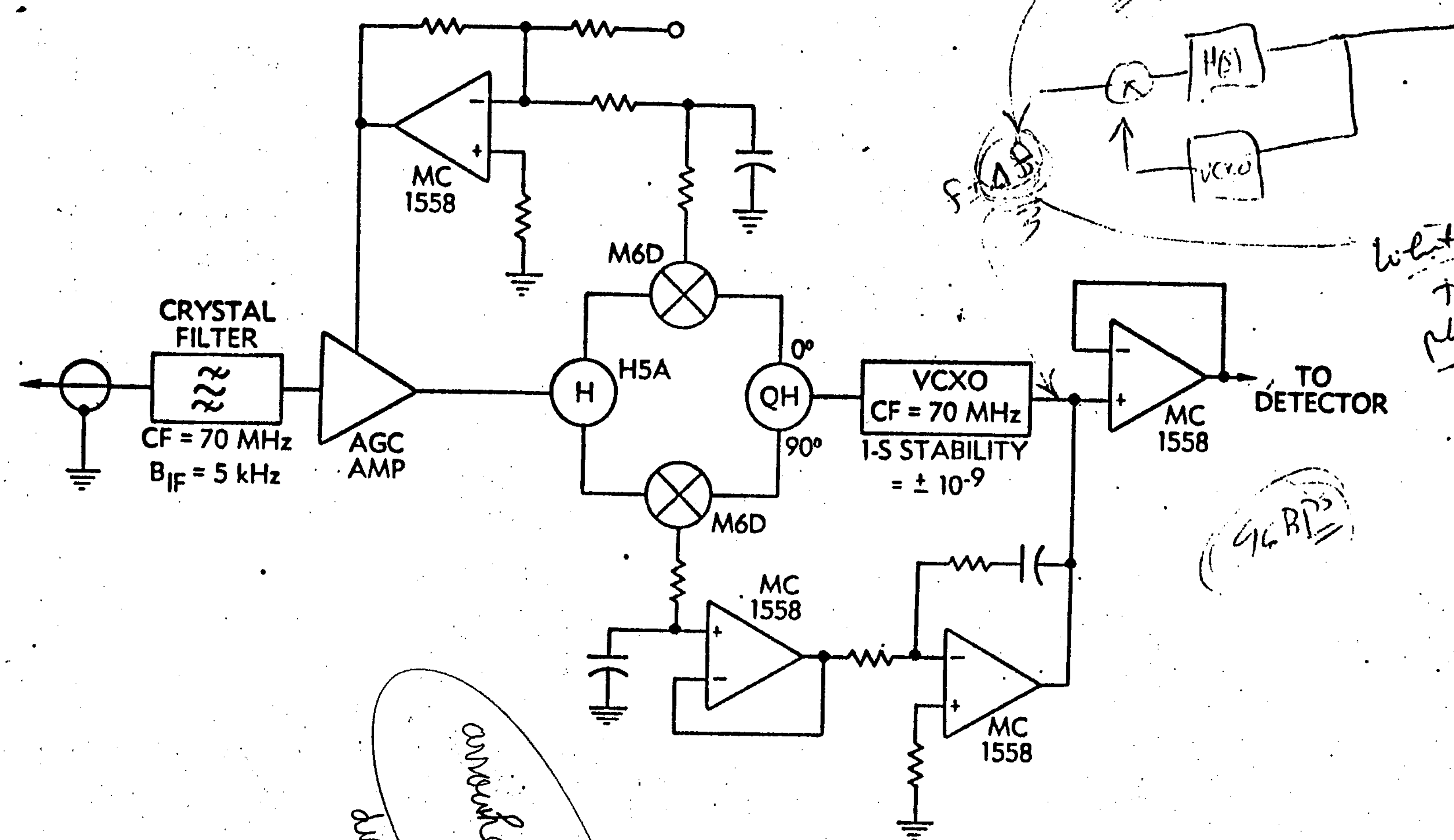



FIG 25: BER OF 2Ø PSK WITH TIMING Errors of variance $\sigma_\tau$ sec.

Figure ~~3.5.4~~ 26: FSK DEMODULATOR



27
Figure ~~3.5.10~~: PLL BLOCK DIAGRAM




phase noise spec, jitter of received signal

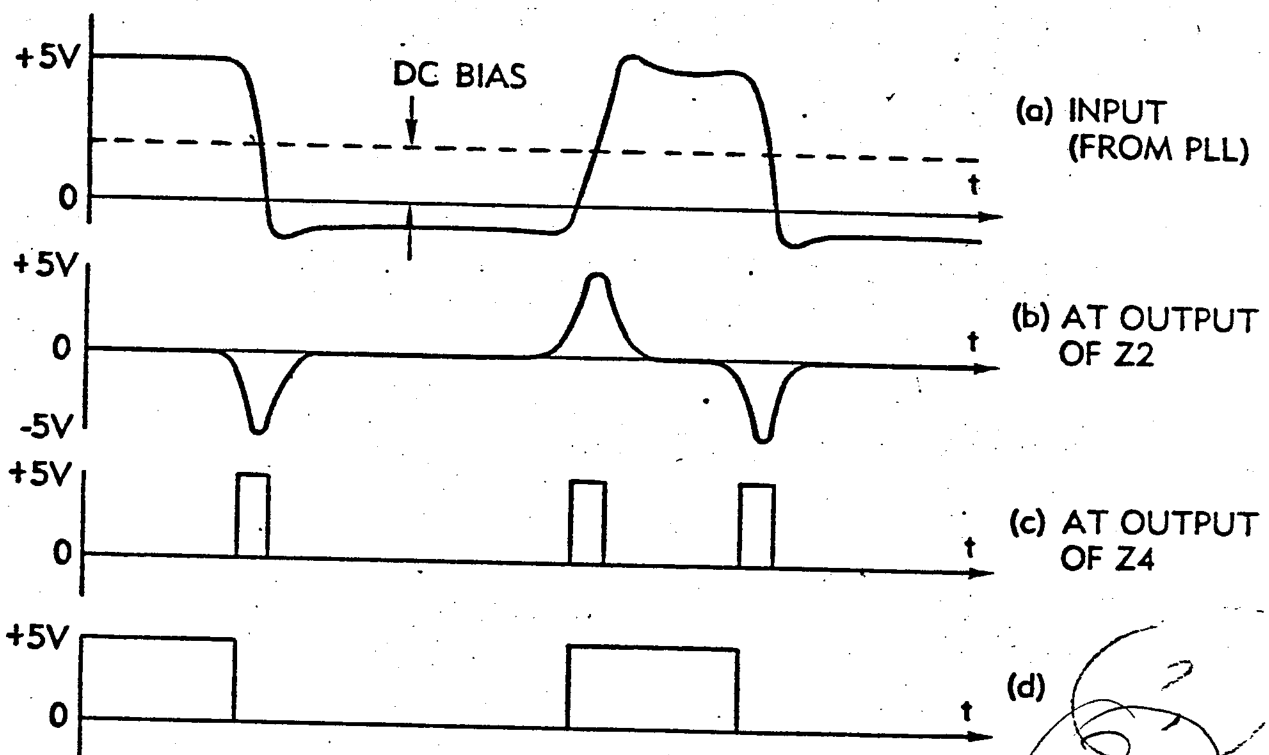
white noise + phase noise

arrowheads!
discussion!

discussion

28

Figure ~~3.5.1~~2: VOLTAGE LEVELS IN FSK DETECTOR




?

## 3.2 Coding

The use of coding in a satellite application is applied when there is a power to bandwidth tradeoff required. The coder takes an input bit stream at some rate $R_o$ and puts onto a coded sequence at a higher rate, $R_s$. Here $R_s$ is called the symbol rate. It is the higher rate that is used to provide input to the modulator. This higher rate across the channel thus requires more bandwidth. The tradeoff is that the $E_b/N_o$ required for a given bit error rate is reduced.

The satellite channel is characterized most appropriately by one that is an additive Gaussian white noise channel. No burst errors occur on this channel as occurs on telephone channels. The noise on the satellite channel is truly white with the Gaussian amplitude distribution. Thus the coding techniques applicable to these types of channels are convolutional codes and maximim likeliehood decoding. Block codes also!

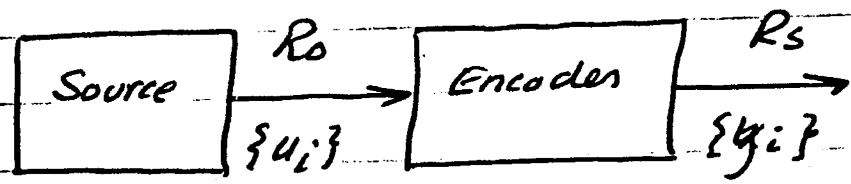
A typical structure for a convolutional encoder is depicted in Fig. 30. Here we have the source generating an infinite sequence $(U_i)$ at a rate $R_o$ per sec or $T_o=1/R_o$ sec/bit. The encoder is a device that outputs the sequence $(Y_i)$ at rate $R_s$. We shall assume the ratio

$$r = \frac{R_o}{R_s}$$

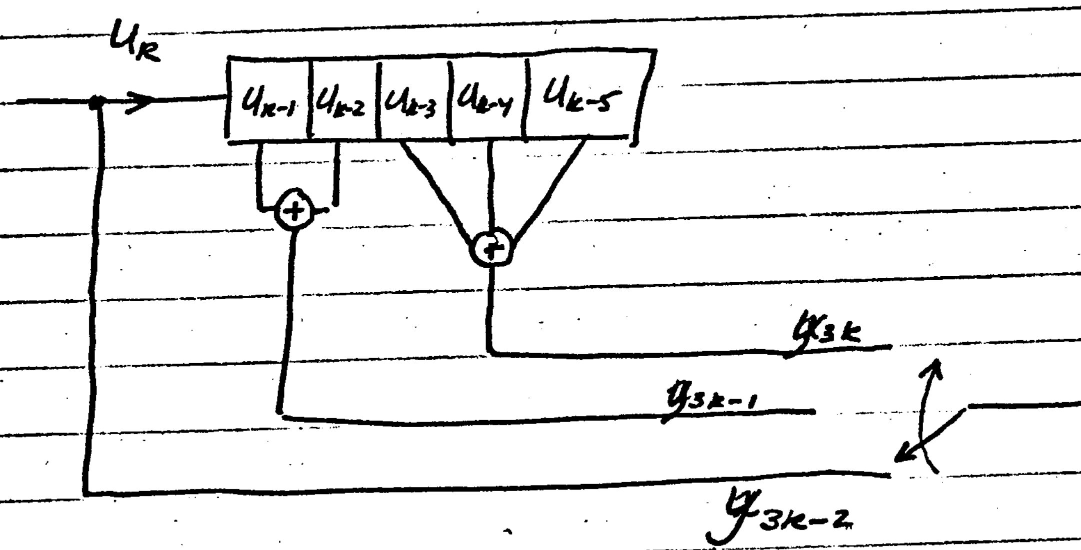
is a rational number and we shall term this the rate of the encoder.

We will also assume that both the $\{U_i\}$ and $\{Y_i\}$ sequences are chosen from the binary numbers 0,1. Fig. 30 b depicts a realization of just such a convolutional encoder. As the bit $U_k$ enters the encoder the previous bits $U_{k-1}$ ... are shifted down and only five below $U_k$ are held at any one time. Thus at time k the bits $U_k$, $U_{k-1}$, $U_{k-2}$, $U_{k-3}$, $U_{k-4}$, $U_{k-5}$ are used to generate three outputs; $Y_{3k-2}$, $Y_{3d-1}$, $Y_{3k}$. For this encoder we have;

3k-2 3k-1 3k




(a)
Rates




(b)
Typical Encoder

Fig 30 : Convolutional Encoder

$$Y_{3k-2} = U_k \qquad y_{3k-2} = u_k$$
$$Y_{3k-1} = U_{k-1} \oplus U_{k-2} \qquad y_{3k-1} = u_{k-1} \oplus u_{k-2}$$
$$Y_{3k} = U_{k-3} \oplus U_{k-4} \oplus U_{k-5} \qquad y_{3k} = u_{k-3} \oplus u_{k-4} \oplus u_{k-5}$$

where $\oplus$ is a binary addition. This is a rate 1/3 encoder with one bit in for every three symbols out. From a channel point of view, this will increase the bandwidth requirements by a factor of 3. As we shall show it will have a much more significant impact on power required.

We can generalize the convolutional encoder to that shown in Fig. 31. Here we have introduced vector notation for both $\underline{U}_k$ and $\underline{Y}_k$. We let for example

$$\underline{U}_k = (\underbrace{1 \ldots 1}_{p})$$

and

$$\underline{Y}_k = (\underbrace{1 \ldots 1}_{q})$$

not clear

p bits input
q symbols output

This then yields a rate R=p/q code. Here f( ) is an arbitrary transformation of the $\underline{U}$'s into $\underline{Y}$'s.

One of the key factors in the choice of a convolutional code is the rate R and the length of the memory, L, called the constraint length. This parameter pair (R,L) characterizes the chosen code to fair degree. The full characterization is the choice of interconnections in the design. That is, what input bits influence what output bits in what manner.

The output of the convolutional encoder depends upon two factors. The first is the present input, $U_k$, and the second is the L sequence $(U_{k-1}, \ldots U_{k-2})$ that is in the encoder. This latter sequence is called the state, $X_k$, of the encoder. Thus we have in the general form

$$Y_k = f(U_k, U_{k-1}, \ldots U_{k-2})$$
$$= f(U_k, X_k)$$

We typically find it useful to call the ~~triple~~ a pair $(X_{k+1}, X_k)$ the transition $\zeta_k$. Thus

$$\zeta_k \triangleq (X_{k+1}, X_k)$$

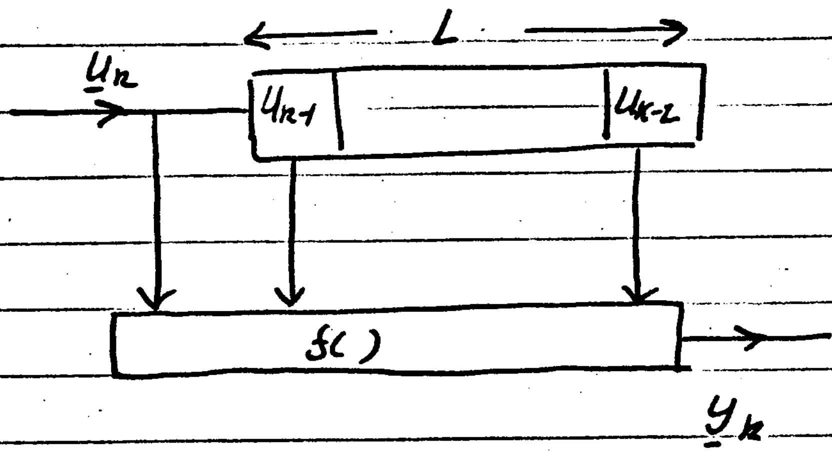



Fig 31 : Generalized Convolutional Encoder

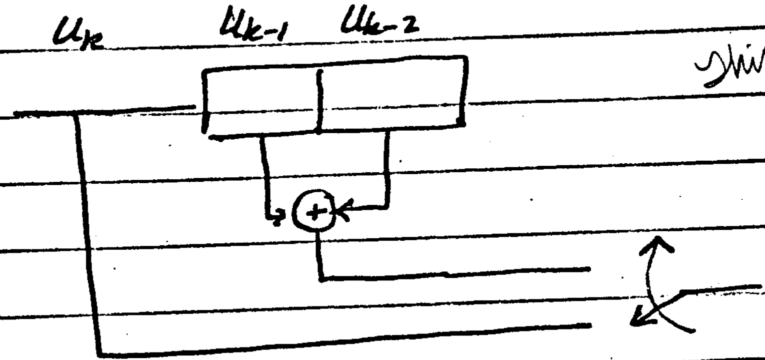
The time characteristics of this encoder can be presented in two ways. Both utilize the concept of state and demonstrate transistion effects on the system due to inputs and the corresponding outputs.

The first approach is called the trellis. At each instant of time a set of points is drawn, one for each state. Then lines are drawn to the states that can be reached at the next instant. These lines are characterized by the input required to make that transition and the output produced.

As an example of the above consider the encoder shown in Fig. 32. The encoder uses binary alphabets and the state is one of $2^2=4$. The trellis diagram is also shown. We assume that the system starts in the all zero state. Then as the input changes the state moves to the right in the trellis with the outputs as shown. Note that states at two consecutive times define a transition which is a 1:1 mapping to input and output.

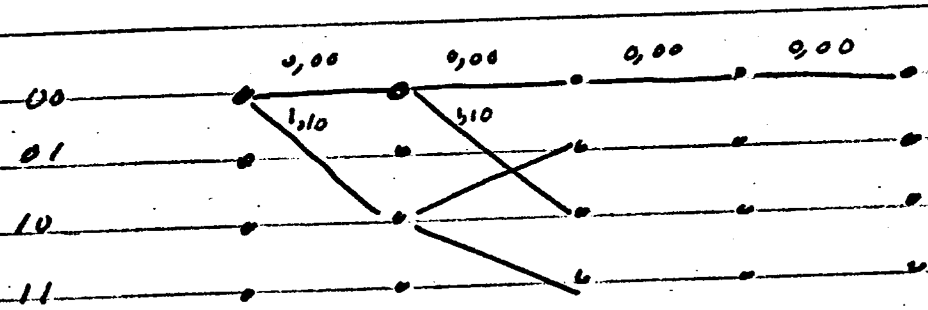
Another approach is to use a finite state machine technique where the states are defined independent of time but transitions from states to states are shown with inputs and outputs. Fig. 33 depicts such a finite state machine architecture. Note that certain states loop on themselves.

We shall return shortly to the choice of codes for the system when we complete the issue of performance analysis. Before doing so let us consider the effects of transmitting the coded sequence across an AGWN channel. The coded sequence gets modulated and then transmitted across the channel and then demodulated. As we noted before, the output of the demodulator, which is either a matched filter or integrator correlator, is a quantity that is used to make a bit decision. We typically would make a hard decision as to what the bit was, however with an encoded sequence this is not optimal due to the correlation of the real data bits. Thus for convolutional encoded systems we keep the full output of the demodulator. We can represent this output as;




This is a systematic encoder! State this fact!

(a)
Encoder




(b)
Trellis

Fig 32: Convolutional Encoder and Trellis

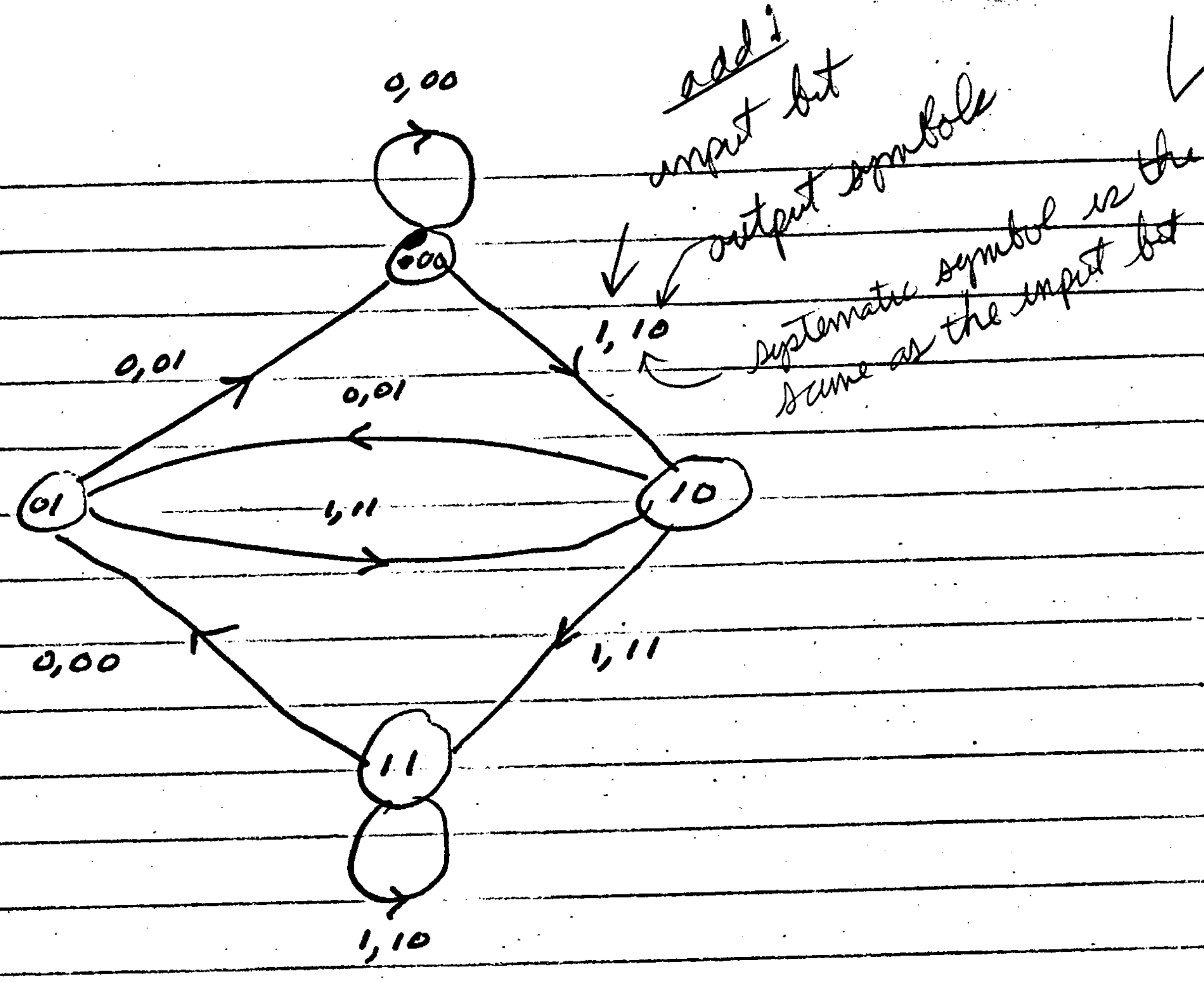



Fig 33 : Convolutional Encoder Finite State Machine

$$z_k = \sqrt{E}\, Y_k + W_k$$

where E is signal energy and $W_k$ is a zero mean white Gaussian sequence with variance $N_o/2$.

The decoding problem can be stated as follows. Let $\underline{X}$ represent the sets of states which in turn define $\underline{U}$ the sets of inputs. Recall that the sets of transitions, $\zeta$, are equivalent to these. Let $\underline{Z}$ represent the set of all outputs corresponding to this input on an AGWN channel as defined above. Now given Z we want to optimally determine $\underline{U}$, $\underline{X}$ or $\underline{\zeta}$.

The optimality criterion is to select the sequence that maximizes the a posteriori probability. That is let p(x/z) be the probability a probability density of the input sequence given what we receive. The decoder then is the processor that finds the x to maximize this measure. As we noted in the case of demodulation, this technique does provide minimum error performance. Now note that

$$\max_{X} p(x/z) = \max_{X} \frac{p(z/x)\,p(x)}{p(z)}$$

$$\hat{=} \max_{x} (p(z/x)\,p(x))$$

where we have eliminated p(z) since it does not depend on $\underline{x}$.

Now let us assume that there are K such sequences of $X_k$, $Z_k$. Then due to the nature of the encoder and the channel we can write

$$p(\underline{x}) = \prod_{k=0}^{K-1} p(X_{k+1}/X_k)$$

which follows from the fact that the encoder is a finite state machine and;

$$p(z/x) = p(z/\zeta) = \prod_{k=0}^{K-1} p(z_k/\zeta_k)$$

which follows from the fact that the channel is memoryless. Thus

$$p(z,x) = \prod_{k=0}^{K-1} p(X_{k+1}/X_k) \prod_{k=0}^{K-1} p(z_k/\zeta_k)$$

To optimally decode we want to find the x to minimize or find the "path" through the trellis that minimizes a "distance" defined by that probability. This minimum distance path is the essence of the Viterbi algorithm (VA).

In order to develop the VA, define a length $\lambda(\zeta_k)$ as

$$\lambda(\zeta_k) = -\ln P(x_{k+1}/X_k) - \ln P(Z_k/\zeta_k)$$

Note that since we wish to "maximize" the probability we want to "minimize" the path length. The total path length is then defined as;

$$-\ln P(x,z) = \sum_{k=0}^{K-1} \lambda(\zeta_k)$$

Then the x that is decoded is that x which minimizes the total distance given z. Now let $X_o^k$ be a segment of x given by $(x_o, x_1 \ldots x_k)$. Let $\hat{x}(x_k)$ be the shortest path to the state $x_k$. This is called the survivor.

Now at level k in the trellis there are M possible $x_k$ values. that is if

$$x_k = (U_{k-1} \ldots U_{k-L})$$

and U is binary item $M=2^L$. So there exist $2^L$ $x_k$ at each depth k in the trellis and there is a survivor to each of them.

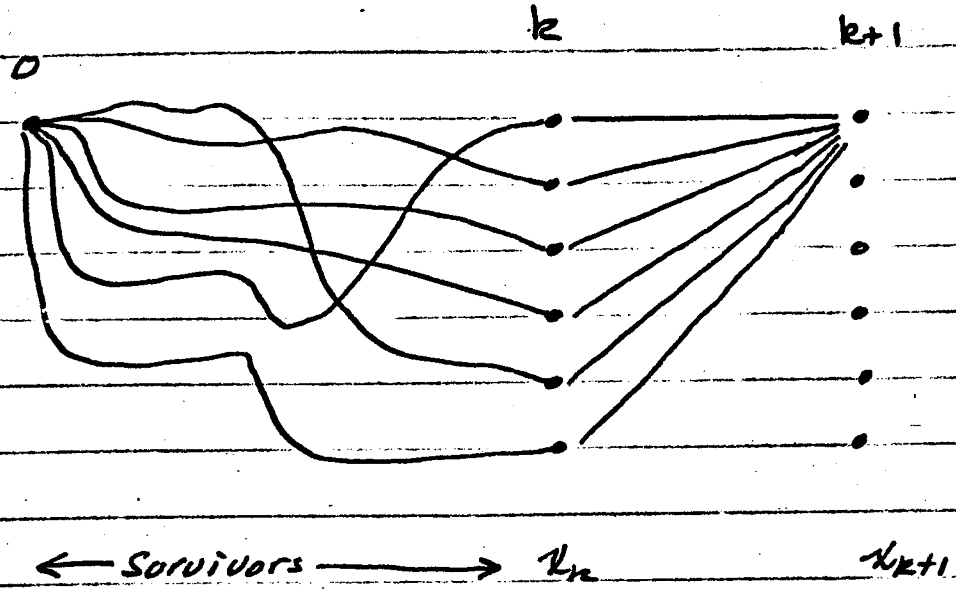
Fig. 34 depicts all the survivors up to k from 0. Note that for each $x_k$ there is a survivor. Now let us consider what happens when we go to k+1. Again we are looking for the shortest path. But note; the shortest path at k+1 must begin with one of the survivors at node k. If it did not then the shortest path at k+1 would not contain the shortest path at k, but some other. This would lead to a contradiction since no other path at k was shorter.

Thus at time k we need only to remember the survivors $\hat{x}(x_k)$ to that point plus the lengths of those survivors $\Gamma(x_k)$ which is

What is this letter?

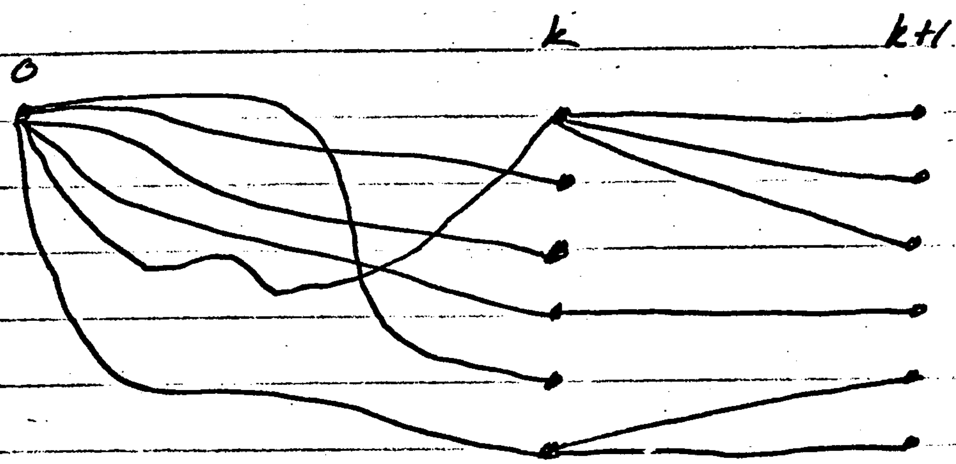
$$\Gamma(x_k) = \lambda(\hat{x}(x_k))$$

Now consider the state at k+1 as shown in Fig. 34 a. We must consider all paths that enter that node. Let $\Gamma(x_{k+1}, x_k)$




(a)

Single Node Survivor Test




(b)

Survivors at k+1

Fig 34 : Survivor Trajectories

be the distance of each path. Then

$$\Gamma(x_k+1, x_k) = \Gamma(X_k) + \lambda(\zeta_k)$$

is computed for each of the possible paths. Then the path which has the minimum is chosen and this is the survivor for that $X_{k+1}$ at k+1. This is then done for all $X_{k+1}$ at level k+1 and we define

$$\Gamma(x_{k+1}) = \min_{x_k} \Gamma(x_{k+1}, x_k)$$

The VA can be summarized as follows:

1. Define k at as the time index
2. Let $X_{k,j}$, j=1,...M be a specific $x_k$ at depth k. Recall that M values exist.
3. Let $\hat{x}(x_{k,j})$ be the survivor to that node.
4. Let $\Gamma(x_{k,j}) = \Gamma(x(x_{k,j}))$ be the survivor length.
5. Let k=0 be the start point.
   Let $\hat{x}(x_o) = x_o$ be a know start point
   Let $\Gamma(x_o) = 0$
6. Compute for each i the following at depth k+1.

$$\Gamma(x_{k+1,i}, X_{k,j}) = \Gamma(X_{k,j})$$
$$+ \lambda(\zeta_{k,i,j} = (x_{k+1,i}, x_{k,j}))$$

Find for each i the j such that

$$\Gamma(x_{k+1,i}) = \min_{x_{k,j}} \Gamma(x_{k+1,i}, x_{k,j})$$

7. Define $\hat{x}(x_{k+1})$ and $\Gamma(x_{k+1})$ for all j=1,...M and store these.
8. Increase k and repeat.

We can see in Fig. 34 b that the survivors are extensions from the previous node. In addition the survivors eliminate previous survivors and create new ones. This algorithm is basically the dynamic programming algorithm that is used in discrete time systems.

The following example considers the VA applied to a specific channel.

Example:

Consider the encoder shown before in Fig. 32 . The state can take on any one of 4 values (00,01,10,11). Let us assume that we are starting in the 00 state. Now using the algorithm recall that

$$\lambda(\zeta_k) = \log p(x_{k+1}/x_k) - \log p(Z_k/\zeta_k)$$

Now note that in this case if the data source is totally random we have

$$p(x_{k+1}/x_k) = 1/2 \text{ or } 0$$

Thus certain transitions are impossible and they have infinite lengths.

Now let us assume that we use a 2Ø-PSK modulator format for the encoder symbols and the output is the matched filter output of the two symbols due to $Y_k$. These are $Z_{k1}$ and $Z_{k2}$. Remember that $U_k$ generates $Y_{k1}$ and $Y_{k2}$ which are binary.

These binary numbers are used to generate signals in a modulator. For this example we have chosen a 2Ø-PSK system where 1 corresponds to a $+\sqrt{E}$ amplitude and 0 to $-\sqrt{E}$. Note that E represents the energy per symbol and not the energy per bit.

With this modulation scheme, using an optimum demodulator such as a matched filter we obtain outputs $Z_{k1}$ and $Z_{k2}$. Thus we have

$$Z_{ki} = \pm \sqrt{E} + W_{ki}$$

where $W_{ki}$ is zero mean Gaussian of variance No/2 and independent. ~~Table~~ Fig 35 depicts the values for the transition probabilities for the state and the amplitudes for the modulation.

Now we can write

$$p(Z_k/\zeta_k) = p(Z_{k1}, Z_{k2}/\zeta_k)$$

and with this system

$$p(z_k/\zeta_k) = C \exp\left[\frac{-(Z_{k1} \pm \sqrt{E})^2}{2N_o}\right] \exp\left[\frac{-(Z_{k2} \pm \sqrt{E})^2}{2N_o}\right]$$

where the sign of $\sqrt{E}$ depends on $\zeta_k$ as in Table 35 b.

Then clearly we have;

$$\lambda(\zeta_k) \simeq (Z_{k1} \pm \sqrt{E})^2 + (Z_{k2} \pm \sqrt{E})^2 + \text{possibility weight}$$

?

Note that the possibility weight is a constant or infinity for an impossible state transition. Now it should be clear how $\lambda(\zeta_k)$ depends upon the states $X_k, X_{k+1}$ and the measurement.

In practical applications the approach in the above example has several problems. Since we are using the actual matched filter output we must carry the $\lambda(\zeta_k)$ around in a sufficient number of significant bits. In addition the comparison process requires comparison between real numbers. A simplification results if we quantize the outputs of the demodulator to several levels so that they are easily stored and compared. As we shall show the difference in performance is insignficant. In the following example we consider such quantizations.

Example: Using the same scheme as in the previous example except that we use a 2 bit or 4 level quantizer. This is shown in Fig. 36 a where 4 levels of output 00,01,10,11 are shown. Now note that $Z_k$ contains two values each quantized to four levels. Thus there are 16 possible $Z_k$ values.

| $x_k$ \ $x_{k+1}$ | 00 | 01 | 10 | 11 |
| --- | --- | --- | --- | --- |
| 00 | 1/2 | 0 | 1/2 | 0 |
| 01 | 1/2 | 0 | 1/2 | 0 |
| 10 | 0 | 1/2 | 0 | 1/2 |
| 11 | 0 | 1/2 | 0 | 1/2 |

(a)

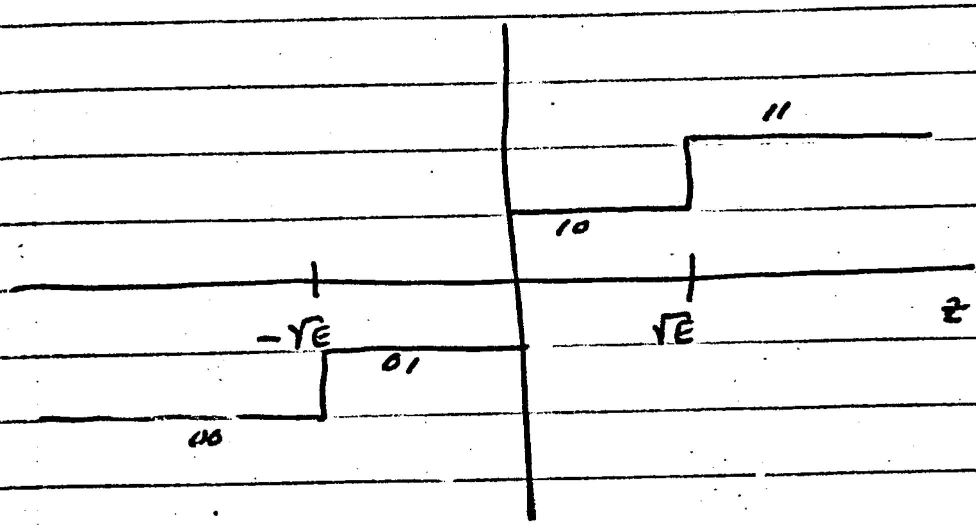
Transition Probabilities $p(x_{k+1}/x_k)$

| $x_k$ \ $x_{k+1}$ | 00 | 01 | 10 | 11 |
| --- | --- | --- | --- | --- |
| 00 | $-\sqrt{E}, -\sqrt{E}$ | — | $\sqrt{E}, -\sqrt{E}$ | — |
| 01 | $-\sqrt{E}, \sqrt{E}$ | — | $\sqrt{E}, \sqrt{E}$ | — |
| 10 | — | $-\sqrt{E}, \sqrt{E}$ | — | $\sqrt{E}, \sqrt{E}$ |
| 11 | — | $-\sqrt{E}, -\sqrt{E}$ | — | $\sqrt{E}, -\sqrt{E}$ |

(b)

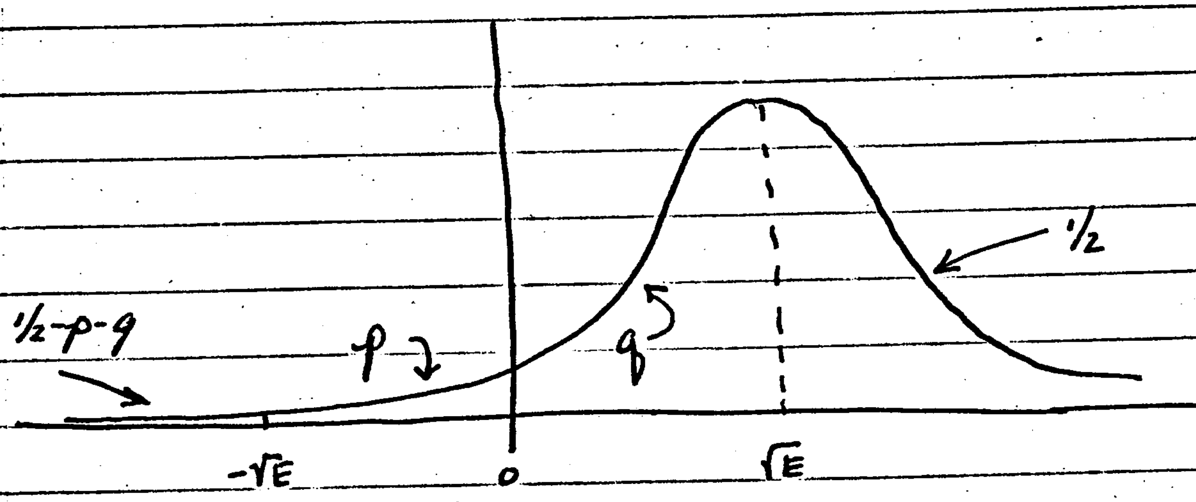
Output Amplitudes

Fig 35: Encoder Channel Output Characteristics




(a)

Quantizer




(b)

Regions and Probabilities

Fig 56: Quantization of Matched Filter Output

The probabilities of these quantized $Z_k$ values, called $Z_{k1}^{l}$ and $Z_{k2}^{l}$ can be given as;

$$p(Z_{ki}^{l}/\zeta_k) = 1/2, q, p, 1/2-p-q$$

depending on $\zeta_k$ and $Z_{ki}^{1}$ as shown in Fig. 36 b. Let us now organize the outputs in the form

$$\left.\begin{array}{c} 00,00 \\ \vdots \\ 11,11 \end{array}\right\} 16$$

where all 16 are shown. Consider first the output 00,00. Given that output we can write the $p(Z_k^{l}/\zeta_k)$ we take the $Z_k^{l}$ pair and the $\zeta_k$ 4-tuple and do a table look up to get $p(Z_k/\zeta_k)$. The table has a page for every possible output pair. Note that the probabilities are as defined in Fig. 37 b as the areas under the Gaussian density.

Now in this particular example there are 10 possible output ~~transistor~~ transition probabilities. These can be grouped in descending order of magnitude and further quantized as follows:

$$\begin{aligned}
10 &= 1/2.1/2 \\
9 &= 1/2.q \\
8 &= 1/2.p \\
7 &= qq \\
6 &= pq \\
5 &= pp \\
4 &= 1/2\ (1/2-p-q) \\
3 &= q\ (1/2\ p-q) \\
2 &= p\ (1/2-p-q) \\
1 &= (1/2-p-q)^2
\end{aligned}$$

Then if the ln operation is taken, the distances resulting can also be quantized in reverse direction. This can then be converted into binary and stored. This allows for simple binary operations on distance metrics.

| $x_k$ \ $x_{k+1}$ | 00 | 01 | 10 | 11 |
|---|---|---|---|---|
| 00 | $1/2 \cdot 1/2$ | — | $1/2 \cdot (1/2-p-q)$ | — |
| 01 | $1/2 \cdot (1/2-p-q)$ | — | $(1/2-p-q) \cdot (1/2-p-q)$ | — |
| 10 | — | $1/2 \cdot (1/2-p-q)$ | — | $(1/2-p-q) \cdot (1/2-p-q)$ |
| 11 | — | $1/2 \cdot 1/2$ | — | $1/2 \cdot (1/2-p-q)$ |

$(11,11)$

16 pages

$z_k' = (00,00)$

Fig ~~3b~~ 37: Output Transition Tables and Pages for Quantized Measurements

It has been shown by extensive simulation that the degradation in this soft decision, output quantized, metric quantized scheme is only 0.25 dB. Ref?

We can now proceed to the performance analysis of these decoders. The following result, as derived in Viterbi Ref and Omura, provides a reasonably tight upperbound to the bit error probability. The derivation is beyond the scope of the present text but its analysis and interpretation is straightforward. For a rate b/n convolutional code, with a binary channel input and symmetric channel output (e.g., symmetric $p(z/\zeta)$) we can show that the bit error probability $P_b$ is bounded by

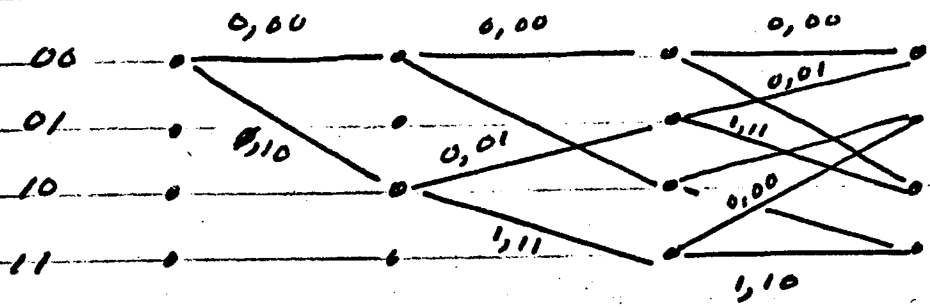
$$P_b \le \frac{1}{b} \frac{\partial T(D,I)}{\partial I} \Big/_{I=1,D=Z}$$

where $T(D,I)$ in the code ~~transitor~~ tim function and

$$Z = \sum_{z_i} \sqrt{p(z_i/o)\,p(z_i/1)}$$

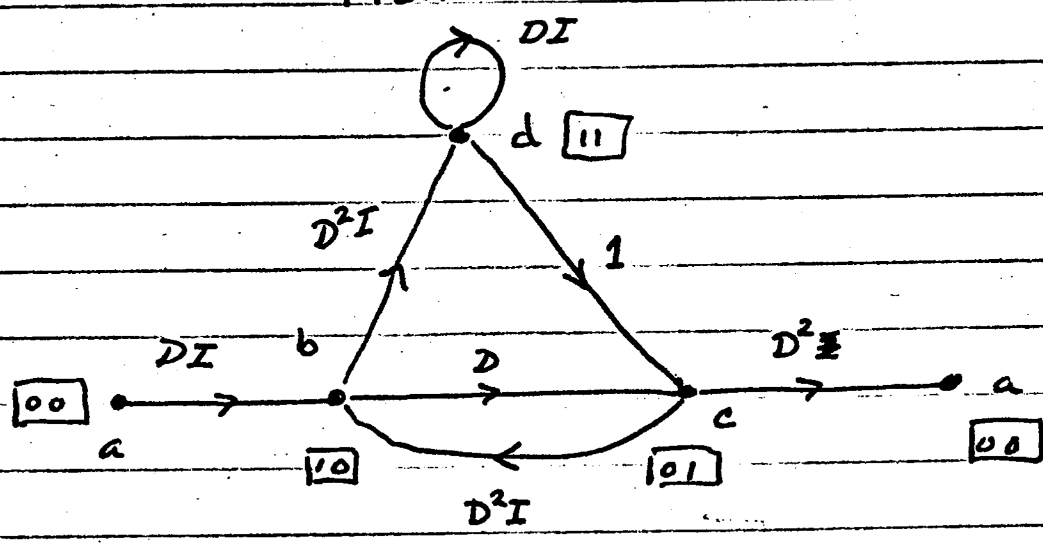
and $p(z_i/o)$, $p(z_i/1)$ are the output probabilities (or densities) of the demodulator output given a 0 or 1 into the channel. Thus $P_b$ depends on the code through $T(D,I)$ and on the channel through Z. Note that we take a symbol at a time in Z and that we evaluated these in previous examples. Note also that these depend also upon the modulation scheme.

Let us now return to an evaluation of $T(D,I)$. To do this we must use the trellis developed previously. Using the encoder in Fig. 32 we have redrawn the trellis in Fig. 38 a. Now with convolutional codes we can choose an arbitrary path and consider deviations from it for performance purposes. Thus we shall choose the all zero path. Two factors must be considered in error events. First is if the all zeros are sent then what paths occur with a 1 being sent. That is what transitions are 1 input dependent. Second we know that the output of the encoder is 00 for the all zero path each time. What is the difference or distance between other paths from the correct one. Also we must note that if an error occurs it will be a path that diverges from and remerges with the all zero path.




(a)

Trellis




(b)

State Diagram

Fig 38 : Calculation of $T(D,I)$

We can address these issues with the help of the state transistion diagrams shown in Fig. 38 b. Here we start in the all zero state and end up back again. Shown are all possible paths, including loops. Now we label each path with an enumerator to count the two events above. For each path that results from a 1 input we place an I. For each path that has an output distance of k from 00 we place a $D^k(1,D,D^2)$. Thus, for example, the transition from 00 to 10 is due to a 1 input and the output is 10 and has distance 1 from 00; therefore it has weight DI. We can obtain T(D,I) as follows. Define a transmission at each node as;

$$T_b = DI + D^2 I T_c$$

$$T_c = DT_b + Td$$

$$T_d = DIT_d + D^2 IT_b$$

$$T(D,I) = D^2 T_c$$

Solving for T(D,I) yields

you can use some details here!

$$T(D,I) = \frac{D^4 I}{1-DI-D^3 I}$$

Note that this code has a minimum distance of 4, that is the shortest path has an exponent of 4 on D. This implies that at high signal to noise ratios where Z is small

$$T(D,I) \simeq D^4 I$$

so that

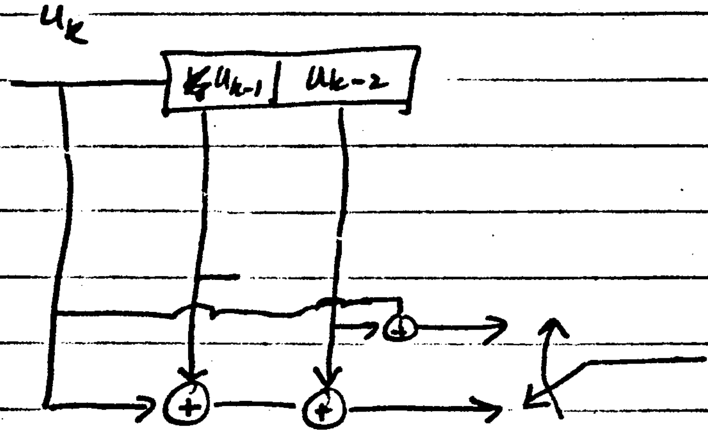
$$P_b \simeq \frac{1}{b} Z^4$$

Now we can ask if for the same constraint length but with a different set of connections to generate $Y_k$, can we improve this. The answer is yes. Such an encoder is shown in Fig. 39 and it has a T(D,I) of

$$T(D,I) = \frac{D^5 I}{1-2DI}$$

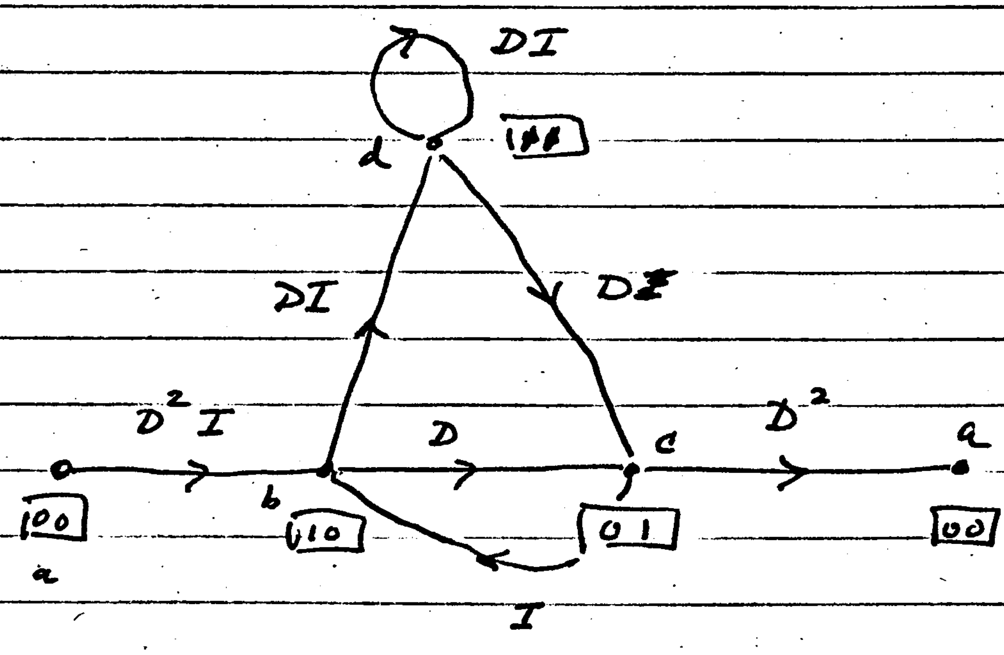
Now note how this differs from the previous code, both rate 1/2 and constraint length 2. Here

$$T(D,I) \simeq D^5 I \qquad (Z<<1)$$




nonsystematic encoder!

(a)
Encoder




(b)
State Diagram

Fig 39 : Encoder-2 and Performance

with

$$P_b \simeq \frac{1}{b} Z^5$$

This code is more powerful than the previous code.

Example: Let us choose a one bit quantizer, called a hard decision quantizer with a 2∅-PSK modulator scheme. Then we can readily obtain two Z values, a Z=0 if the output of the matched filter is <0 and Z=1 if the output is >0. Thus if the input is 1 and this corresponds to $+\sqrt{E}$ and the output is $-\sqrt{E}$ we have

$$p(0/0) = 1 - Q\left(\sqrt{\frac{2E}{N_o}}\right) = 1-p$$

$$p(0/1) = Q\left(\sqrt{\frac{2E}{N_o}}\right) = p$$

$$p(1/0) = p$$

$$p(1/1) = 1-p$$

This then yields

$$Z = 2\sqrt{p(1-p)}$$

Clearly as $E/N_o$ increases p goes to zero and so does Z. Comparing the two codes already discussed we see that for a binary symmetric channel of this type we have

$$P_b \simeq 2^4 (p(1-p))^2 \quad \text{(Code 1)}$$

$$P_b \simeq 2^5 (p(1-p))^{5/2} \quad \text{(Code 2)}$$

Actual performance results are typically plotted as $P_b$ versus $E_b/N_o$. For a coding system as we have described, the symbol energy is E or $E_s$. The rate is R in bits per symbol so that the $E_b/N_o$ is given by

$$\frac{E_b}{N_o} = \frac{1}{R} \frac{E_s}{N_o}$$

Table 1 presents a summary of coding performance for various codes and modulation formats. The coding gain is defined as the difference between the $E_b/N_o$ required with the code and that of an uncoded scheme to achieve the give bit error probability.

Table ~~2-4~~ 1 CODING GAIN FOR SOFT-DECISION VITERBI DECODING AND HARD-DECISION SEQUENTIAL DECODING

You haven't defined sequential decoding!!!

| | | Soft Decision Viterbi Decoding | | | | | | | | | Hard Decision Sequential Decoding | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $E_b/N_o$ | R | 1/1 | | 1/2 | | | 2/1 | | 1/4 | | 1/2 | |
| Uncoded (db) | $P_b$ \ K | 7 | 8 | 5 | 6 | 7 | 6 | 8 | 6 | 9 | 41 (1) | 47 (2) |
| 6.8 | $10^{-3}$ | 4.2 | 4.4 | 3.3 | 3.5 | 3.8 | 2.9 | 3.1 | 2.6 | 2.6 | 1.5 | 3.0 |
| 9.6 | $10^{-5}$ | 5.7 | 5.9 | 4.3 | 4.6 | 5.1 | 4.2 | 4.6 | 3.6 | 4.2 | 3.8 | 4.2 |
| 11.3 | $10^{-7}$ | 6.2 | 6.5 | 4.9 | 5.3 | 5.8 | 4.7 | 5.2 | 3.9 | 4.8 | 4.8 | 6.5 |
| ----- | Upper Bound | 7.0 | 7.3 | 5.4 | 6.0 | 7.0 | 5.2 | 6.7 | 4.8 | 5.7 | 7.4 | 7.4 |

This is typically given in dB. Analytically, for example, the BSC, hard decision, constraint length two, rate 1/2 code (Code 2) discussed before with 2Ø-PSK has a coding gain that can be defined as follows. Let $P_b$ be given. Then let $E_b/N_o$ uncoded be the solution of

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o}}\right)$$

what about decoder memory to make a bit decision - typically five constraint lengths see figure 40!

Then $E_b/N_o$ coded is the solution of

$$P_b = 2^5 \; (p(1-p))^{5/2}$$

where

$$p = Q\left(\sqrt{2 \quad \frac{1}{R} \frac{E_b}{N_o}}\right)$$

The coding gain then is

$$G = 10 \log_{10}\left(\frac{E_b}{N_o}\right)_{\text{uncoded}} - 10 \text{ Log}_{10}\left(\frac{E_b}{N_o}\right)_{\text{coded}}$$

Figs. 40 to 42 present $P_b$ versus $E_b/N_o$ for a wide variety of codes. The use of soft decision typically gives a 3 dB performance improvement. The loss of performance with 3 soft decisions as compared to no quantization is less than 0.25 dB with only 3 bits. Thus most systems use 3 bit soft decision.

One final point is worth noting. This is that the phase errors in the demodulation loops tend to limit ever decreasing $P_b$ curves as $E_b/N_o$ increases. This can be seen by noting that the unquantized matched filter output is

$$r = \pm \sqrt{\frac{2\ E_s}{N_o}} \cos Ø + n$$

define $\alpha = \frac{P_c}{N_o B_L}$

where Ø is the phase error. This phase error was discussed in Section 4.3. If we define α as the PLL signal to noise ratio then $P_b$ vs $E_b/N_o$ is shown in Fig. 43. Note the large losses that occur due to this increased error. Therefore in operational systems we must take care in loop design and maintain loop performance.

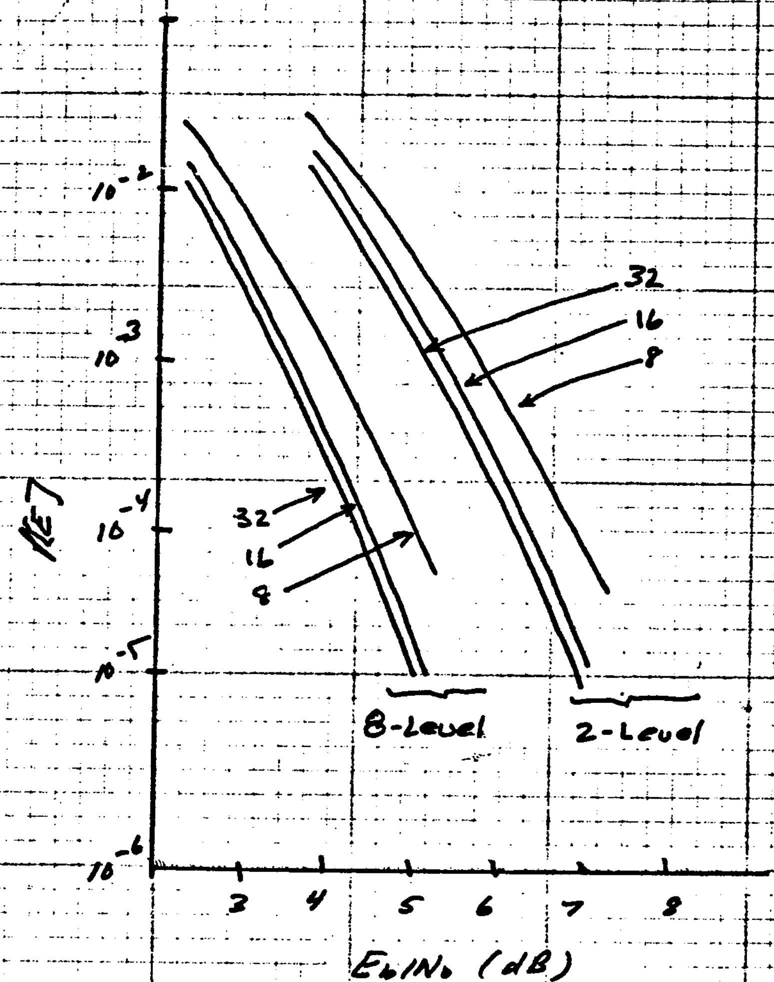



Fig 40: Performance of Decoder for K=5, r=1/2 Encoder with 8, 16, 32 bit path lengths and 2 and 8 level quantization

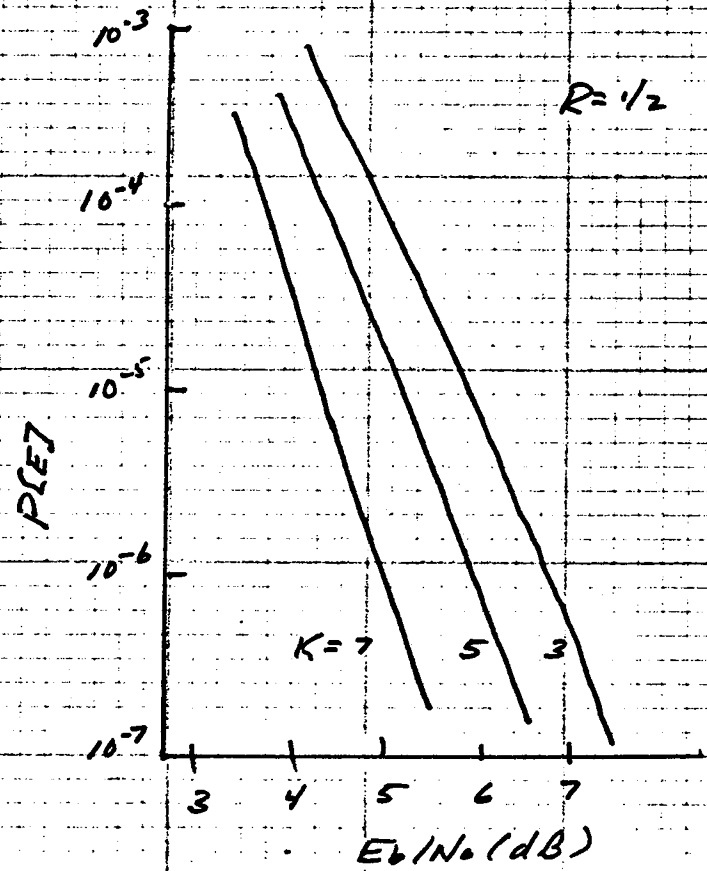



FIG 41: Performance as a function of Constraint Length

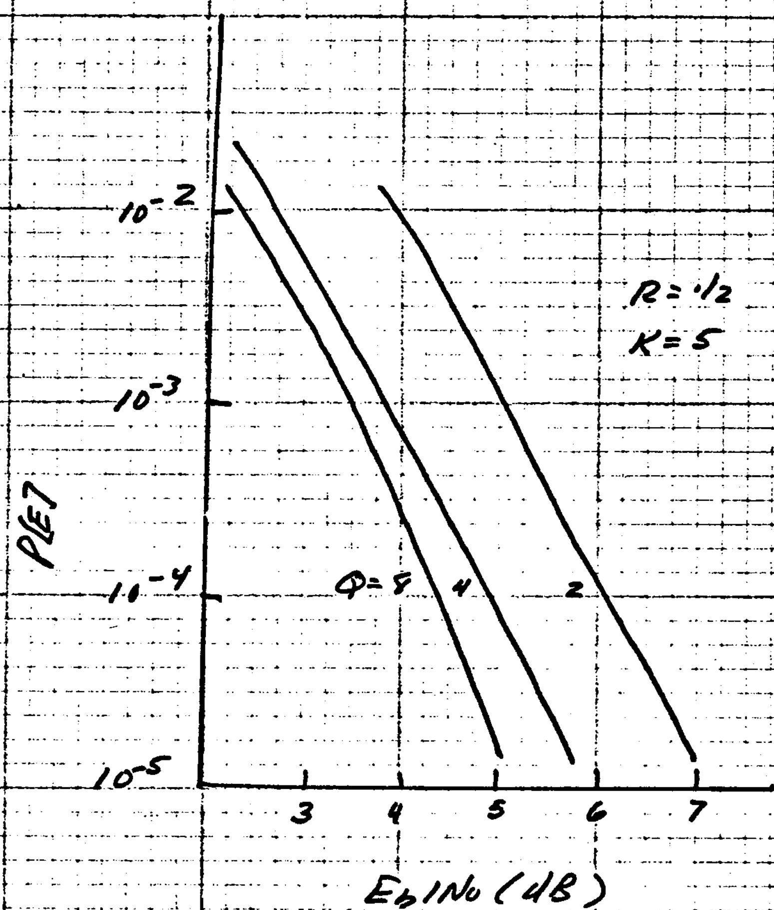



Fig 42: Performance as a function of ~~Constraint length~~ Quantization Level Q.

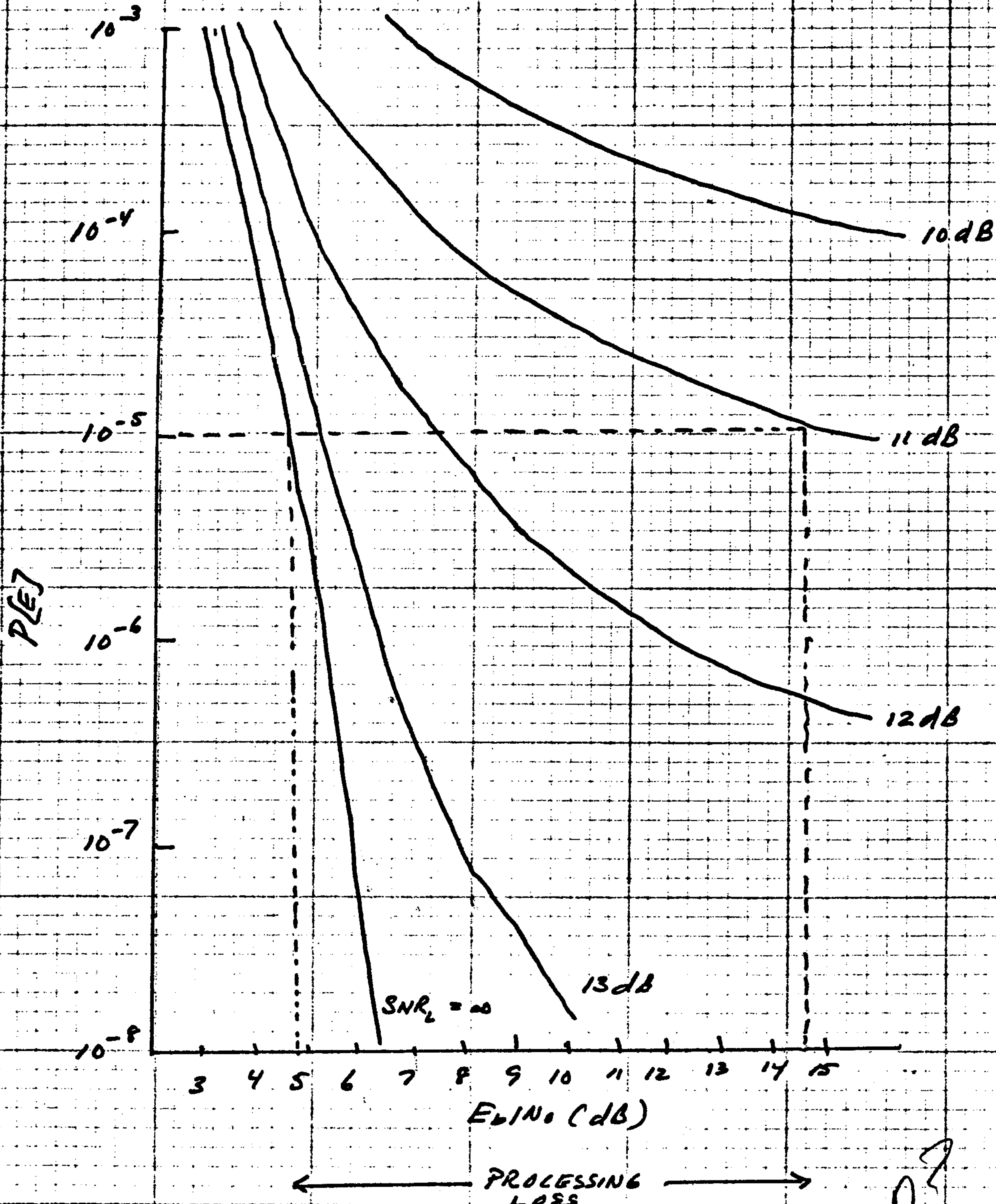



FIG 43: Performance for R=1/2, K=7, with loop degradation.

### 3.3 Data Encryption

The data that comes from the source frequently contains sensitive information that the user may wish to keep. Thus the user would require a secure means of transmitting this data. This is performed by an encryption scheme where the data from the source is altered before it is even encoded. Then at the receiving end, after decoding the data is decrypted.

A second area of importance is that of authentication. Here one is not so much concerned about transmitting data securely but rather to ensure that it came from the correct source. In this case the data is authenticated.

We can compare these two concepts as follows. Let s be a block of data. For encryption purposes we generate a block $s^1$ such that

$$s^1 = f_k(s)$$

At the receive end we have $f_k^{-1}(\ )$ so that the receiver performs the operation

$$f_k^{-1}(f_k(s)) = s$$

The authentication problem allows transmission of s in the clear but requires that an authentication code be placed to the message. Thus we generate an authentication code from s, say m as

$$m = g_k(s)$$

which is smaller in bits than s. We then send

$$(\underline{s},\underline{m})$$

At the receiver we do not decrypt but just encrypt s again to generate an $m^1$ and compare it to m.

Encryption requires both the forward and inverse operations; authentication just the forward.

In this section we will be concerned with several issues. First we discuss the two key commercial encryption schemes. We then recognize that if errors are present on the channel then if we send the encrypted message $s^1$, what is received is s", which may have bit errors in it. This can be represented as

"2
S ↓

$s' = s' + n$

where n is a binary noise sequence. Then at decryption we have

$$s''' = f_k^{-1}(f_k(s)+n)$$

which also contains errors. Note that decryption does occur but the errors may now form patterns. ?

A similar problem may occur with the authentication scheme. In this case we may be interested in the miss rate, that is the probability that data is not authenticated due to channel errors. This miss rate is P(N) the miss probability and we investigate for a crude encryption technique its performance.

### 3.2.1 Commercial Schemes

3.2.1

For the majority of encryption/authentication schemes, the strength of the scheme is dependent upon the secrecy of both the algorithm and the key. Most are special purpose designs with minimal public documentation and validation. Major drawbacks include undetermined reliability and manual key distribution through a secure courier. Such schemes will not be examined further in this paper since the necessary information is unavailable.

The National Bureau of Standards [3] has published an encryption/authentication scheme in which the strength of the scheme is dependent only upon the secrecy of the encryption/decryption key.

A further scheme has been proposed in which both the algorithm and the encryption key are public and only the decryption key remains secret. This system, known as a Public-Key Cryptosystem, [4] has only been experimentally implemented in hardware, but will be examined further here for its unique properties of key management and one-way authentication.

The Data Encryption Standard (DES) developed in conjunction with IBM, was issued in November 1976 by the National Bureau of Standards. It has been adopted as the universal encryption scheme for non-classified government data. It has also found widespread use in the area of Electronic Fund Transfer (EFT) and non-government applications.

The DES is a recirculating block product cipher of block size 64 bits, using an encryption key of length 64, 8 bits of which are used for error detection. The decryption key is identical to the encryption key. A block product cipher is one in which the basic transformations of permutation and substitution are combined and operate on a group of data bits to produce a group of cipher bits simultaneously. In this manner, each bit of cipher is a function of all input bits and all key bits.

The sequence of operations comprising the algorithm begins with an initial permutation of the 64 bit input. The input is then divided into two 32 bit data vectors, L and R. The following sequence is then performed 16 times. The right vector, R, is combined with a subset of the encryption key ($K_n$) using an enciphering function F. This value is added modulo 2 to the left vector (L), giving T.

$$T = L \oplus F(R,Kn)$$

Where $\oplus$ indicates modulo 2 addition.
Setting L=R, R=T, this sequence is repeated 16 times. The inverse of the initial permutation is then performed on the full 64 bits, giving the final 64 bit output cipher. This algorithm is illustrated in Figure 44.
44

The function F uses permutations and substitutions to change both the position and content of the input. Figure
45 45. illustrates this procedure. E is an expansion operation which duplicates selected bits of R(32 bits in length) giving a 48 bit vector. KS is a Key Schedule function which iteratively selects a 48 bit subset $K_n$ of Key. These two vectors are added modulo 2, giving a 48 bit vector. Using the eight different substitution tables $S_1$, $S_2$,...$S_8$, each six bit sub-block is replaced by a four bit sub-block, reducing the total number of bits to 32. This vector is then permuted, by a permutation $P_c$, to give $F(R,K_n)$.

Decryption is performed using the same algorithm, but reversing the order of $K_n$. The identical key must be used initially. In this manner, data that is encrypted twice will not yield the clear message, and the message may be deciphered first and then enciphered to give plain text.

High reliability is based on the fact that there are $2^{56}$ possible keys, and no method currently exists for finding the key from given plaintext and ciphertext except trying all possible keys. A key management system must be used to ensure that the sender and receiver have a common key known only to

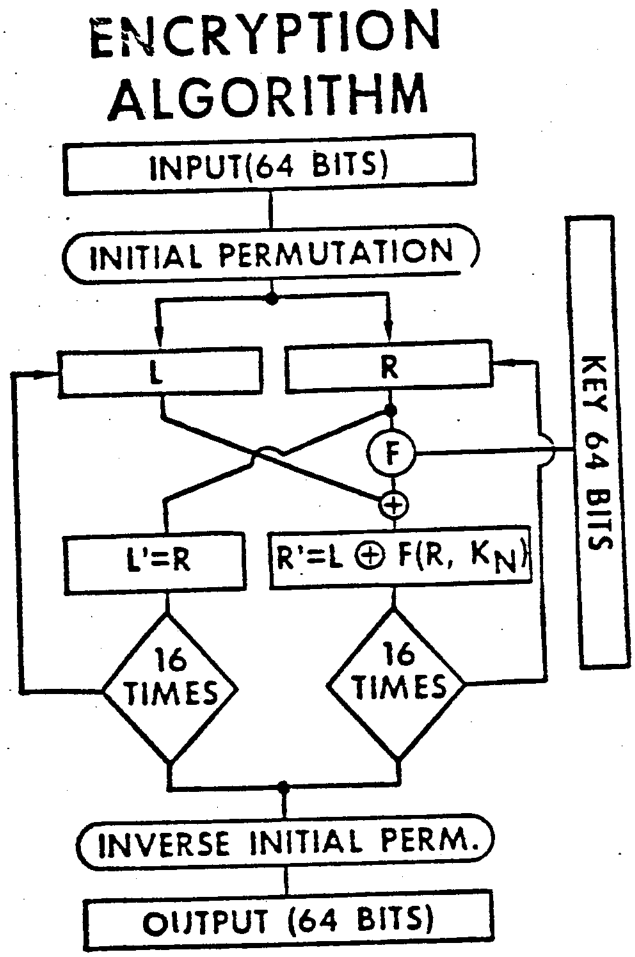



Figure 44 DES Encryption Algorithm

44

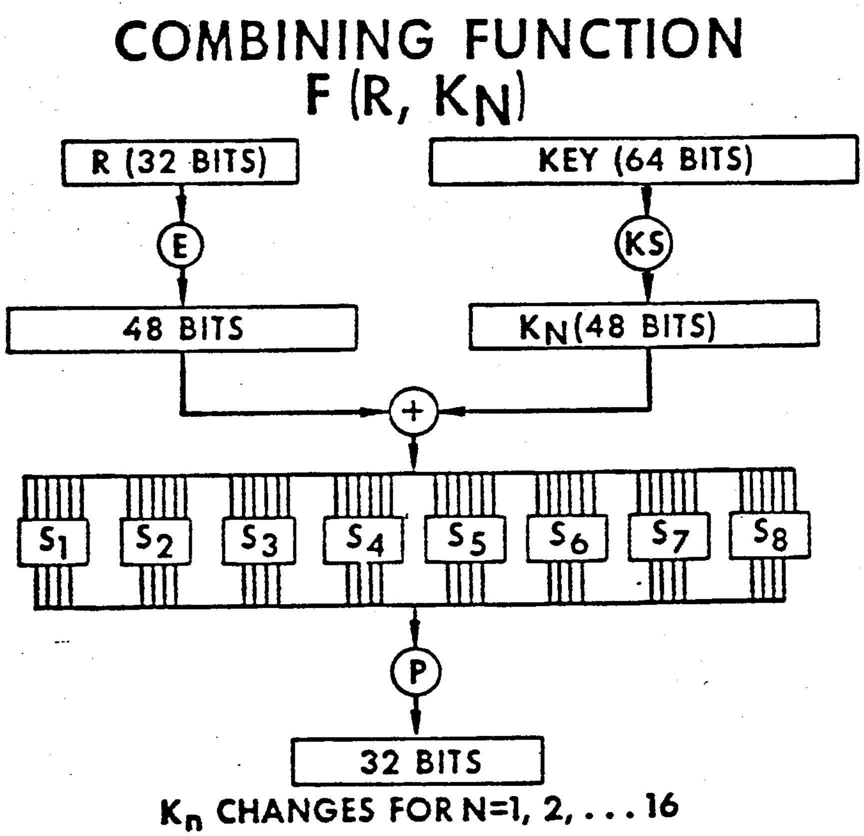



Figure 45 DES Combining Function

45

them. Key distribution must be handled manually, and may create problems in some applications.

The DES has three basic modes of operation: the electronic codebook mode, the cipher feedback mode, and the cipher block chaining mode.

The electronic codebook (ECB) mode produces 1 block of code for exactly 1 block of plaintext. Each block of code is independent of all others. This method becomes less efficient when the message is not exactly 64 bits, and padding characters must be used. If the data bits are generated in a stream, the first bit cannot be encrypted until the 64th bit is generated, imposing a delay in communications.

The cipher feedback (CFB) mode uses the DES as a binary stream pseudo-random number generator. The input to the algorithm is the last 64 bits of ciphertext generated. This is encrypted, and each bit is added modulo 2 to the corresponding bit of cleartext to produce a bit of ciphertext. The message can be of any length from 1 to 64 bits, and the corresponding number of ciphertext bits are produced. Thus each byte of cipher depends on the previous 64 cipher bits. The algorithm must be initialized, and both the encrypting device and decrypting device must use the same initial values. This is referred to as crypto synchronization.

The cipher block chaining (CBC) mode encrypts a 64-bit block of plaintext, after it has been added modulo 2 to the previous ciphertext block. The same initializing vector must be used for encrypting and decrypting the first plaintext block. This mode has similar disadvantages to the electronic codebook mode.

The DES has been implemented as a self-contained hardware device, available commercially from several manufacturers. All such devices have been certified and validated by the National Bureau of Standards, assuring a uniform level of performance. Table 2 gives presently available information on such devices implementing DES, as well as devices implementing several other encryption/authentication algorithms. Only those models which are capable of operating at speeds in the range 2400 bps to 64K bps are listed.

Implementations which comply with the standard include Large Scale Integration (LSI) chips in individual electronic packages, Medium Scale Integration (MSI) electronic devices, and micro-processors using Read Only Memory (ROM).

Table 2 Commercially Available Security Devices

| Manufacturer | Model/Series | Type | | | Security Level | Encryption Scheme | Code Combination | Communications Channel Requirements |
|---|---|---|---|---|---|---|---|---|
| | | Voice | Data | Fax | | | | |
| AB Trans-Vertex | TD-200 | | x | | 6 | Bit stream ciphering. Programmed by means of a standard 80 column punched card. | $10^{140}$ | Data Rates: 45.45 to 9600 bps. Alphabets: CCITT No. 2 & No. 5. Transmission modes: Asynchronous, half duplex Synchronous, half duplex |
| Crypto AG | MCC 314 | x | x | x | 6 | Digital information is processed in a ciphering computer along with a key chain. Basic key is stored in a magnetic core memory | $10^{68}$ | Full duplex,on-line. Data rate from 10K to 2M bps. Matches with PCM or DM systems. |
| COM TECH Systems, Inc. | SECRE/DATA 102 Series | | x | | 5 | Auto-key cipher with cryptogram as part of the key | $2 \times 10^8$ in a system out of 7 billion available codes | Data rates: synchronous up to 1 Mbps; asynchronous up to 19,200 bps |
| | SECRE/DATA 1102 Series | | x | | 6 | Same as 102 series | $4 \times 10^{16}$ | Same as 102 series |
| Computation Planning, Inc. | Cryptopak[tm] | | x | | 6 | DES-software implementation and other simulation programs | | Data rate: 6400 bps using 370/155. |

* Security Level:

2 basic privacy level

3 general or industrial privacy level

5 tactical privacy level

Table 2 Commercially Ava...able Security Devices (cont.)

| Manufacturer | Model/Series | Type: Voice | Type: Data | Type: Fax | Security Level | Encryption Scheme | Code Combinations | Communications Channel Requirements |
|---|---|---|---|---|---|---|---|---|
| Motorola, Inc. | DES 3100 NSM Network Security Module | | x | | 6 | DES | $2^{56}$ | Data rate 110 to 9600 baud special design, up to 200 Kbps. |
| | MGD 3100 NSM | | x | | 6 | DES | $2^{56}$ | Data rate: 110 to 9600 baud. Bi-sync network. Interface: EIA RS-232-C |
| | MGD 1300 HSM Hybrid Security Module | | x | | 6 | DES | $2^{56}$ | |
| | MGD 6800 DSM Data Security Module | | x | | 6 | DES | $2^{56}$ | |
| | MGD 8080 DSM | | x | | 6 | DES | $2^{56}$ | |
| | DES 1100 DSM | | x | | 6 | DES | $2^{56}$ | |
| IBM | 3845/3846 | | | | 6 | DES | $2^{56}$ | Data rates: asynchronous 110-9600 bps. Synchronous up to 19,200 bps. |
| Datotek, Inc. | DS-138 | | x | | 6 | Modulo 2 addition of data and key digits. Key digits are generated by a complex multi-register device using multistep dynamically programmed circuits providing non-linear outputs. Mark II key | $10^{52}$ | Data rate: up to 9600 bps Interface: EIA RS232C |

There are three approaches to implementing the DES in a communications system: link-by-link encipherment, node-by-node encipherment, and end-to-end encipherment.

Link-by-link encryption is the most common type of implementation. It is carried out by placing the encryption devices in series with the circuit between data terminal equipment and data communications equipment. The major disadvantage of this method lies in the fact that the message passes through the CPU of any node in the clear.

Node-by-node encryption is designed to compensate for this problem. Each link uses a unique key, and a security module carries out the task of switching from one key to the next. Thus, the plaintext is only found in the security module, and not in the CPU.

End-to-end encryption becomes more complicated, requiring a key control center to generate a tempory session key. This session key is then encrypted with each end-user's key (held in security at the key control center) and transmitted to the respective end users. Upon decryption, they both hold the session key, and can communicate using it.

---

There are two basic ways the DES can provide authentication of data and source of data. The first method provides encryption and authentication simultaneously, while the second provides authentication only.

For a given block of 64 bits, define subfields within the block for authentication information, reducing the number of bits available for data. Such fields could include sequence number, a random number to ensure identical blocks resulted in different cipher blocks, error detecting information, time and date field, or a portion of the last received message. The whole block would then be encrypted in the usual way. If any party attempted to modify one of the encrypted authentication fields by even one bit, all other bits in the block would be affected; the ensuing decryption would yield totally changed authentication fields and message, allowing for detection of the modification.

This authentication technique prevents the disclosure of plaintext, and verifies for the receiver that the legitimate sender transmitted the message. It prevents the retransmitting of an older recorded message by an unauthorized party.

The second method providing authentication uses the DES to generate a unique Message Authentication Code (MAC) which is appended to the message, and is dependent on the message. The plaintext message is added modulo 2 to the previous ciphertext message, and then encrypted. The first two bytes of this ciphertext become the MAC, and are appended to the message. This block is then transmitted in the clear. If any party modified data in the message, it would be detected upon decrypting the MAC.

Neither authentication method can easily provide "signed messages". The advantage of a signed message is that the recipient can prove not only to himself, but also to any other third party that the message came from the authorized sender. This property is essential for the future of the EFT and electronic mail concepts.

The performance of the DES has been analyzed extensively by the National Bureau of Standards. To date, no method exists for breaking the algorithm except trying all of the $2^{56}$ possible keys. Even with the fastest computers available today, this is totally infeasible. For added protection, the key should be changed at regular intervals.

Table 3 gives performance characteristics and costs, when available, for commercially available devices implementing either DES, or a similar algorithm. In most cases, cost is dependent on the communication channel speed.

### 3.3.2 Public Key Cryptosystem

In general, a public key cryptosystem, as defined by Diffie and Hellman in [4], is one in which two different keys are used for encryption and decryption, the algorithm is known, the encryption key is known and only the decryption key is secret. In addition, the encryption and decryption operations should commute. This property is required to implement digital signatures, i.e. authentication by signed messages.

The algorithm described here is due to Rivest, Shamir and Adleman [6]. It is probably the best known public key algorithm to date.

Let n be the block size and M be the message, expressed numerically. Encryption is performed by computing

$$C = M^e \pmod{n}$$

where C is the resulting ciphertext, and (e,n) is the enciphering key.

Deciphering is performed by calculating

$$M = C^d \pmod{n}$$

where (d,n) is the deciphering key, of which d is secret.

Table 3
Performance Characteristics of
Commercially Available Security Devices

| Manufacturer | Model/Series | Reliability | Physical Size | | Power Supply | Operating Temperature | Humidity | Special Feature | Price ($) |
|---|---|---|---|---|---|---|---|---|---|
| | | MTBF | WxHxL $cm^3$ | Weight Kg | | | | | |
| Crypto AG | MCC 314 | | 16.5x 64.5 x 30 | 17.5 | 110/220 VAC $\pm 20\%$ 45-65 Hz or 18-30 VDC | $-25^oC$ to $60^oC$ | | One-line | |
| AB Transvertex | TD-200 | | 22x24x43 | 17 | 110 or 220 V $\pm$ 10% 48-65 Hz | $0^oC$ to $50^oC$ | 98% at $50^oC$ | On-line, off-line | 12,000 to 15,000 |
| Datotek, Inc. | DS-138 | 10,000 hours | 41.2x 14x41.2 | 16 | 115/230VAC, 50/60 Hz, 20 watts (60 watts while loading key variable.) | $0^oC$ to $50^oC$ | 95% | On-line synchronous | |
| IBM | 3846 | | 19" standard rack | | Battery power backup for reliability | | | One or 2 lines, synchronous, asynchronous | 2,200 |

Performance Characteristics of
Table 3 . Commercially Available Security Devices (cont.)

| Manufacturer | Model/Series | Reliability | Physical Size | | Power Supply | Operating Temperature | Humidity | Special Feature | Price ($) |
|---|---|---|---|---|---|---|---|---|---|
| | | MTBF (hr) | WxHxL $cm^3$ | Weight Kg | | | | | |
| Motorola, Inc. | DES 3100 NSM Network Security Module | | (19" standard rack) | | 110 VAC, 60 Hz, 120 watts (220 VAC available) | $0^oC$ to $50^oC$ | 95% | Data encryption time 160 ps at 2 MHz clock | 3,000 |
| | MSD 3100 NSM | | | | | $0^oC$ to $70^oC$ | 85% | | |
| | MCD 1300 HSM Hybrid Security Module | | | | | $0^oC$ to $70^oC$ | | | |
| | MSD 6800 DSM Data Security Module | | 25x14.6 x 0.16 | | | $0^oC$ to $70^oC$ | | For M6800 Exorciser and micromodule system | |
| | MGD 8080 DSM | | | | | $0^oC$ to $70^oC$ | | For Intel MDS developed system and SBC module | |
| | DES 1100 DSM | | | | | | | Handle data transfers using PDP-11 NPR control | |

He then encrypts S with Y's encryption key, $E_y$, (available in a public file) to get ciphertext, C, ensuring privacy.

$$C = E_y(S) \tag{2.5}$$

Upon receiving the ciphertext, Y decrypts it using his secret deciphering key, $D_y$. Presumably, no one else can decipher it, including X. This yields the signature S.

$$S = D_y(C) \tag{2.6}$$

Y then uses X's encryption key, also available in a public file, to extract the message

$$M = E_x(S) \tag{2.7}$$

Y holds a unique pair, M and S, which must have come from X, since no one, including Y, could have generated S.

Cryptosystems having the above property are called one-way authentication schemes. They are designed to protect against disputes between sender and receiver.

### 3.3.3 Comparison of DES and PKC

3.3.8

The Data Encryption Standard has the advantage of being established, tested and implemented previously. Devices exist and have been validated by the NBS. The public key cryptosystem, as developed by Rivest, et al. is currently undergoing testing for cryptographic strength, and is not yet fully established and accepted as a valid cryptosystem. Nonetheless, PKC has several very desirable properties for our application. Key distribution and management is simplified, eliminating the need for a private courier to distribute keys whenever a new key is generated. Presumedly, new keys will be issued periodically even if there is no suspicion of compromise. Considering the global distances involved, and the relative isolation of array sites, this property could be very important.

One-way authentication is an additional property of PKC not readily available with DES. Due to the political sensitivity of the issues involved, this property would be desirable for a system of checks and balances on both countries involved. Any third party or overseeing agency (for example, the United Nations) would be able to make a clear decision on message authentication with such a system.

Table 4 4 gives a computational comparison of the two types of encryption/authentication schemes. As projected, the PKC is much slower than DES. This speed may improve as special purpose hardware devices are developed.

## TABLE 4: COMPUTATIONAL COMPLEXITY

| MEASURE | CC (DES) | PKC (RIVEST) |
|---|---|---|
| SOFTWARE THROUGHPUT | 230 KB/S (MOTOROLA) ----- 1.6 MB/S (COLLINS) | 6 KB/S (PROTOTYPE PROJECTION) |
| HARDWARE THROUGHPUT | 200 KB/S (KL - 10) ----- 900 KB/S (370/168) | 1.8 KB/S (EST) (370/168) |
| EQUIVALENT KEY-SIZE | 56 BITS | ≈ 320 BITS |

3.3-1

### 3.3.4 Direct Schemes

The results presented in this section are based on the following authentication algorithm. data is generated and produces an L bit block of binary data, $\underline{s}$. A function is applied to the text $\underline{s}$ to produce an authentication codeword $\underline{m}$ of length M bits. The authentication word $\underline{m}$ is a function of $\underline{s}$, $f(\underline{s})$. Then $\underline{s}$ and $\underline{m}$ are transmitted over a binary symmetric channel with crossover probability p, equal to the bit error rate. At the other end, the noisy versions of $\underline{s}$ and $\underline{m}$, denoted $\underline{s}'$ and $\underline{m}'$, are received. A function identical to the one on the transmit side is then applied to $\underline{s}'$ to produce $\hat{\underline{m}}$. A comparison is made of $\underline{m}$ and $\hat{\underline{m}}$, and the seismic data, $\underline{s}$, is authenticated if there are no differences between them. If there are any differences, $\underline{s}'$ is not authenticated. For a diagram of this procedure, see Figure 46 .

This section presents three functions, in order of increasing complexity, and examines how they affect the probability of a miss. In addition, the possible advantages of the transmission of coded bits instead of uncoded bits are examined.

The first authentication function considered is defined by $f(\underline{s}) = A\underline{s}$ where A is a MxL binary matrix, which performs a linear matrix operation on $\underline{s}$, namely, binary addition of k components of $\underline{s}$ at a time. For example, if k = 3, the matrix A that represents f would be

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & \cdot & \cdot & \cdot & & & & & & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \cdot & \cdot & \cdot & & & & & & 0 \\ \cdot & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & \\ \cdot & & & & & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Since the same function is used on the receiving side, the algorithm generating the vector $\hat{\underline{m}}$ is

$$\hat{\underline{m}} = f(\underline{s}') = A\underline{s}' = A(\underline{s} + \underline{n}_s) \qquad (2.8)$$

$$= \underline{m} + A\underline{n}_s$$

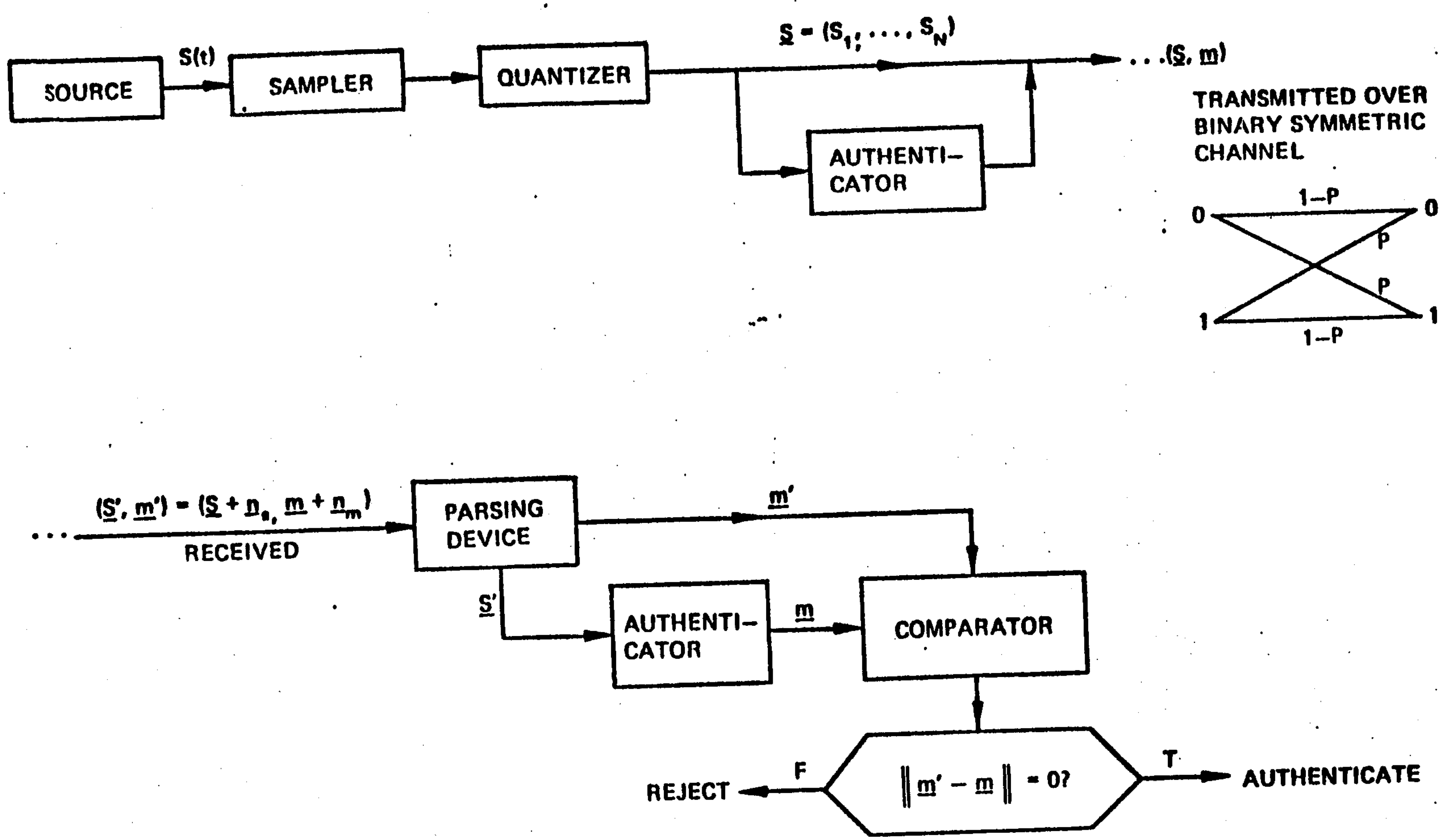



Figure 46: DIAGRAM OF TRANSMISSION AND AUTHENTICATION PROCESS

46

G–19190

Now the received noisy authentication word is $\underline{m}' = \underline{m} \oplus \underline{n}_m$, and it will be sufficient to compare $A\underline{n}_s$, and $\underline{n}_m$ to compute the probability of a miss, rather than comparing $\hat{\underline{m}}$ and $\underline{m}'$.

To ensure that each component of $\underline{m}$ is independent, we will assume that A is constructed in such a way that each component of $\underline{s}$ appears in no more than one sum. Similarly, on the receiving side, this constraint on the nature of the function, f, ensures that each component of $\hat{\underline{m}} = \underline{m} + A\underline{n}_s$ is independent, and therefore each component of $A\underline{n}_s$ is independent. This condition of independence is important to the computation of the probability of a miss.

Since $\underline{s}$ has L components, and we are adding them k at a time, $f(\underline{s}) = \underline{m}$ has M components, where

$$M = \left[\frac{L}{k}\right] = \text{the greatest integer less than or equal to } L/k$$

There, the noise vector that is added to $\underline{m}$ during transmission, $\underline{n}_m$, has M components:

$$\underline{n}_m = n_{m_1}, n_{m_2}, \ldots, n_{m_M}$$

For each component, $n_{m_i}$, of $\underline{n}_m$:

$$P(1) = p$$

$$P(0) = 1 - p$$

The noise vector added to $\underline{s}$, $\underline{n}_s$, has L components. Therefore, $A\underline{n}_s$ has M components:

$$A\underline{n}_s = \begin{bmatrix} n_{s_1} + n_{s_2} + \ldots + n_{s_k} \\ n_{s_{k+1}} + n_{s_{k+2}} + \ldots n_{s_{2k}} \\ \cdot \\ \cdot \\ \cdot \\ n_{(M-1)k+1} + \ldots + n_{Mk} \end{bmatrix}$$

For each component of $A\underline{n}_s$:

<u>k even</u>

$$P(0) = (1-p)^k + \binom{k}{2} (1-p)^{k-2} p^2 + \ldots + p^k$$

$$P(1) = \binom{k}{1} (1-p)^{k-1} p + \binom{k}{3} (1-p)^{k-3} p^3 + \ldots + \binom{k}{k-1} (1-p) p^{k-1}$$

<u>k odd</u>

$$P(0) = (1-p)^k + \binom{k}{2} (1-p)^{k-2} p^2 + \ldots + k (1-p) p^{k-1}$$

$$P(1) = (1-p)^{k-1} p + \binom{k}{3} (1-p)^{k-3} p^3 + \ldots + p^k$$

Where $\binom{k}{n}$ denotes the number of combinations of n objects Selected from a set of k objects, and $\binom{k}{n} = \frac{k!}{n!\ (k-n)!}$. Therefore,

P (exact match of all M components of $A\underline{n}_s$ and $\underline{n}_m$) =

<u>k even</u>

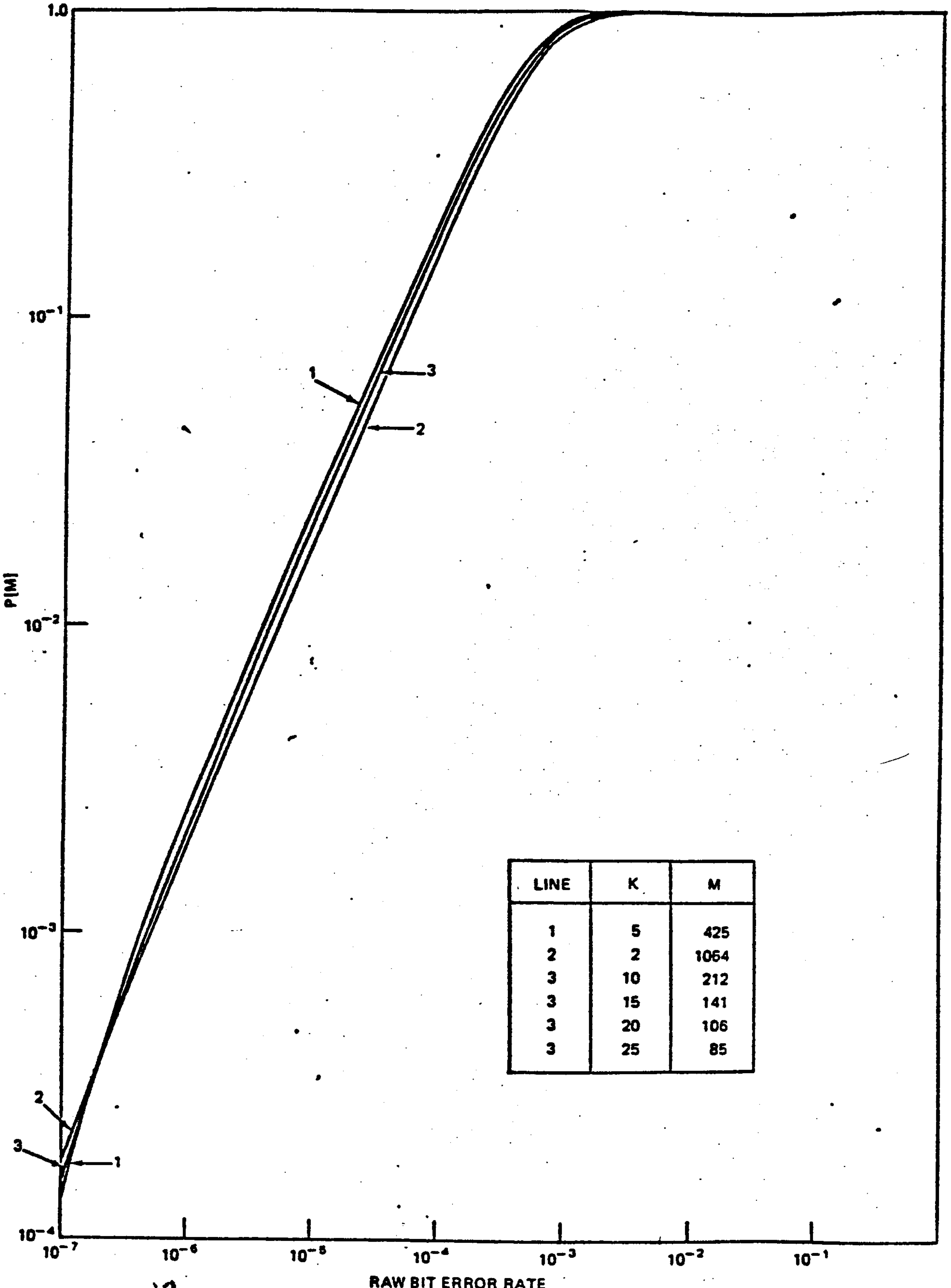
$$= \left[ (1-p) \left[ (1-p)^k + \binom{k}{2} (1-p)^{k-2} p^2 \ldots + p^k \right] + p \left[ k(1-p)^{k-1} p + \binom{k}{3} (1-p)^{k-3} p^3 + \ldots + k(1-p) p^{k-1} \right] \right]^M$$

$$= \left[ (1-p)^{k-1} + \left[ \binom{k}{1} + \binom{k}{2} \right] (1-p)^{k-1} p^2 + \left[ \binom{k}{3} + \binom{k}{4} \right] (1-p)^{k-3} p^4 + \ldots + \left[ \binom{k}{k-1} + \binom{k}{k} \right] (1-p) p^k \right]^M$$

<u>k odd</u>

$$= \left[ (1-p) \left[ (1-p)^k + \binom{k}{2} (1-p)^{k-2} p^2 + \ldots + k(1-p) p^{k-1} \right] + p \left[ k(1-p)^{k-1} p + \binom{k}{3} (1-p)^{k-3} p^3 + \ldots + p^k \right] \right]^M$$

$$= \left[ (1-p)^{k-1} + \left[ \binom{k}{1} + \binom{k}{2} \right] (1-p)^{k-1} p^2 + \left[ \binom{k}{3} + \binom{k}{4} \right] (1-p)^{k-3} p^4 + \ldots + \left[ \binom{k}{k-2} + \binom{k}{k-1} \right] (1-p)^2 p^{k-1} + p^{k+1} \right]^M$$




| LINE | K | M |
|---|---|---|
| 1 | 5 | 425 |
| 2 | 2 | 1064 |
| 3 | 10 | 212 |
| 3 | 15 | 141 |
| 3 | 20 | 106 |
| 3 | 25 | 85 |

47

Figure 47 : P[M] VS. BIT ERROR RATE FOR VARIOUS VALUES OF K

G–19230

k odd

$$P[\text{authentication}] = \left[ (1-p)^{k+1} + \left[\binom{k}{1} + \binom{k}{2}\right](1-p)^{k-1}p^2 + \left[\binom{k}{3} + \binom{k}{4}\right](1-p)^{k-3}p^4 + \ldots + \left[\binom{k}{k-2}+\binom{k}{k-1}\right](1-p)^2p^{k-1} + p^{k+1} \right]^I$$

Since there were I independent bits of $\underline{m}$, i.e. I independent sums, each containing k bits from $\underline{s}$, then the total number of bits used from $\underline{s}$ is kI. Therefore, there are L-kI bits of $\underline{s}$ left for the dependent sums. For each of these dependent components,

$$P[\text{authentication}] = P[\text{no change of any bits out of L-kI total available}] = (1-p)^{L-kI} \qquad (kI \le L)$$

Therefore, the net probability of authentication, taking into consideration both the independent and dependent parts of $\underline{s}$, is

Net P[authentication] =

k even

$$= \left[ (1-p)^{k+1} + \left[\binom{k}{1} + \binom{k}{2}\right](1-p)^{k-1}p^2 + \ldots + \left[\binom{k}{k-1} + \binom{k}{k}\right](1-p)p^k \right]^I (1-p)^{L-kI}$$

k odd

$$= \left[ (1-p)^{k+1} + \left[\binom{k}{1} + \binom{k}{2}\right](1-p)^{k-1}p^2 + \ldots + p^{k+1} \right]^I (1-p)^{L-kI}$$

And, Net P[M] = 1 - P[authentication].

The limiting situation, when I = 0, will be discussed separately in the next section.

Figure 48 represents P[M] versus bit error rate, assuming L = 2128, for all k and for all I. The explanation for the fact that these curves are identical is verified by the following reasoning: for small p, the first part of the equation for the probability of authentication is dominated by

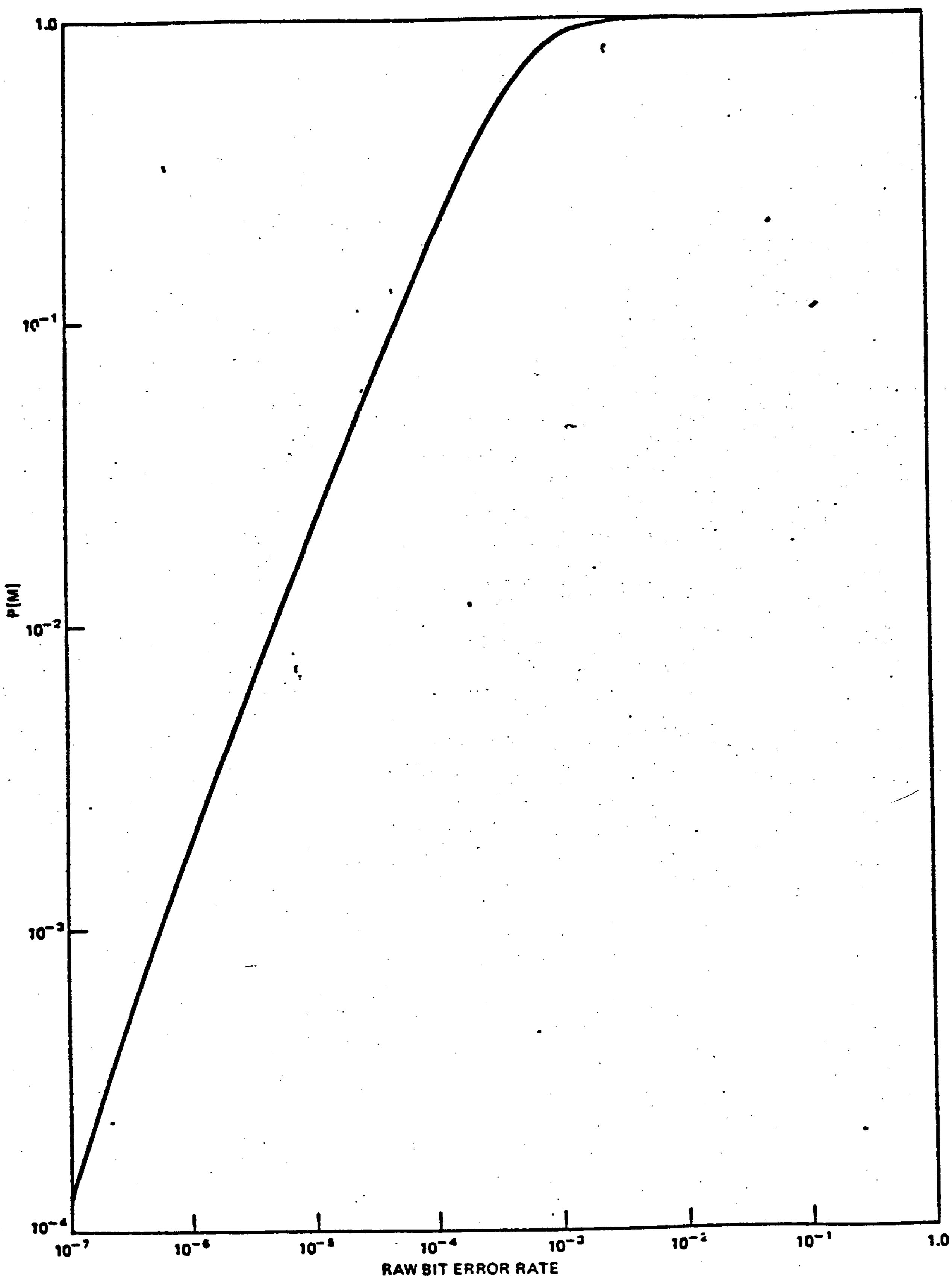



Figure 48 : P[M] VS. RAW BIT ERROR RATE

14

G-19230

maxtrix that was being used, the time involved would be prohibitive, even using a high-speed computer. Therefore, the probability that false data will get through the authentication process is negligible. This analysis applies to all three functions, since it deals only with the size of the maxtrix A.

To determine the possible advantages of transmitting coded bits instead of uncoded bits, the influence of bit coding on the bit error rate is examined.

The convolutional encoder operates in the following fashion: informaton bits are fed into a k bit shift register d bits at a time. The register has $b = k/d$ such blocks. For each block of d bits, v bits are sent to the modulator. The rate of the encoder, R, is defined as the ratio of input data bits d to transmitted channel bits v:

$$R = \frac{d}{v} \tag{2.22}$$

49

Figure 49 illustrates the bit error rate P(E), for a coded bit, versus $E_b/N_o$, the energy per bit to Noise Spectral Density Ratio, where

$$E_b/N_o = E_b/N_{o\ \text{uncoded}} - \text{Coding gain} \tag{2.23}$$

BPSK

scheme of interest is coherent phase shift keying (~~CPSK~~), in which the demodulator assumes knowledge of the phase error that has been introduced into the signal. This data is graphed in Figure 50 .

50 50

Figure 50 is then used in conjunction with Figure 49 to produce the desired conversion from the bit error rate, P(E), for a raw bit to the bit error rate for a coded bit, represented in Figure 51.

51

The probability of a miss for the three functions versus bit error rate is represented in Figure 52 . It is clear from these curves that channel encoding greatly decreases the probability of a miss, thereby providing an improvement in system performance.

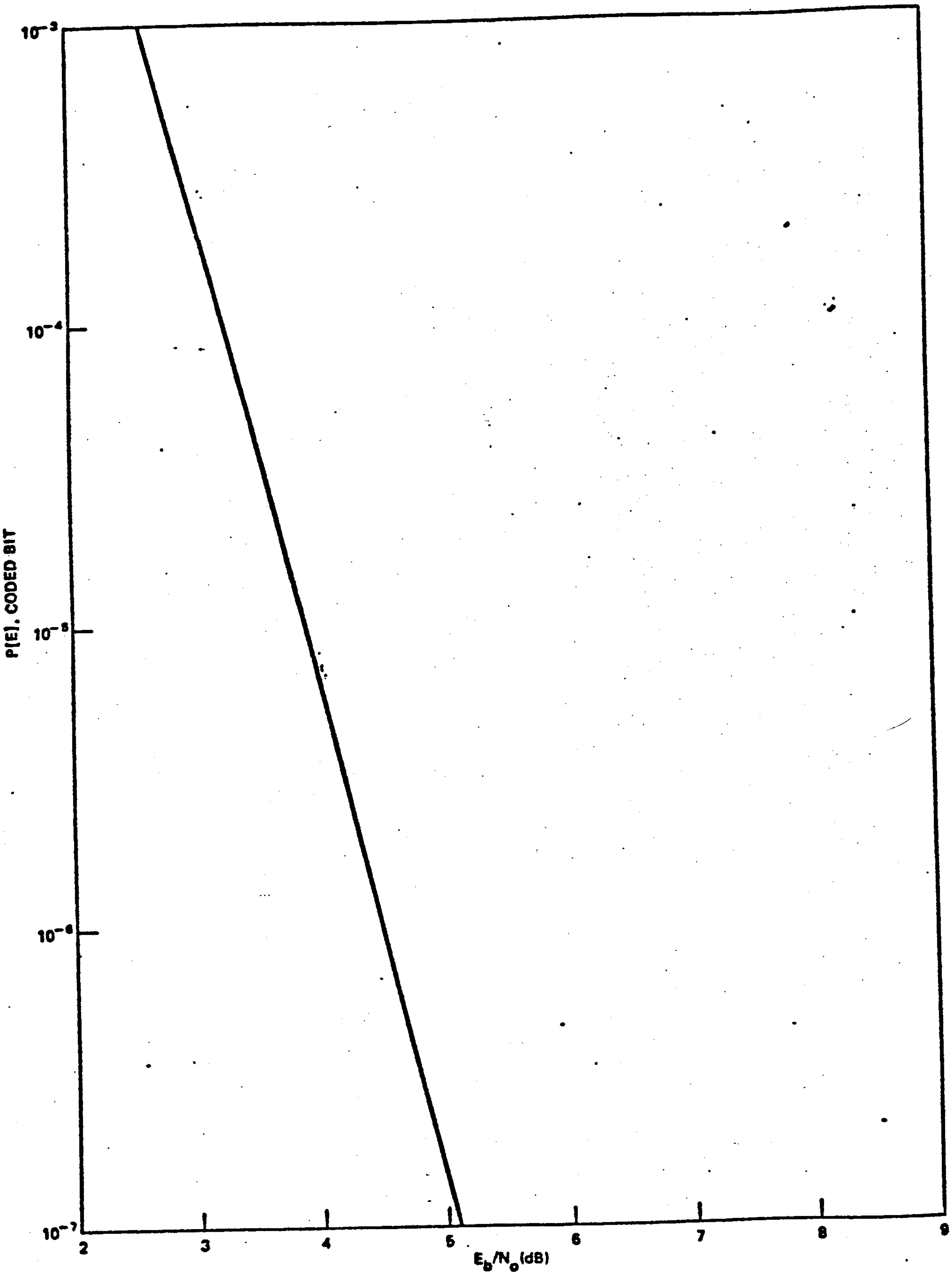



Figure 49 : P[E], CODED BIT VS. $E_b/N_o$

G-19230

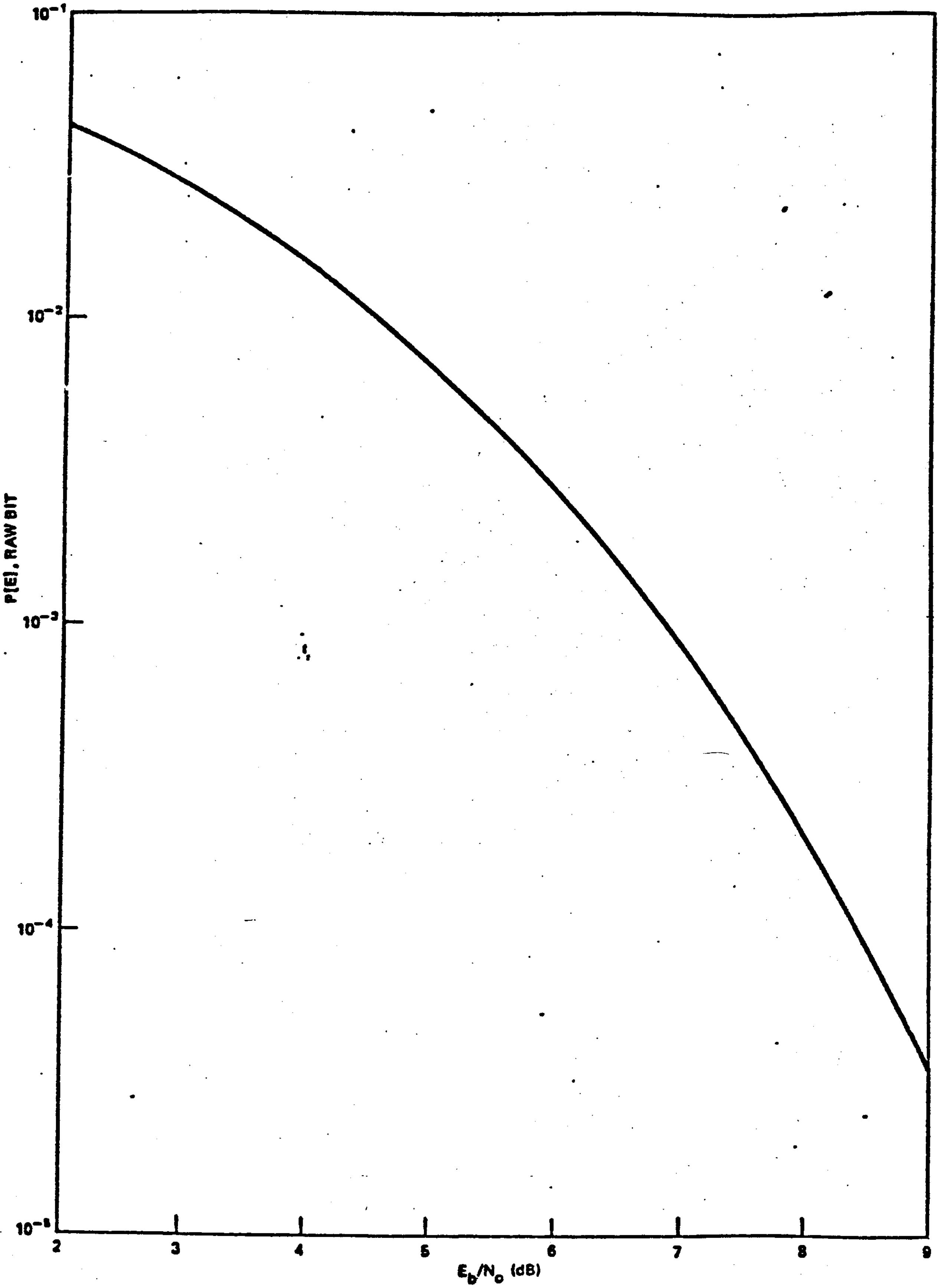
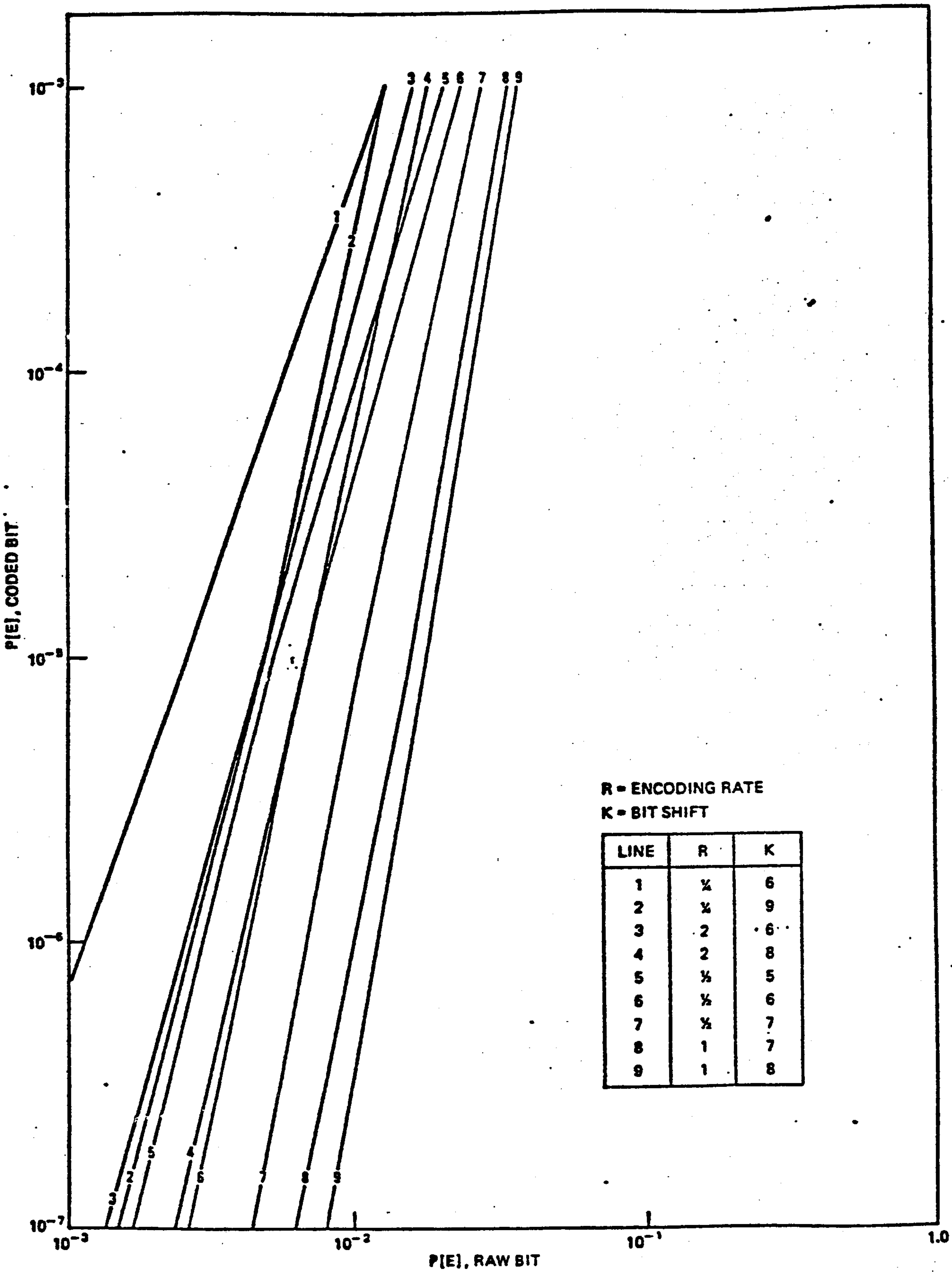



Figure 50: P[E], RAW BIT VS. $E_b/N_o$

50

G-19230




| LINE | R | K |
|---|---|---|
| 1 | 1/4 | 6 |
| 2 | 1/4 | 9 |
| 3 | 2 | 6 |
| 4 | 2 | 8 |
| 5 | 1/3 | 5 |
| 6 | 1/2 | 6 |
| 7 | 1/2 | 7 |
| 8 | 1 | 7 |
| 9 | 1 | 8 |

Figure 51 : P[E], RAW BIT VS. P[E], CODED BIT

G-19230

51

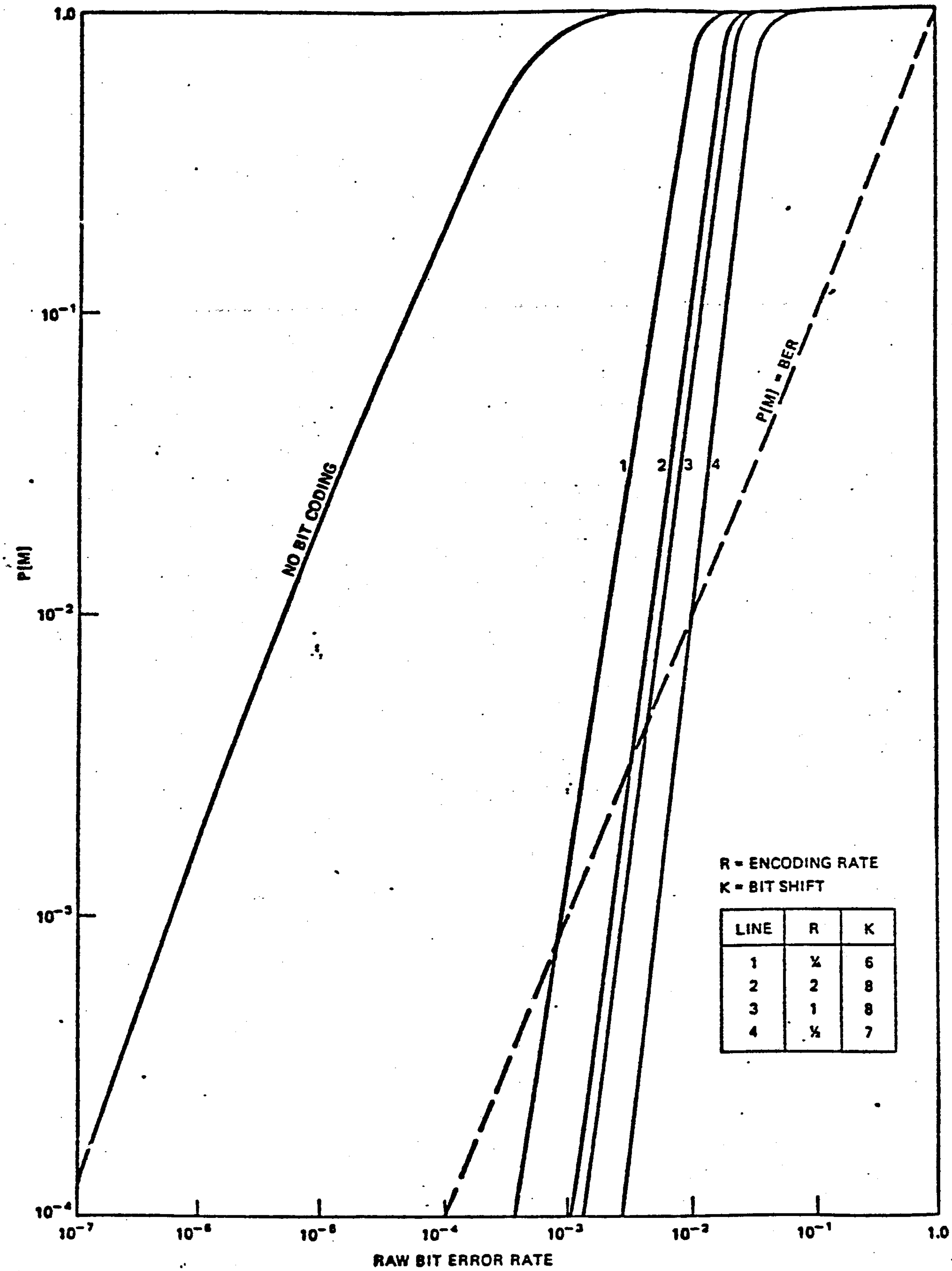



Figure 52 : P[M] VS. RAW BIT ERROR RATE

G-19230

## 3.4 Data Link Control

The final element of the data interface of a satellite link concerns the inter-relationship between the actual data source and the transmission equipment. The data source is called the data terminal equipment (DTE) which may be a computer mainframe or other data generating element. As we previously discussed, the modem functions to translate a binary baseband data stream into a waveform suitable for channel transmission. Encryption and coding may be applied prior to modulation to secure the link and improve performance. However as we shall discuss in this section, there are other performance factors that impact overall systems performance that lie at a level closer to the DTE.

In the transmission of data, the link performance has been specified in terms of error rate. Typical error performance with coding (called forward error correcting codes, (FEC)) is on the order of $10^{-6}$ to $10^{-8}$. For many computer communication needs this performance may not be adequate and $10^{-11}$ or $10^{-12}$ may be required. These rates are not achievable by FEC techniques. In addition, with certain types of data transmission, such as packet or other types of computer communications, routing information is also provided along the way. Thus the data from a DTE is packaged inside a set of layers, the layers providing control information as the message moves through the network. The modems at both ends are then made smart enough either by themselves or working in conjunction with the DTE to provide positive control on the data link. This is called the data link control (DLC) layer of computer communications.

In this section we will describe both the logical and physical interface of the link to the DTE, specify and analyze the performance factors and provide the reader with both design and analysis tools. It should be noted that this area is structured within a set of standards, some conflicting and some actually incompatible. It is not the purpose here to describe those standards but to provide a framework in which to evaluate them for future use.

describe

DLC functions by having the modems talking to one another, transforming control information and taking actions on that data. This dialogue can be accomplished on either a half-duplex or full-duplex mode. For half-duplex one terminal talks at a time. In full duplex both terminals talk simultaneously. As we have noted, a satellite link can support both modes.

A key driving factor in performance of the system is the delays that exist in transmission. There are several sources of delay which we identify below;

1. Modem Transmit Delay ($D_m$) which results from the amount of time it lacks for the modem to respond.

2. Propagation Delay ($D_p$) which results from both terrestrial and satellite links. Recall that synchronous satellites are about 23,000 mi in orbit and at the total round trip delay (about 50,000 mi) is 0.25 sec. Thus to send a message and receive a response could easily take 0.7 sec! This delay will tend to dominate performance.

3. Transmission delay ($D_T$) results from the amount of time it takes to transmit the data block. For example if a block is 2.4Kb and the channel rate is 2.4Kbps then this delay is 1 sec. At higher data rates the delay is less, at lower, longer. Unlike the $D_p$ delay, $D_T$ is somewhat controllable in system design.

Thus for a single transmission from one modem to another and return, the total delay is

$$D = 4D_M + 4D_p + 2D_T + D_o$$

where $D_o$ is all other delays which could account for overhead, protocols, etc.

The modem to DTE interface consists of both a physical and logical level. At the physical level there are standards that have been developed by EIA and CCITT. In the U.S., the common standard is the RS-232C interface which is a 25 pin connector that provides both electrical and data link information. Fig. 53 depicts a typical RS232C interface for an IBM modem to DTE. In this example pins are shown for separate functions. Pin 1 is the ground pin and pins 2 and 3 are data (transmit and receive). The other pins are for such functions as synchronization, timing, testing, acknowledgements, etc. The DTE requires the modem to partition this information in this parallel fashion. Thus it can be seen that the actual modem is more complex than what we discussed in section 3.1. In addition note that the encoding and encryption must follow this physical interface and precede the actual modulation. Thus the full modem configuration is fairly complex. A more recent DTE to DCE (data communications termination equipment, e.g., modem) standard called RS-449 has been issued that is a 37 pair interface with increased complexity. Its use in systems is still in its early stages and will most probably supplant RS232C.

The logical interface is what is called the data link control layer (DLC). It functions by packetizing the data and providing overhead control bits to ensure the integrity of transmission. There are several DLC formats; the most common being HDLC (High Level DLC of ISO), ADCCP (Advanced Data Communication Control Procedure of ANSI), BDLC (Burroughs DLC), DDCMP (Digital Data Communications Message Protocol), UDLC (Univa or DLC of Sperry Univac) and SDLC (Synchronous DLC of IBM). In this chapter we concentrate on SDLC. All of these are bit oriented, code independent DLC protocols that are widely used. They are rapidly replacing the older character oriented protocols that evolved from manual telex systems.

Fig. 54 depicts a typical SDLC format. It is comprised of three major elements.

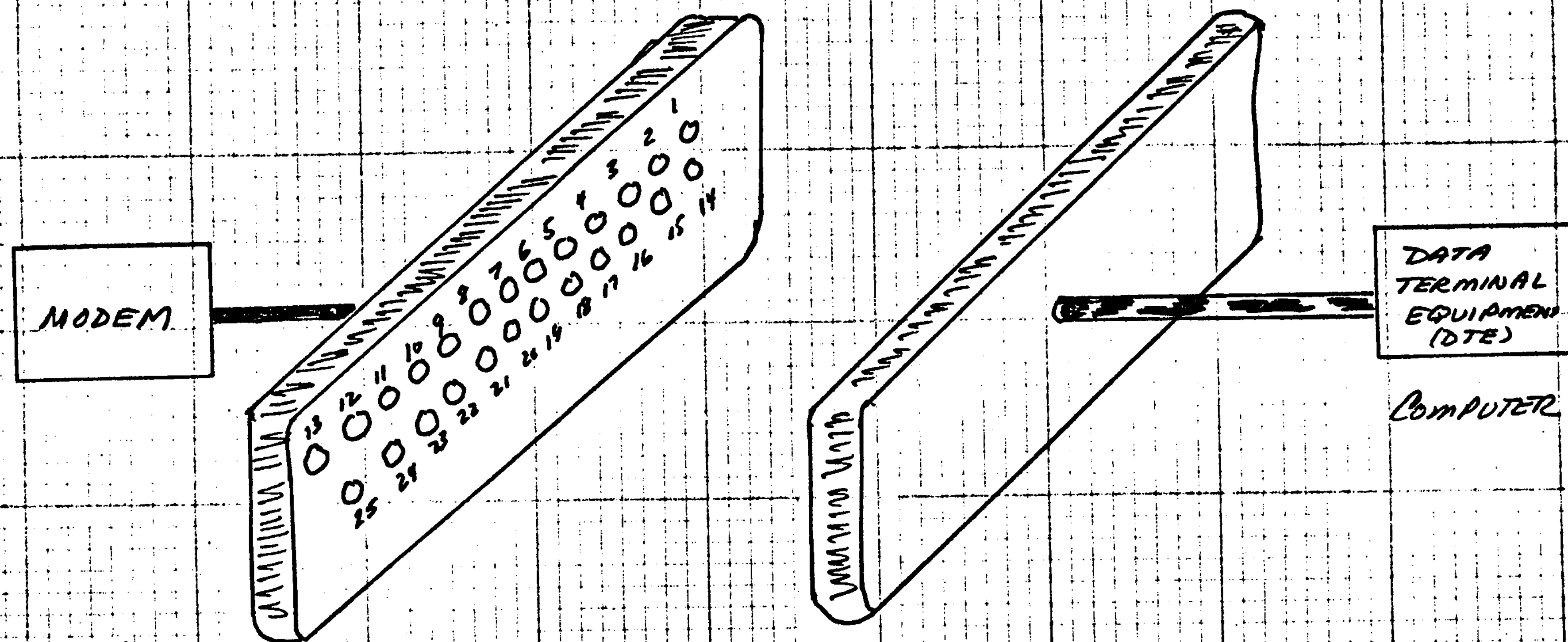



FIG 53: RS 232 C physical Interface Level

**Nonswitched Line**

| Pin | Signal Name |
|---|---|
| 1 | Ground |
| 2 | Transmitted Data |
| 3 | Received Data |
| 4 | Request to Send |
| 5 | Ready for Sending |
| 6 | Data Set Ready |
| 7 | Signal Ground |
| 8 | Receive Line Signal Detector |
| 14 | New Sync |
| 15 | Transmit Signal Element Timing |
| 17 | Receive Signal Element Timing |
| 18 | Test Control |
| 23 | Data Signaling Rate Selector |
| 24 | Transmit Signal Element Timing (DTE clock) |
| 25 | Test Indicator |

**Switched Line**

| Pin | Signal Name |
|---|---|
| 1 | Ground |
| 2 | Transmitted Data |
| 3 | Received Data |
| 4 | Request to Send |
| 5 | Ready for Sending |
| 6 | Data Set Ready |
| 7 | Signal Ground |
| 8 | Receive Line Signal Detector |
| 14 | New Sync |
| 15 | Transmit Signal Element Timing |
| 17 | Receive Signal Element Timing |
| 18 | Test Control |
| 20 | Data Terminal Ready |
| 22 | Calling Indicator |
| 23 | Data Signaling Rate Selector |
| 24 | Transmit Signal Element Timing (DTE clock) |
| 25 | Test Indicator |

*Fig. 53 Contd.*

| F | A | C | DATA | FCS | F |
|---|---|---|---|---|---|
| 8bits | 8bits | 8bits | | 16bits | 8bits |

Fig 54 SDLC Format for Logical Link Control

1. Flag (F) a unique 8-bit sequence indicating start and stop of a packet. This sequence is (01111110) and it must appear nowhere else in the transmission.

2. Address (A): a field of eight bits that indicates one of 254 stations that is a distination. One must be careful in SDLC which is part of IBM's SNA (System Network Architecture) since there are multilayer addressing available.

3. Control Field (C) which is used to send information, acknowledge the flow and poll stations. It is this field that we concentrate on from a communications viewpoint.

4. FCS: An error detection field similar to the authentication word appended in Section 3.3. Note here we take the message vector s and generate a 16-bit error detection message m. This is then used at the received node to determine if an error has occured. Suffice it to say that this functions identically to the linear authentication in Section 3.3. When the block is transmitted a check is made and error can be detected. The chance of an undetected error is less than $10^{-12}$ in most cases.

The SDLC formats for the control field are shown in Fig. 55. The information sequence is used to transfer data and is the one we shall concentrate on. The supervisory sequence is for control and other supervisory functions. The nonsequential format is used for setting certain operating modes and miscellaneous operations. In the information format we have a single 0 bit indicating the format and a poll bit (bit5) indicating whether this is a poll from a primary station. It should be noted that SDLC is part of the SNA hierarchial structure and polling is done centrally. A multidrop feature is possible but we shall not discuss it in detail. The two counts that are important are the send and receive counts used to enumerate the number of free blocks transmitted. For SDLC this is 8 blocks although this can be increased to 128 for satellite usage. Each station in a session (a computer dialog) keeps track of $N_r$ and $N_s$.

| | Send Sequence Count | Poll | Receive Sequence Count |
|---|---|---|---|
| 0 | $N_s$ | P/F | $N_r$ |
| 1 | 3 | 1 | 3 |

(a) INFORMATION

| | Supervisory | | |
|---|---|---|---|
| 1 0 | S | P/F | $N_r$ |
| 2 | 2 | 1 | 3 |

(b) Supervisory

| | Modifier Function | | |
|---|---|---|---|
| 1 1 | M | P/F | M |
| 2 | 2 | 1 | 3 |

(c) Nonsequenced

Fig. 55 DLC Control Field

With this DLC structure we can now consider how additional error control is conducted. With the use of the FCS field, each block of data can be checked for errors. If an error is detected at the receive station, a message can be returned indicating that an error occurred and that a retransmission is required. This technique is called automatic repeat request (ARQ). Since the errors are then dominated by the factor of the FCS check and not just FEC on the link, the error performance and only what? indirectly on the FEC. The net result is error performance on the order of $10^{-11}$ to $10^{-13}$ or better. The recognition of an FCS failure is done automatically by comparing the counters that are received. For example if at the sender we counted seven sent and the block from the receiver, containing data, indicates only three, we know we lost messages. Thus a repeat is required. ?

There are several types of ARQ systems (see Gatfield) that have been found useful. These are;

1. Stop and Wait ARQ - here one block is sent at a time and then an acknowledgement (positive or negative) is sent and received before the next transmission.

2. Continuous ARQ - here all the data is sent continuously. When a negative response (NAK) is received the transmitter must back up and retransmit all data from that point.

3. Go-Back-N ARQ - in this system the transmitter, upon an acknowledgement of a NAK, completes transmission of its current block and goes back N blocks, where N and the block length are large enough to exceed the round trip delay.

4. Selective Repeat ARQ - in this technique, short blocks are used and only the block with an error is repeated.

It should be noted that for 1-3 of the above, no memory is assumed at the receiving node, so that if an out of sequence block is received, it discards it and the rest. If we permit memory, which is getting less expensive, then we can perform a more complex task. However, the data sequencing may be done at DLC layers and not higher in the computer architecture and thus sequencing must be retained. It is important for the communication designer to have an understanding of the computer systems architecture, in order not to create problems in highly coupled layered structures.

We can now analyze the efficiency of these various schemes. The efficiency, $\eta$, is defined as

$$\eta = \frac{\text{total bits transferred}}{\text{total bits that could be transferred}} = \frac{B_A}{B_T}$$

The numerator is the number of bits (or blocks) that are actually transferred in T sec. The denominator is the number that could be. For example, if the channel rate is R bps and T was the duration, then this is RT. Here $B_A$ is the actual bits transferred and $B_T$ the total. Note that if R is the data rate then $\eta$ also equals

$$\eta = \frac{T_A R}{T_T R} = \frac{T_A}{T_T} = \frac{\text{actual information transmission time}}{\text{total transmission time}}$$

Let us now divide $\eta$ into three factors. Let

$$\eta_O = \frac{\text{number of data bits}}{\text{number of bits in block accepted}}$$

Here a block contains n bits, only k of which are data and n-k are overhead. Thus

$$\eta_O = \frac{k}{n}$$

The second is

$$\eta_R = \frac{\text{number of blocks accepted}}{\text{number of blocks transmitted}}$$

This is a measure of how much repetition is required. Finally

$$\eta_D = \frac{\text{number of blocks transmitted}}{\text{total number of blocks that could be transmitted}}$$

Thus we have;

$$\eta = \eta_O \eta_R \eta_D$$

Now if the bit error rate is p, then since the bits are independent, we have the probability of there being no errors $(1-p)^n$ and the probability of there being ~~some~~ at least one block error~~s~~ is

$$B = 1-(1-p)^n$$

Thus errors occur at the rate of one every 1/B blocks. Now for ARQ systems if an error is obtained there are M additional blocks that are transmitted due to the delay of the NAK. Those are don't care blocks that must eventually be repeated also. There are N blocks that must be repeated. Thus the number of blocks transmitted between errors is 1/B+M while the good blocks are 1/B+M-N. Therefore,

$$\eta_R = \frac{1/B+M-N}{1/B+M}$$

Let us now consider the waiting factor, $\eta_D$. This is

$$\eta_D = \frac{1}{d+1}$$

where d is the round trip delay in blocks and this applies for stop and wait transmissions. For continuous transmissions, d is zero.

We can now consider several examples.

1. Stop and Wait

   For this case N=1 since only one block is repeated and M=0 since only block is sent. Thus

$$\eta = \frac{k}{N}\,\frac{(1/B)-1}{1/B}\,\frac{1}{d+1}$$

2. Go-back-2

   Now we note that the block is chosen to equal the round trip delay. Thus if a block is in error only one additional block is received (M=1) and two are retransmitted (N=2). No dead space exists so d=0. Thus

$$\eta = \frac{k}{n}\,\frac{(1/B)-1}{(1/B)+1}$$

insert

Figs. 56 to 59 depict performance of these schemes as a function of p for several cases.

We will now discuss how this is applied in SDLC. In SDLC we have a capability for eight blocks of data to be stored. This is called the MAXOUT. We let RTD be the normal trip delay. In all cases we request a duplex operation to receive answers to polls. Poll bits are set in certain blocks in transmit,and responses to those polls are observed or receive. There are three modes of SDLC that we shall describe and this is based on the results of Trayham and Stien.

1. Half Duplex - In this case we transmit the eight blocks and set a poll on the last. We then wait for a response before continuing. If we get a reject due to a block error we then retransmit that block plus seven new ones. This scheme is shown in Fig. 10

2. Normal Response Mode, Full Duplex (NRM/FD) - This is shown in Fig. 61. Here we poll more frequently since we can receive a poll response at the same time. Thus if the RTD is small we obtain a continuous transmission. If RTD increases we get an intersequence delay ISD waiting for a response to take poll of the MAXOUT block. If the RTD is very large then we get an ISD before even sending the last block plus an additional ISD before the new sequence starts.

3. Asynchronous Response Mode, Full Duplex (ARMIFD) - This is shown in Figure 62. Here we receive a response to each block. If RTD exceeds MAXOUT then we get an ISD. When an error recurs we note the loss of a response and backup and repeat from that loss.

This scheme is shown in Fig 60.

2. <u>Normal Response Mode, Full Duplex (NRM/FD)</u>:
This is shown in Fig 61. Here we poll more frequently since we can receive a poll response at the same time. Thus if the RTD is small we obtain a continuous transmission. If RTD increases we get an inter sequence delay ISD waiting for a response to the poll of the MAXOUT block. If the RTD is very large then we get an ISD before even sending the last block plus an additional ISD before the new sequence starts.

3. <u>Asynchronous Response Mode, Full Duplex (ARM/FD)</u>.

This is shown in Fig 62. Here we receive a response to each block. If RTD exceeds MAXOUT then we get an ISD. When an error occurs we note the loss of a response and backup and repeat from that loss.

(14)

With SDLC we have 8 bits per byte, and six bytes of overhead. If the data field has DBY bytes of data then the ratio $y_o$ is

$$Y_o = \frac{DBY}{DBY+6}$$ ?

derivations!!! show them or reference them!

We can further show that

$$y_W M_R = \frac{(1/B-1)}{1/B + NPF+TD}$$ ?

where NPF are the nonproductive frames and TD is the delay. Now note that $M_o$ increases as the block size increases (DBY+6) while $y_W y_R$ decreases since 1/B gets smaller. Thus if we were to plot y versus the block size there would be a maximum. Figures 63 to 65 are from Trayham and Steen Ref? and present a summary of SDLC performance under a wide variety of circumstances.

Note in particular the efficiency as a function of p, the bit error rate. At a BER of $10^{-4}$ with a block size of 2,000 bytes, we see a 20% efficiency compared to 90-95% at higher link performance. This means that a 56Kbps line can only handle 10 Kbps of data since most of the time is spent retransmitting! This effect has been a major problem on satellite links. Note also the drop in efficiency with increasing delay. It shows a 20% degradation between optimized terrestrial and satellite links.

Conclusions?

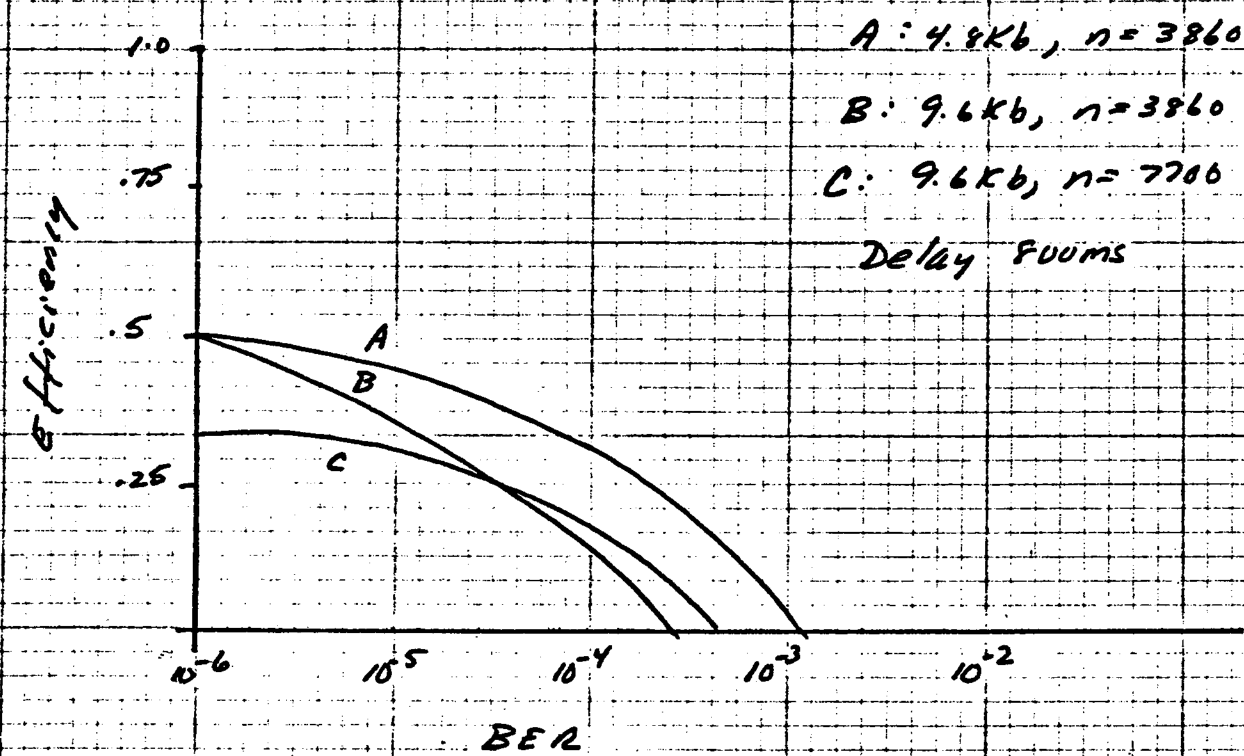



Fig 56: Efficiency for Stop and Wait ARQ
(from Gatfield) Ref 2.

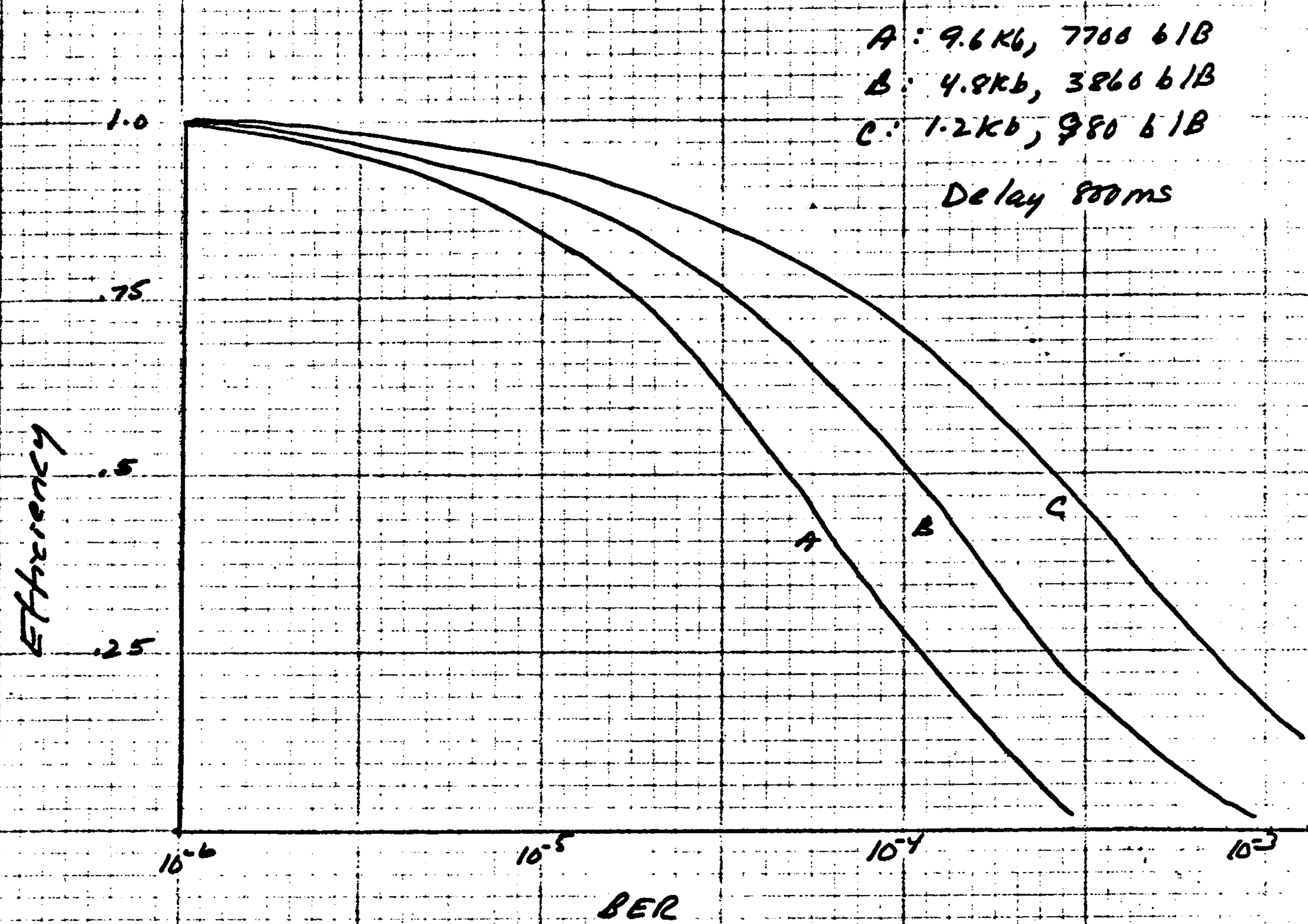



Fig 57: Go Back 2 Efficiency
(from Gatfield) Ref ?

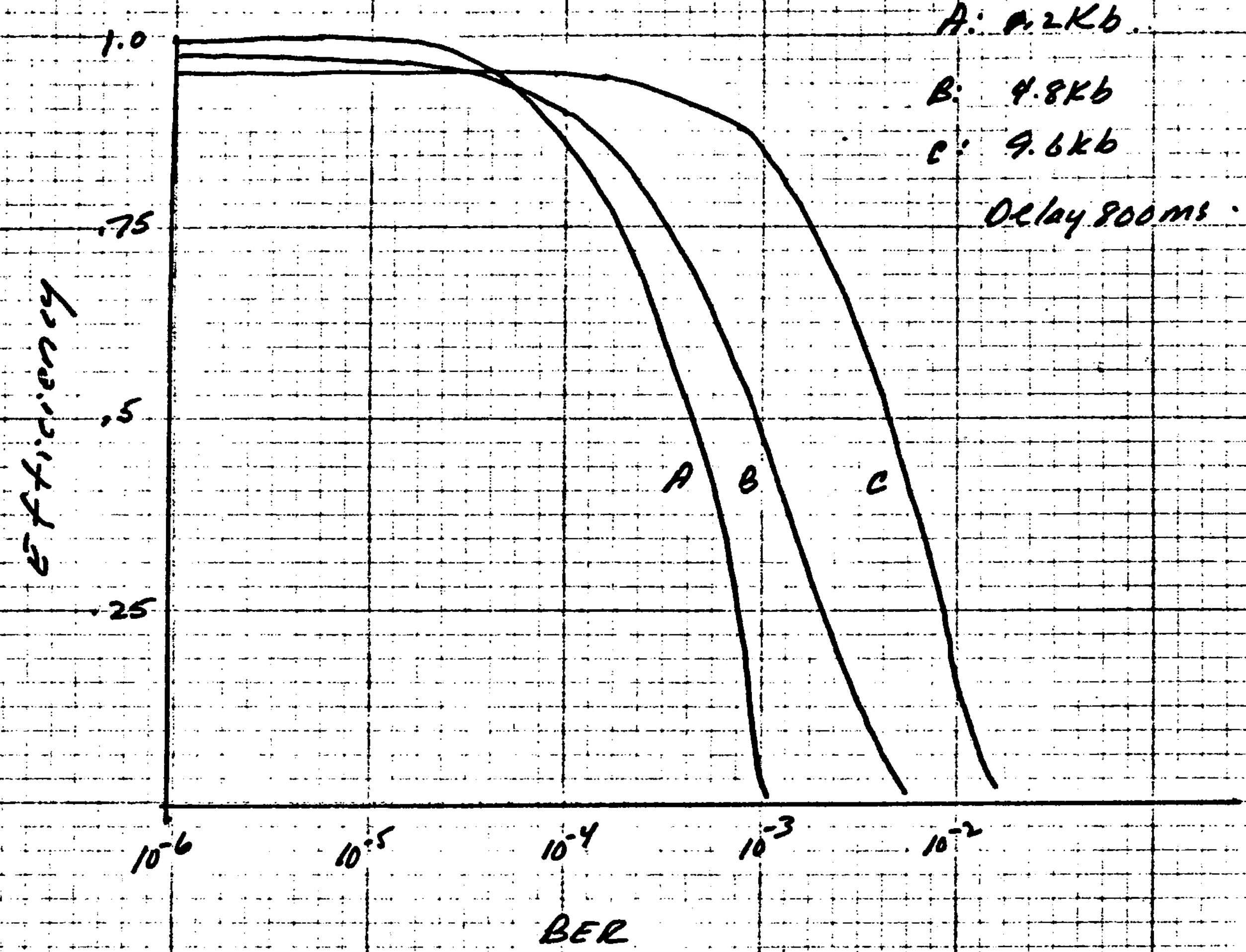



Fig 58: Efficiency for Go-back-N ARQ
(From Gatfield) Ref ✓

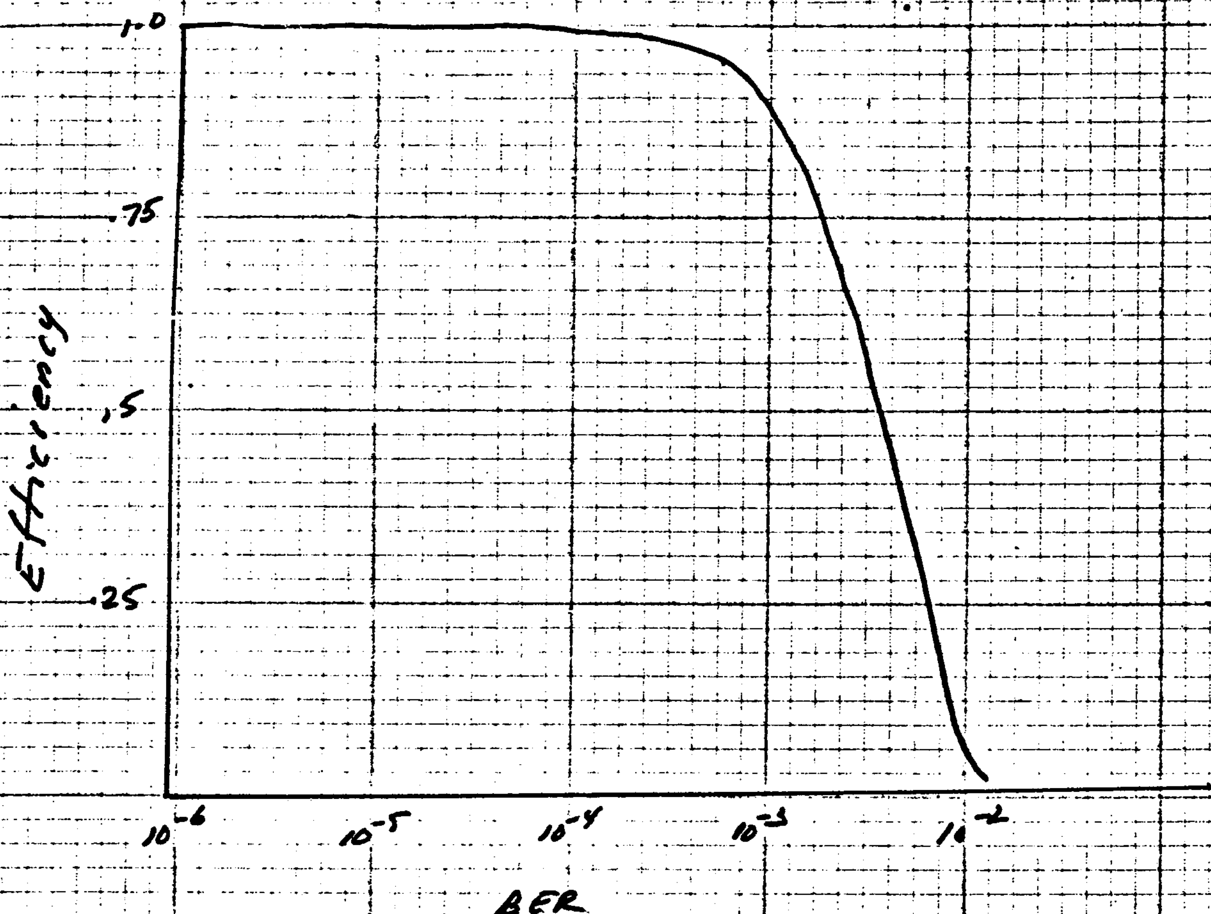
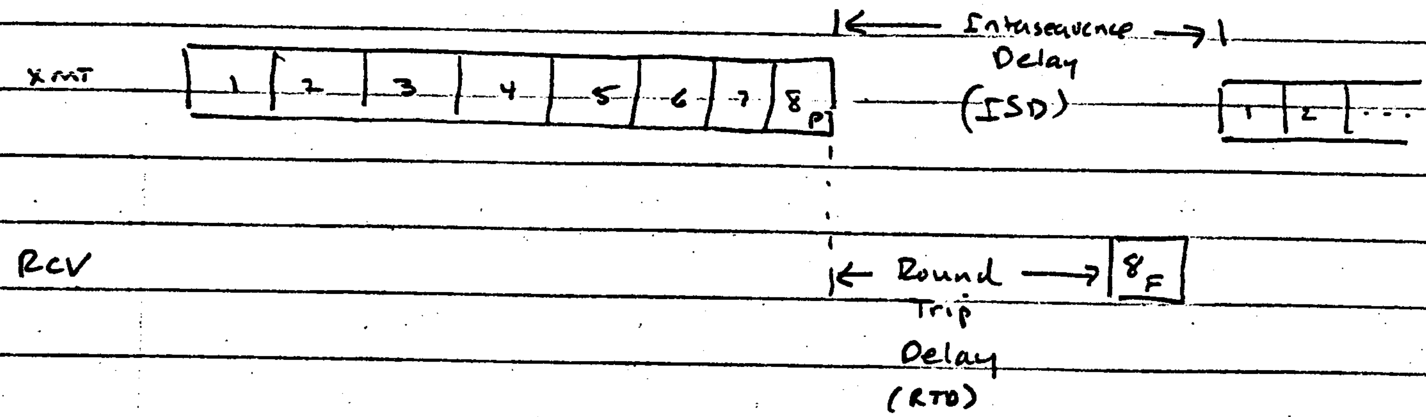
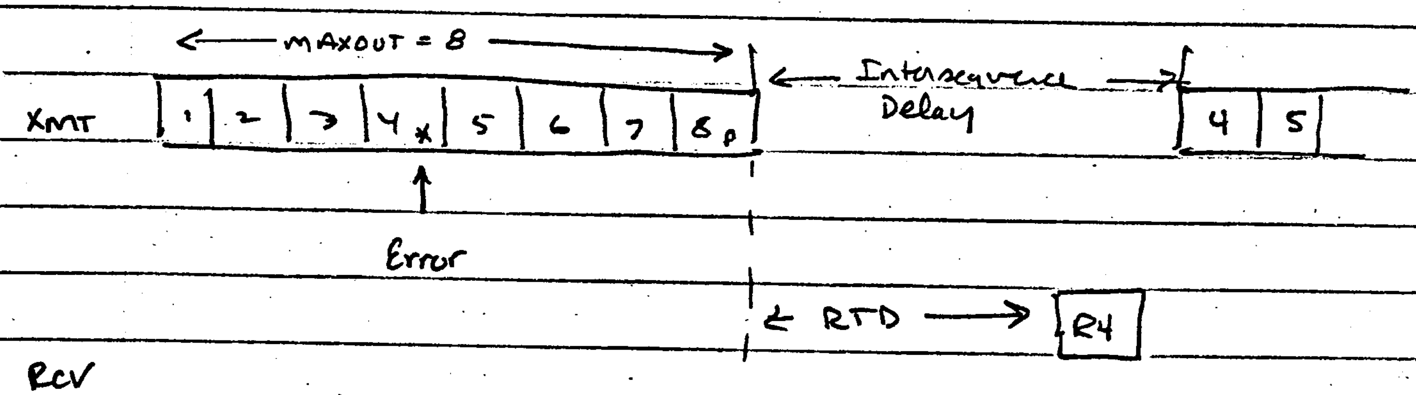



Fig 59: Efficiency for Selective Repeat ARQ (from Gatfield) Refs




(a)

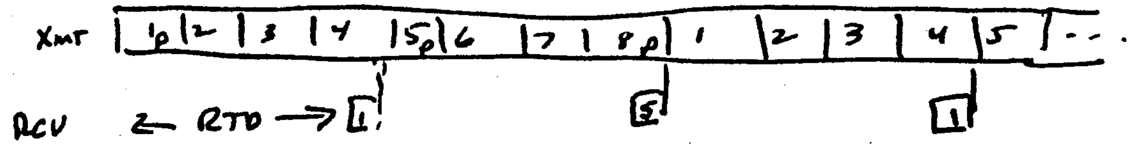
Half Duplex - No Errors



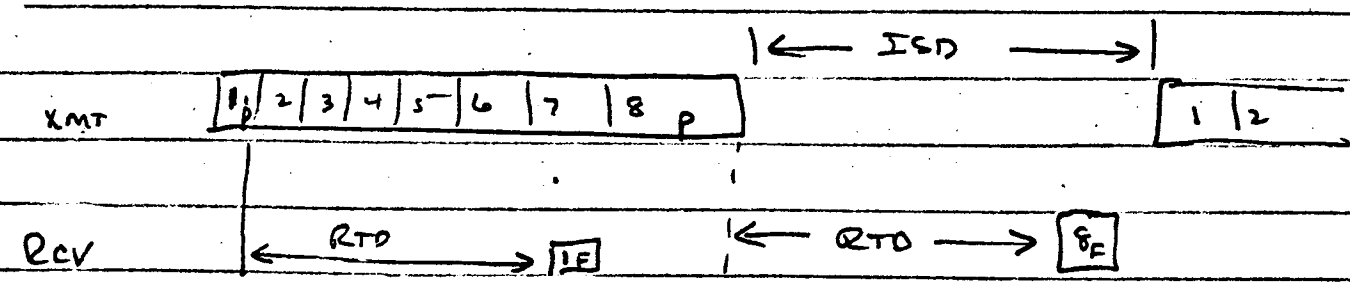

(b)

Half Duplex - Error

Fig 60 : Half Duplex Operation. Here we send message blocks as numbered. MAXOUT is the maximum number that we send at any one time.

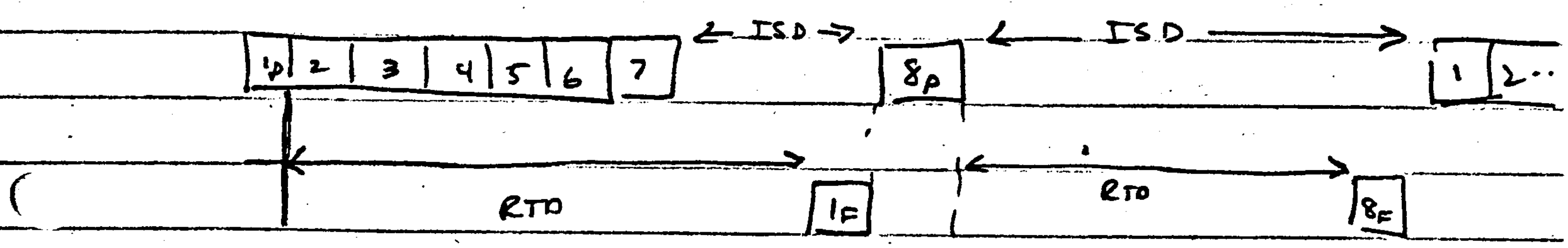



(a)




(b)

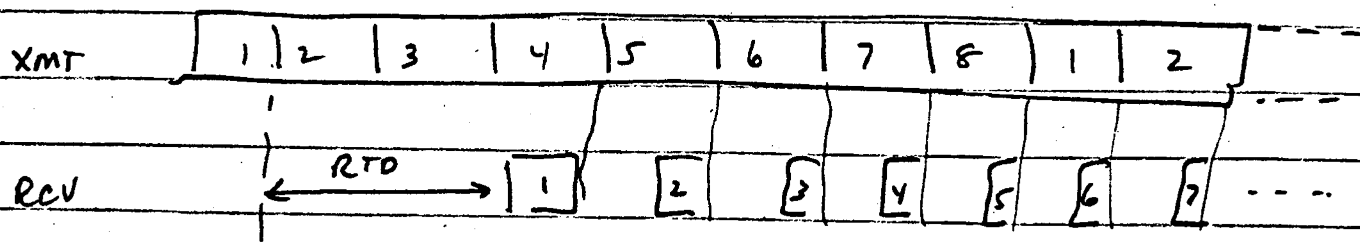
NRM Full-Duplex with ISD (Intersequence Delay)




(c)

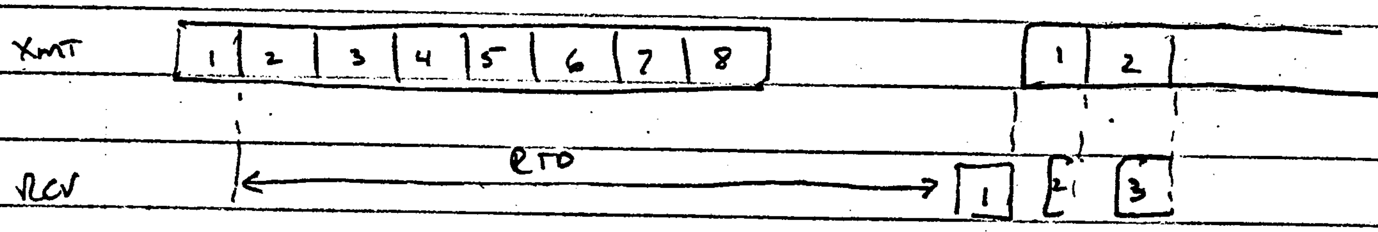
NRM Full Duplex with ISD and MAXOUT less than RTD

Fig 61 : NRM Full Duplex



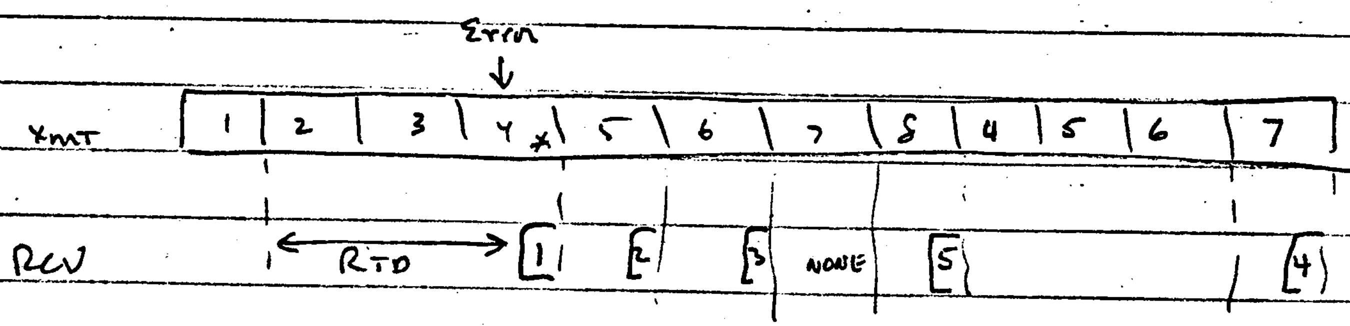

(a)

ARM Full Duplex




(b)

ARM/FD with RTD Greater Than Maxout




(c)

ARM/FD Error Recovery

Fig 62: ARM Full Duplex

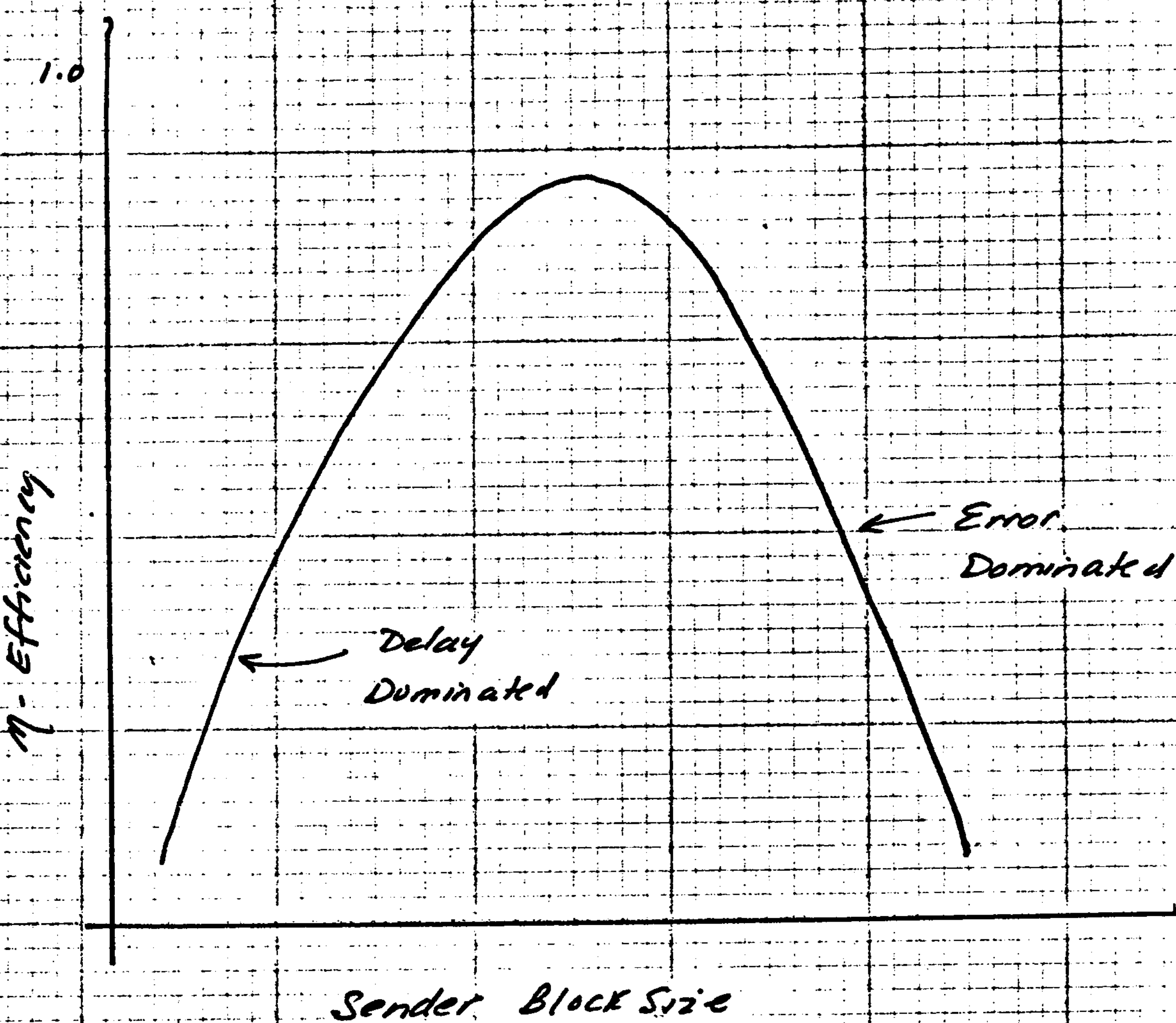



Fig. 63 Efficiency as a function of Block size. At the low end with short blocks, efficiency decreases due to high overhead ~~At the~~ high end, channel errors cause retransmission

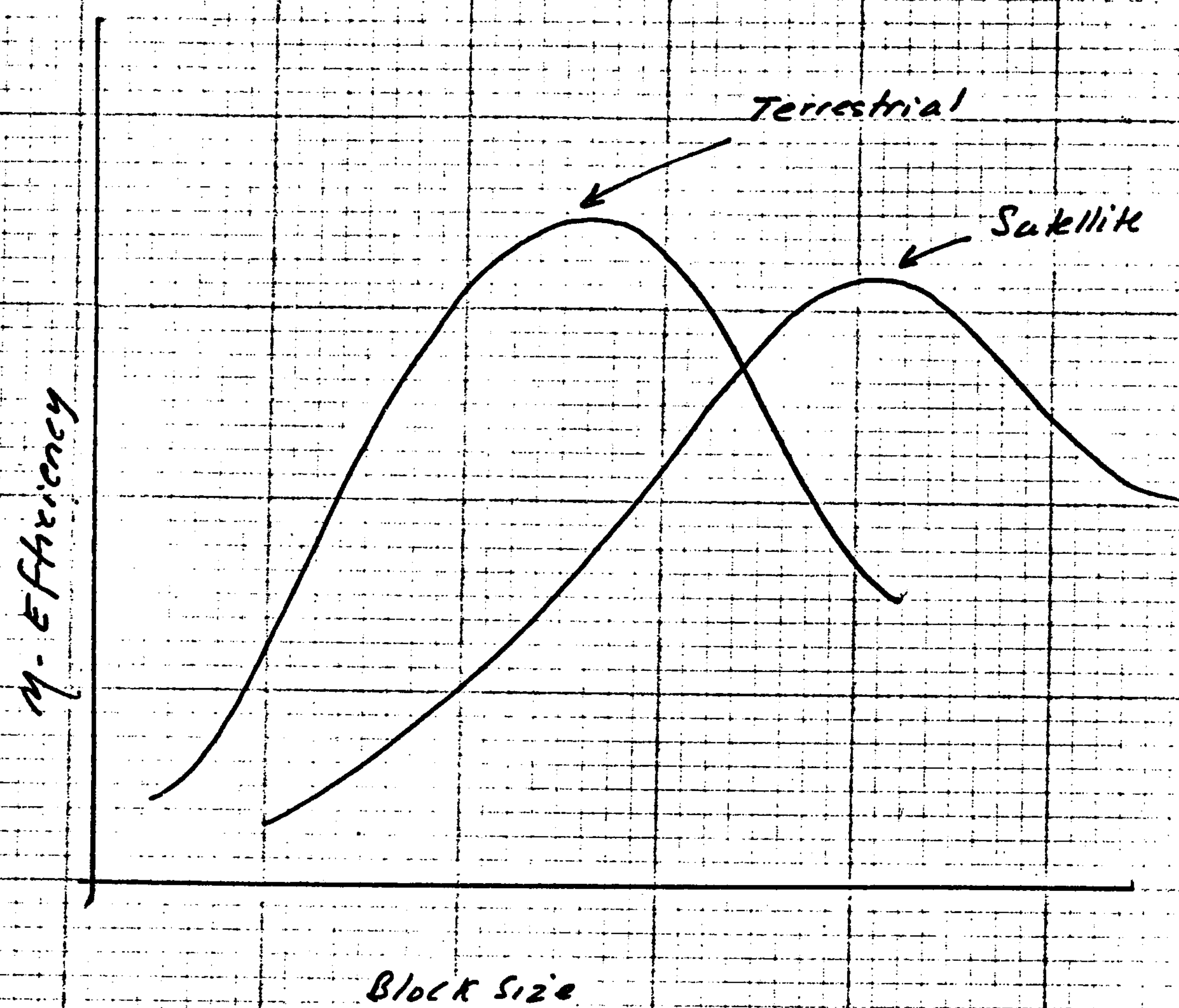



Fig 64 : Comparison of Terrestrial and Satellite Links. Reduced efficiency due to satellite delay

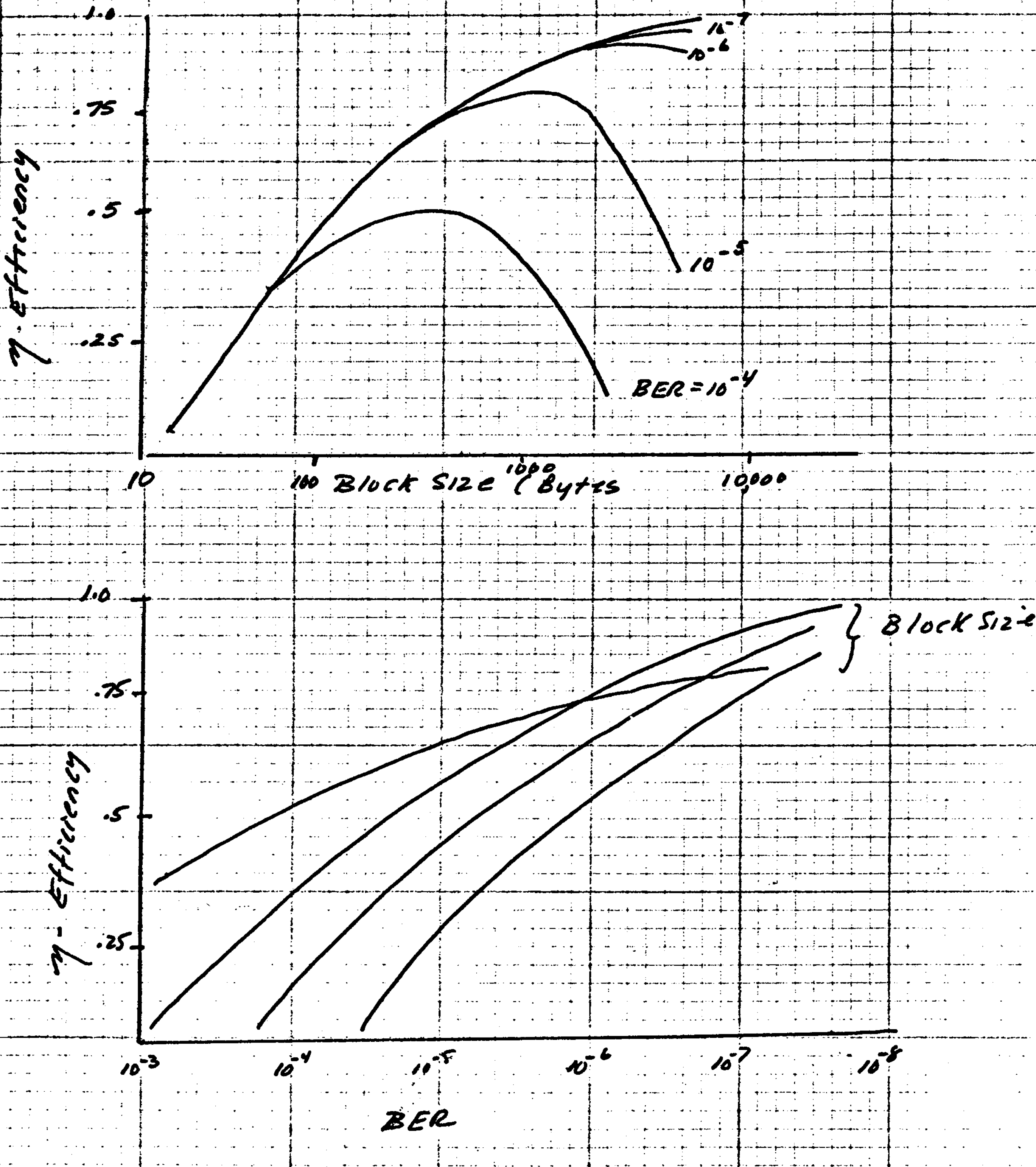



Fig 65: Efficiency and BER

# CHAPTER 4
# EARTH STATIONS

This chapter provide an overall system design for a wide variety of earth stations that are used for commercial satellite communications. The function of the earth station is to act as the interface on both transmit and receive for the digital signals emerging from the modems and the freespace propagation path to the satellite. In this section we develop first the overall model for the earth station and then proceed to consider the basic elements. We pay careful attention to the interaction of performance and design because it is in the earth stations design that the most significant cost savings can be amassed.

## 4.1 Structure

The function of the earth station is to transmit and receive RF signals and provide for the translation to an intermediate frequency (IF) band in some suitable fashion signals for the demodulators or demultiplexors. The earth stations can be viewed on the transmit and receive sides respectively. On transmit the intermediate frequency (IF) interface with the modulator or multiplexor provides a signal of a bandwidth adequate for assigned use. For example if an SCPC use is contemplated, the IF output would be a single channel waveform of a bandwidth equal to the basic channel bandwidth. In INTELSAT an SCPC channel would be 45 KHz. In contrast with SBS the IF is at a 50 Mbps burst modem which occupies a 40 MHz transponder.

Thus the choice of the IF will depend first upon the data bandwidth. However, as we shall see later, this will also depend upon IF converters and their phase characteristics. Typical IF frequencies range between 20 MHz up to 150 MHz and beyond. A typical digital IF is 70 MHz.

The earth station on transmit then takes the IF signal, filters it to prevent out-of-band omissions, up-converts it by means of an oscillator mixer pair to the carrier frequency and then amplifies and transmits the signal out the antenna.

On receive a similar process occurs with reception at the antenna, front end amplifications, down-conversion to IF, filtering and interfacing.

66?

Figure 3.1 depicts a typical single channel earth station design. In this design the interfaces at IF are the modulator and demodulator. The station is composed of six basic elements.

1. <u>Up-Converter System</u>

   This is composed of a mixer which multiplies the IF signal with a local oscillator that is centered at the desired transmit frequency. The output of this is filtered to eliminate any out-of-band signals that may be generated by the oscillator. As indicated the filter is separated by isolators that are used to match impedances to minimize power losses due to reflection mismatches.

2. <u>IPA/HPA Subsystems</u>

   This system, the intermediate and high power amplifier system again amplifies and filters the signal now at RF. Several stages of amplifications are typically required since, at higher powers, the gains per stage are not very large. Filtering is again used to eliminate any out of band distortions. or interferences ?

3. <u>Antenna Subsystem</u>

   The antenna used are used for both transmit and receive. The earth station antennas are generally parabolic or outgoing radiation. This radiation is usually polarized (linear or circular) in two directions on transmit and receive. Since both transmit and receive use the same antenna an orthomode transducer (OMT) or some other form of high isolation splitting device is used. In addition since the transmit is high power and receive is low, and the low noise amplifier may still respond to power in the

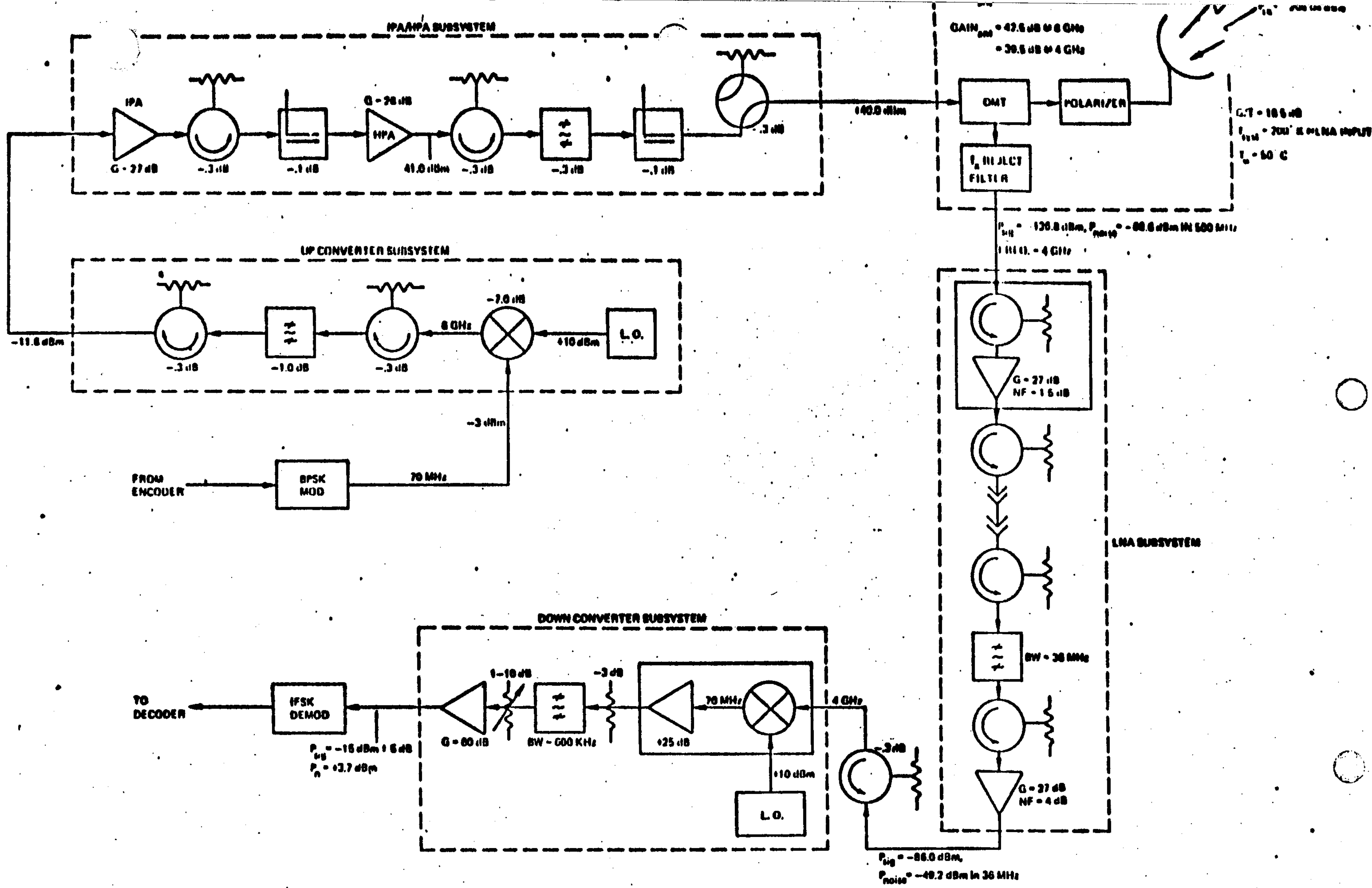



Fig. 1: TYPICAL REMOTE TERMINAL BLOCK DIAGRAM

G-17200

(?)

transmit band, an additional transmit reject filter is used. This insures that saturation does not occur. The antenna can be realized in many forms and it may or may not track the satellite.

4. LNA System

On receive, a low noise amplifier (LNA), is used to amplify the signal. This LNA is typically pla~~s~~ced close to the antenna to minimize losses due to long coaxial cable runs or waveguide runs. Several stages of gain are used to insure as high a signal to noise ratio as possible.

5. Down Converter

This is the mirror image of the upconverter system. It provides for the conversion if the RF signal to the IF level (e.g. 70 MHz). Typically a single conversion is adequate but in some applications dual conversion is used. Both up and down converters run off of a reference crystal that is tuned to the center transmit signal (possibly using a sideband).

## 4.2 Antennas

The antenna is the device used to radiate or receive radiation that links the earth station to the satellite. The radiation results from a current density that is excited on a radiating surface and it in turn generates a field at a distance from the radiator. What we want to do in this section is to develop the basic concepts of the antenna and present the elements that make it up. As throughout the text, we shall develop the basic system issues and not concentrate on the technology per se.

Let us first consider the problem of radiation from an antenna. Due to the reciprocity of such systems, the results for reception are identical. Consider the aperture in Fig. 2 . The aperture lies in the x, y plane and is bounded. A current density, K (x,y), is induced on the aperture by a means that we shall discuss. This current density causes a field to be radiated out into space at all points. Consider the point $\underline{r}_o$ at a distance $\underline{N}$ from a point $\underline{r}'$ on the aper~~ture~~. Let $K(\underline{r}')$ be the current at $\underline{r}'$. Let $k_o$ be the wave number defined as $2\pi/\lambda$ where $\lambda$ is the wavelength. We are assuming that the current density is at a constant rf frequency of $f_o$ Hz. Extension to multifrequency is done by superposition. Now the electric field at $\underline{r}_o$ can be shown to be (see Silver) Ref ?

$$E(\underline{r}_o) = \int_s K(\underline{r}') \frac{\exp(-jk_o r)}{r} \, ds$$

where K is suitably normalized and

$$r = |\underline{r}| = (|\underline{r}_o - \underline{r}|^2)^{1/2}$$

Substituting we can show that this equals at the far field $(|r_o| \gg |r'|)$

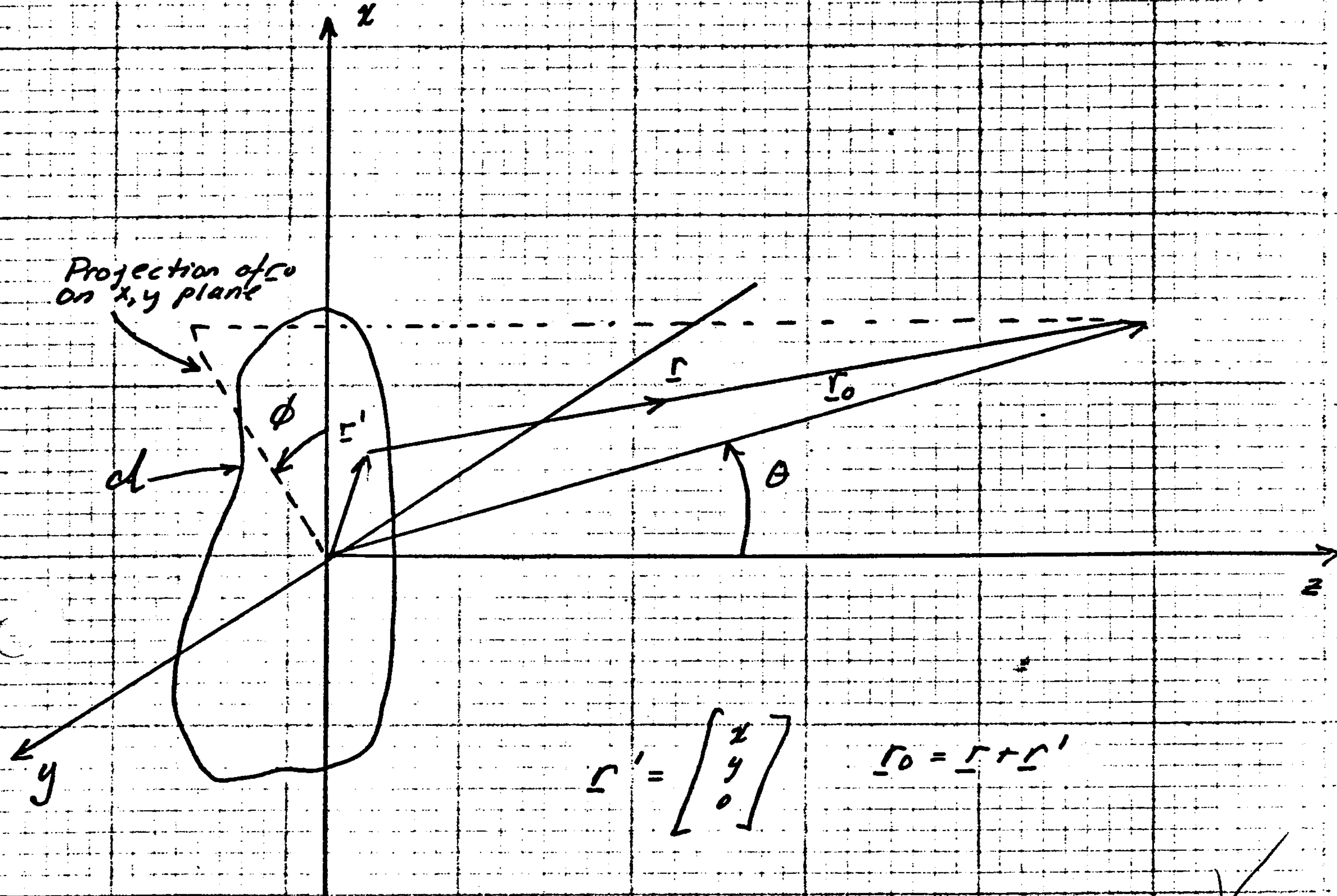
$$E(r_o) = \frac{1}{|r_o|} \int_s K(\underline{r}') \exp(+j\underline{k}\cdot\underline{r}')\, ds$$

where

$$\underline{k} = k_o \begin{bmatrix} \cos\phi \sin\theta \\ \sin\phi \sin\theta \\ \cos\theta \end{bmatrix}$$

and

$$\underline{r}' = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$




$$\underline{r}' = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \qquad \underline{r}_0 = \underline{r} + \underline{r}'$$

Fig ~~2~~ A-2: Basic Antenna Radiation Geometry

Here $\emptyset$ is elevation and $\theta$ is asimuth. Thus $E(\underline{r}_o)$ depends on the effective current density vector, distance, $|r_o|$, and azimuth and elevation.

Using the results of diffraction theory, we can determine the diffraction field as we did the electric field. The latter is then readily derived from the former. For angles close to beam center and for fields with constant phase the diffraction field can be shown to be; ?

$$u(\underline{r}) = \frac{j}{\lambda r} \exp(-jk_o r)$$

$$\cdot \int_A F(x,y) \exp(j\underline{k}\cdot\underline{r}')\,dxdy$$

where now $F(x,y)$ is the field density over the aperture. The power radiated at distance r in a unit solid angle can be written as**

$$P(\theta,\emptyset) = \frac{1}{2}\left(G/M\right)^{\frac{1}{2}} r^2|u(r)|^2$$

The total power radiated, $P_o$, can be determined by integrating over any surface, and a convenient one is the aperture. Thus

$$P_o = \frac{1}{2}\left(G/M\right)^{\frac{1}{2}} \int_A |F(x,y)|^2 dxdy$$

The antenna gain is then the rates of these elements;

$$G(\theta,\emptyset) = \frac{4\pi}{\lambda^2} \frac{\left|\int_A F(x,y)\exp(j\underline{k}\cdot\underline{r}')\,dxdy\right|^2}{\int_A |F(x,y)|^2 dxdy}$$

The peak antenna gain is defined at $\theta=\emptyset=0$ or

$$G = \frac{4\pi}{\lambda^2} \frac{\left|\int_A F(x,y)\,dxdy\right|^2}{\int_A |F(x,y)|^2 dxdy}$$

$$\triangleq \frac{4\pi}{\lambda^2} A_{eH} \quad A_{eff}$$

where $A_{eH}$ is defined as the effective antenna area. Note that antenna gains always relate to power levels. This will be important to the calculation of link budgets where power is the key factor.

** Note that G is the dielectric constant and M the magnetic ____.

We now wish to consider two key factors in the definition of the antenna pattern, its beamwidth and its sidelobe behavior. To do this we consider two apertures each with constant density on the aperture.

Example: Let the aperture be a square of length d on a side and assume that F(x,y) is constant on the aperture. Then clearly on the aperture

$$G(\theta,\phi) = \frac{4\pi}{\lambda^2} \frac{\left| \int_{-d/2}^{d/2} \int_{-d/2}^{d/2} \exp(j\cdot k\cdot r)\,dx\,dy \right|^2}{A}$$

where A equals $d^2$.

Substituting we obtain;

$$G(\theta,\phi) = \frac{4\pi}{\lambda^2 d^2} \left| \int_{-d/2}^{d/2} \exp(jk_o x \cos\phi \sin\theta)\,dx \right|^2 \cdot \left| \int_{-d/2}^{d/2} \exp(jk_o y \sin\phi \sin\theta)\,dy \right|^2$$

$$= \frac{4\pi}{\lambda^2/d^2} \frac{\left|\sin\left(k_o\frac{d}{2}\cos\phi\sin\theta\right)\right|^2}{\left|k_o\frac{d}{2}\cos\phi\sin\theta\right|^2} \cdot \frac{\left|\sin\left(k_o\frac{d}{2}\sin\phi\sin\theta\right)\right|^2}{\left|k_o\frac{d}{2}\sin\phi\sin\theta\right|^2}$$

Clearly the peak gain is $4\pi/(\lambda^2/d^2)$. Note that the gain depends on the ratio of the aperture size to the wavelength. For a fixed aperture the gain increases with frequency. Now the beamwidth can be determined by looking for the first zero. This occurs at

$$k_o\frac{d}{2}\sin\theta\cos\phi = \pi$$

The antenna gain pattern is depicted in Fig. 3 where $u = \frac{d\pi}{\lambda}\sin\theta$

Note the peak sidelobe is down about 14dB from the peak gain.

Example: Consider now a circular aperture of radius d with a uniform density on the surface. Now we have

$$x = d\cos\phi'$$
$$y = d\sin\phi'$$

where $\phi'$ is the integral over the surface. The integral of concern is

$$g(\theta,\phi) = \int_0^{2\pi} \int_0^{d} \exp\left(jk_o r \sin\theta \cos(\phi-\phi')\right) r\,dr\,d\phi'$$

The resulting antenna pattern is shown in Fig. 4. Note two effects; one is a decrease of the level of the first sidelobe and the second is the broadening of the main beam. This type of tradeoff is typical in such a configuration.

Other factors can change the antenna pattern. Several of these are;

a. The change of distribution of the density

b. A phase change in the density

c. A concavity of the reflector     Concavity

d. Secondary reflectors such as struts, supports, sub-reflectors and other elements.
   struts

e. The presence of radomes, buildings, and other absorbers or reflectors.

The density of the current on the surface of the reflector is usually generated by a feed system. When we discuss satellite antennas we will discuss complex feed structures that are used for sophisticated antenna patterns. For earth station antennas however, the feeds are generally simple. The desired patterns are symmetric with high gain, narrow beam width and low side lobes. The typical feeds are shown in Fig. 5,6. The feed element is a feed horn or radiator which has a wide pattern to cover the main reflector. The main reflector is a parababoid with high directionality and low sidelobes.

Fig. 6 depicts the details of the feed systems. It must be capable of use at both the transmit and receive frequencies and to do this a device called an orthomode transducer is used. This is basically a unidirectional microwave filter placed at the junction of the transmit and receive waveguide. The actual feed horn is shown in Fig. 7. Fig. 8, and 9 depict the radiation patterns of these feed horns at the transmit and receive frequencies which in turn are used to calculate the overall radiation patterns.

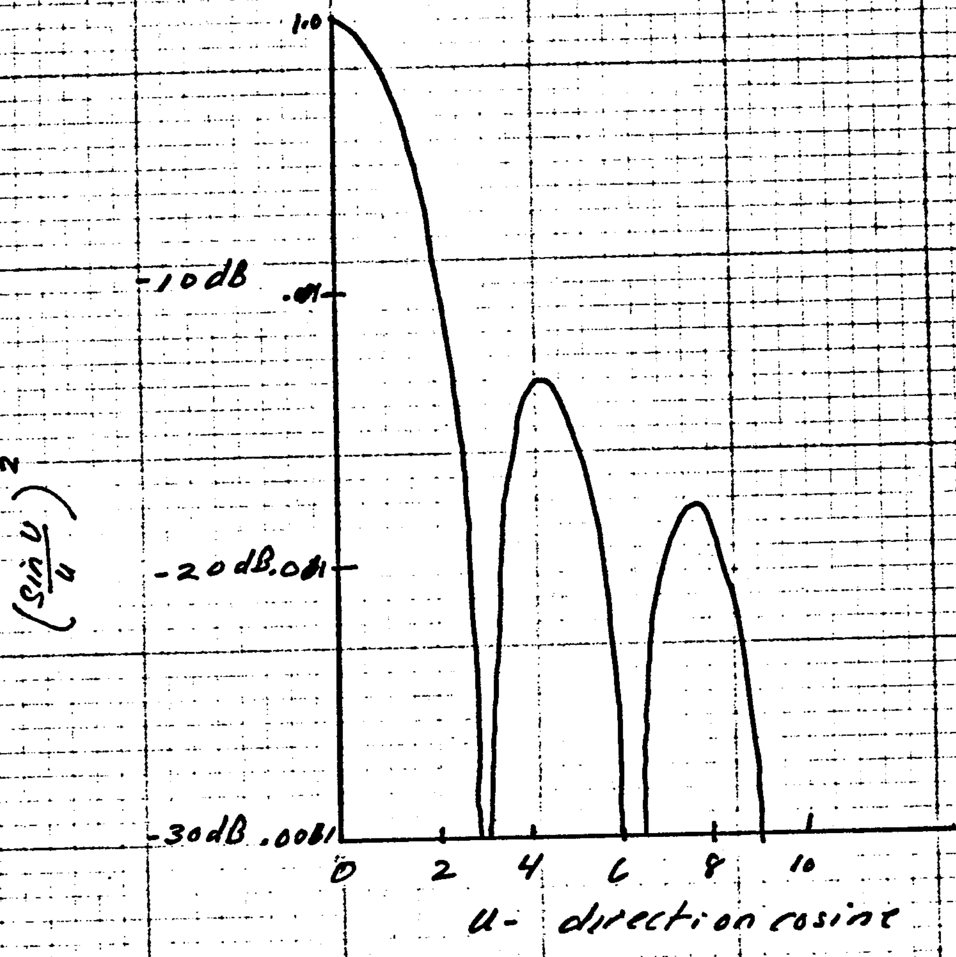



Fig 3: Antenna Pattern for Square Aperture (from Silver) Ref ?

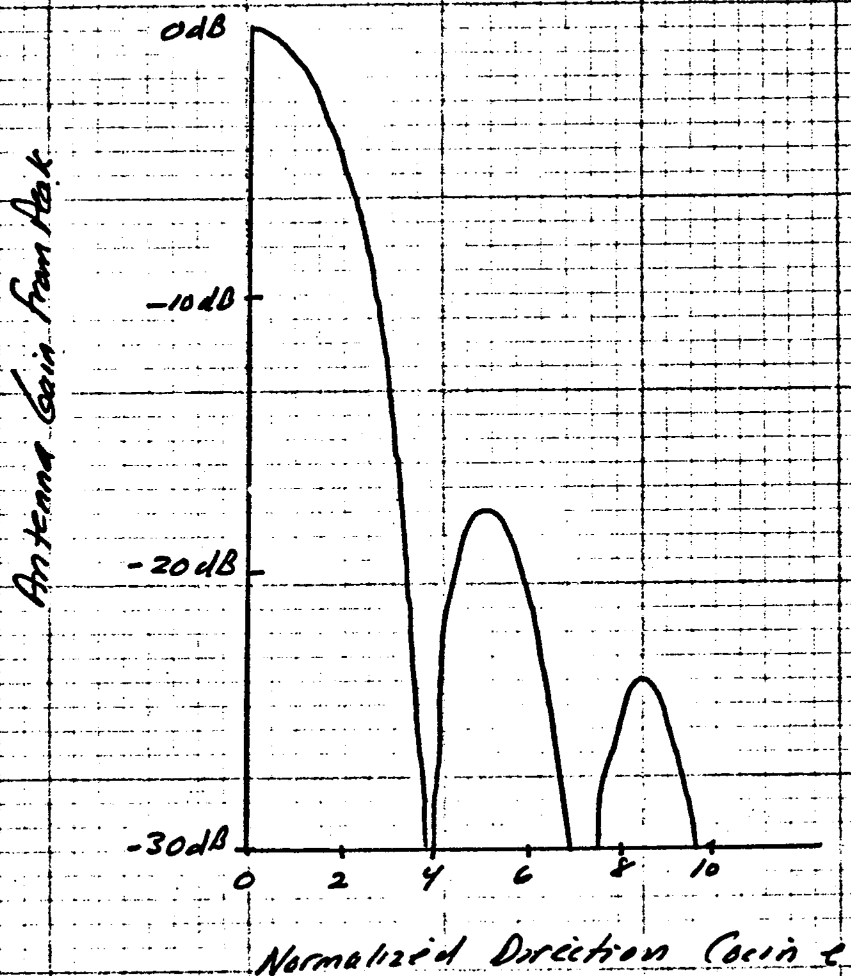



Fig 4: Uniform Circular Aperture Radiation Pattern (from Silver)

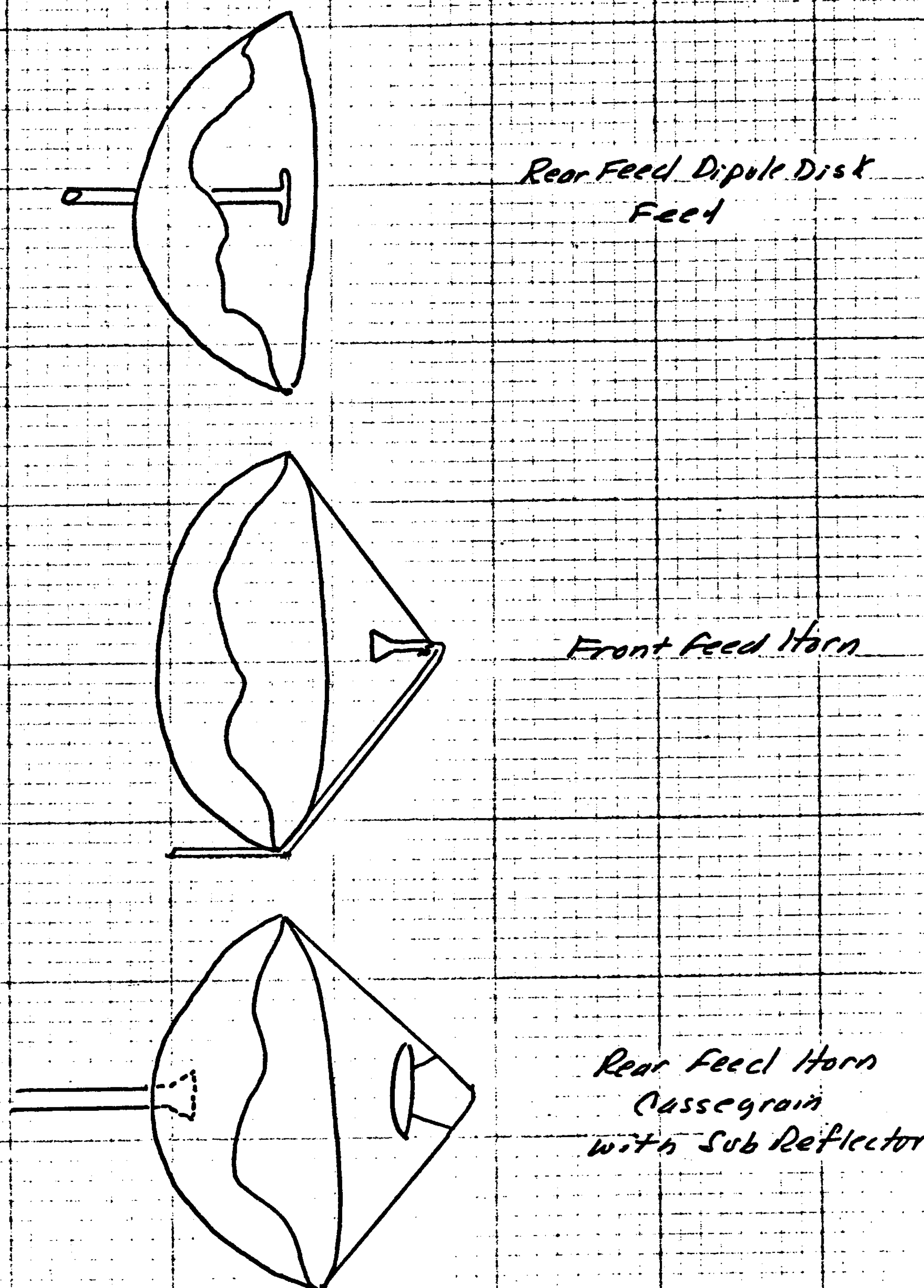



Fig 5: Antenna Configurations

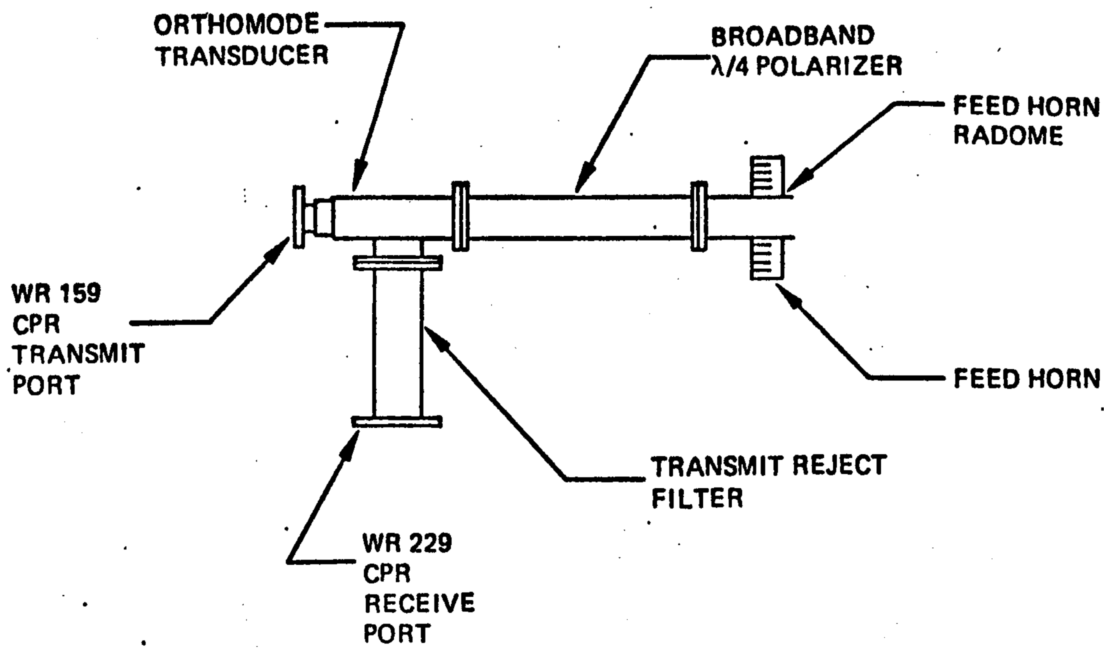



Figure [redacted] 8 FEED SYSTEM

G–17389

Figure ~~3.10.4~~ 7: FEED HORN DETAIL

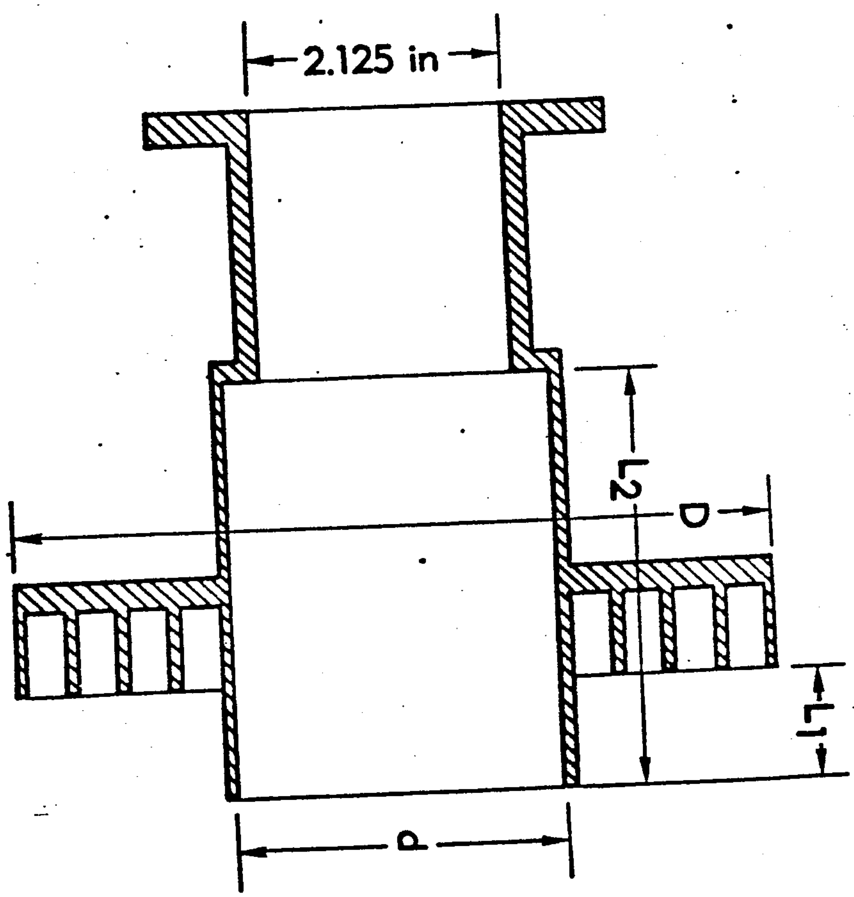


Figure 8: FEED HORN AMPLITUDE AND PHASE PATTERNS AT F = 6175 MHz

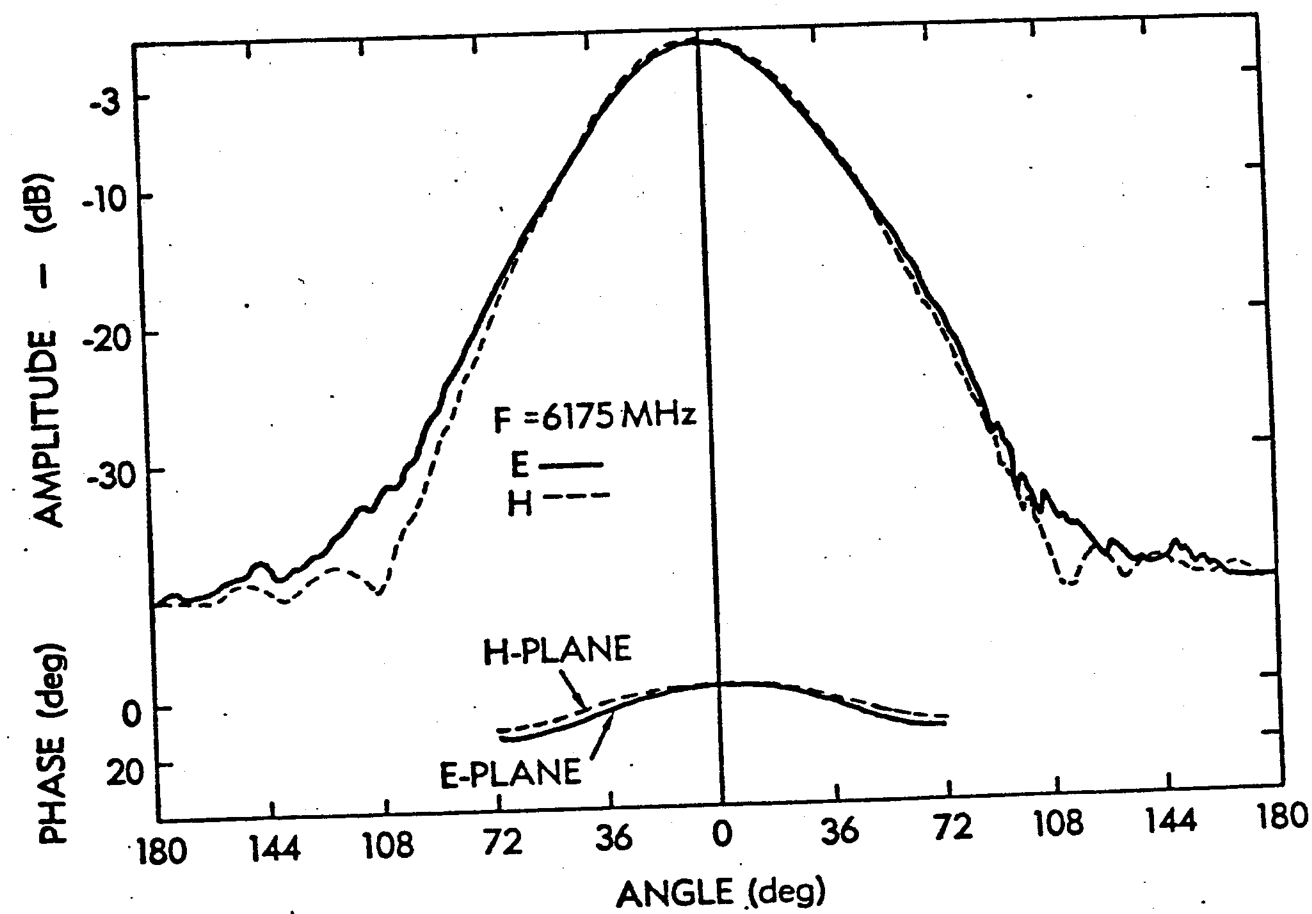
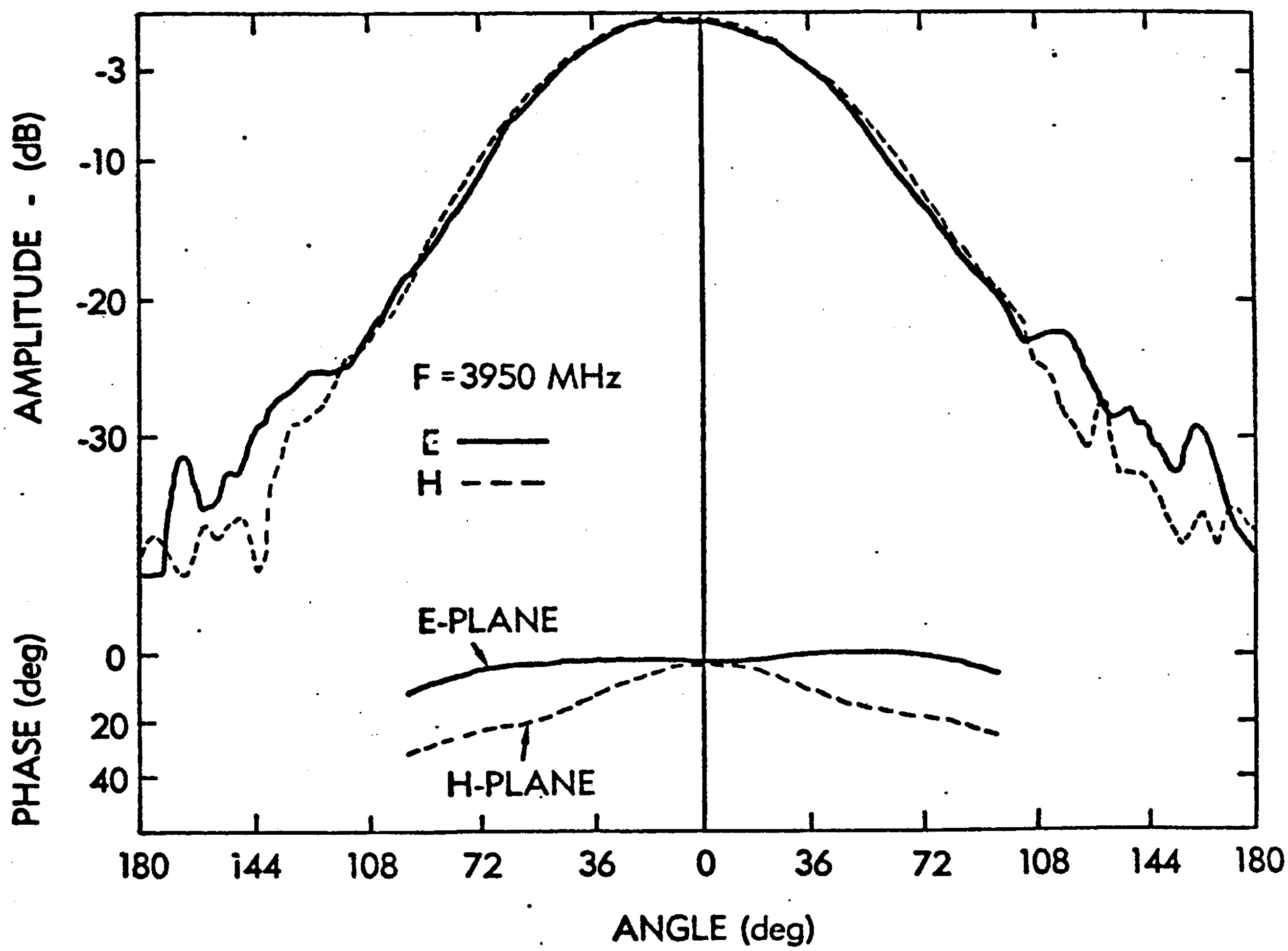


Figure ~~3.2.~~ 9 4: FEED HORN AMPLITUDE AND PHASE PATTERNS AT F = 3950 MHz



Cassegranian

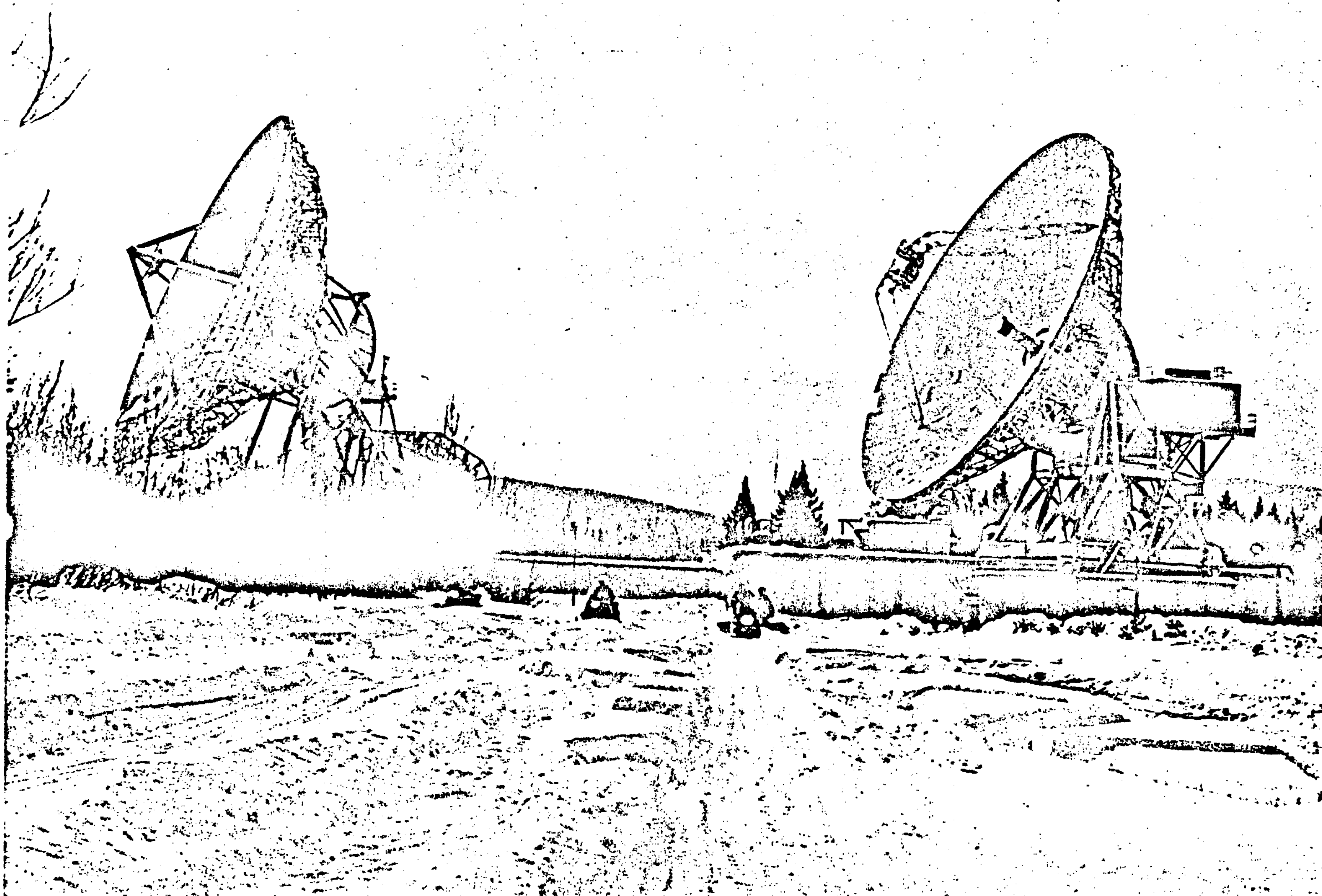
There are several other methods of radiating current densities on antennas. The Cassegranian antenna, typically used for very large apertures, has a subreflector which reflects the feed radiation onto the main reflector (see Fig. 10 ). An offset fed antenna has feeds which are off center. We shall return to this when we discuss the satellite antenna. A torus antenna is a recent development that allows multiple beams at angles that are tens of beamwidths. It is effective for use with multiple satellite systems.

The earth station antenna may or may not track the satellite. The beamwidth of the antenna gets narrower as the gain increases. However the satellite moves in orbit and is station kept in a box of $\Delta\theta$ degrees on side. Thus if the antenna is fixed and the satellite moves, the effective gain can change. Fig. 11 presents an analysis of this effect. What is plotted is the minimum antenna gain versus $D/\lambda$, the diameter to wavelength for a parabolic reflector. The curves are parametized on the basis of the satellite uncertainty box of size $\theta$ on the side. Note that at $\theta=0$ the satellite is fixed and the antenna points directly at it with no loss. At a $1^{\circ}$ box, typical of an INTELSAT-IV satellite, at $D/\lambda$ of 60 a decrease appears. Plotted also are the ranges in diameter for a 6/4 GHz satellite.

Thus for large antennas and not well enough stabilized satellites, we have to track the satellite with a device such as a monopulse tracking system. Fortunately most recent systems do not require tracking.

Other factors enter into the antenna design specifications. The first is the type of polarization required such as linear or circular. Most US domestic systems use linear polarization whereas INTELSAT uses circular. A second factor is antenna efficiency. A rough rule of thumb is that 70 dB is the greatest attainable antenna gain. That is, for large antennas, efficiency decreases as rapidly as peak gain increases. Most small antennas have an efficiency of 50 percent or greater. A third factor is sidelobe constraints. INTELSAT requires a limit in power emission of less than

Fig 10 Typical Cassegrain Antennas



F.10

$$(42 - 25 \log \phi) \text{ dbw } 2.5^0 < \theta < 45^\circ$$

where as the FCC requires an antenna gain constraint

$$(32 - 25 \log \theta) \text{ dbi } 2.1 < \theta < 45^\circ$$

The latter constraint can be difficult to meet with some small antennas. Figs 12 and 13 depict typical antenna pattern measurements for a 3m. antenna with the FCC standard superimposed. Table 3.1 depicts a typical antenna specification sheet.

FCC standard superimposed. Table 3.1 depicts a typical antenna specification sheet.

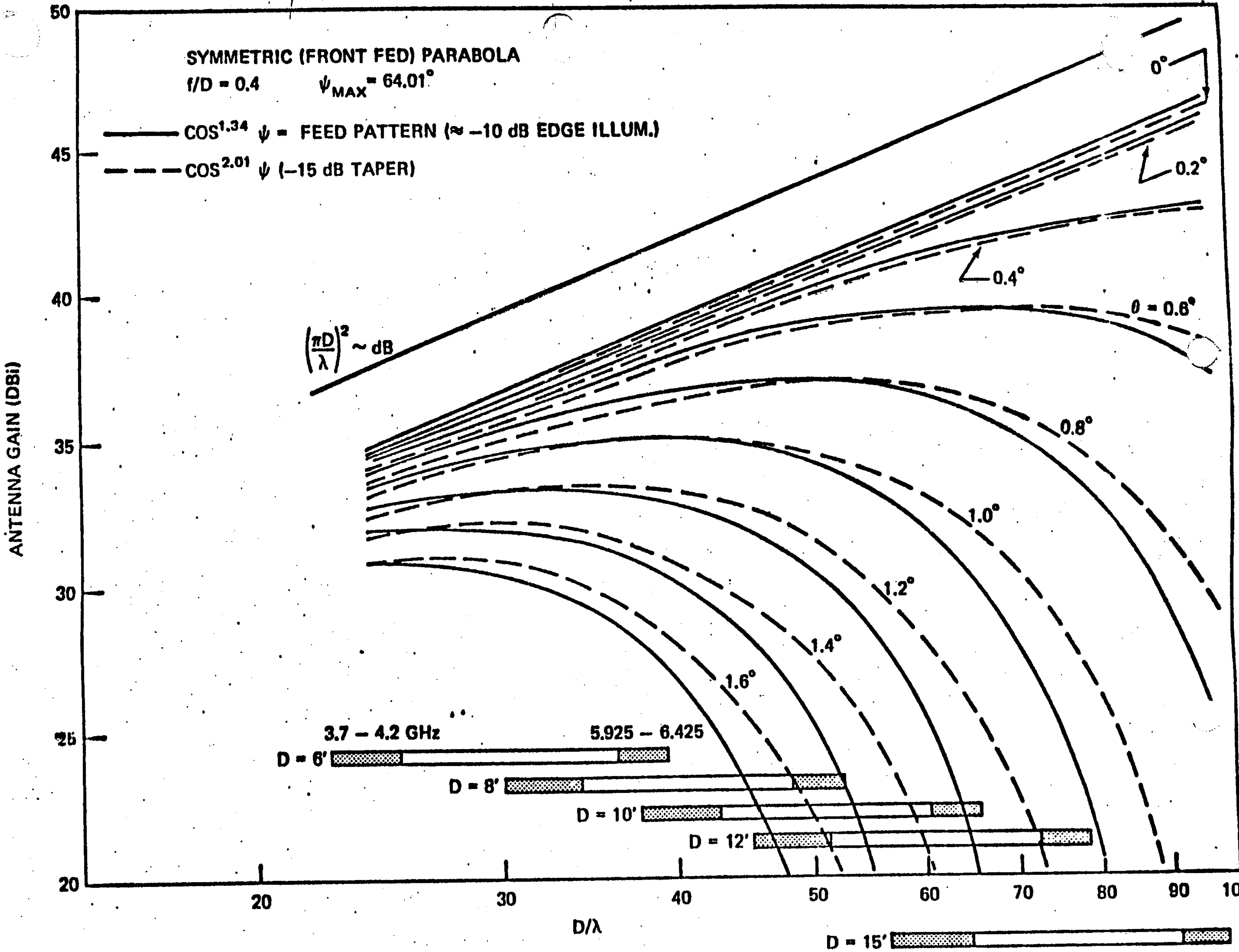



Fig. 1: ANTENNA GAIN VS D/λ AND OFF–AXIS ANGLE, θ

G–17389

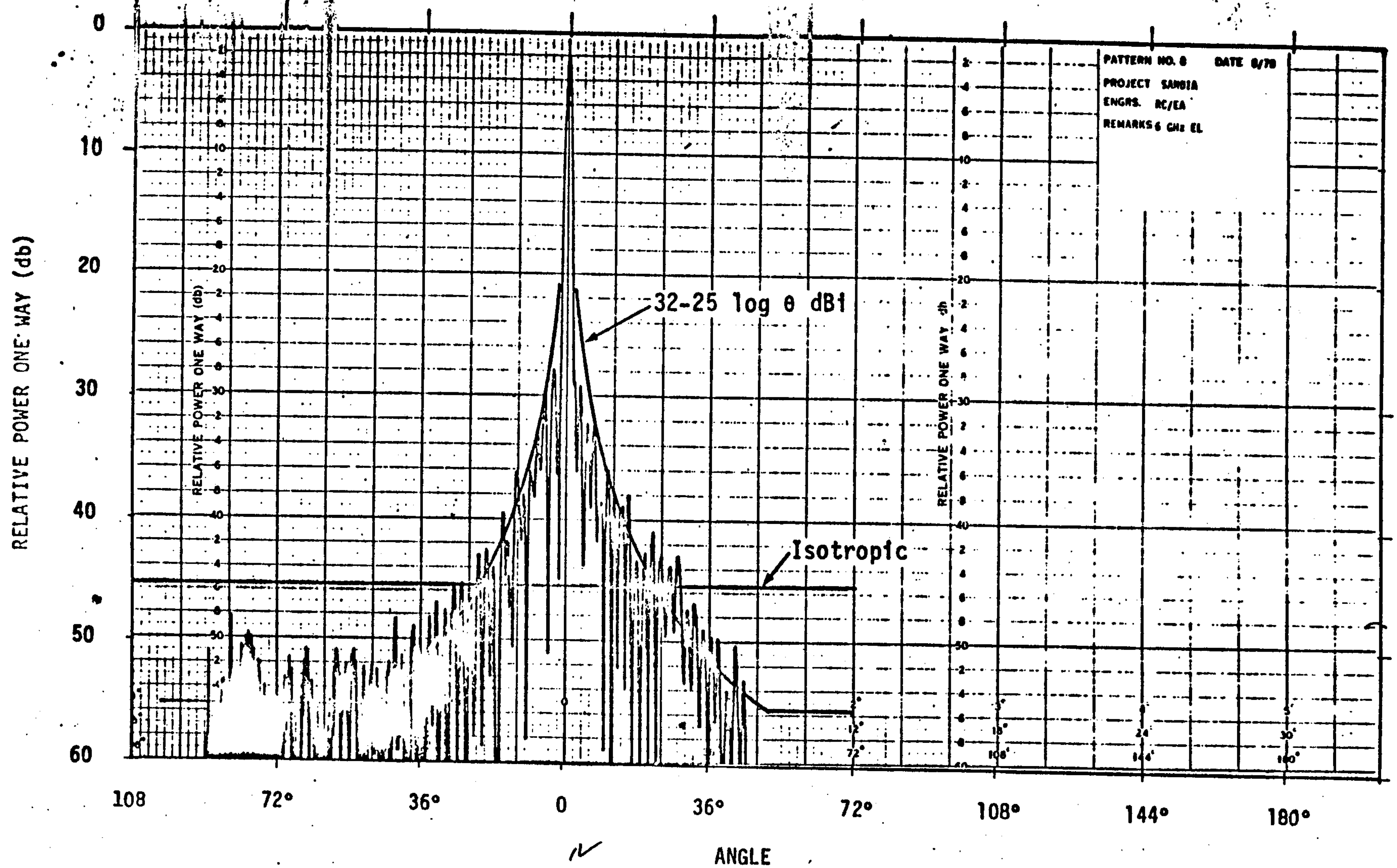



Fig. 8: Antenna Pattern: Elevation

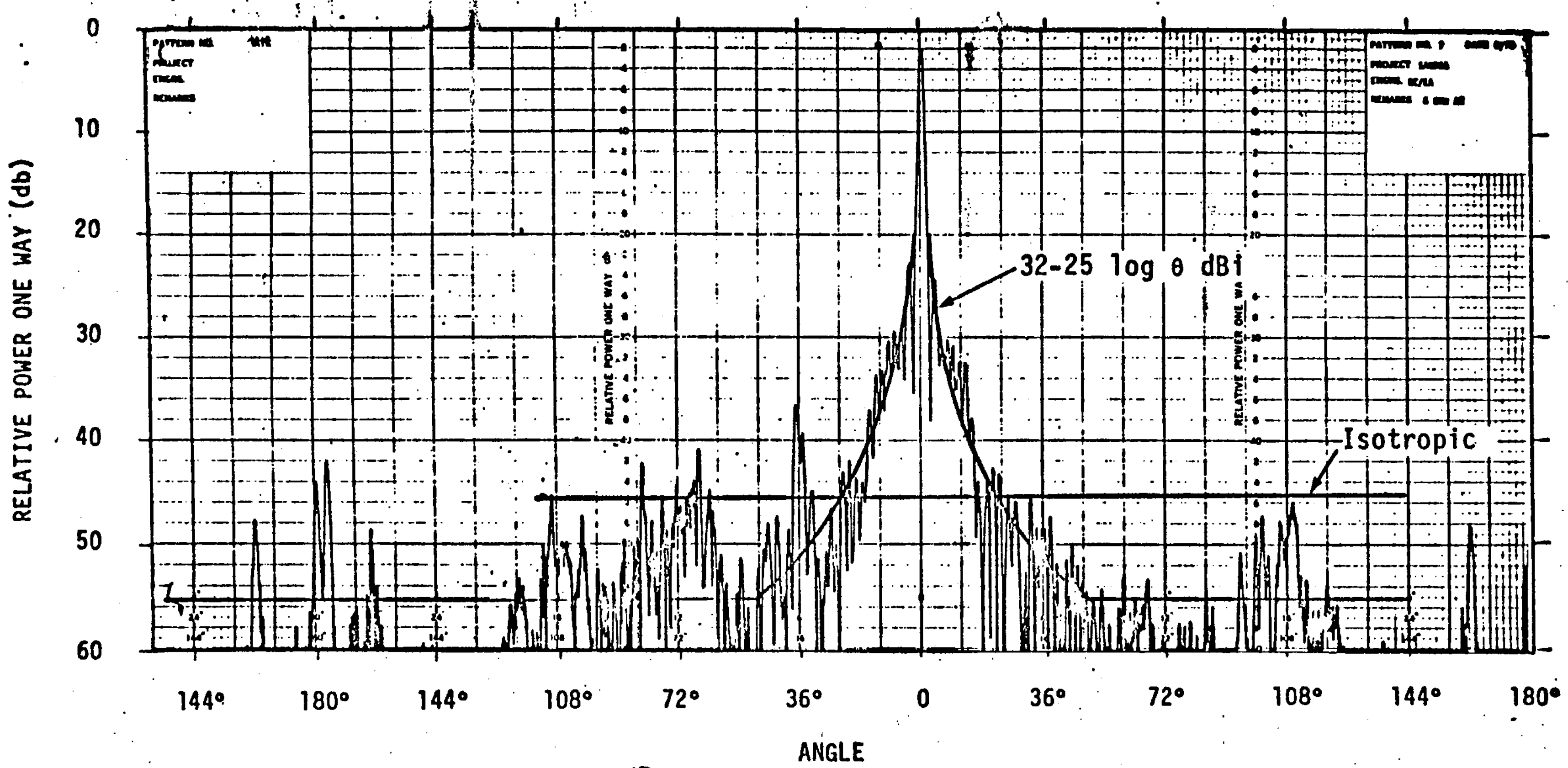



Fig. ~~10~~ 13: Antenna Pattern Azimuth

Table ~~3.2.1~~ 3.1: Specifications of A TYPICAL Feed Assembly

| | Band | |
|---|---|---|
| | 4 GHz | 6 GHz |
| Frequency | 3700 - 4200 MHz | 5925 - 6425 MHz |
| VSWR | 1.15 max | 1.10 max |
| Axial Ratio | 2.5 dB max | 2.5 dB max |
| Insertion Loss | .10 dB max | .05 dB max |
| Transmit to Receive Isolation | ----------- | 80 dB max |
| Efficiency | 60 - 65% | 55 - 60% |
| Gain (10' dish) | 39.5 - 40.6 dBi | 43.1 - 44.0 dBi |
| Waveguide Inputs | WR229 CPR | WR159 CPR |

## 4.3 HPA

The high power amplifier (HPA) provides for the final amplification stage in the earth station before entry into the antenna feed assembly. It is usually preceeded by an intermediate power amplifier (IPA) which is used to drive the HPA.

The key design factors that impact on an HPA choice are:

a. Peak Power - this represents the maximum power that can be provided out of the device.

b. Bandwidth - this is to total operational bandwidth that can be used. It should be noted that most of these devices are tuned to operate about some given frequency. The bandwidth then is determined in terms of this frequency without retuning.

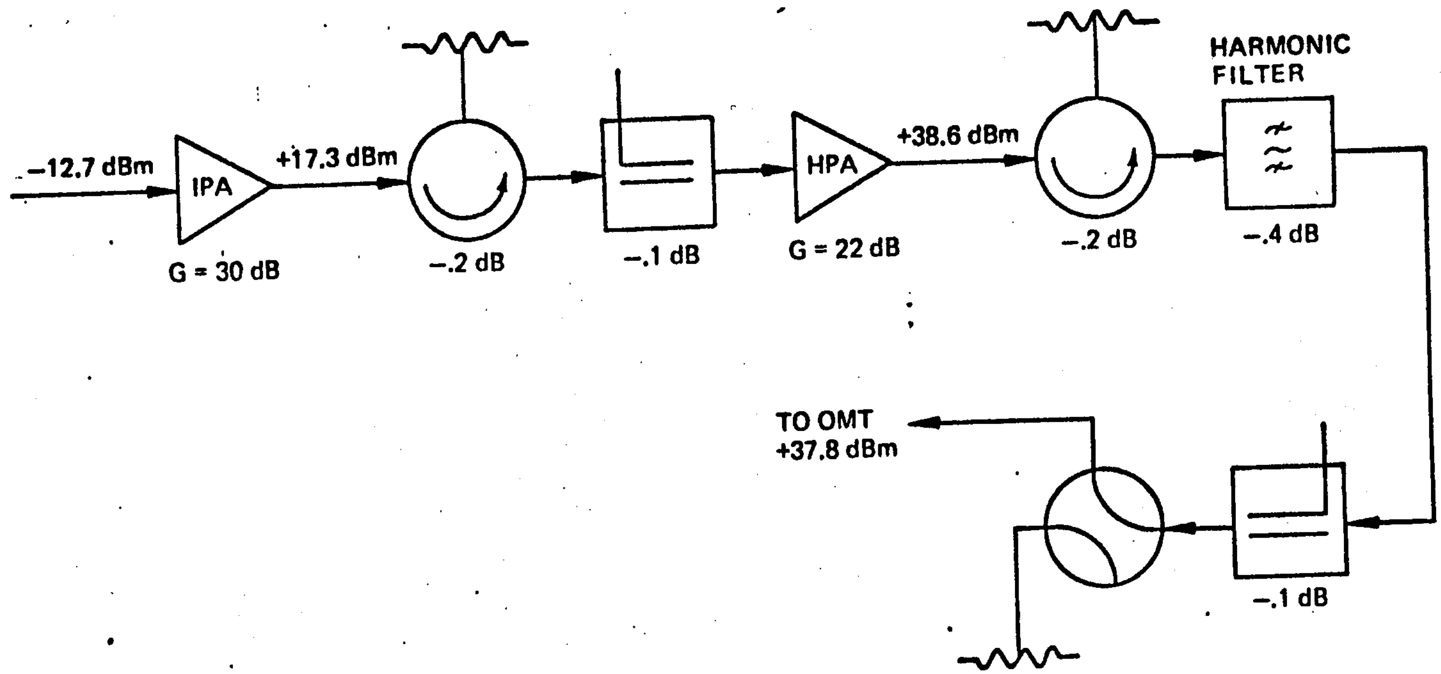
c. Efficiency - this is the ratio of RF power delivered to d.c. power required to operate the device at this point.

d. Gain - this represents the overall gain from input power level to output power level. The gain level is important since it impacts on the total number of IPA stages required. On the other hand if the gain for the HPA is too large, instabilities could occur.

e. Linearity - The more linear the device the better capable it is to carry multiple traffic loads without distortions.

f. Noise Figure - The HPA, like any amplifier, has noise that is injected into the data stream along the amplification process. This level is given in a noise figure (see LNA discussion).

We shall consider several of the more typical amplifier devices that are used. Our emphasis is directed toward briefly describing the devices and relating them to the specific performance factors.




12

Figure 12 : 6 GHz IPA/HPA SUBSYSTEM

G-17389

Galium Arsenide

There are several HPAs that can be used. They are travelling wave tube amplifiers (TWTA), Klystron, GaAs (Galium Arsenide) FETs (Field Effect Transistors), IMPATT diodes, and bipolar transistors. The first two are tube type resonators or tuned devices and the latter three are solid state.

1. TWTA - this device uses a tube to transmit the RF signal and uses a coil wrapped around its tube to modulate the RF signal with the data message. It tends to be a very wide band amplifier (500 MHz or more) that requires a minimal amount of tuning. It can provide power over a wide range, from tenths of watts to over kw. (Fig. 13)

   The TWTA has a broad range of gains, thus allowing for significant front end gain. The major problem with the TWTA is that it becomes nonlinear as the output power approaches saturation. Fig. 14 depicts the TWTA characteristics. These are plotted as a function of the ratio of input power to the saturated input power. Note first that output power increases almost linearly with input power until $P_{iN}$ nears $P_{iN,SAT}$. At this point the $P_{out}$ value does not increase as rapidly. This implies a peak gain up to this point with a sharp decrease thereafter. The efficiency on the other hand reaches a peak at or near saturation. A fourth factor is pulse shift. It remains faily constant at low input power levels but drops off rapidly at those approaching saturation.

   If a single carrier of constant envelope is used, there is no problem. However if multiple carriers are used the output mixes them to generate third order harmonics. We shall discuss these later. In addition there is an AM/PM (AM to PM (Phase modulation)) effect that converts amplitude changes into phase changes. That is, as the input power level changes by adding or subtracting new carriers, the phase changes. This then generates an

How does a TWT work?

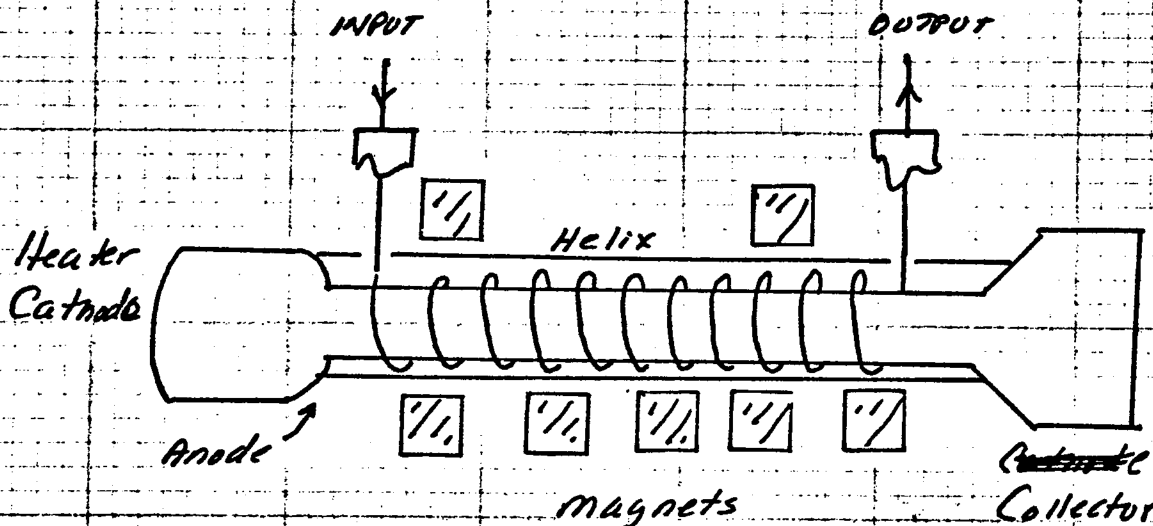



Fig 13 : TWT structure

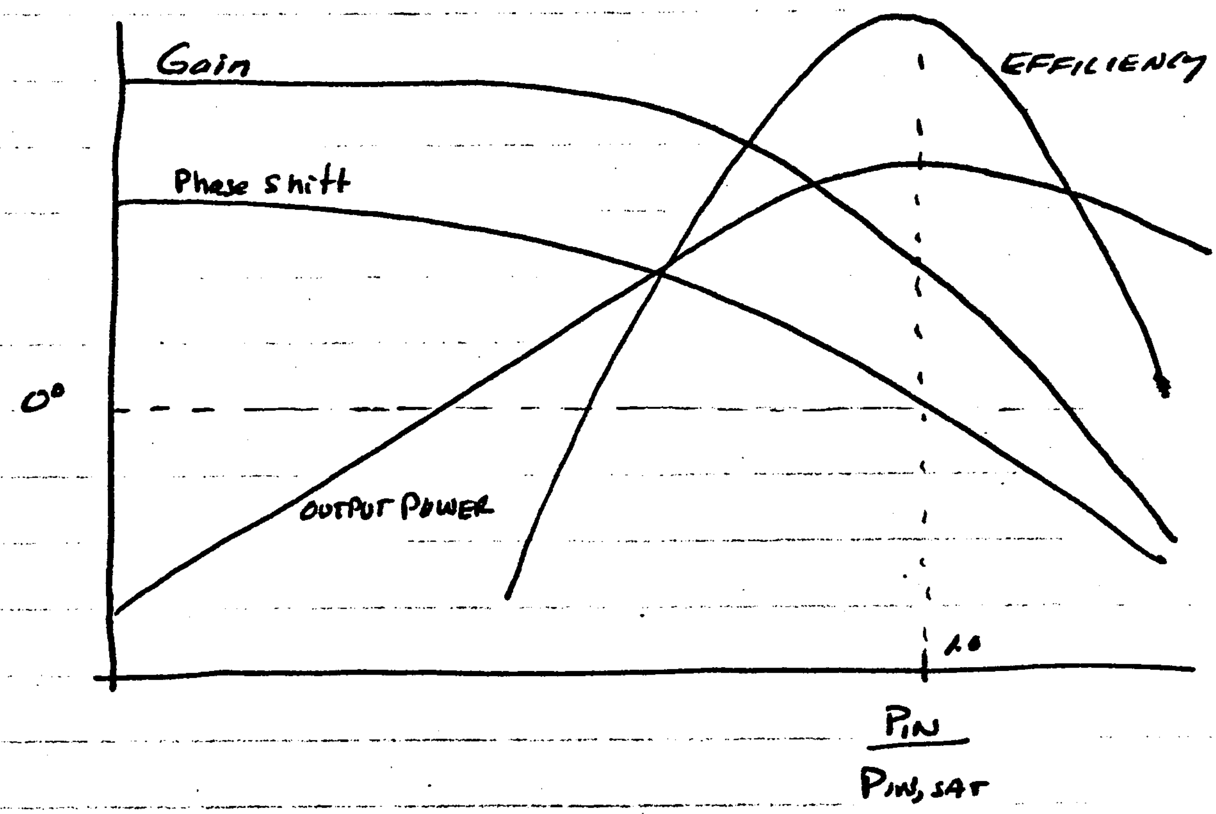



Fig : TWTA Characteristics

induced phase error onto the signal. The closer to saturation the greater the effort. The net result of these factors is an interference signal that looks like additional white noise. This is called the interference, I, and could be related to an effective interference noise temperature or noise figure. As the power approaches saturation, in a multiple carrier case, the interference increases.

operating point

Thus, the TWTA can be optimized in terms of operating point. If C is the carrier power, then $C/N_o$ is the carrier to front end noise ratio. Since $N_o$ is independent of the operating point, as we increase the power in, we increase C and thus $C/N_o$. On the other hand, as we increase the power in we increase I more rapidly than C thus C/I decreases. The total C/N is given by;

$$C/N = ((C/N_o)^{-1} + (C/I)^{-1})^{-1}$$

as we show in Chapter 6. Thus in Fig. 15 we show that there is an optimum operating point for the TWTA. optimum ⇒ maximize $\left(\frac{C}{N}\right)$!

If we use a single carrier the interference does not exist and the tube can be operated at saturation. Such is the case for some SCPC (single channel per carrier) systems and for TDMA. Thus in multicarrier operations we backoff the TWTA by 10 dB to avoid the problem.

The most useful performance measure of a TWTA is its overall efficiency that can get as high as 40 percent. This is a key factor in the use of these devices in satellites. Environmentally they can withstand various extremes but not too severe.

How much?

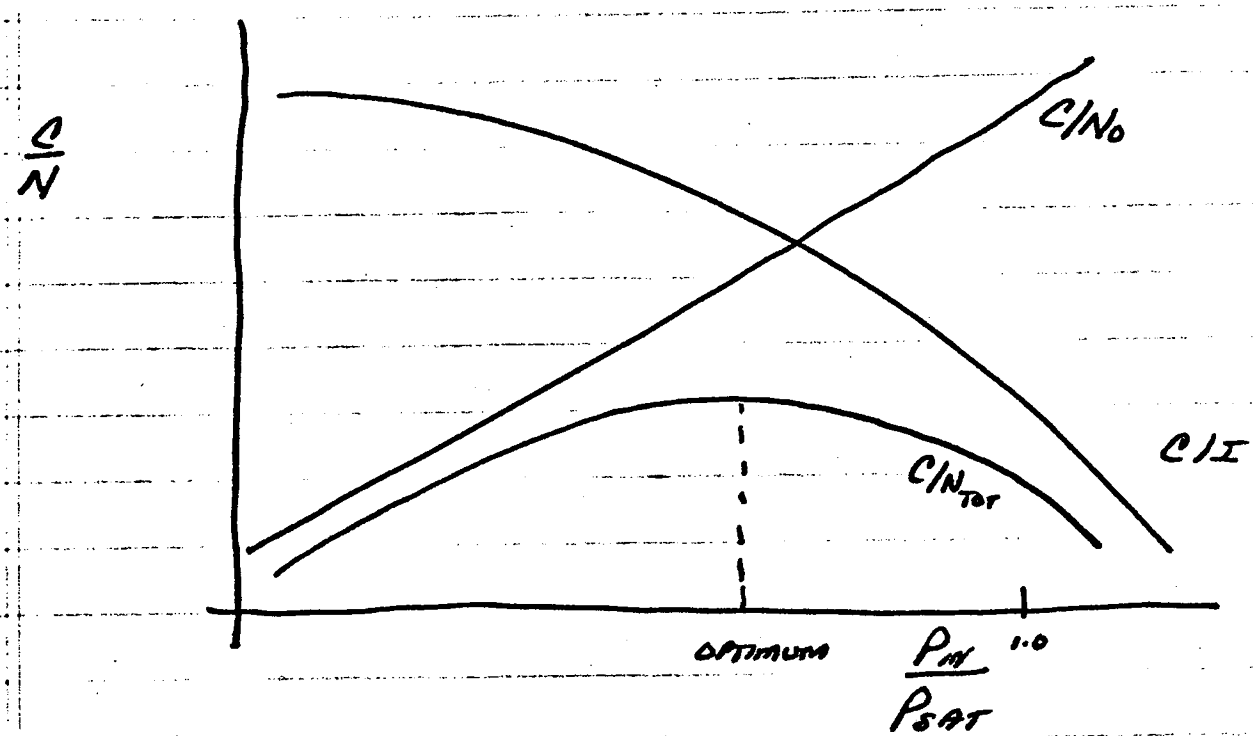



Fig 15: Optimization of TWTA Operating Point

2. Klystrons - These devices are typically used when high power and narrow bandwidth are required. The bandwidth available runs about 1% of the center frequency. This is because the device is a tuned cavity and as much has a narrow bandwidth and high Q. These devices are used in special applications and are not frequently used in most systems.

Table 2 presents a summary of available TWTA and Klystron operating a various frequencies and useful for a wide assortment of satellite communication applications.

The next selection of HPAs are solid state and represent a growing trend. Typically the solid state devices are lower powered and tend to occupy less space. In addition it is expected that they will also have a higher reliability.

3. GaAs FET - This device is a FET structure with GaAs as a substrate. The RF signal is passed through it and the low power signal modulates the high power RF signal thus providing amplification. These devices operate from 1-2 GHz and up and can on a single device deliver up to 10 W. The limiting factor is junction temperature and breakdown.

junction

The GaAs FET is fairly wide band (10% of carrier is not uncommon) and functions well at high frequencies. They have similar nonlinear distortions as do TWTAs and thus intermod is a concern.

The efficiency of these devices is quite high being 20-30%. Combined with this is a reliability that is at least an order of magnitude greater than TWTAs. Unlike a TWTA which has a natural aging effect, the GaAs FET has no such phenomenon. As such failure rates on the order of one per 500,000 hours are not unattainable.

## High-Power Amplifiers (HPAs) for 5.9 - 6.4 GHz

| Manufacturer | Tube Type | Power Level (Watts) | |
|---|---|---|---|
| Thomson-CSF | TWT | 4 | |
| Hughes, Thomson-CSF | TWT | 10-50 | 5.9-6.4 |
| LogiMetrics | TWT | 25 | |
| Varian | TWT | 35 | |
| LogiMetrics | TWT | 50 | |
| Thomson-CSF | TWT | 70 | |
| NEC, Varian, LogiMetrics, Thomson-CSF | TWT | 75 | |
| Varian | TWT | 100 | |
| Varian | TWT | 125 | |
| NEC | TWT | 150 | |
| AEG Telefunken | TWT | 200-600 | |
| Hughes | TWT | 300-600 | |
| Comtech | TWT | 300-700 | |
| NEC, Varian | TWT | 400 | |
| Varian, NEC | Klystron | 400 | |
| Varian | TWT | 600 | |
| Varian, NEC | TWT | 700 | |
| Varian, NEC | TWT | 1000 | |
| Varian | TWT | 1200 | |
| Comtech | Klystron | 1500-35,000 | |

## High-Frequency HPAs

| Manufacturer | Tube Type | Power Level (Watts) | Frequency (GHz) |
|---|---|---|---|
| NEC | Klystron | 200-300 | 27.5-31 |
| Toshiba | TWT | 300 | |
| NEC | TWT | 800 | |
| Varian | Klystron | 2000 | |
| Siemens | TWT | 5000 | |

## High-Power Amplifiers (HPAs) for 5.9 - 6.4 GHz

| Manufacturer | Tube Type | Power Level (Watts) |
|---|---|---|
| Thomson-CSF | TWT | 2500 |
| Varian, NEC, Thomson-CSF | Klystron | 3000 |
| Varian, NEC | TWT | 3000 |
| Hughes, Siemens | TWT | 3000-10,000 |
| Varian | Klystron | 3350 |
| NEC, Toshiba | TWT | 8000 |
| NEC | TWT | 12,000 |
| Varian | Klystron | 12,000 |

## High-Power Amplifiers

| Manufacturer | Tube Type | Power Level (Watts) | Frequency (GHz) |
|---|---|---|---|
| NEC | TWT | 500 | 35 |
| Varian | TWT | 30,000 | |
| NEC | TWT | 500 | 36-38 |
| Siemens | TWT | 1000 | |
| Hughes | TWT | 200 | 40-50 |
| Siemens | TWT | 400 | |
| Hughes | TWT | 17 | 60 |
| Hughes | TWT | 60 | |
| Hughes | TWT | 400 | |

## HPAs for 14 - 14.5 GHz

| Manufacturer | Tube Type | Power Level (Watts) |
|---|---|---|
| Varian | TWT | 5 |
| Varian | TWT | 25 |
| Varian | TWT | 40 |
| LogiMetrics | TWT | 100 |
| Logimetrics, Varian | TWT | 200 |
| Varian | TWT | 250 |
| NEC, Hughes | TWT | 200-400 |
| Varian | Klystron | 500 |
| Hughes | TWT | 500 |
| Varian, Thomson-CSF | TWT | 800 |
| NEC | TWT | 1000 |
| Varian | Klystron | 1000 |
| [illegible]n | Klystron | 1200 |
| [illegible]an | Klystron | 1500 |
| NEC, Varian | Klystron | 2000 |
| Toshiba | TWT | 2000 |
| Hughes | TWT | 5000 |
| Varian | TWT | 10,000 |

Table 2: TWTA's that are commercially available

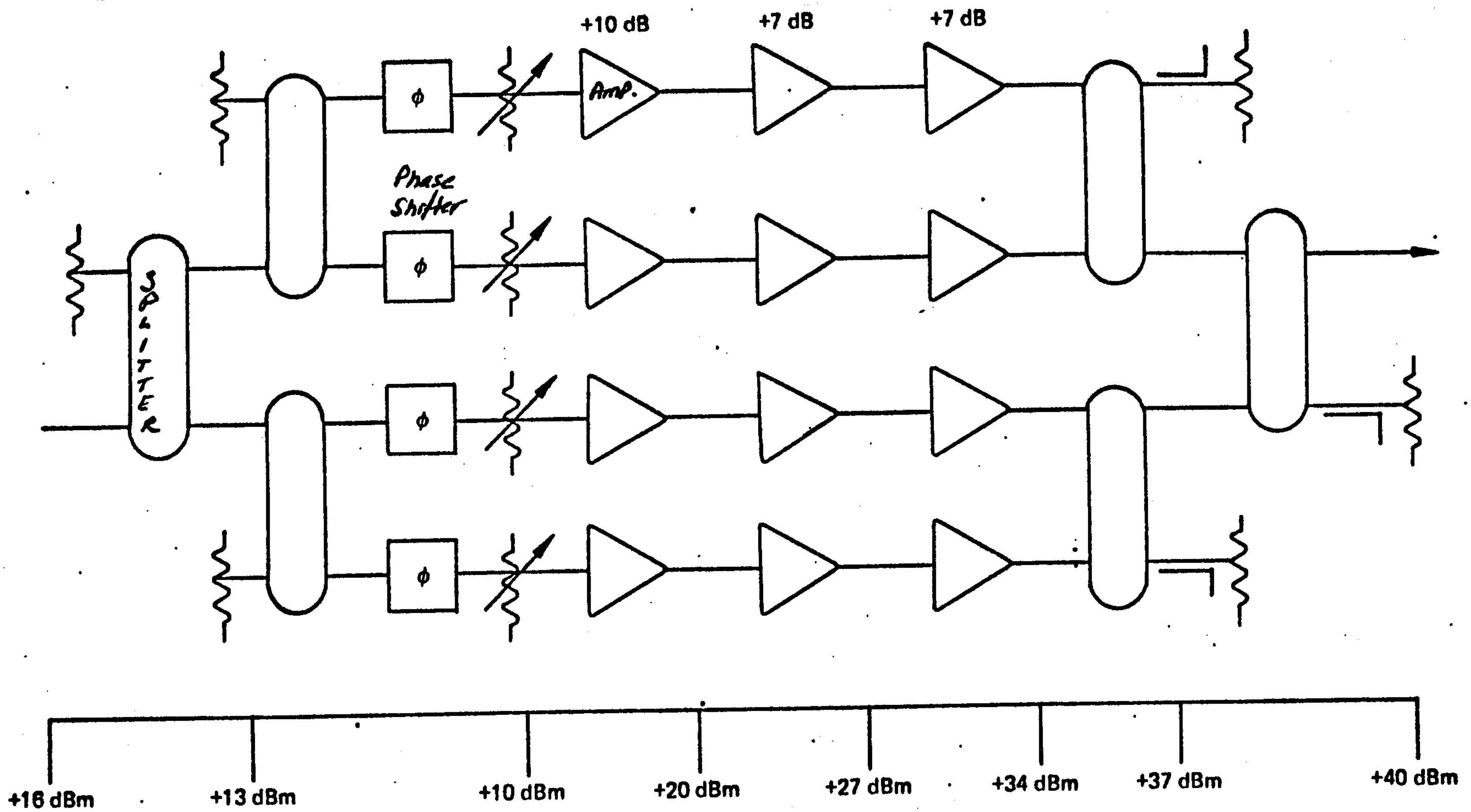
To attain higher power with GaAs FET devices it is possible to use coherent combiners to use the outputs of several devices. Figs. 15 and 16 depict two ways in which this can be accomplished. The output power level is to be 10W (40 dBM or $10^4$ above a milliwatt). Combiners or splitters are used along with phase shifters. We use 2.5W peak GaAs FET in the final stage. We must note that losses occur in the combiners and phase shift devices. The power levels throughout the circuit are shown. Large gain and output power levels are attainable with these techniques, however, the ability to control phase and stability combined with device failures could cause problems.

These devices are in a rapid state of transition and are being used as TWTA replacements in a wide set of areas. It is expected that power levels will increase more as time goes by and that more sophisticated combining techniques will be developed.

4. <u>Impatts</u> - These are the solid state equivalent of the Klystron. They have narrow bandwidth, low efficiency, moderate power, poor reliability and can be easily combined. They are generally being replaced by GaAs FETs.

5. <u>Bipolar Transistor</u> - below 4 GHz this device provides for a capability of class-C amplifiers with high efficiency, moderate power (20W) and adequate gain (6-10dB). In general, they are not used at 4GHz and above and thus would not fall within the normal satellite bands. However, they do provide options in the mobile satellite bands at L-band and below.

Table 3 presents a typical spec sheet for an HPA. The key performance factors have already been discussed.




$P_{out_{max}}$ = 2.5 WATTS/OUTPUT FET

Figure ~~9.4.7~~ 15: HPA CONFIGURATION

G–17389

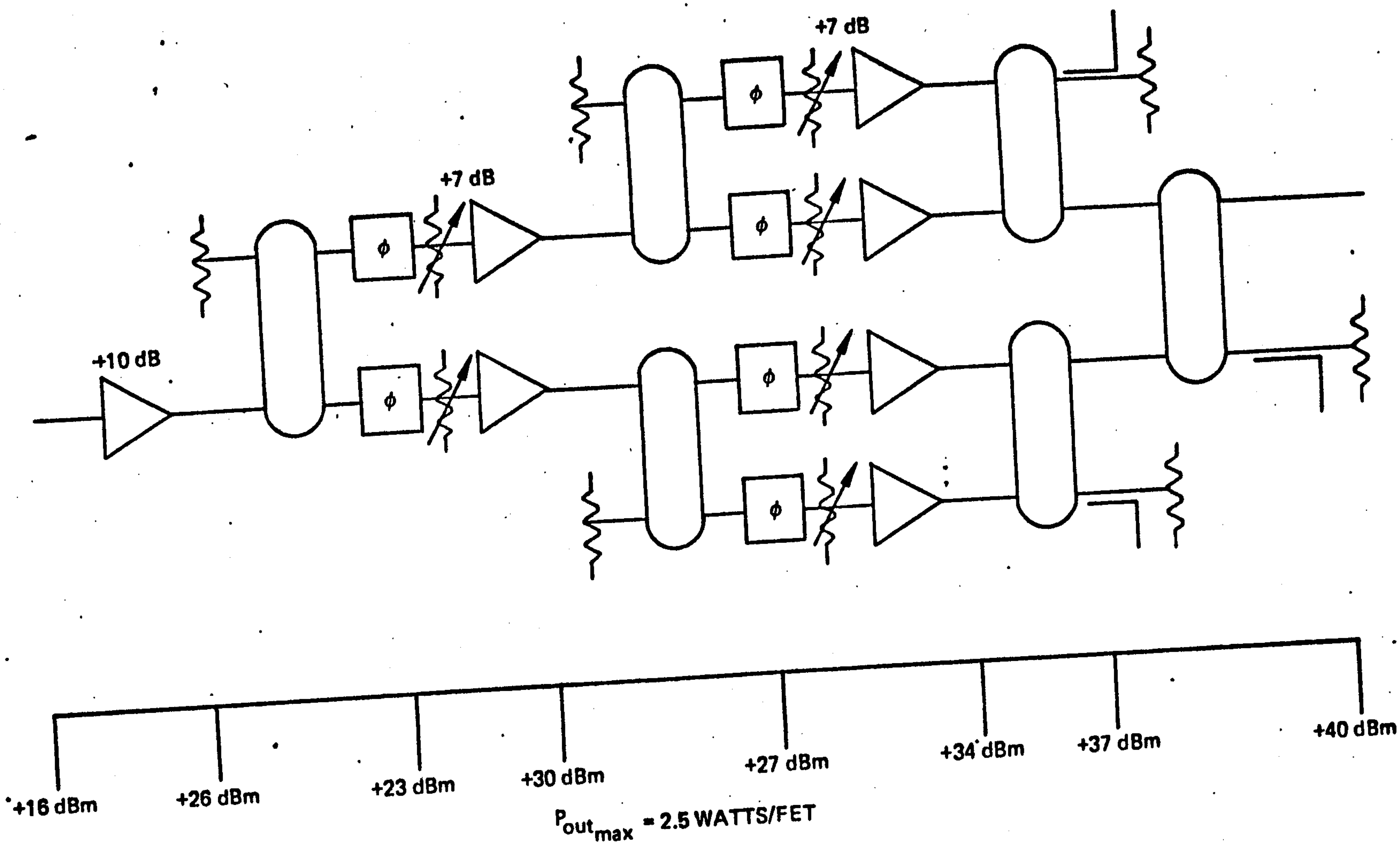



Figure ~~3.4.8~~ 16: ALTERNATIVE HPA CONFIGURATION

G–17418

Table 4

3 ~~4.2.2~~ High Power Amplifier Specification

| | |
|---|---|
| Type | ~~Solid State - FET~~ |
| Frequency | 5.925 - 6.425 GHz |
| Bandwidth (3 dB) | 500 MHz min. |
| Power Out (Saturated) | +40 dBm (10 watts) min. |
| Gain | 30 dB ±2 dB |
| Gain Flatness | ±1.0 dB max. over any 10 MHz |
| Power Stability with a constant input power | ±0.2 dB/12 hrs. max.<br>±0.5 dB/week max.<br>±1.0 dB over the operating temperature range |
| Noise Figure | 15 dB max. |
| Input VSWR | 1.5:1 max. (50 ohms) |
| Load VSWR | 1.3:1 max. (50 ohms) |
| AM-to-PM Conversion | 10° max. over an input drive level from Pin (saturated) to I3 dB below Pin (saturated) at any frequency in the band. |
| Harmonic Power | The total harmonic power generated by the amplifier shall be at least 20 dB below the carrier when the amplifier is operating in saturation (10 watts output) |
| Connectors (in/out) | SMA-Female |

| | |
|---|---|
| Efficiency | 30% min. |
| Prime Power | 30 watts max. |
| Operating Ambient Temperature Range | 0°C to +60°C |

Possible Sources:

At present there are no amplifiers of this type available on the market. Possible Device sources are:

Fujitsu
Raytheon
Microwave Semi-conductor Corp.

## 4.4 LNA

The function of the low noise amplifier (LNA) is to provide for a rapid increase in signal level while at the same time introducing a minimal amount of noise. Before discussing the options and characteristics of the devices we first will discuss the general concept of noise and its sources.

Noise is generated by several sources. Internal to the receiver itself there is noise generated by the random motion of the electrons that make up various resistances. These electron vibrations generate a noise current through the resistance which can be measured by a power meter. The noise power is measured over a bandwidth B and we call the resulting noise power $P_n$. It is found experimentally that the noise power per bandwidth B does not depend on the frequency at which the filter is centered but is constant for all frequencies. This type of noise is called white noise and is defined by a spectral density $S_n(f)$ which is

$$S_n(f) = \frac{N_o}{2} \quad (\mathrm{w/Hz})$$ two sided (all f)

Thus

$$P_n = \int_{-B/2}^{B/2} S_n(f)\,df = \frac{N_o B}{2}$$

The constant spectral height is also defined in terms of a noise temperature, T, which is a measure of the activity of the electrons. Thus

$$\frac{N_o}{2} \triangleq kT$$

where k is Boltzmanns constant which is -228.6 db/°K or $1.38 \cdot 10^{-23}$ jonle/erg. Thus front ends are defined in terms of their noise temperatures. This represents how much noise is introduced into the circuit.

There are other sources of noise that are introduced into the systems. These include cosmic noise ($T_c$), Atmospheric noise ($T_a$) and antenna noise ($T_B$) due to back-lobe and reflective sources. The first two of the previous noise temperatures depend on frequency and have a minimum in the 4 to 6 GHz range and increase above that to about 100$^\circ$K. The minimum is around 10$^\circ$K.

Another way to define the noise behavior is to use the noise figure F which is defined relative to a fixed temperature $T_o$, equal to 290$^\circ$K. That is

$$F = 1 + \frac{T_c}{T_o}$$

or

$$T_e = (F-1)\ T_o$$

Fin ? ?

where $T_e$ is the effective noise temperature. We typically represent Fin dB (10 Log$_{10}$ (l) so that a 3dB noise figure implies that $T_e$ equals $T_o$. Clearly a 0db noise figure equals 0$^\circ$K.

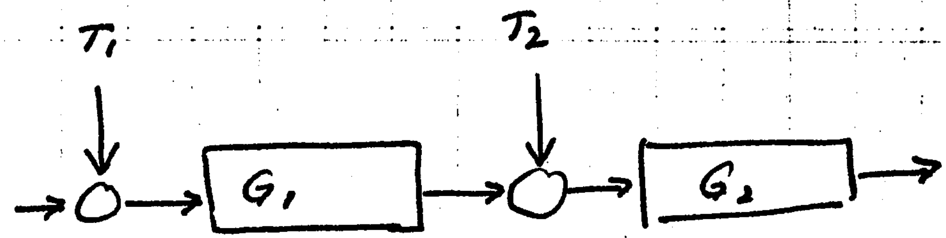
We now want to briefly consider noise in a network with gain $G_i$ or loss $L_i$ ($G_i$<1). Consider the circuit in Figure 17 a. Noise enters at two places. To reflect all the noise to the input we note that power $kT_2B$ occurs at the output of $G_1$. At the input this would be $kT_2B/G_1$. Thus the effective noise temperature is

$$T_e = T_1 + \frac{T_2}{G_1}$$

The effective noise figure is

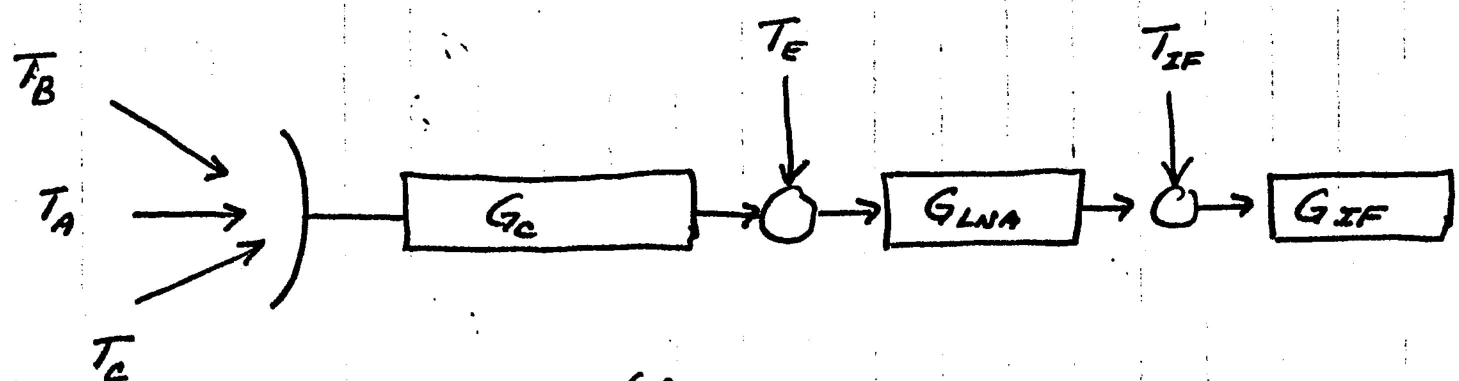
$$F = 1 + \frac{T_e}{T_o}$$

$$= 1 + \frac{T_1}{T_o} + \frac{1}{G_1}\frac{T_2}{T_o}$$




(a)

Cascaded System




(b)

Earth Station Front End

FIG 17: Noise Figure Evaluation

But we define

$$F_1 = 1 + \frac{T_1}{T_o}$$

and

$$F_2 = 1 + \frac{T_2}{T_o}$$

as noise figures of the separate devices. Therefore

$$F = F_1 + \frac{F_2 - 1}{G_1}$$

This is easily expanded. Now consider this applied to an earth station front end. Recall that $G_c$ is a cable gain but it is less than 1. Therefore (Fig. ___b)

$$T_e = T_A + T_B + T_C + \frac{T_E}{G_c} + \frac{T_{IF}}{G_c G_{LNA}}$$

Note that if $G_c$ is a large loss, it multiplies $T_E$ by a constant greater than one. On the other hand if $G_{LNA}$ is large, it reduces all other IF noises accordingly.

What this then suggests is that we want to minimize cable losses (e.g., short cables) and minimize the LNA temperature. This is usually accomplished by mounting the LNA at the feed horn of the antenna (see Fig. 18 )

The performance factors of the LNA are similar to the HPA. They are:

1. noise temperature
2. gain
3. bandwidth
4. reliability
5. frequency response.

There are basically two types of LNAs in use; paramps and GaAs FETs. The paramp is a parametric amplifier that amplifies using the nonlinear characteristics of a transducer that is pumped by a reference oscillator. This device may be cooled or uncooled.

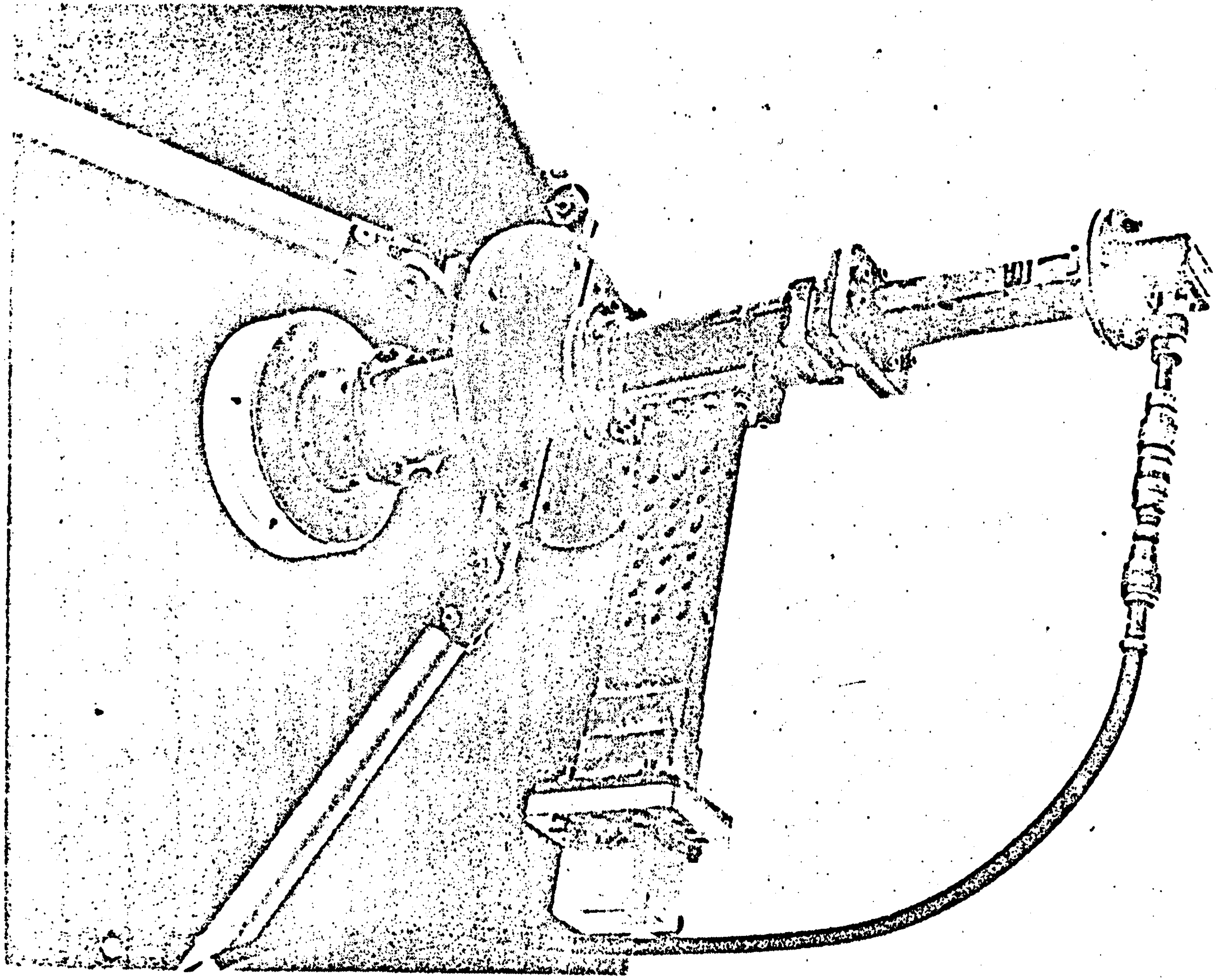


Figure ~~14~~: 18 ~~Reflector OMT~~, Close-up View LNA MOUNT

It has a relatively narrow band response and has reliability problems. It also is costly.

The $G_aA_s$FET is the now generally accepted device and has a broad set of ranges. Table 5 presents a summary of these devices. In addition Table 6 provides a typical spec sheet for the LNA system.

## Low-Noise Amplifiers (LNAs)

| Manufacturer | Type | Noise Figure (dB) Noise Temp. (K) | Frequency (GHz) |
|---|---|---|---|
| AIL | Cryogenic up-converter | 7K | 0.5-2.2 |
| AIL | Cryogenic paramp | 10-20K | |
| Alpha, Aydin | GaAs FET | 100-200K | |
| Plessey | GaAs FET | 120K | |
| Gardiner | Bipolar | .5dB | |
| Gardiner | GaAs FET | 1.2-1.5dB | |
| Amplica | GaAs FET | 1.25-2.2dB | |
| Western Microwave | Bipolar | 2.5-4.0dB | |
| Watkins-Johnson | GaAs FET | 2.5-4dB | |
| Comtech | Cryogenic paramp | 15K | 3.7-4.2 |
| AIL | Cryogenic paramp | 20K | |
| AIL, LNR | Thermoelectrically cooled paramp | 30K | |
| AIL, LNR | Thermoelectrically cooled paramp | 31-40K | |
| Comtech | Thermoelectrically cooled paramp | 30-35K | |
| GTE Tele-communicazione | Paramp | 33K | |
| GTE Tele-communicazione | Paramp | 40K | |
| Comtech | Thermoelectrically cooled paramp | 40K | |
| Comtech | Thermoelectrically cooled paramp | 55K | |
| LNR, AIL | Paramp | 45-65K | |
| LNR, AIL | Cooled paramp | 45-65K | |
| GTE Tele-communicazione | GaAs FET | 48K | |
| Amplica | GaAs FET | 80K | 3.7-4.2 |
| Comtech | Paramp | 90K | |
| Gardiner | Paramp | 75-90K | |
| Gardiner | GaAs FET | 120-150K | |
| Scientific-Atlanta, Avantek | GaAs FET | 80K | |
| Comtech | GaAs FET | 100K | |
| Alpha | GaAs FET | 120-150K | |
| Scientific-Atlanta | GaAs FET | 90K | |
| Micromega, Scientific-Atlanta | GaAs FET | 100K | |
| Aydin | GaAs FET | 150-200K | |
| Scientific-Atlanta | GaAs FET | 120K | |
| Plessey | GaAs FET | 140K | |
| Dexcel | GaAs FET | 1.12-1.30dB | |
| Systron-Donner | GaAs FET | 1.2dB | |
| Dexcel, Western Microwave | GaAs FET | 1.5dB | |
| Alpha | GaAs FET | 1.7dB | |
| TRAK | GaAs FET | 2-10dB | |
| Watkins-Johnson | GaAs FET | 3dB | 3.7-4.2 |
| Microwave Power Devices | GaAs FET | 6dB | |
| Microwave Power Devices | GaAs FET | 30dB | |

*Table 5: LNA Suppliers and Characteristics (1980)*

## Low-Noise Amplifiers (LNAs)

| Manufacturer | Type | Noise Figure (dB) Noise Temp. (K) | Frequency (GHz) |
|---|---|---|---|
| AIL | Cryogenic paramp | 10-50 K | 11.7-12.2 |
| AIL, LNR | Paramp. | 100-160 K | |
| LNR | Thermoelectrically cooled paramp | 80-200 K | |
| Comtech | Thermoelectrically cooled paramp | 80-120 K | |
| AIL | Cryogenically cooled FET amp | 120-140 K | |
| Amplica | FET amp | 280-400 K | |
| LNR | Paraconv. | 205-240 K | |
| AIL | Thermoelectrically cooled GaAs FET | 300-400 K | |
| AIL, Gardiner, Aydin | GaAs FET | 400-600 K | |
| Plessey | GaAs FET | 425 K | |
| Micromega | GaAs FET | 260 K | |
| Micromega | Paramp | 125 K | |
| Avantek | GaAs FET | 300 K | |
| Alpha Semiconductor | GaAs FET | 620 K | |
| LNR | Schottky-barrier mixer | 600-800 K | |
| TRAK | GaAs FET | 2-10 dB | 11.7-12.2 |
| Systron-Donner | GaAs FET | 3.5 dB | |
| Alpha | GaAs FET | 5 dB | |
| Watkins-Johnson | GaAs FET | 5.5-6.5 dB | |
| Aydin | GaAs FET | 4.0 dB | |
| LNR | Paramp | 200-400 K | 17.7-20.2 |
| LNR | Schottky-barrier mixer | 600-1000 K | |
| Systron-Donner | GaAs FET | 5 dB | |
| LNR | Paramp | 400-600 K | 34-37 |

Table 5 (contd)

Table 6

~~xxxxx~~ LNA Specifications

| | |
|---|---|
| Frequency Range | 3.7 - 4.2 GHz min. |
| Bandwidth | 500 MHz min. |
| Noise Figure at 50°C | 1.7 dB max. |
| Gain | 27 dB min. |
| Gain Flatness | ±0.5 dB max. |
| Gain Slope | 0.01 dB/MHz max. |
| Power Output (1 dB gain comp.) | +5 dBm min. |
| Intercept Point | +15 dBm min. |
| Gain Stability | ±0.2 dB/12 hrs.<br>±0.5 dB/week<br>±1.0 dB over the operating temperature range |
| Input VSWR | 1.3:1 max. |
| Input Connector | CPR 229 Flat Flange |
| Output VSWR | 1.5:1 max. |
| Output Impedance | 50 ohms |
| Output Connector | Type N - Female |
| Input Power | +15 VDC ±0.5 VDC<br>1.2 watts max. |
| Operating Temperature Range | -72°C to +50°C |
| Storage Temperature Range | -80°C to +60°C |

Unit and connectors shall be hermetically sealed.

| | |
|---|---|
| Possible Sources: | Amplica Inc., Westlake Village, California<br>Avanetk, Santa Clara, Calif.<br>Watkins-Johnson, Palo Alto, California |

### 4.5 Up/Down Converters

The up/down converter system is used to translate to or from the IF frequency to carrier. The system is comprised of three major elements. They are: (see Fig. 18)

1. Oscillator - this is the primary frequency reference source that generates a constant frequency.

2. Multiplier - this device takes the reference tone and converts it to the carrier frequency level. This function can be performed in a variety of ways, but we will represent it as a direct multiplication.

3. Mixer - in this unit the input frequency is multiplied by the reference carrier to produce the desired output center frequency. We will include in this element any filters needed for sideband rejection.

In this section we shall discuss the key systems issues to select the proper up/down converter systems. We shall not discuss the detailed technology nor shall we discuss options such as dual conversion or direct down conversion on receive. We shall however provide the references to understand these options. Our major interest is in developing an understanding of how this sub-system will impact communications systems performance.

The oscillator is most commonly generated using a crystal in a feedback circuit. Fig. 19 a depicts the crystal (XTAL) placed in a feedback loop and the output is used to provide the reference signal. The equivalent circuit of the crystal is shown in Fig. 19b. This is a high Q circuit that oscillates in a very narrow range. The crystal is usually a quartz material that is coupled to the electrical circuit. The center frequency of the crystal is chosen by cutting the crystal to a certain size. The crystal has several modes of vibration and by filtering a desired one is chosen.

The output voltage of the oscillator can be modelled as:

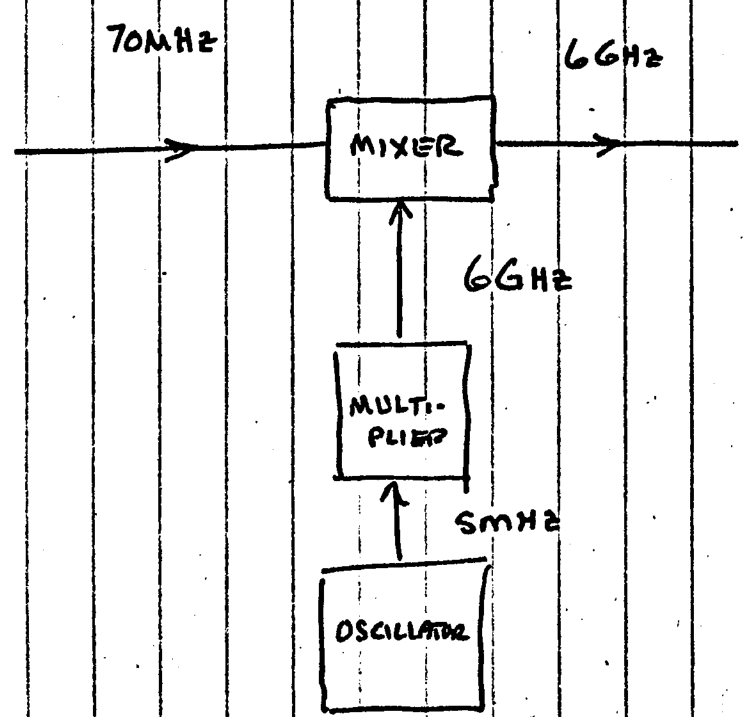
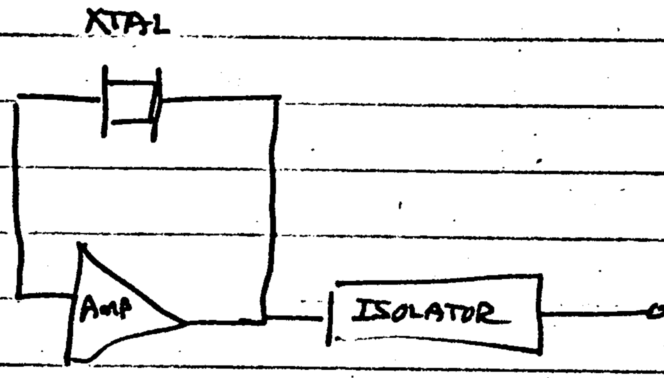
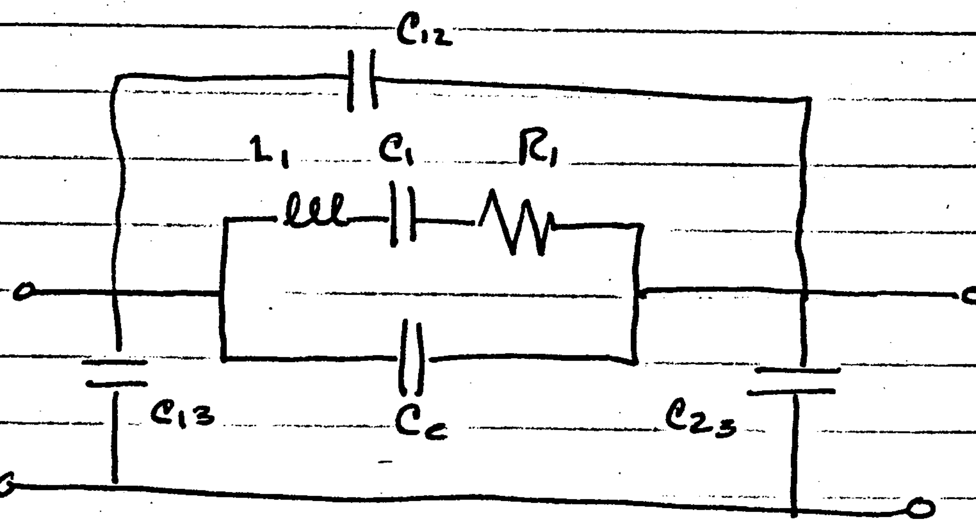



FIG 1f: UP CONVERTER ELEMENTS




(a)

Oscillator Circuit




(b)

Crystal (XTAL) Equivalent Circuit

Fig 19: Crystal Oscillator

$$C_o(t) = \cos\,(2\pi f_o t + 2\pi(\Delta f)t + \phi(t))$$

Here $f_o$ is the center frequency, $\Delta f$ is the drift rate and $\phi(t)$ is called phase noise. A common problem with such oscillators is drift represented by $\Delta f$. This results from aging or temperature effects. We typically represent it in parts per million per month (year,etc.). In an unattended mode, these oscillators will slowly move from their desired center. Careful selection and aging will help minimize this effect.

The second factor in error is the phase noise which is a random process. The phase noise term is a zero mean random process which is usually characterized by its power spectrum ($S_\phi(f)$). Here

$$S_\phi(f) = \int_{\infty}^{\infty} R_\phi(\tau)\exp\,(-j2\pi f\tau)\,d\tau$$

where $R_\phi(\tau)$ is the correlation function given by;

$$R_\phi(\tau) = E[\phi(\tau)\phi(t+\tau)]$$

The phase noise ~~spectrum~~ spectral density typically appears as in Fig. 20. Here we have five basic regions. The first three relate to frequency since the frequency noise spectrum is

$$S_{\dot\phi}(f) = (2\pi f)^2 S_\phi(f)$$

They are called random walk, flicker and white frequency noise and decrease proportionate to $f^{-4}$, $f^{-3}$, $f^{-2}$, respectively. The last two levels and flicker and white phase terms. Physically such a spectrum has a problem since it integrates to infinity. However in practice, down even to tenths of Hz, this pattern holds. What may change are the intercepts and locations of slopes, but the general shape is consistent. The frequency in Fig. 21 represents the offset from the carrier. Also the figure is symmetric about the carrier. In addition the reference levels are always referred to the carrier in terms of power density.

The multiplier circuit provides for conversion of the oscillator output up to the desired corner frequency. This is done by a multiplication factor m such that the multiplier output is

$$C_m(t) = \cos\,(2\pi m f_o t + m\,(2\pi\Delta f t)t + m\dot\phi(t))$$

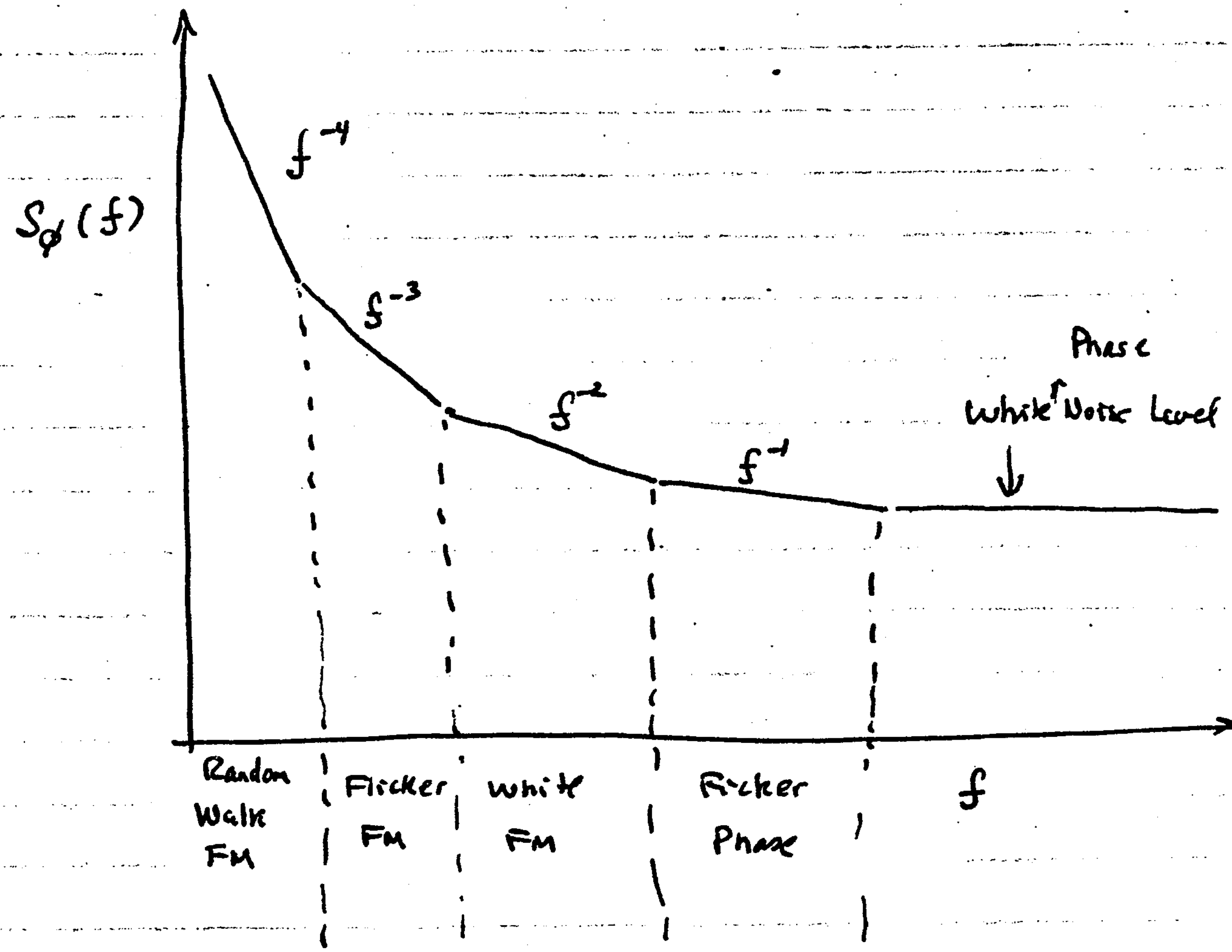
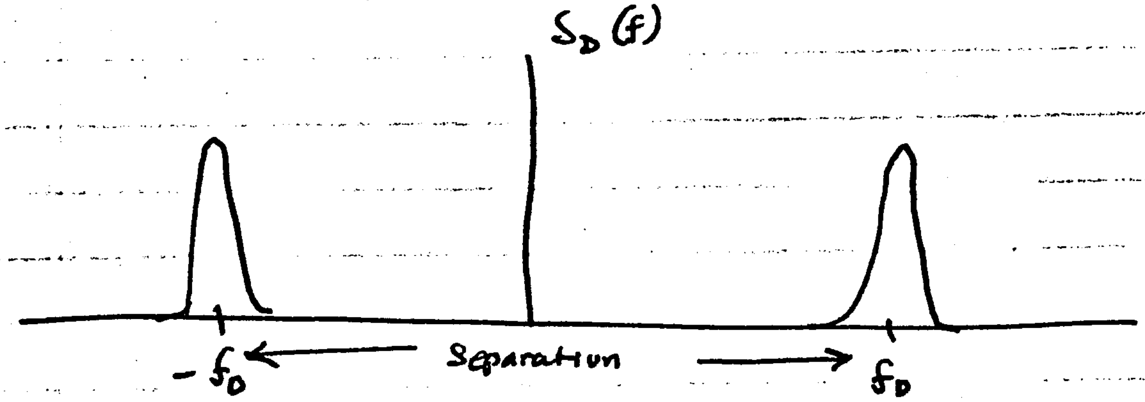
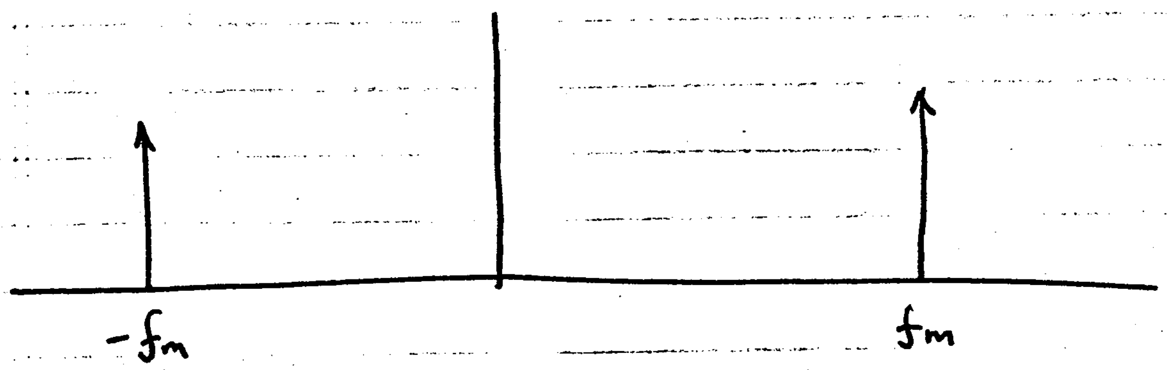



Fig 20 : Phase Noise Spectrum



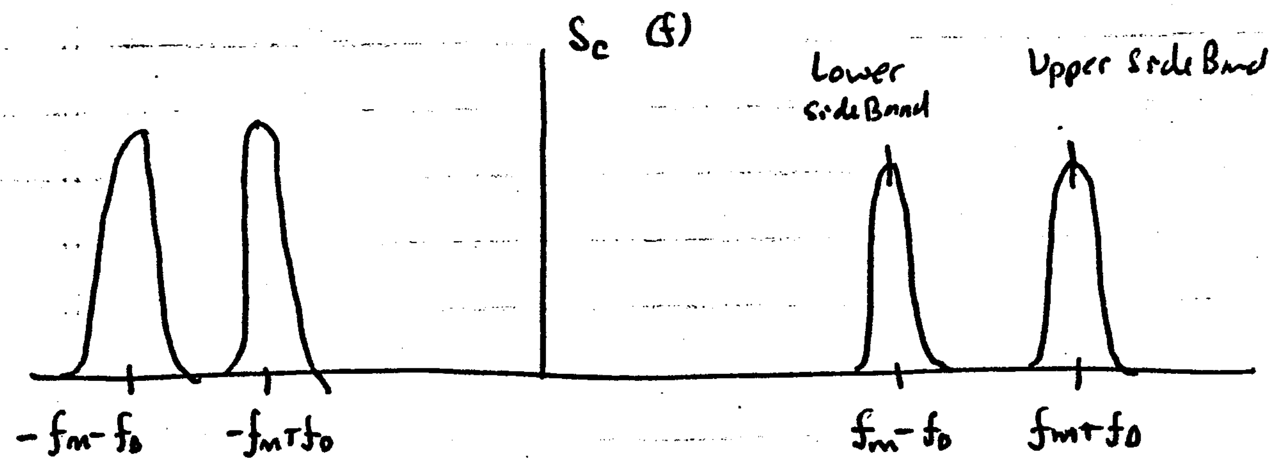

(a)

Data Spectrum (IF)




(b)

Carrier Spectrum




(c)

Carrier Spectrum

Fig 21 : Multiplier Effects

Note that this multiplication also increases the absolute drift (however the relative drift is constant) and it multiplies the phase noise, or its spectrum by $m^2$. Thus by choosing a low noise LO that may be at a low LO frequency, when multiplying we may signficantly increase its effect.

The mixer is used to multiply the oscillator signal with the carrier signal on down convert or the IF signal on up convert. Fig. 22 depicts a typical down convert operation.Note that the multiplication results in four spectra, two on each side. These are called upper and lower sidebands. One of these must be chosen to be used. The other is eliminated by filtering at the RF level after the mixer. Note that if the upper sideband is used then the center frequency is $f_D = f_M$ where $f_o$ is the IF center and $f_m$ is its multiplier center.

The down-convert spectrum is shown in Fig. 23. Here we assume $f_m$ $f_o$ so that $f_m - f_o$ is in the IF band. The upper side band is at $f_m + f_o$ which is twice the RF carrier and is filtered out. Also note that this mixer splits power into upper and lower sidebands thus resulting in a 3dB loss.

Table 7,9 and Fig. 24 presents a typical set of specifications for a down-converter assembly. The phase noise spectrum provided is a mask for a 5MHz reference crystal oscillator.

We now want to briefly analyze the impact of the phase noise on the overall system performance. To do this recall that in our discussions of modulation we found that with PSK, the error probability was given by

$$P[E|\tilde{\phi}] = \tfrac{1}{2}\,\mathrm{erfc}\left[\sqrt{\frac{E_b}{N_L}}\cos\tilde{\phi})\right]$$

note: assumes that phase error is constant over a bit time!

where $\tilde{\phi}$ is the phase error. Here we are assuming that the phase in remaining fairly constant per bit. Also erfc is a short hand for the error co-function in terms of the Q functions developed in Chapter 3. Now if $\tilde{\phi}$ is a random variable, which it is, we have

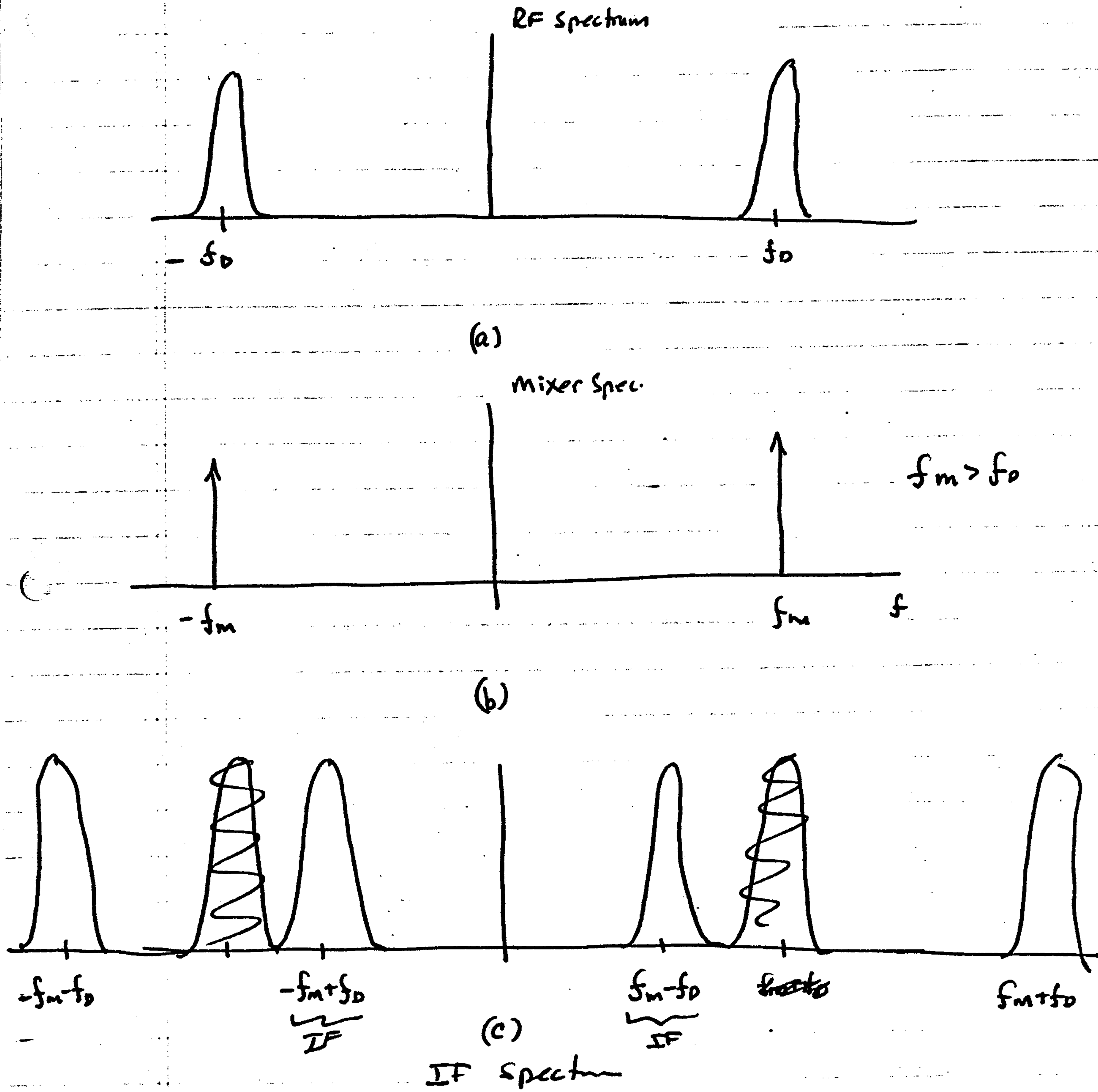



Fig 22: Down Converter Spectra

## Table 7 ~~3.3.2.1~~ Local Oscillator Specification

| | |
|---|---|
| Type | Ovenized crystal controlled |
| Frequency | 4032.2925 MHz |
| Frequency Adjustment | $1 \times 10^{-6}$ minimum adjustment range shall be provided. The adjustment range shall be sufficient to maintain the specified frequency and performance for a period of at least one year. The control shall be of the grounded type to minimize the effects of a metallic tuning tool. |
| Accuracy | The frequency shall be accurate to within $\pm 1 \times 10^{-9}$ after warm-up. Warm-up time shall be defined as one hour after initial turn-on. |
| SSB Phase Noise | See Figure 3.3.8 |
| Frequency Stability | $\pm 1 \times 10^{-10}$/sec.<br>$\pm 1 \times 10^{-10}$/hour<br>$\pm 1 \times 10^{-9}$/24 hours<br>$\pm 4 \times 10^{-7}$/year<br>$\pm 2 \times 10^{-8}$ over operating temperature after warm-up |
| Output Level | + 10 dBm minimum |
| Output Level Stability | ± 1.0 dB over the operating temperature range after warm-up<br>± 0.5 dB/24 hours<br>± 1.0 dB/year |
| Output Impedance | 50 ohms with a 20 dB minimum return loss |
| Output Connector | SMA Female |
| Load VSWR | 1.5:1 min. |
| Spurious Outputs | All harmonics and subharmonics at the output shall be -35 dBc min.<br>All spurious signals within ±20 MHz of the desired frequency shall be -80 dBc min. |

| | |
|---|---|
| | All other spurious signals shall be -60 dBc min. |
| Operating Ambient Temperature (without degradation) | 0 to +60°C |
| Prime Power | + 15 VDC ± 0.5 VDC<br>10 Watts max. |
| RF Monitoring | (a) Frequency output jack to enable a counter to monitor output frequency and power when unit is in operation.<br>(b) Phase lock adjustment monitor if phase locked source used. |
| Humidity | 10% RH to 98% RH |
| Altitude | 0 to 10,000 ft. |
| Mechanical | Units shall be hermetically sealed |
| Possible Sources: | Frequency Electronics Inc.<br>New Hyde Park, New York<br><br>Frequency Sources<br>Massachusetts<br><br>Frequency and Time Systems<br>Danvers, Mass.<br><br>Erie Frequency<br>Pennsylvania<br><br>Vectron Labs<br>Norwalk, Conn. |

Table 8

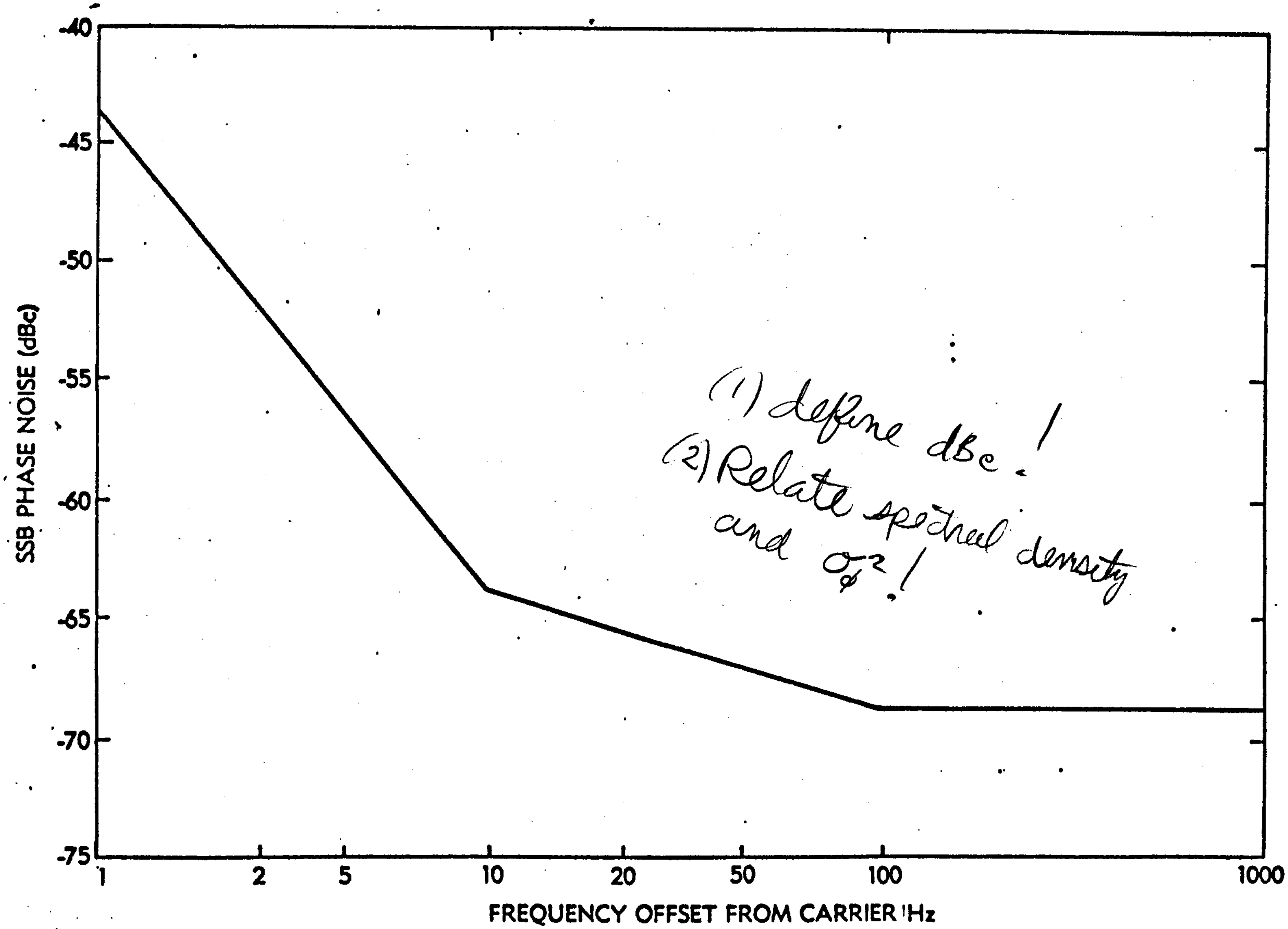
## ~~3.3.2.~~2 Mixer Amplifier Specification

| | |
|---|---|
| RF Frequency | 3.7 - 4.2 GHz |
| IF Frequency | 70 MHz |
| IF Bandwidth | 20 MHz min. |
| Noise Figure | 8 dB max. |
| Gain (RF-IF) | 25 dB min. |
| VSWR (Input) | 2.5:1 max. |
| L.O. Injection Level | + 10 dBm $\pm$ 3 dB |
| IF Output (1 dB Gain Compression) | 0 dBm |
| Impedance All Ports | 50 ohms, SMA Female Connectors |
| Operating Ambient Temperature Range without Degradation | $0^{o}C$ to $60^{o}C$ |
| Prime Power | +15 VDC $\pm$0.5 VDC<br>3 Watts Max. |
| Possible Source: | RHG Laboratories<br>Deer Park, New York |

3.3.2.4 70 MHz IF Bandpass Filter

| | |
|---|---|
| Frequency | 70 MHz |
| Bandwidth (3 dB) | 700 KHz nom. |
| Insertion Loss at $f_o$ | 6 dB max. |
| Rejection: $f_o \pm 1.5$ MHz | 40 dB min. |
| $f_o \pm 2.25$ MHz | 60 dB min. |
| Impedance | 50 ohms, SMA Female Connector |
| VSWR at $f_o$ and 50% of 3 dB BW min. | 1.5:1 max. |
| Operating Ambient Temperature (without degradation) | $0^oC$ to $60^oC$ |
| Possible Source: | CIRQTEL Kensington, Maryland |

Figure ~~3.4.2~~ 2.31: PHASE NOISE, 6 AND 4 GHz LOCAL OSCILLATORS



$$P[E] = \int P[E/\tilde{\phi}]\, p(\tilde{\phi})\, d\tilde{\phi}$$

where here $p(\phi)$ is the probability density of the phase error.

Recall also that we used a phased-locked loop (PLL) to track these phase errors. It has been shown by Warner that given a phase noise spectrum, an optimal PLL can be designed to minimize the phase ~~inn~~ error and that the density of the phase ~~inn~~ error is   who?

$$p(\tilde{\phi}) = \frac{\exp(\gamma \cos\tilde{\phi})}{2\pi I_0(\gamma)}$$

This is only correct for a first order PLL and approximate for a second order PLL
$\tilde{\phi} \in (-2\pi, 2\pi)$ and zero otherwise

Bessel

$I_0(\gamma)$

where $I_0(\gamma)$ is a modified Bessel function and $\gamma$ depends on the slope and breakpoints of $S_\phi(f)$. As $\gamma$ goes to ~~zero~~ $\infty$ $p(\tilde{\phi})$ goes to an impulse at zero and thus gives the case for no phase noise. when $\alpha \to 0$ $p(\phi) \to 1/2\pi$ A similar result holds for other modulation forms. Thus a loss in performance results when phase noise is present. The loss can be calculated. It is a higher loss at lower frequencies in most cases since the phase noise spectrum has more power there. Thus low bit rate transmissions with satellites must be carefully analyzed.

It also depends on loop transfer function!

Author look at Lindsey and Simon or Holmes and clean this section up!

One final problem area arises and that is the problem of tracking long term drifts. As we noted the frequency changes so many parts per mission per year. This can cause a drift from the center frequencies and can thus result in a skewing of the data out of the IF band of the demodulator. This problem is circumvented by using a frequency pilot loop, where a very stable reference oscillator (not a crystal) is used and checked periodically. The stations can then use this to offset their center frequencies. A similar problem could occur with the satellite offset oscillator which would also need drift compensation.

## 4.6 Example Systems

In this section we will go through several examples of earth station design and configuration. The purpose is to tie together all the elements that we developed thus far.

Consider the earth station block diagram in Fig. 25. This was the structure that we used at the beginning of the chapter to introduce the basic concepts. We now want to go through this design in detail. Let us first consider the transmit leg. The design calls for a 2Ø-PSK modulation scheme and requires an eirp (effective isotropic radiative power) of 52.5 dB W. Note that eirp equals

$$\text{eirp} = 10\text{Log}_{10}G_E + 10\text{Log}_{10}P_E$$

where $G_E$ is the antenna transmit gain and $P_E$ the power. Note also that these gains and powers are effective values which means that losses and less than perfect efficiencies have been considered. Also note that dBw is watts in dB and dBm is milliwatts in dB. We are using a 10' antenna with a gain of 42.5 dB at 6GHz the transmit frequency and 39.5 dB at 4GHz the receive frequency.

This design uses an INTELSAT satellite so that left and right handed polarization are used and the use of a polarizer is necessary behind the antenna. This device is lossless Preceeding this is an orthomode transducer which is used to split signals. This device works as follows. We number the parts in a clockwise fashion from 1 to 3. What goes in 1 comes out 2, in 2 out 3, and in 3 out 1. In general this is lossless. However, what goes into port 1 may leak out port 3 but attenuated 50 dB. As we shall see on the receive side, this must be further attenuated.

We are now at the point of feeding 40 dBm (10dBw or 10 watts) with the OMT. This occurs over the IPA/HPA subsystem. Here we use two 6 GHz amplifiers, the final stage having an output of 41 dBm. It also has a gain of 26 dB so the input level is (41-26) dBm. We also use isolators (circulators) to provide for isolation or impedance matching. These are the devices with the arrows. They typically have losses of 0.3 dB. A third device, a power probe is used to monitor the power levels.

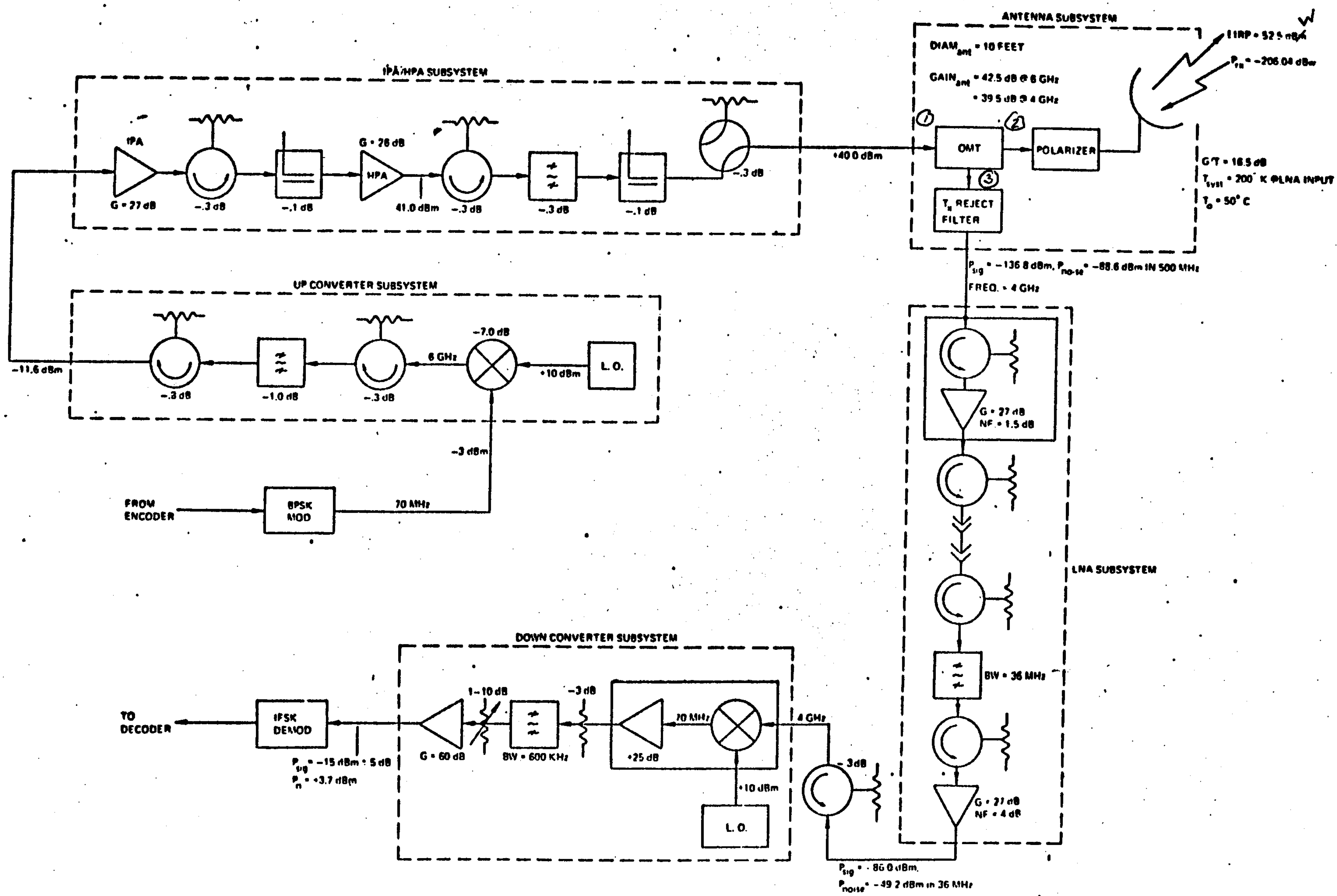



Figure ~~3-1~~ 2-8. Remote Terminal Block Diagram

The IPA has a gain of 27 dB so that the input to it is only -11.6 dBm. Note that we used a filter following the HPA to filter out any extraneous harmonics. We are using this system for only one channel.

We now concentrate on the up converter systems. The signal level from the modem at IF (70 MHz) is at -3dBm. The L.O. level is 10 dBm. These two enter the mixer and the mixer suffers a 7 dB loss at 6 GHz. A filter is used to select the upper sideband and it has a 1 dB loss. It is between two isolators to ensure no mismatch. Again they each have 0.3 dB loss.

This describes the typical uplink chain. The logic used requires first to start at the antenna output with the required eirp. Then go back to the modem and determine the IF power level. You must then provide enough gain to match the difference. A careful choice of LO level is also essential to ensure that saturation in the mixer does not occur.

On the receive side the analysis becomes a bit more complex because now we must account for signal and noise. Recall that the noise power in dBw is given by

$$10\mathrm{LOG}_{10}\ kT_eB$$

where B is the bandwidth of the appropriate sections. For example at RF, B is 500 MHz, so that the noise power is quite large. In this example the data bandwidth is only 100 Hz so that we will eventually filter all of that noise out.

On receive the receive signal is -206.0 dBw. The LNA noise plus all others reflected to the front end are 250$^\circ$K. In the full receive band of 500 MHz this yields a total noise power of -88.6 dBm.

The antenna gain is 39.5 dB on receive so that the effective receive power is -206.0 dBw + 39.5 dB or -166.5 dBw or -136.5 dBm. Note that the transmit reject filter is used to ensure that no 6 GHz signals come through the OMT. If this were not there then these signals could be at the -10 dBm level which would saturate the LNA front end. The important thing to note now is that the noise power is 50 dB higher than the signal. Yet remember that the data bandwidth is only $10^2$Hz but we are at this point using $5\times10^8$Hz. Thus the bandwidth difference is $5\times10^6$Hz or about 67 dB. This in effect yields a 17 dB (67-50 dB) signal to noise in the data band!

The LNA system now includes the low noise high gain $G^aA^s$FET and the isolators. After the LNA we include a filter to narrow it to 36 MHz. This is done because we know that the data was in a certain 36 MHz band. We did not do this before since such a filter has a high Q and could be lossy and difficult to mount with the LNA.

We follow this filter with another amplifier and use this in the down converter system. Now note that the signal level is at -86.0 dBm and the noise if -49.2 dBm in 36 MHz. The difference now is about 37 dB.

The down converter includes the LO, mixer and sideband reject filter. That filter can be fairly wide as we discussed. Its width here is dominated more by the need to slow over frequencies in the demodulator than to reject sidebands. The net signals are then fed to the demodulator.

As with the uplink, the process is fairly straight forward. It starts at what is received and works backward. Care must always be taken to ensure that noise saturation does not occur.

In more complex earth stations with multiple modems, the signals can be combined on transmit and divided on receive. Figs. 26 and 27 show how this can be accomplished. In both cases we use two stages of conversion. Various other schemes are also possible.

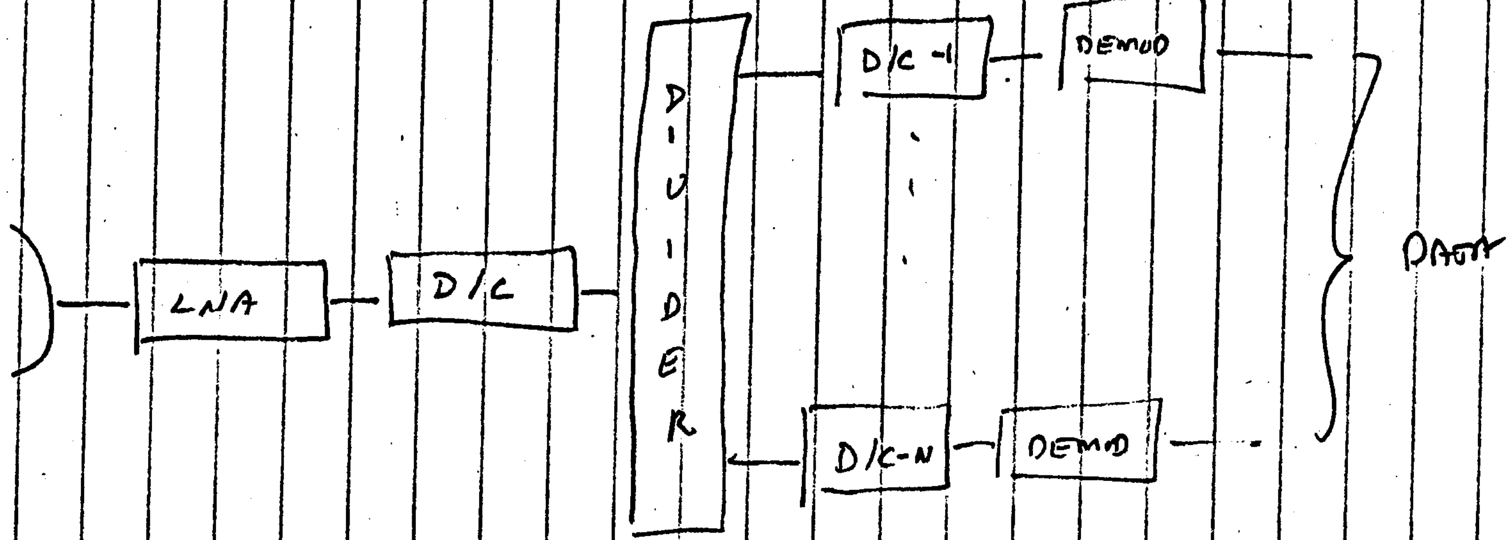



FIG 28: MULTIUSER STRUCTURE RECEIVE

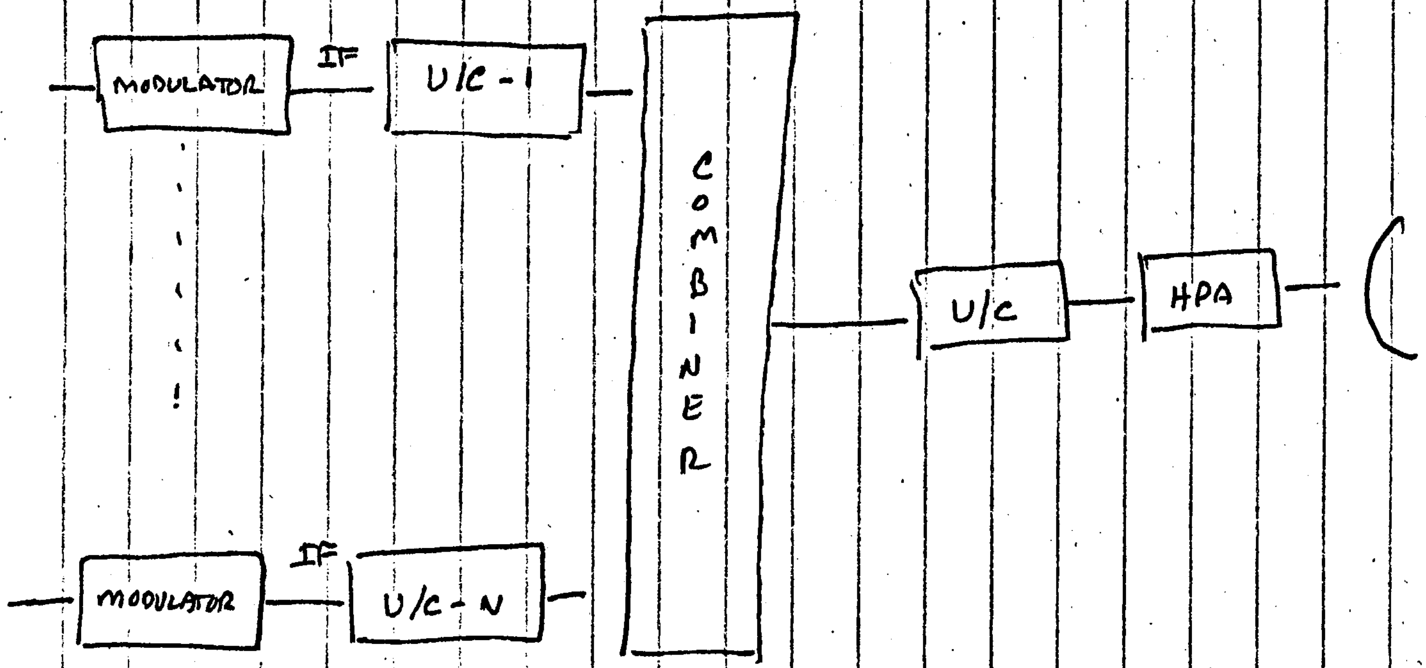



FIG 27 : MULTIUSER STRUCTURE TRANSMIT

## CHAPTER 5
## SATELLITES

In this chapter we develop the basic system elements as they relate to the satellite itself. We do not intend to develop a presentation of satellite design but only concentrate on the communications package and show how it relates to the overall system design. We shall consider only the antenna, receiver/transmitter portion, on-board processing, and inter-satellite links. These are the key elements that will provide significant changes in capability to future satellite generations.

As we shall see, the satellite has acted in the past as merely a repeater; amplifying and relaying the signal. In future designs it will play a more integral role in the overall network. Thus what we shall emphasize are those key elements that will be germane to expansion of capability and service.

### 5.1 Basic Concepts

Fig. 1 depicts the basic elements of the communication portion of the satellite. It is very similar to an earth station in that it has a receiver (LNA), transmitter (HPA), antennas, frequency converter, and some form of interface. The up link frequencies are usually not the same as the down link so that the frequency conversion ~~has~~ must as a minimum, ~~to~~ convert from the uplink to the downlink.

The interface may be as simple as a direct translation or as complex as a computer to process the digital data. The detailed structure of the satellite depends upon many factors. The first is the types of antenna patterns and coverage. If the antenna has several disjoint beams then the satellite must be structured to connect traffic from one beam to another. Each beam can then be considered as a separate antenna, and indeed they are isolated to be treated as such. Let us consider the simple case of a two beam satellite and let us further assume that we have available 500 MHz of bandwidth. Now if the two beams do not overlap we can use the 500 MHz twice, once in each beam.

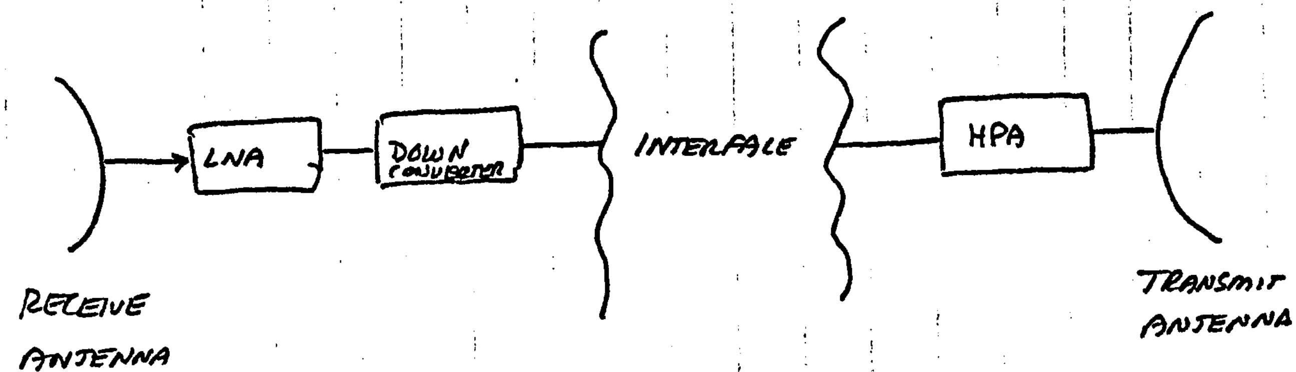



FIG 1 : BASIC SATELLITE COMMUNICATIONS DESIGN

This simple case can now be considered as in Fig. 2. We take the signal from beam-1 (B-1) and split it in half. This is done by using filters (F-1 and F-2). F-1 is 250 MHz wide and covers the lower band and F-2 is 250 MHz wide and covers the upper band. Thus from B-1 to B-1 we use F-1 and connect. From B-1 to B-2 we use F-2. On the bottom we reverse. Here B2 to B1 uses F2 and B-2 to B2 uses F-1. The figure shows what is called the transponder plan, four transponders (HPAs) with two in each beam and with non-overlapping frequencies. This process is called channelization of the satellite.

Let us now extend this concept to what is used in INTELSAT V. Here we have the following beams.

1. Global (F) - this covers almost one third of the earth and is 6 GHz up and 4 GHz down.

2. Hemi - there are two hemi (hemisphere) beams and these do not overlap. They are called East and West (~~WH~~ EH, WH). They are at 6/4 GHz.

3. Zone - these beams overlap the hemis but are of opposite polarization. They are orthogonal to EH and WH. ~~we~~ They are called ~~these~~ Z-1 and Z-2. They too are at 6/4 GHz.

4. Spots - these are at 14/11 GHz and thus are orthogonal to the others by frequency and to themselves by space. We call these S1 and S2.

We see that frequency reuse can be accomplished by one of three ways, different beam locations, different frequencies and different polarizations. In this case we use all three.

Clearly the global beam would overlap EH, WH, Z1, Z2 unless we do not use the same band of frequencies. Fig. 3 depicts the transponder plan from the seven beams. We use about 80 MHz band-widths to occupy the almost 500 MHz band. It is divided into 12, 40 MHz slots called channels. Thus WH has been channelized into ~~4~~ four 80 MHz slots using 8 channels. G uses four 40 HMz slots or 4 channels. The spot beams use two 80 MHz slots plus a

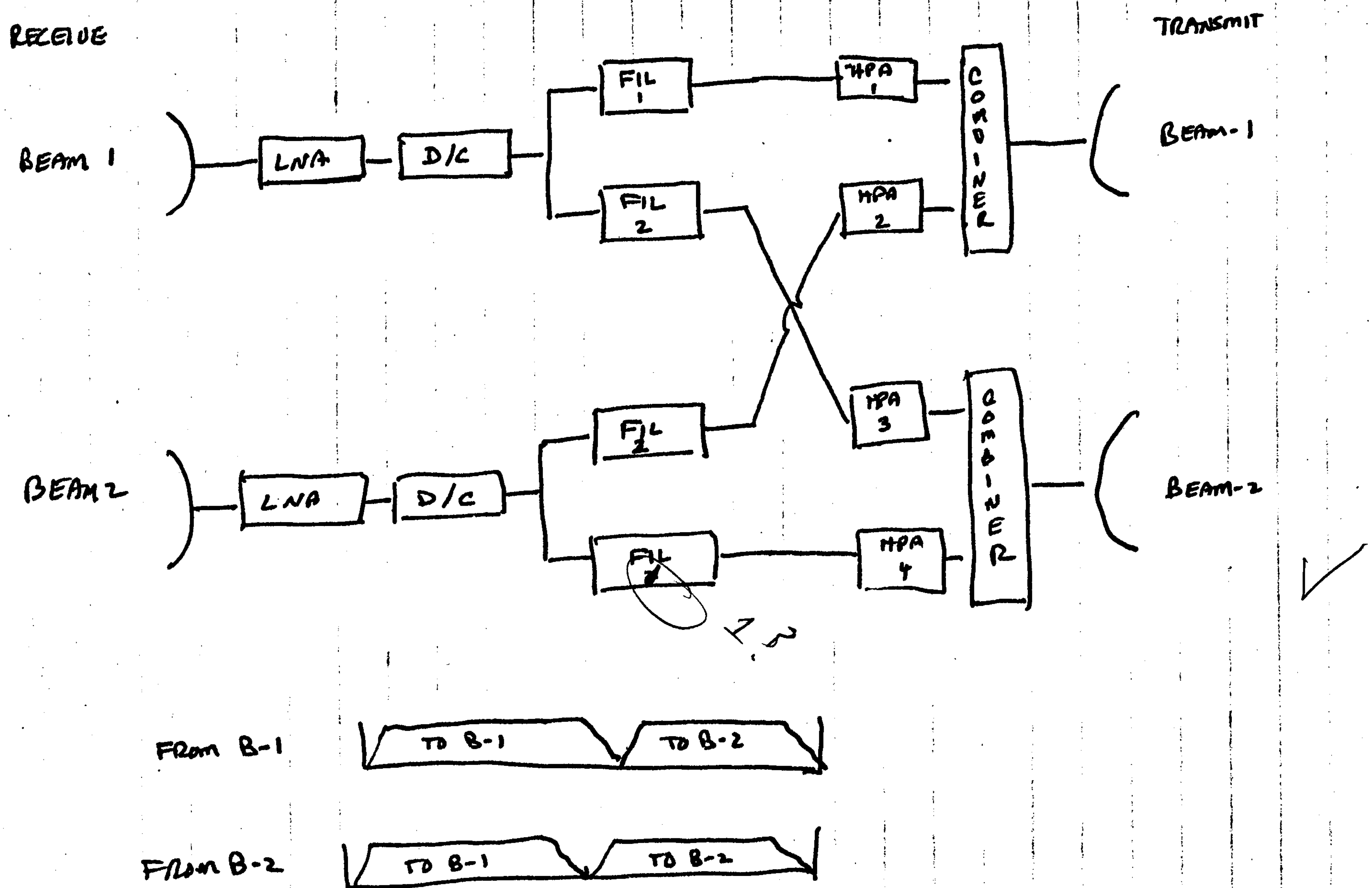



FIG 2: SAMPLE CHANNELIZATION SCHEME

24

~~240~~ MHz slot. We must now be able to connect these "from" slots to "to" slots. This is accomplished by a switch matrix on the satellite. To each one of these slots is assigned an HPA on transmit. We cannot for example assign the WH slot 3-4 to WH and at the same time assign the EH slot 3-4 to WH. This would be an overlap. Fig. 4 depicts the INTELSAT V channelization. There are seven beams and 26 channels. each with a down converter and a TWTA. The beams use a low loss combiner to join together the channels. Figs. 5 and 6 depict the antenna coverages for these beams for the Atlantic and Indian Ocean regions respectively.

The choice of bandwidth, number of channels, beam coverage, etc. are dependent upon traffic and cost considerations and we discuss these issues in the following two chapters. But to generalize, the beams could go to other satellites and as such would be an ISL. The channelized data could be brought to baseband and thus be an onboard processor. We shall consider all of these issues in this chapter.

Au Have you defined ISL before?

Orthogonal in space?

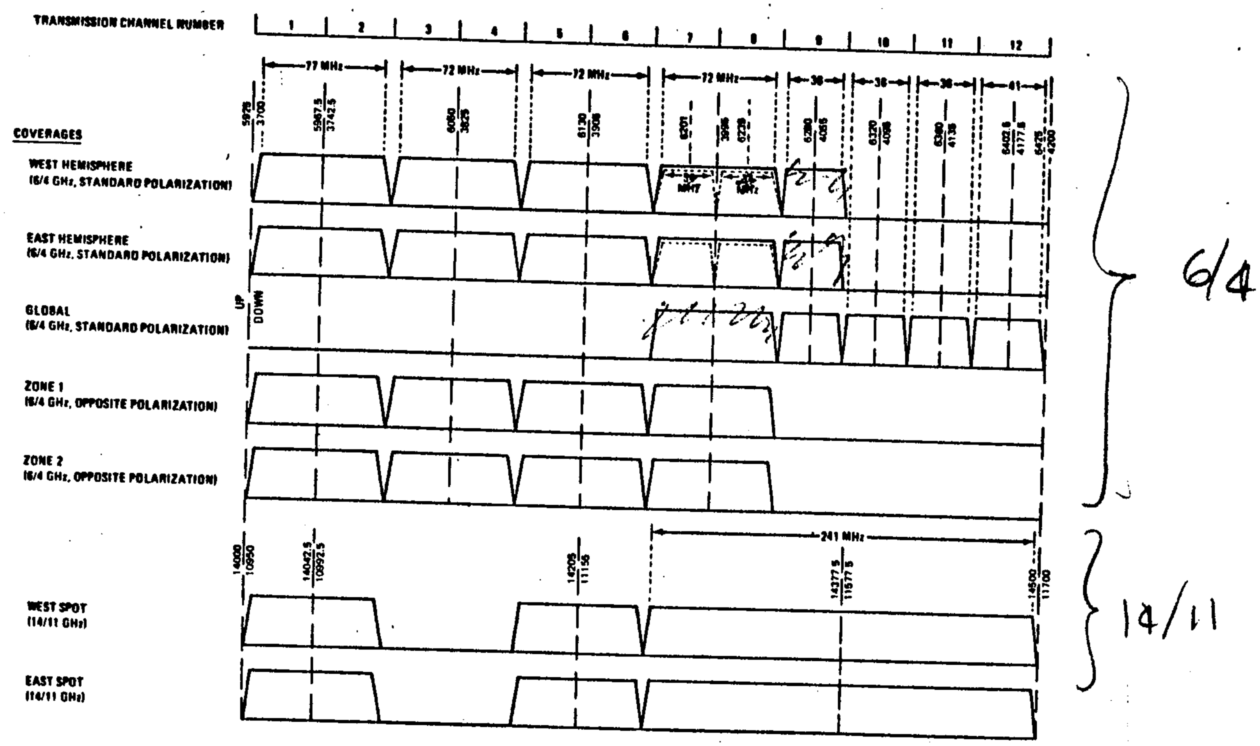



FIG 3: INTELSAT V Channel Allocation

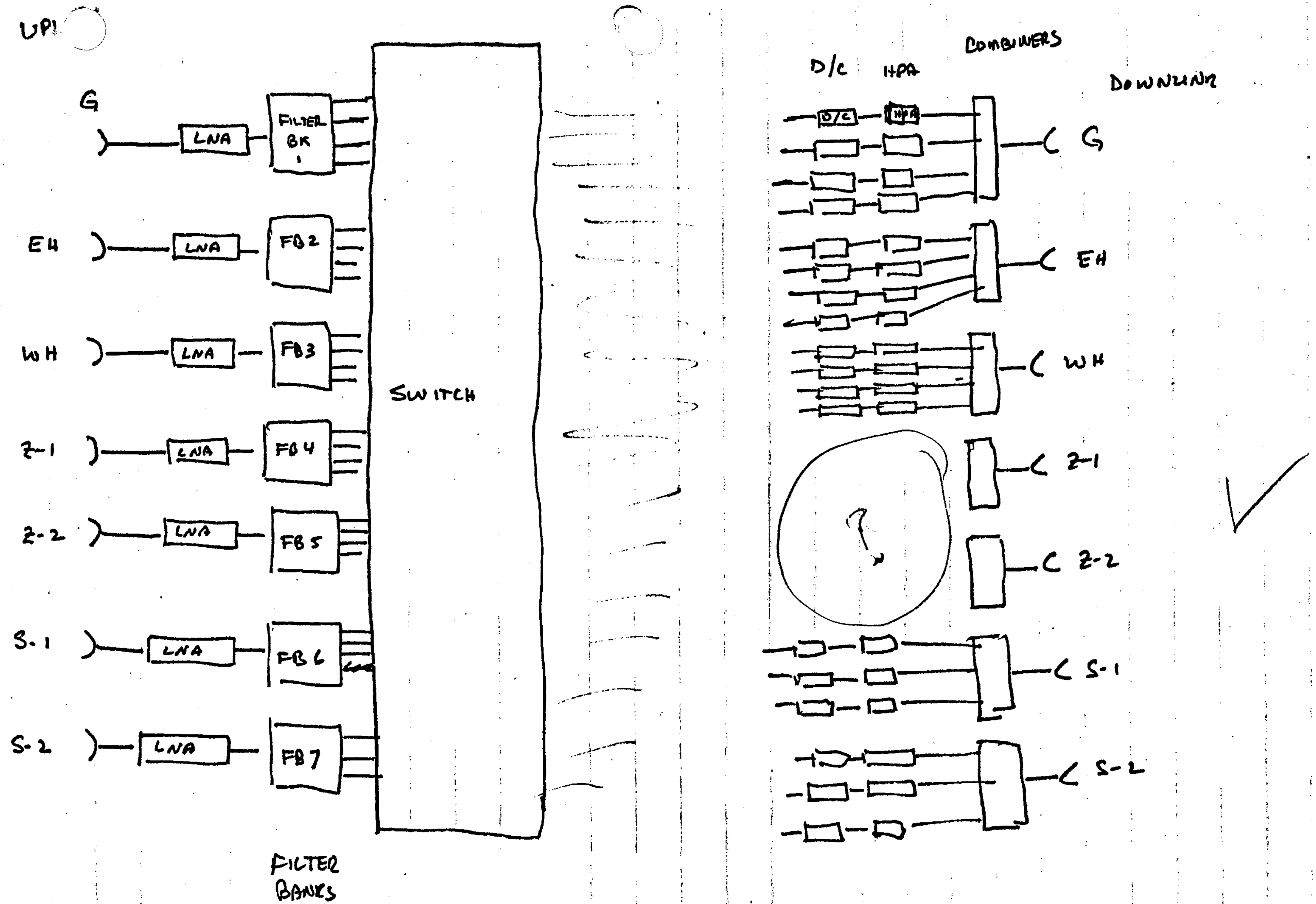



FIG 4 : INTELSAT V CHANNELIZATION

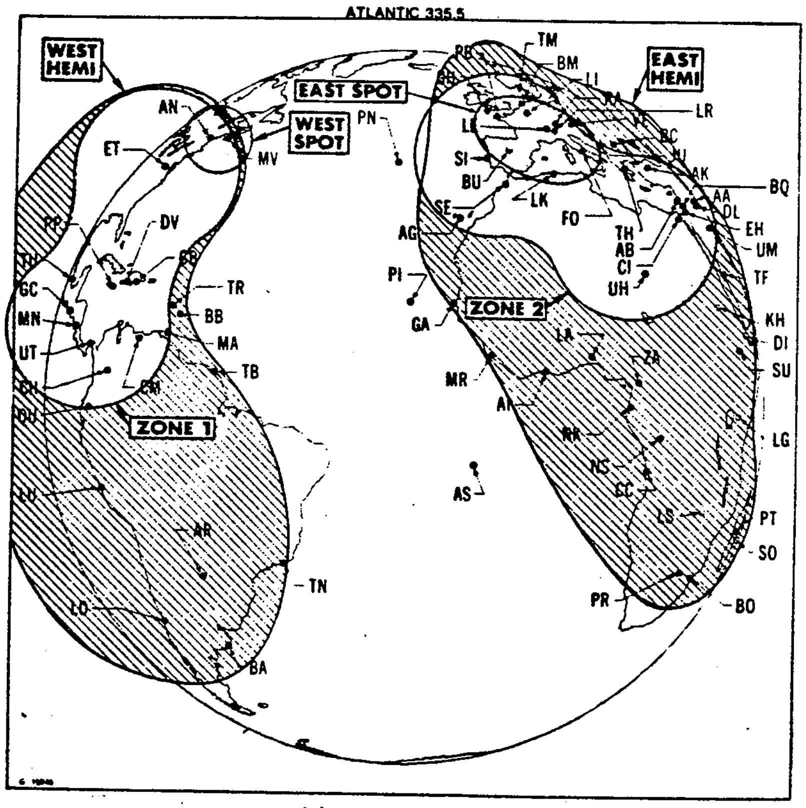



FIG 5: INTELSAT V Atlantic Ocean Coverage.
Note that the two letter combinations denote earth station locations.

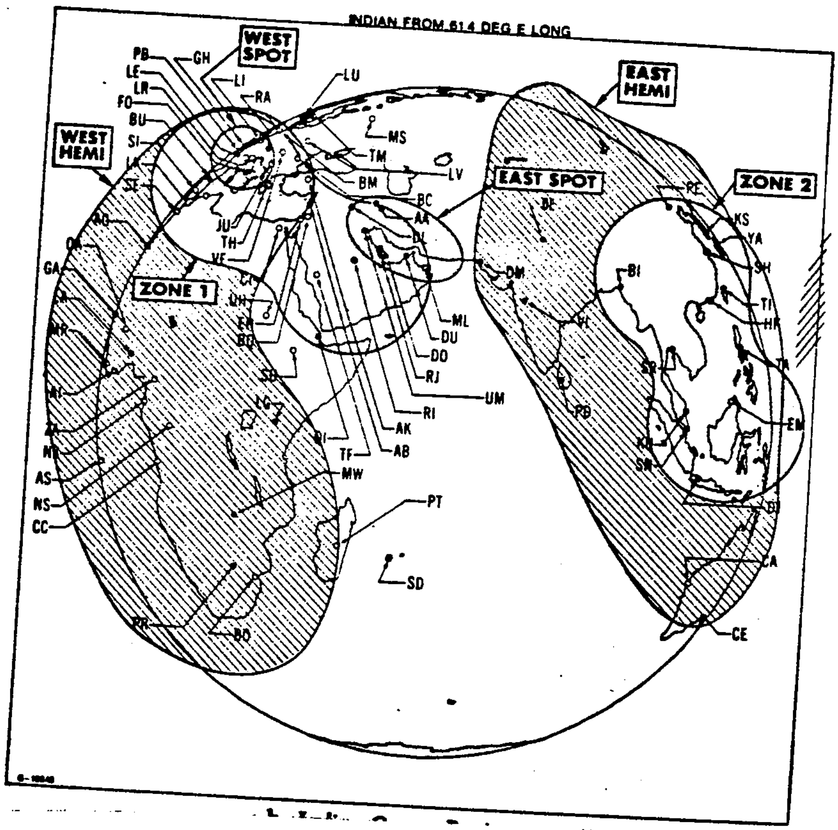



Fig 6: Intelsat V Indian Ocean Coverage

## 5.2 Satellite Antennas

The satellite antenna differs significantly from the earth station antenna in its overall objective. The satellite antenna must be capable of covering defined portions of the earth with a certain gain and also must not interfere with other coverages on the earth if operated in a multiple beam mode. Thus we must combine gain and low side lobes with a possibly irregular shape. In the case of the earth station we saw that the objective was to generate one single beam with some side lobe limitation. For the earth station size, structure and weight were not very important. For the satellite these issues are critical. What is common however is the basic principle. In the earth station case we generated only one beam. For the satellite we generate several narrow beams that add up to the desired level of performance.

Let us begin by considering the general geometry. The satellite is kept at a stationary location at a distance of 22,300 miles from the earth. This is about five earth radii distance. From a geometrical perspective this allows coverage by a cone with apex angle of $7.5^{\circ}$. In Fig. 7 we show the geometrical constraints. We can see that the peak latitude is $82.5^{\circ}$. Above or below that we cannot see the satellite.

If we now have an antenna with a 3dB beamwidth of $17.5^{\circ}$ across then that just covers the earth. Such a beam is called the earth coverage beam. Consider now the two beams that are shown in Fig. 8. In both beams we want the gain to be stable to $\pm 1.5$dB across the main beam and to have sidelobes that are 30 db down when they cross the adjacent beam. This requirement is typical of a multibeam design.

What performance does this yield?

To see how an antenna is designed to provide for this coverage we must return to the analysis developed in Chapter 4. Consider the rectangular aperture in Fig. 9. The aperature is 2d on each side. θ represents the angle from Boresight (boresight is 0=0) and Ø is the rotation angle about the aperature. Now we have shown that the antenna gain* G(θØ) can be given by

$$G(\theta,\emptyset) = |g(\theta,\emptyset)|^2/(4\pi/\lambda^2)$$

* See Chapter 4.

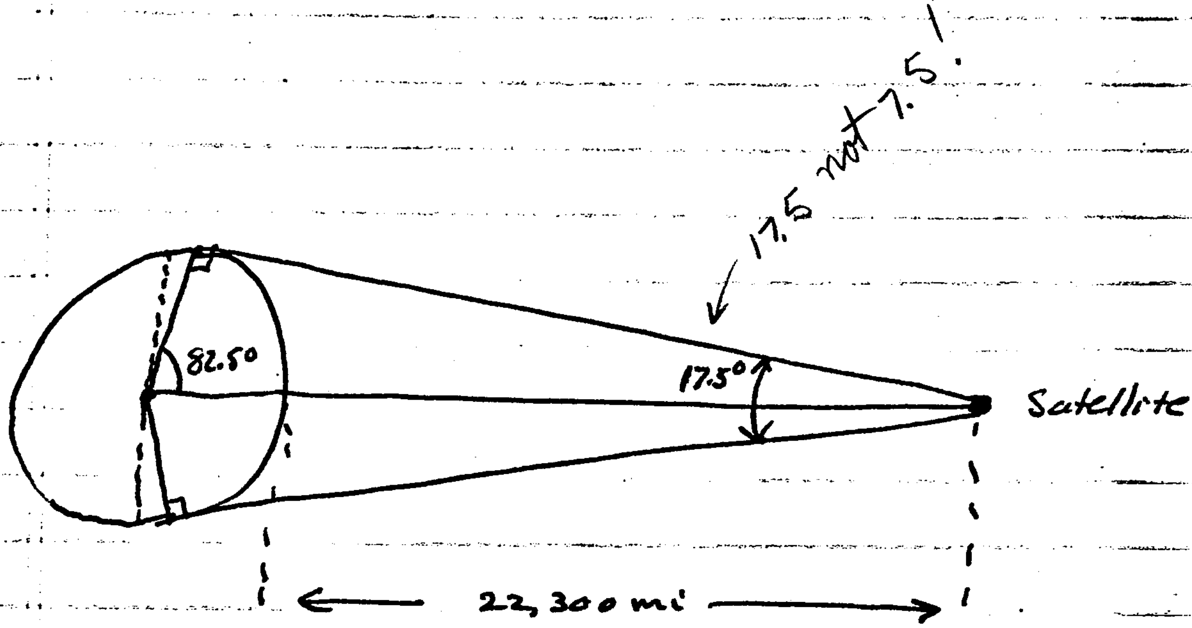



Fig 7 : Geometrical Constraints

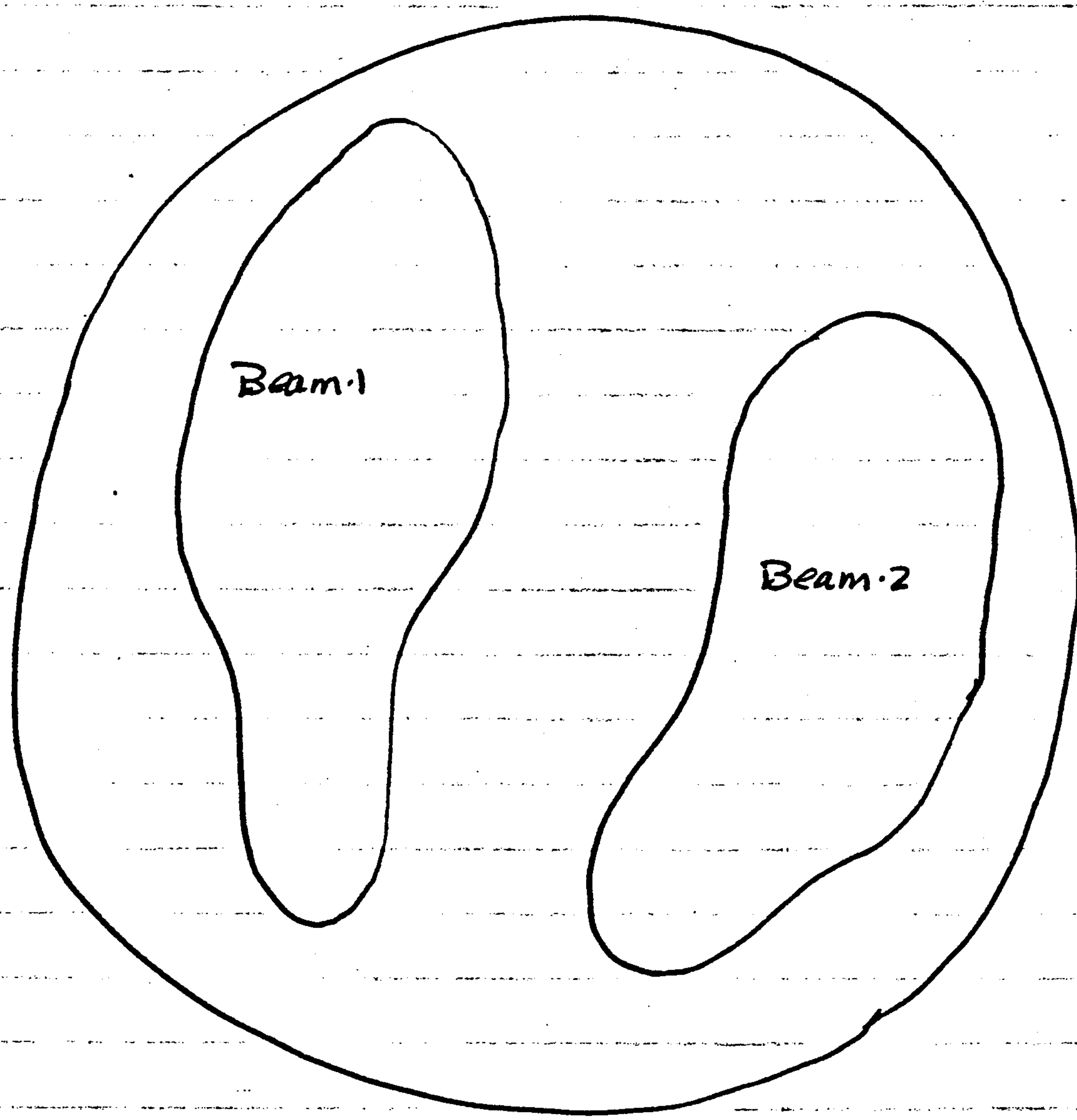



Fig 8 : Example Coverages

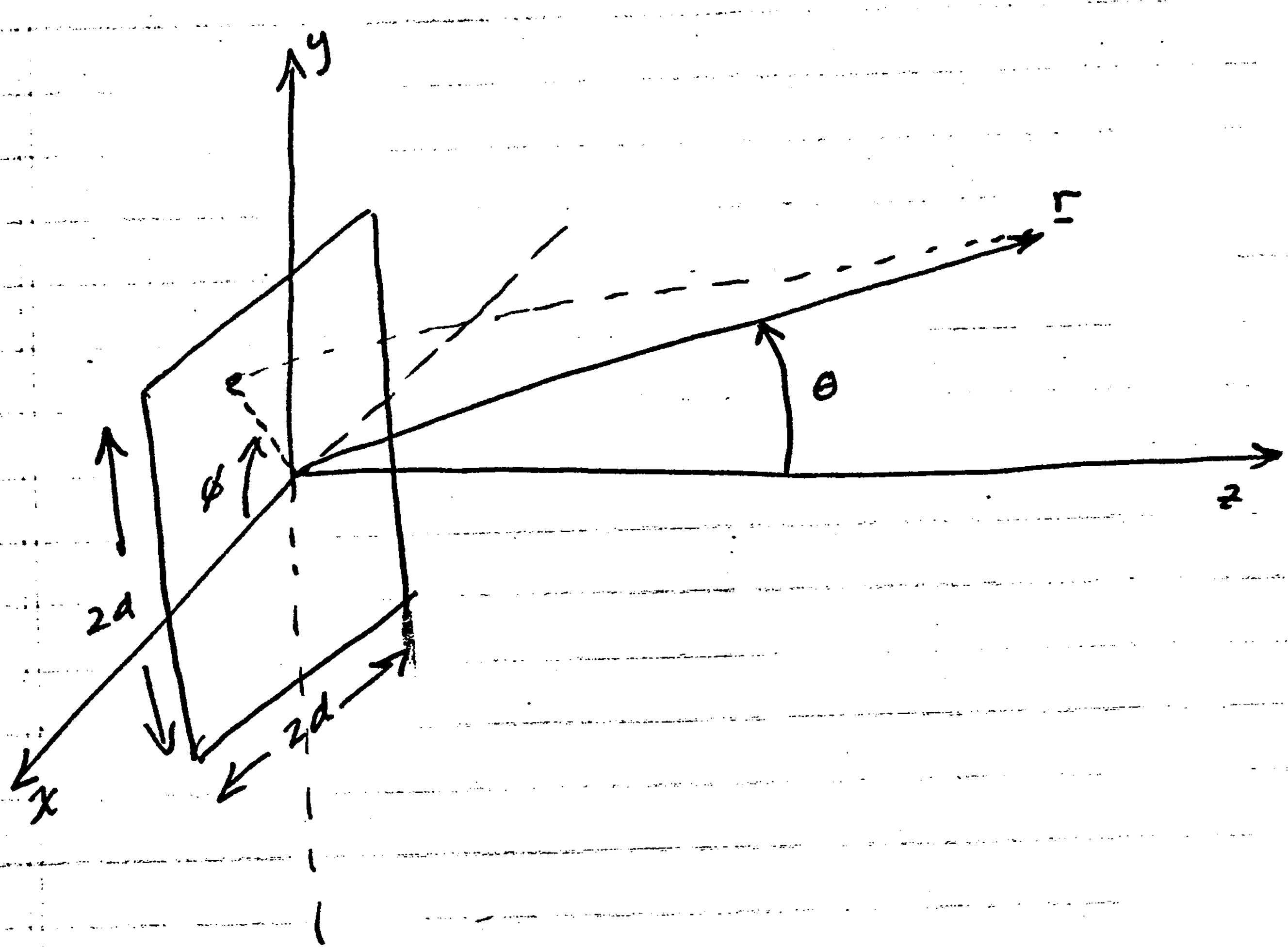



Fig 9 : Rectangular Antenna Aperature

where

$$g(\theta,\emptyset) = \int_A I(x,Y)\exp(j2\pi k_o x \cos \emptyset \sin \theta + j2\pi k_o y \sin \emptyset \sin \theta)\, dxdy$$

where I is the effective current density on the aperature. When I is constant we have (cite a reference)

$$g(\theta,\emptyset) = \frac{\sin (2\pi k_o d \cos \emptyset \sin \theta)}{2\pi k_o d \cos \emptyset \sin \theta} \cdot \frac{\sin (2\pi k_o d \sin \emptyset \sin \theta)}{2\pi k_o d \sin \emptyset \sin \theta} \cdot 4d^2 \left(\frac{4\pi}{\lambda^2}\right)$$

Clearly the peak gain is $4d^2(4\pi/\lambda^2)$. The 3dB beamwidth is also readily obtained by determining the points where $G(\theta,\emptyset)$ is down by that factor. If we stay in the x plane ($\emptyset=0$) we have

$$g(\theta,0) = G^{1/2}(0,0) \frac{\sin (2\pi k_o d \sin \theta)}{2\pi k_o d \sin \theta}$$

so that the first zero occurs at

$$\frac{d}{\lambda} \sin \theta = 1/2$$

I get First zero: $\cancel{2}\cancel{\pi} k_o d \sin\theta = \cancel{\pi}$

$\sin\theta = \frac{1}{2k_o d}$ ?

or $\theta \cong \frac{1}{2k_o d}$

or

$$\theta \simeq \frac{\lambda}{2d}$$

Let us now introduce a different current density $I(x,y)$. Let

$$I(x,y) = \exp(-jdx)$$

This is a linear phase shift across the x axis. This does not change the y axis integral but it does the x. The result is give reference

$$g(\theta,\emptyset) = G^{1/2}(0,0) \frac{\sin (2\pi d/\lambda \cos \emptyset \sin (\theta-\alpha))}{(2\pi d/\lambda \cos \emptyset \sin (\theta-\alpha))} \cdot \frac{\sin(2\pi d/\lambda \sin \theta \sin \theta}{2\pi d/\lambda \sin \emptyset \sin \theta}$$

$-\phi$?

add paren's!

This linear phase shift now has the effect of moving the peak of the beam along the x axis ($\emptyset=0$) to the point

$$\theta_o = \sin^{-1}\left(\frac{\alpha}{2\pi \frac{d}{\lambda}}\right)$$

The greater the phase shift the more the offset. If it were a positive phase shift then we would get a negative offset. Fig. 10 depicts three such beams. We can do the same with phase shifts across the y axis. In fact we can position the beam almost anywhere in space.

From the previous analysis we see that antenna peak gain is

$$G = C_1 \frac{d^2}{\lambda^2}$$

that is proportional to $(d/\lambda)^2$ and beamwidth is

auth?

$$BW = C_2 \cdot (\lambda/d)$$

proportional to $(\lambda/d)$ and $C_1$ and $C_2$ do not depend on d and $\lambda$. These hold for any type of aperature and only the ~~choice of~~ constants change.

Let us now return to our example of two arbitrary beams. The first question we ask is how large does the antenna have to be. There are two constraints on the answer. First what gain do we want and second how low do the sidelobes have to be to ensure the two beams do not interfere. We see from the above that $d/\lambda$ determines this.

The next question we ask is how do we make such an odd shaped as in figure 8 beam if all we know how to make are the symmetric patterns we just discussed. The answer is quite simple; use a weighted collection of these simple beams. These are called basis beams. Fig. 11 shows how to do this. Choose eight beams in this example and center them using phase shifts on current densities ~~as~~ of the form

$$\exp(j\alpha x+j\beta y)$$

such that they appear as shown. The actual number of beams depends on two factors. First if $G_o$ is the desired gain of the beam and $G_n$ is the gain of each of the small beams then

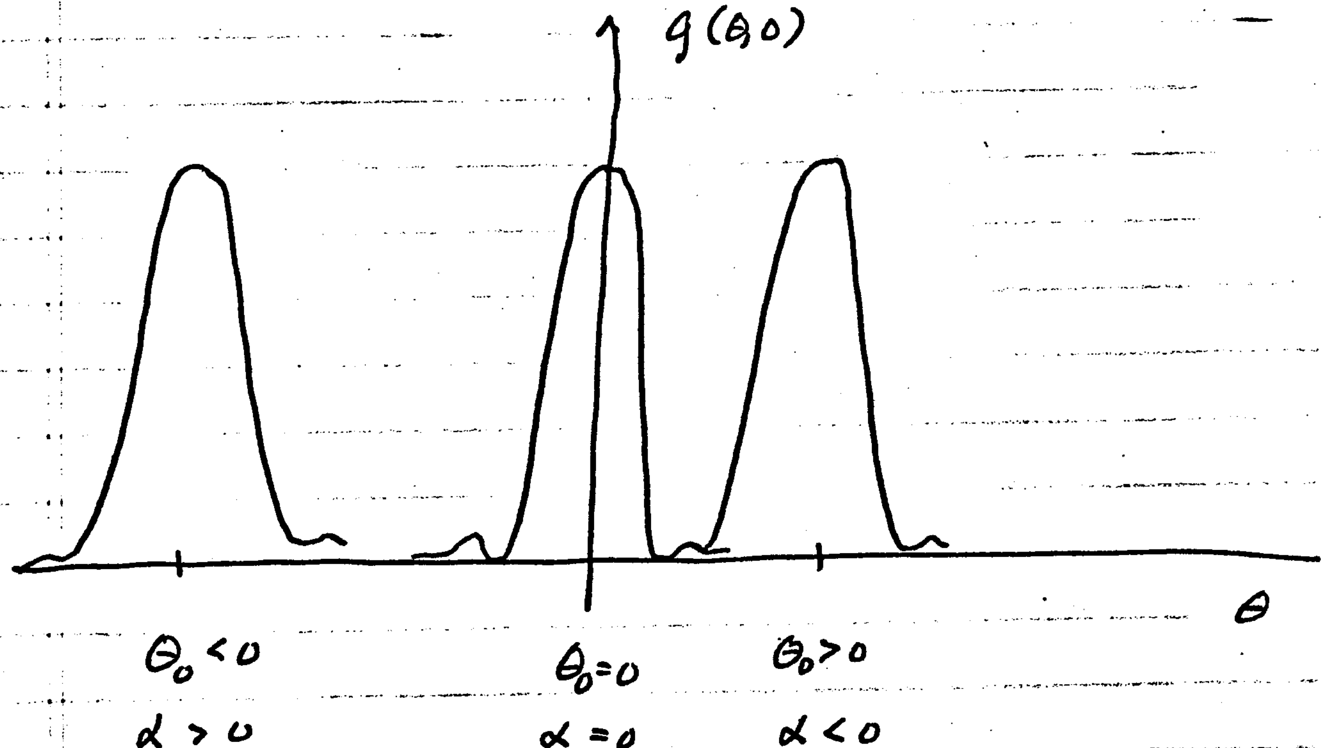
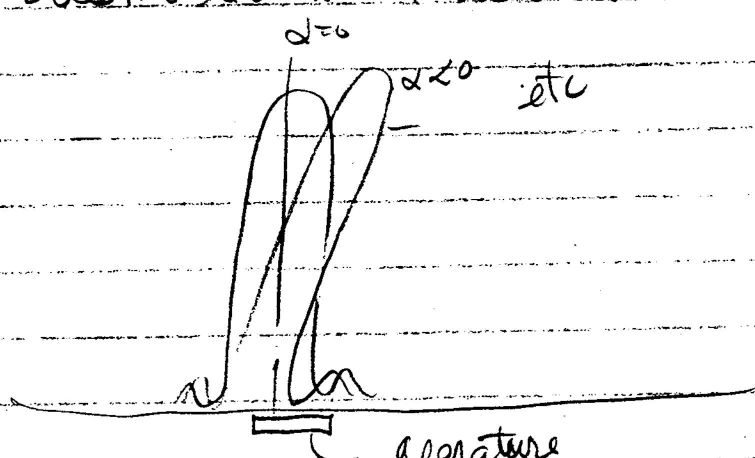



Fig 10: Movement of Shaped Beams.

auth:
Doesn't the beam tilt like so?



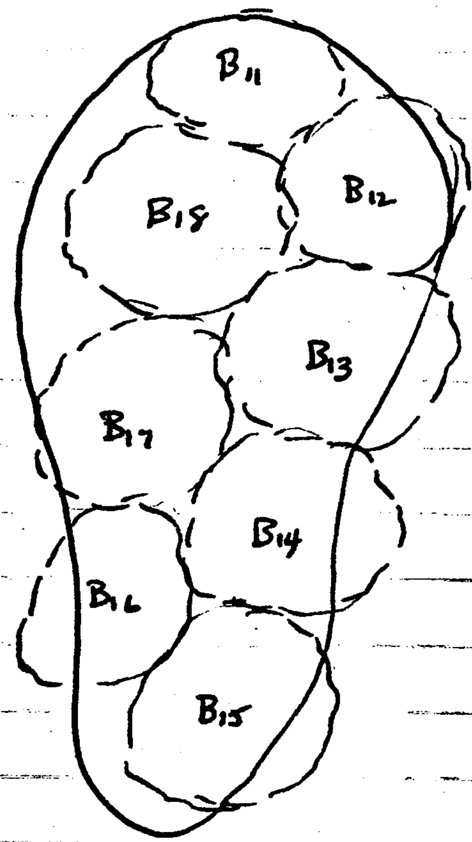



Fig 11 : Use of Basis Beams

Gain at the edge of the basis beam?

$$\frac{G_n}{n} \geq G_o$$

or

$$G_n \geq nG_o$$

This is because each beam gets only 1/nth of the power. Given the beamwidth as a measure of area we note that the area of B-1 is proportional to its $(BW)^2$. The area of each beam is proportional to its $(BW)^2$. Thus be need

$$n(BW_n)^2 \simeq (BW)^2$$

This is consistent with

$$\frac{G_n}{N} > G_o \quad \text{lc } n$$

so we are assured that there is no incompatibility.

Thus all ~~we must~~ that there is left to consider is the sidelobe constraint. To satisfy that we look at B-2 as also generated by a similar set of beams. Then look at the two closest beams and see if the sidelobes are proper. If not increase the antenna size until this is satisfied. ? Define

Now this approach still provides a rough approximation. We can improve this by choosing proper phasing. This can be done as follows. Let $G(\theta,\emptyset)$ be the desired beam patterns. Let

$$g_i(\theta,\emptyset) = A_i \iint \exp(j\alpha_i x + j\beta_i y) \cdot \exp\left(j2\pi \frac{x}{\lambda} \cos\emptyset \sin\theta + j2\pi \frac{y}{\lambda} \sin\emptyset \sin\theta\right) dxdy$$

be the pattern of a single beam. Then define an error function as

$$\varepsilon^2 = \iint \left| G(\theta,\emptyset) - \left| \sum_{i=1}^{n} g_i(\theta,\emptyset) \right|^2 \right|^2 d\theta d\emptyset$$

and choose $(A_i, \alpha_i, \beta_i)$ to minimize this. This is the optimal feed design. Various optimization schemes can be used to determine the separate amplitudes and phases. References?

What we have thus done is shown that to generate any pattern we can use sets of basis beams and coherently add them together. Physically this is done in a similar fashion. Given an aperture this defines the shape of a basis beam. The feeds are then amplitude and phase shifted to generate the constituent beams, with a feed for each beam.

Fig.s 12 to 14 show how this beam forming is doen for an off-set fed para bola, a lens antenna, and an array antenna. We shall consider these in some further detail.

There are six basic factors that are used in deciding the type of antenna to be used. The relative importance of each of these factors can depend strongly on the use, specifically whether as an earth segment or a satellite antenna. For specific antenna designs, several of these factors can be subdivided into specific design tradeoffs that are used in actual implementations. The factors are:

Delete space

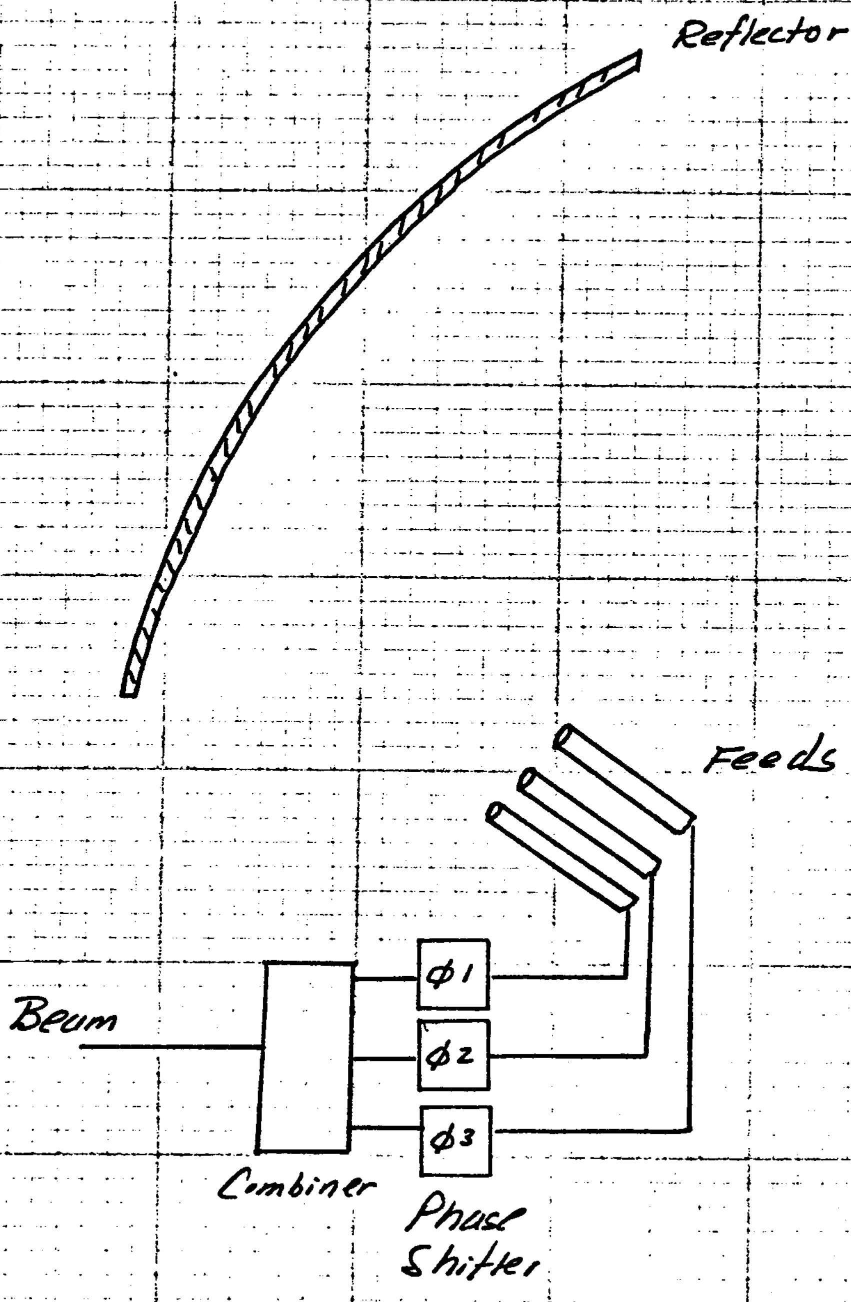



Fig 12: Off Set Fed Parabola

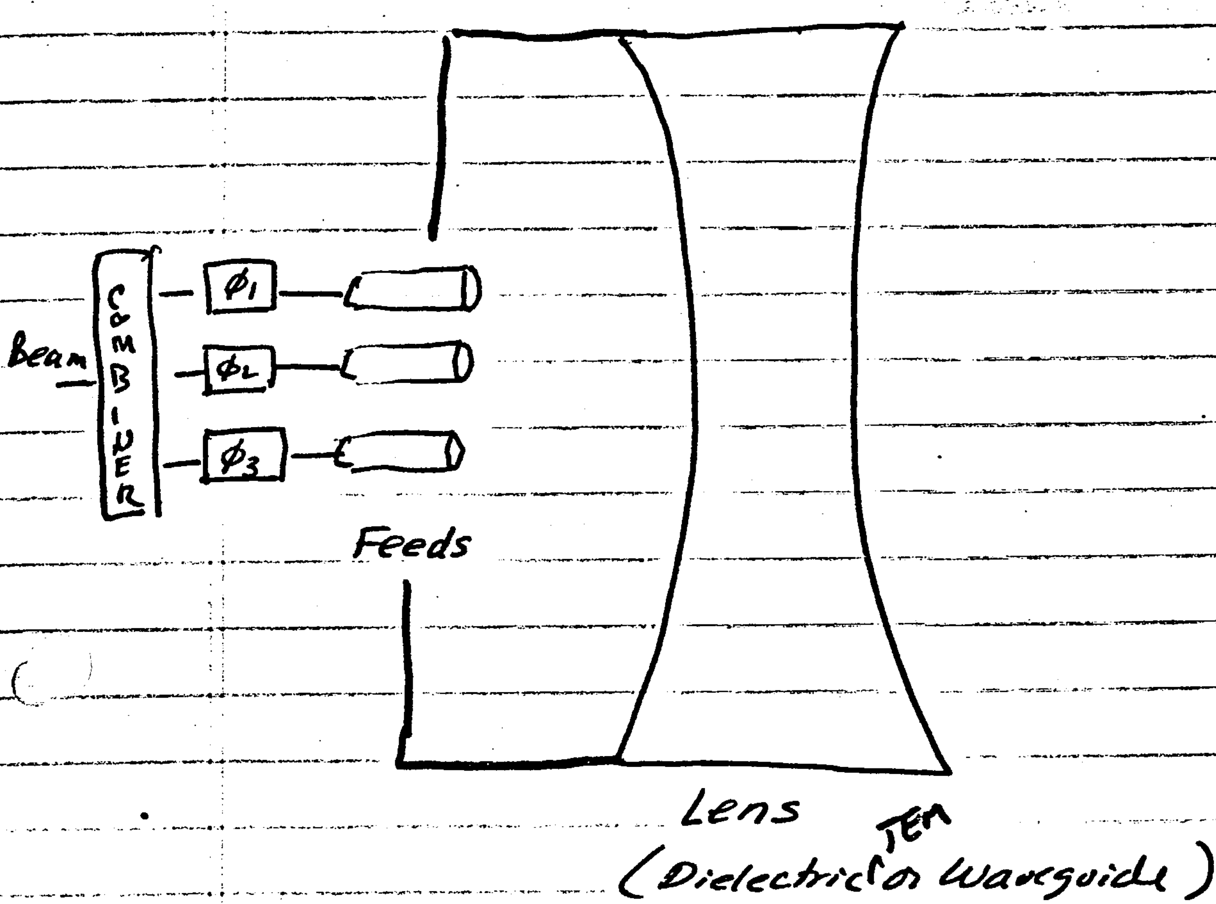



Fig 13 : Lens Antenna

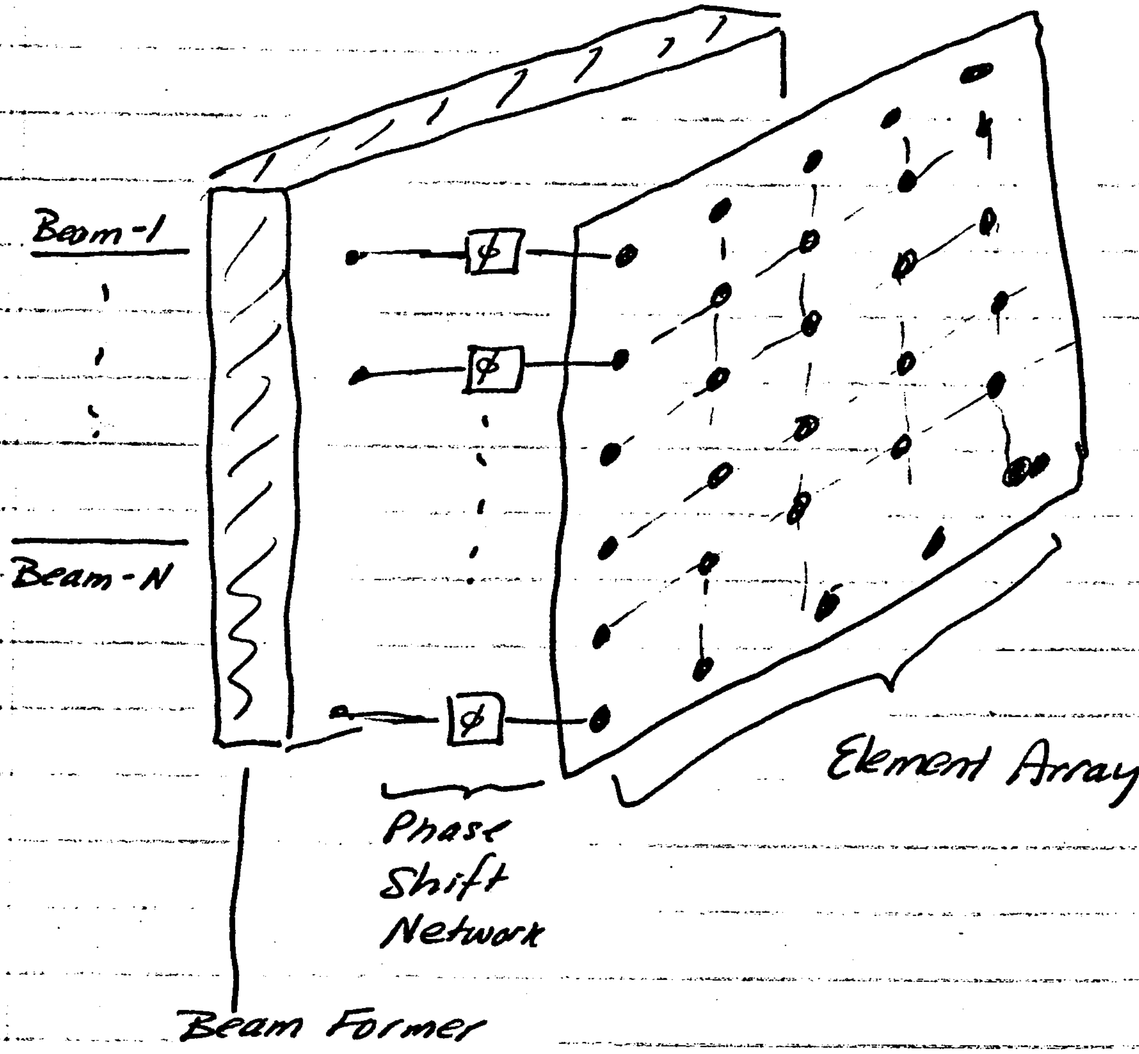



Fig 14 : Array Antenna

1. Weight: For both spacecraft and mobile use, lightweight antennas are required. The antenna weight is determined by the feeds, phase shifters, power dividers, antenna materials and supports. Each antenna design has different tradeoffs between these basic components.

2. Efficiency: The antenna r.f. efficiency can play an important role in spacecraft antennas. High efficiency is required in the power-limited space environment. In the terrestrial environment, cost considerations impact on the desired level of efficiency.

3. Frequency Limit: The choice of antenna will depend upon the desired operating frequency. Such parameters as aperture size, surface tolerances and antenna pointing will be critical factors.

4. Beam Steering: The ability to steer beams off boresight is required for narrow beam antennas that are also required to cover large areas. For example, in a satellite antenna, good beam steering characteristics are required if system flexibility is to be achieved through the ability to reconfigure the coverage of the satellite antenna in orbit. For air mobile applications, the antenna is used to increase gain and eliminate multipath. However, it must be capable of pointing to the satellite continuously. Thus, rapid beam steering is essential.

5. Beam Forming: Beam forming is concerned with the proper shaping of the beam or sets of beams to provide the desired energy distribution as well as low side lobes, good cross polarization isolation and interference rejection. This factor depends strongly upon the number and nature of the feed elements and the complexity of the control of these elements in amplitude and phase.

6. Bandwidth: The bandwidth of an antenna is a difficult factor to define. For a single beam antenna, the bandwidth could be considered to be the set of frequencies that the antenna can operate over without significantly degrading the signal. This is typically determined by the nature of the feed and its cutoff characteristics. However, in a multiple beam configuration, the effects of signal bandwidth on sidelobe levels and polarization isolation become important. These can, in turn, be the limiting factors in terms of system performance and thus define the operational bandwidth.

There are many forms of antennas that can be used for the three specific application areas. There are three generic types that are expected to dominate future applications--arrays, lenses, and reflectors. Each of these types has specific implementations that may be of use for specific applications.

<u>Arrays</u>. Array antennas are composed of three basic components: (1) array elements, (2) phase shifters, and (3) the feed system.[69]

(1) The array of elements is the interface with free space and is composed of the individual radiating elements. These elements may be individual dipoles, helical elements, open-ended waveguides or any other set of radiating elements. The element spacing may or may not be uniform.

(2) Phase shifters feed each element and are used to point and form the desired beam(s). The phase shifters should not unduly attenuate the signal and values of 1.0 dB are commonly attained.

(3) The feed system distributes the energy to the phase shifters and provides amplitude weighting across the array for beam-shaping.

The phase shifters and feed network can be adaptively adjusted to form beams of arbitrary shapes. Techniques of this type are called adaptive beam-forming (ABF).[70] The phase shifters can be digitally or analog-controlled, with respectively discrete or continuous phase shifts. The power distribution network can use variable power dividers to change the amplitude distribution.[68] This gives the array antenna a great deal of flexibility, both as a primary radiator and as a secondary radiator. Control can be maintained over sidelobes so that, in a multiple beam configuration, the cross beam interference can be minimized.[72]

The array has been considered for possible spacecraft operation; however, it is not presently used in its most complete form in any commercial satellite. It has in general been too big and too inefficient (e.g., the feed and phase shifters are too lossy). The MARISAT satellite does use a simple four element helix array at L-Band and a three element helix array at UHF. In both cases the phase shifters and feed elements are fixed. The potential of arrays may be reassessed as the role of the satellit becomes more complex.

For mobile earth segment uses, the array shows a great deal of promise. This is particularly true in aeronautical applications where the desire to minimize protuberances is strong. Conformal arrays have been designed and tested at <u>L and X band for aircraft</u>.[72] Other forms of planar arrays

References?

have been used for aircraft and have been designed to provide some discrimination against multipath while simultaneously providing gain. ~~73-74~~ It is expected that similar arrays can be developed for land and maritime mobile applications. The arrays for mobile applications must usually be small; however, by employing ABF techniques, they can electronically steer, track and null out destructive interference. Less sophisticated techniques, such as beam switching, can also be employed by using separate tone-driven monopulse-activated tracking systems.

For fixed earth station purposes, the array may, in most cases, be too costly a solution at present. As orbit utilization and conservation become more critical, its nulling capability may be of some importance.

Reflectors. Reflector antennas are composed of (1) reflecting surfaces, (2) feed systems, and (3) distribution networks.

(1) The reflecting surface can have various shapes such as parabolic, toroidal, or spherical. In addition, there can be subreflectors as in the Cassegrain antenna. The size of the reflecting surface determines the antenna gain at a given frequency. The surface tolerances required are also related to the operating frequencies, with the higher frequencies requiring more control of the surface tolerances.

(2) The feeds for the reflector can be any set of radiating elements--horns, arrays, lenses, etc. The feeds will normally be essentially located in the focal plane of the effective optical system, although they may be offset from the optical axis; or the optical axis may be folded through use of a secondary reflector. Moveable feeds can be employed to steer the separate beams, or the beams may be electronically steered.

(3) The distribution network contains phase shifters and provides power distribution to the radiating structures at the feed. Beam shaping and beam steering are performed in this network. For multiple beam, dual polarization antennas, this distribution system is a critical element in system performance.

The reflector is most commonly used as a fixed earth segment antenna. The 30 m INTELSAT antennas provide high gain and performance in the telephony area. Many of these antennas are open Cassegrain parabolic designs. For the developing domestic systems, parabolic reflectors are being considered because of their light weight, low cost and easy deployment. Three, 5 and 10 m reflector antennas are being used at 4 - 6 GHz, 7 - 8 GHz and are proposed for the 12 - 14 GHz bands. With the increase in the number of satellites that may provide service to a particular country or satellite communication application, there is a need for a multiple beam earth station antenna. Recent efforts in this area have led to the development of the toroidal reflector with a moveable feed structure.

The reflector is also quite common as a spacecraft antenna. It has been used on the INTELSAT IV and IV-A satellites, the ANIK satellites, WESTAR and SATCOM. On ATS-6, a 10 m deployable center-fed paraboloid was used. The reflector is also being considered for use as a satellite antenna in the high frequency bands at 20 - 30 GHz. This is particularly true for multiple beam frequency reuse purposes. At these higher frequencies, it is relatively easy to obtain 1° spot beams. It is possible to provide 35 db beam to beam isolation with spherical center-fed, or parabolic offset-fed reflectors.

For mobile earth segment antennas, the reflector has limited use. The present COMSAT General MARISAT system uses a 1.2 m four-axis tracking parabola. However, installation and basic costs are high. For land mobile and aeronautical mobile, such reflectors would be too cumbersome.

**Lens.** The lens antenna is a transmission device, in contrast to the reflector antenna. The lens acts as a focusing medium for the electromagnetic energy generated by the feed structure behind the lens. In addition to the feed structure, a distribution network of phase shifters and power dividers is also required. There are three generic types of lens structures: (1) metallic, (2) dielectric, and (3) TEM designs.

(1) The metallic lens is composed of an array of waveguides of varying length, with the resulting difference in propagation times accounting for its beam shaping property. The waveguides are hollow, and such a lens can be made lightweight.[68] However, these metallic lens antennas have small fractional bandwidths on the order of 5 percent.

(2) The dielectric lens is a broadband lens composed of dielectric material. In general, it is quite heavy, but recent work on lightweight dielectrics indicates that this weight factor can be reduced.

(3) The TEM lens design consists of TEM lines that do the required phasing of the electromagnetic field. Figure 15 illustrates a typical TEM lens antenna. A TEM lens antenna is lighter than a dielectric lens

and broader band than a metallic lens.[82] The lens can also be used as a feed for a reflector antenna.

As a spacecraft antenna, the lens has received attention recently for its proposed use on the DSCS-III satellites. For this application, the antenna must provide flexible antenna profiles and be capable of nulling out interference sources on the earth. It does this by controlling the various pencil beams that are formed by the feed network via the amplitude variable power dividers and output phase shifters. The lens has the capability of scanning off boresight quite easily; and, with sufficiently complex feed structures, large nulls can be attained at arbitrary locations.

As a fixed earth segment antenna, the lense has potential because of its beam steering and nulling capabilities. There are, however, currently no lens earth station antennas in operation. For mobile earth segment use, the lens does not appear to be a viable alternative for the 1980s.

Others. There are many other antennas that may be of use for specific applications.

As mentioned earlier, the helix antenna is being employed, in a simple array structure, for the MARISAT satellite because it provides reasonable gain and good circular polarization, which in turn, gives multipath protection.

Horn antennas have been extensively used for global coverage, but they do not have the capability to provide the narrow beams and high gain required in advanced systems.

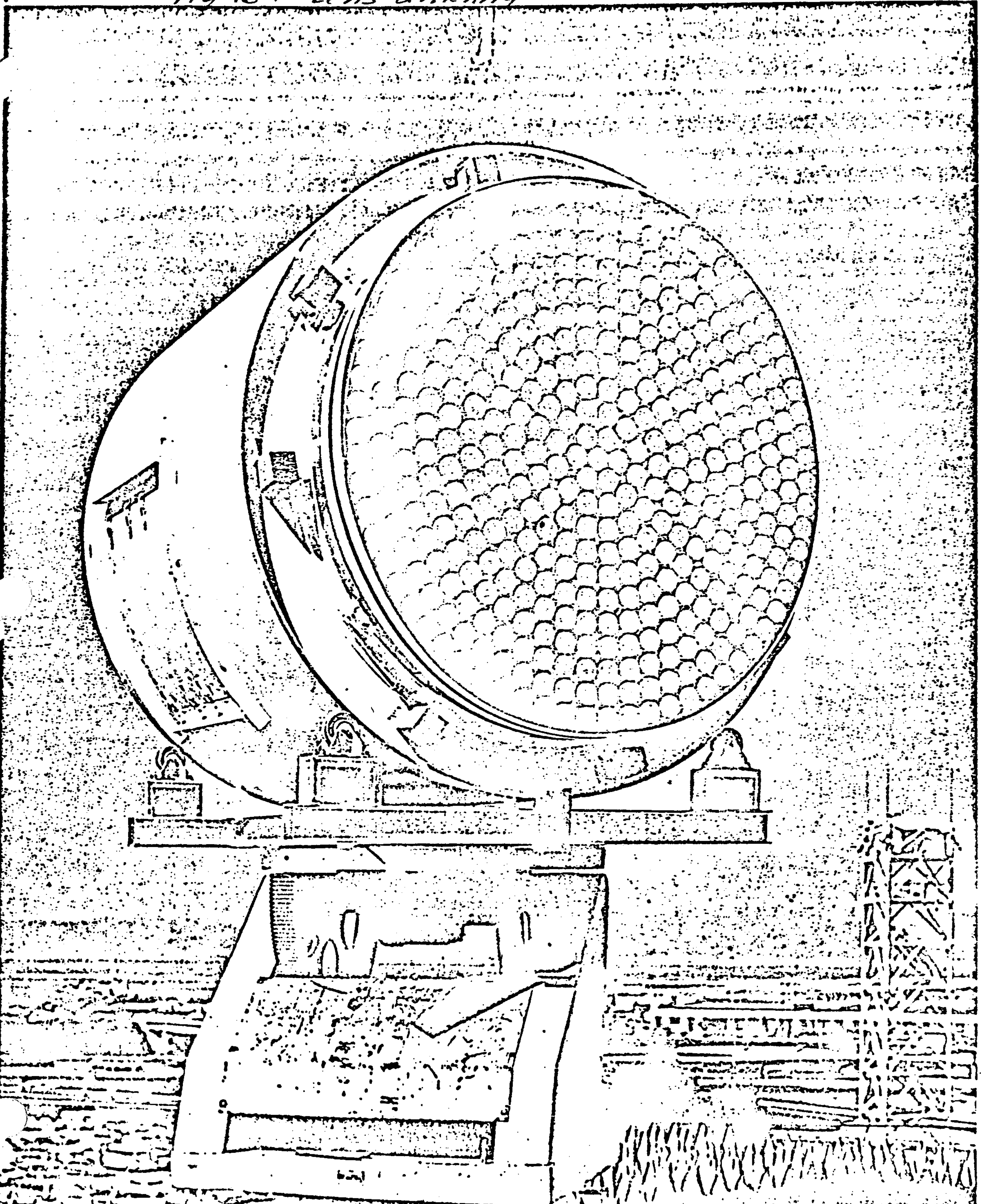
Single element devices, such as blade antennas, crossed dipoles, crossed slot and cavity-backed slot-dipole antennas may be useful for mobile applications.

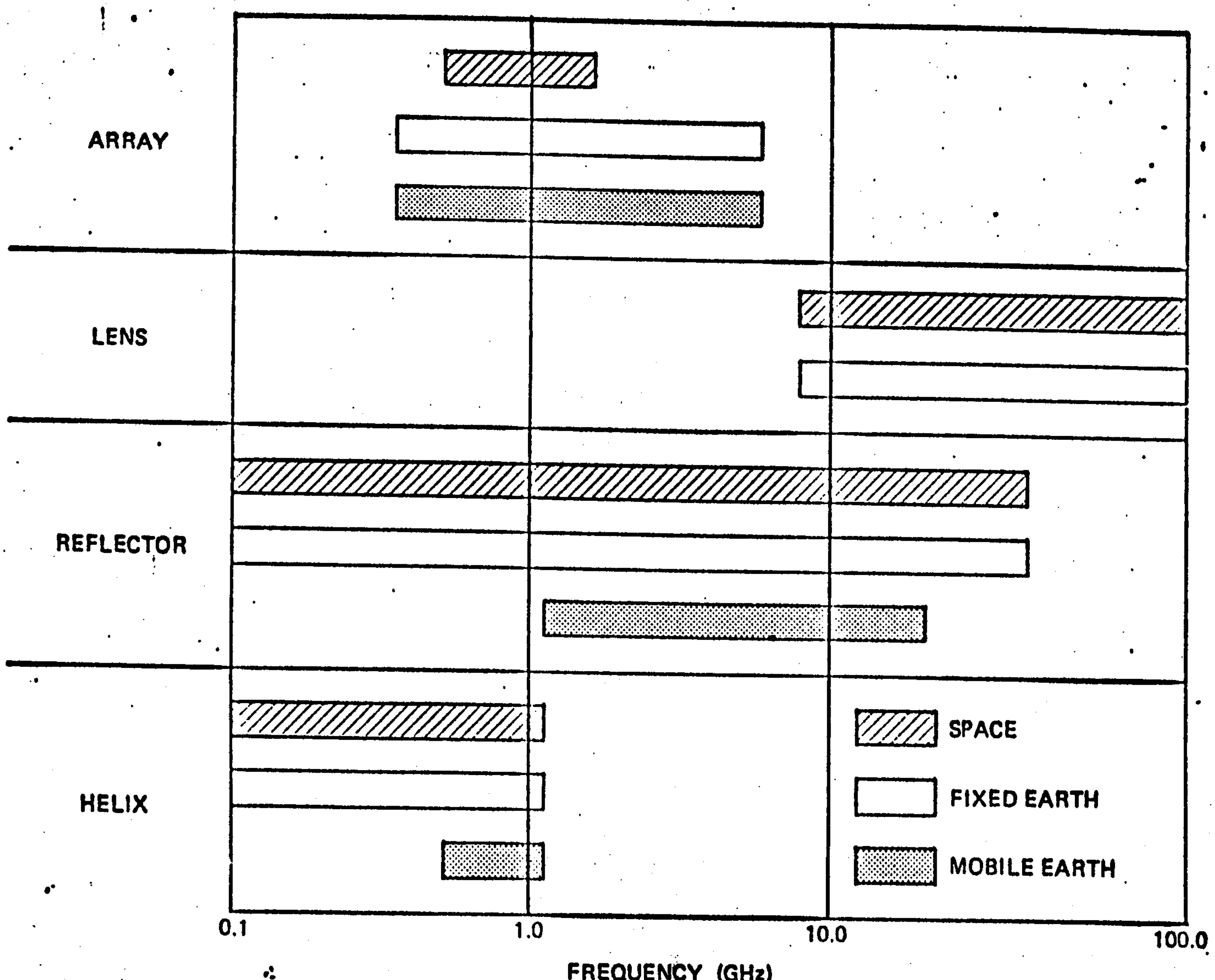
Figure 15 presents a set of operating ranges for various antenna technologies as a function of the operating frequency and the specific implementation (space, fixed earth, and mobile earth). The frequency range covered is from VHF up through V-band. The ranges indicated are based upon the use of existing technology that is commercially available. At most frequencies, there are only two alternatives available, which makes the choice for the system designer fairly straightforward.

Figure 16 shows a projection of the antenna technology ranges for 1990. The array range has been extended, based upon improved efficiencies, using strip line and solid state technologies. The lens range has been extended by developing lightweight dielectric materials which in turn also increase bandwidth. The reflector range has been extended by improving surface tolerances and achieving increased feed efficiencies.

Tables ~~21~~ to ~~23~~ present, for the three classes of antenna applications, some specific performance goals for the 1980s and relate these goals to system design characteristics as discussed in the introduction to this section. The tables also indicate a possible application area for each performance goal and, for that application area, list both present technologies that can be used to satisfy the requirement and projected technological capability that may be achievable in the 1980-1990 time period. In some cases, improvements in the existing technology is foreseen; in others, new technologies are considered as a replacement.

Fig 15: Lens Antenna






G–13257

Fig 16: Present Use of Antennas

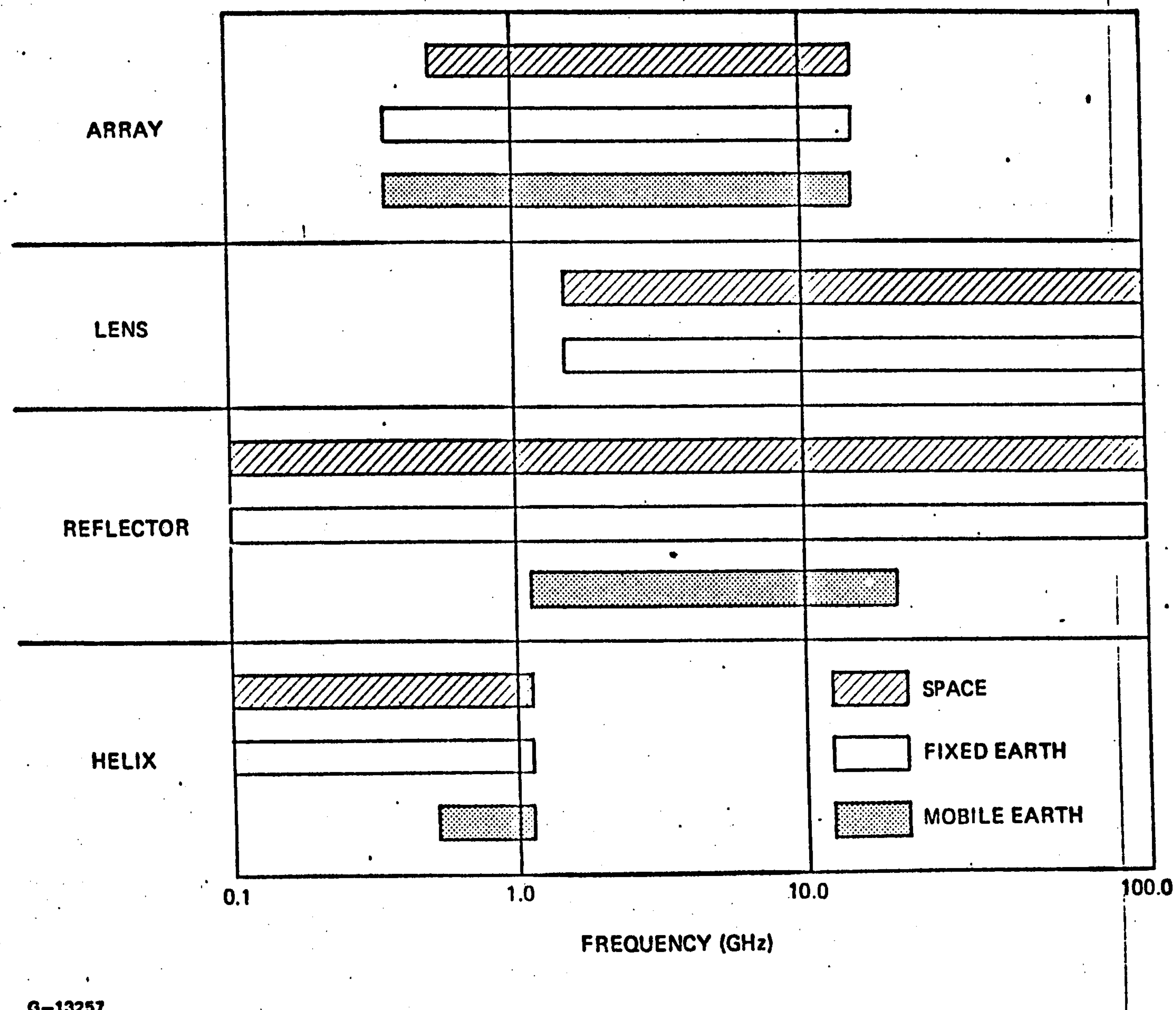



Fig 17: Possible Extent of Antennas by 1990

## 5.3 On-Board Processing

The present satellite communication systems provide a transparent repeater channel to the users of the system. The channel characteristics are static and play a passive role in the communications link. Such links are only reconfigurable by the activation of a non-real time switching operation. On-board processing represents a set of generic techniques, involving real or near-real time processing or switching, that can be used on the satellite to provide an improved performance level in the total communications system. As we will see below there are several types of on-board processors. A satellite that contains the most general form of on-board processor can be viewed as a digital satellite, which has the capability of acting as a processing point in the communications system. The processing can be used to dynamically reallocate the satellite's resources to maximize system efficiency; it can decouple the up-and downlinks and make their performance independent. At the present time, there are no non-military satellites that have on-board processing capabilities.

The use of an on-board processor will impact upon several of the system design characteristics discussed previously. The following four will be most significantly affected:

1. Connectivity: In a large multi-user distributed system with multiple beams on the satellite, the ability to interconnect all the users in an effective fashion becomes a difficult task. With a static satellite switch the interconnectivity is fixed, but with an on-board processor, rapidly changing connectivity can be achieved. This can be accomplished by changing the transmission paths, dynamically rerouting bits or actually changing the access format. In addition, more sophisticated network control can be developed to efficiently use demand-access systems or incorporate certain forms of random access.

2. Capacity: One of the major factors that reduces system capacity is a reduction of the fill factor or effective transponder utilization. With static coverages, the satellite's resources cannot be dynamically shared to maximize satellite usage. This deficiency can be

eliminated by using an on-board processor. In addition, the accommodation of several different multiple access schemes, such as TDMA and FDMA can be facilitated, either as mixed modes during system operation or during periods of system evolution.[83]

The on-board processor also allows for a decoupling of the up-link and the down-link, and thus potentially improves the overall link performance.

3. Cost: On-board processing can reduce satellite costs by providing the flexibility to effectively use the basic capacity associated with a satellite design, i.e., to achieve a high satellite fill factor. In addition, in some cases reduced earth segment costs may result because simpler and less costly modulation schemes can be used without significant performance penalties. On-board processing allows the possibility of a chan[ge] at the spacecraft in the modulation or access scheme to achieve a reduced cost earth segment while retaining the desire[d] system capacity. Earth segment costs can also be reduced by the use of a dynamically interconnected multiple-beam satellite which could result in a cost reduction in the earth segment high power amplifier (HPA) and antenna costs.

4. Launch Costs: Even in the shuttle era, mass will remain a constraint and cost factor. Depending on the nature and function of the on-board processor, LSI chips can be used, combined with other microprocessor technologies, to produce a lightweight component. It should be possible to introduce significant processing capability ~~in the 1980~~s without suffering a significant launch cost penalty.

The processor will also impact upon other System design characteristics ~~SDC~~'s, such as orbit utilization and flexibility, in indirect ways. In all cases, however, the resultant overall impact should lead to an improved and more versatile system design.

There are three generic types of processors--RF, bit stream, and f[ull] baseband. Each serves a different role, has different design factors and provides increasing levels of flexibility. The applications to be served will play a significant role in determining the type of processing that is most suitable for a particular system design.

1. RF Processor: An RF processor deals directly with the uplink RF signal and does all processing at RF. No translations of the signal to a low frequency IF or baseband need be performed, although a simple frequency translation may be used to achieve operation at a convenient frequency. In its simplest form, an RF processor would be a satellite switch matrix having n inputs and m outputs with the capability of connecting any input to

any output. Such a switch could be used in a multiple-beam satellite configuration and provide full connectivity for a TDMA scheme.

The basic design factors for such a processor are:

i. Frequency--The frequency at which the RF signal enters and leaves

ii. Bandwidth--This is related to the data rate or bandwidth of the signal being switched

iii. Mass--A lightweight design is desirable

iv. Reliability--Such a processor should have a long life and should have a fail-safe mode; that is, if a failure occurs, it should not cause major degradation in system performance.

v. Power--The process should use a minimum of power and be active only when processing the signal.

vi. Speed--This relates to the rate at which the switch can change state.

Such processors would still provide a transparent link--one that does not differentiate on the mode of access or modulation.

2. Bit Stream Processor: The bit stream processor has the capability of taking the RF signal, translating it to IF, demodulating the digital bit stream, and then using the demodulated bit stream to remodulate a carrier at IF or RF. The remodulation operation would be done on a symbol by symbol basis using hard decision logic. No use would be made of the signal structure to control the processing and a bit stream processor has essentially no storage capability. Such a scheme could, for example employ a DPSK uplink and a coherent PSK downlink in an attempt to reduce the cost of the earth station's transmitting equipment. The design factors discussed above, such as frequency, mass, power and bandwidth would be of concern. The major difference between this form of processor and the RF processor would be that now the link would not be transparent. The uplink and downlink would be uncoupled. The TWTA would be using its power to transmit just the remodulated signal and not the satellite front end noise. The noise effects would be exhibited in the form of hard decision processor errors. In addition, the processor would probably be matched to a signalling format that would be fixed for the lifetime of the satellite. In this sense, the satellite would become a "digital" satellite.

3. Full Baseband Processor: The most complicated form of processor would be a full baseband processor. In the general case of this realization, assuming digital transmission, the RF signal would be demodulated, decoded using soft decisions and then re-encoded and remodulated. In addition, the processor could use information in the decoded signal to route the messages in a dynamic fashion or otherwise specify, at least in part, the processin

   A potential application of this form of processor would be in a mobile satellite system, where there are many users with low gain antennas and low power transmitters. In this case, the uplink could be a random access mode with low carrier power to noise spectral density ($C/N_O$). By using the processor, the efficiency of the access scheme could be increased and an optimal use of uplink $C/N_O$ could be made. In addition, if the downlink is to a few large earth stations, TDM may be employed via multiple beam antennas. The processor would provide the routing and provide the required change in modulation.

   Another scheme involves the use of FDMA on the uplink and TDMA on the downlink. This would simplify the earth station complexity, make optimum use of the satellite transmission capability, and timing could be centrally provided by the satellite.

   A third area of application would be in computer communications. In this case, the processor would act as a message switch or processor for a packet-type network.

At the present time the only existing, but non-operational, processor is a 16 x 16 RF satellite switch that has been developed for possible use in the INTELSAT system.[84] Figure 17 is a picture of this switch. However, with the rapid developments in LSI technology, microprocessors, and surface acoustic wave (SAW) and charge coupled devices (CCD), an increased technology base will be available to develop both bit stream and full baseband processors.The following is a list of example performance goals for the three types.

1. RF Processor: A high data rate switch would be useful. The switch should have the capability of having a high degree of input-output interconnectivity and be reliable and fail safe.

2. Bit Stream Processors: Presently under development are DPSK to PSK converters. Another potential application would be FSK to PSK coversion.

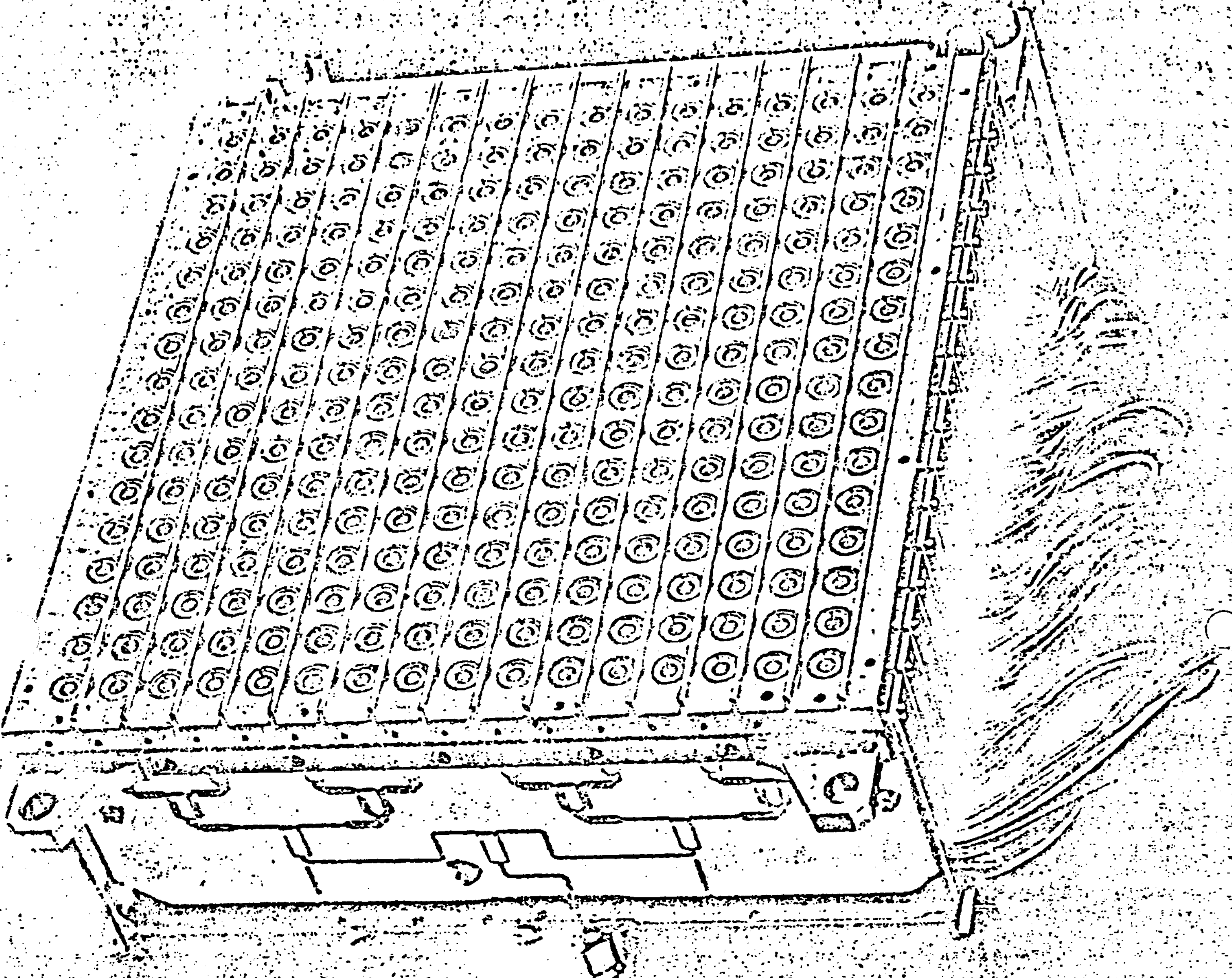


Fig 18: 16x16 RF Satellite Switch

3. Full Baseband Processors: Suitable soft decision decoders -- such as a Viterbi decoder, with variable code and data rates and reprogrammable, to accommodate dynamically varying link characteristics--should be developed. Also, a fully switchable satellite switch to individually route messages or bit streams appears to have a high potential system payoff. Other applications include FDM to TDM conversion and reformating of TDMA bursts to change rates or provide error protection for small stations.

We now will discuss the implementation of these three differing on-board processors. For the RF processor, the technology requires a simple RF switch system which can connect inputs to outputs. This is usually done by employing diode switch matrices with control circuitry to activate the switch. Fig. 19 depicts a possible configuration. Here an N beam satellite uses an NXN switch to connect the outputs to different beams. In this case $B_i$ can communicate to $B_j$ at some instant. Shown is the diode switch matrix. Interlocks are used to ensure that only one input connects to a single output.

This first approach has some limited use. As we shall see in the next chapter, the processor requirements depend strongly on the traffic data matrix. A more flexible RF switch is shown in Fig. 20. Here the switch is large enough, NKxNK, to permit transponder to transponder connection. As is shown for a 12 channel per beam, 12 beam satellite, this is a 144 x 144 switch. This would require 20,736 diodes. Such a switch is commercially feasible for earth use but the state of the art is still progressing in the satellite area. What this switch buys compared to the first is increased flexibility. Again this must be matched against the traffic requirements. Also one must design the system for a fail soft mode so that if the switch fails, the entire satellite is not made inoperative.

Fig. 21 depicts a bit stream processor. In this case we have taken each beam and obtained by filtering each channel. Then each channel is demodulated, switched, remodulated, and transmitted. Baseband configurations of this type can include both modulation and encoding. Fig. 22 depicts four possibilities. We will not consider them in detail. In Fig. 22a we have what is commonly called a regenerative repeater. The data may or may not be encoded. This technique allows for two modulation formats on uplink and downlink. This may be useful from cost points of view or from performance also. A second use for this demod/remod is to decouple up link and down link performance. As an example let us consider a PSK system which is uncoded with a BER of

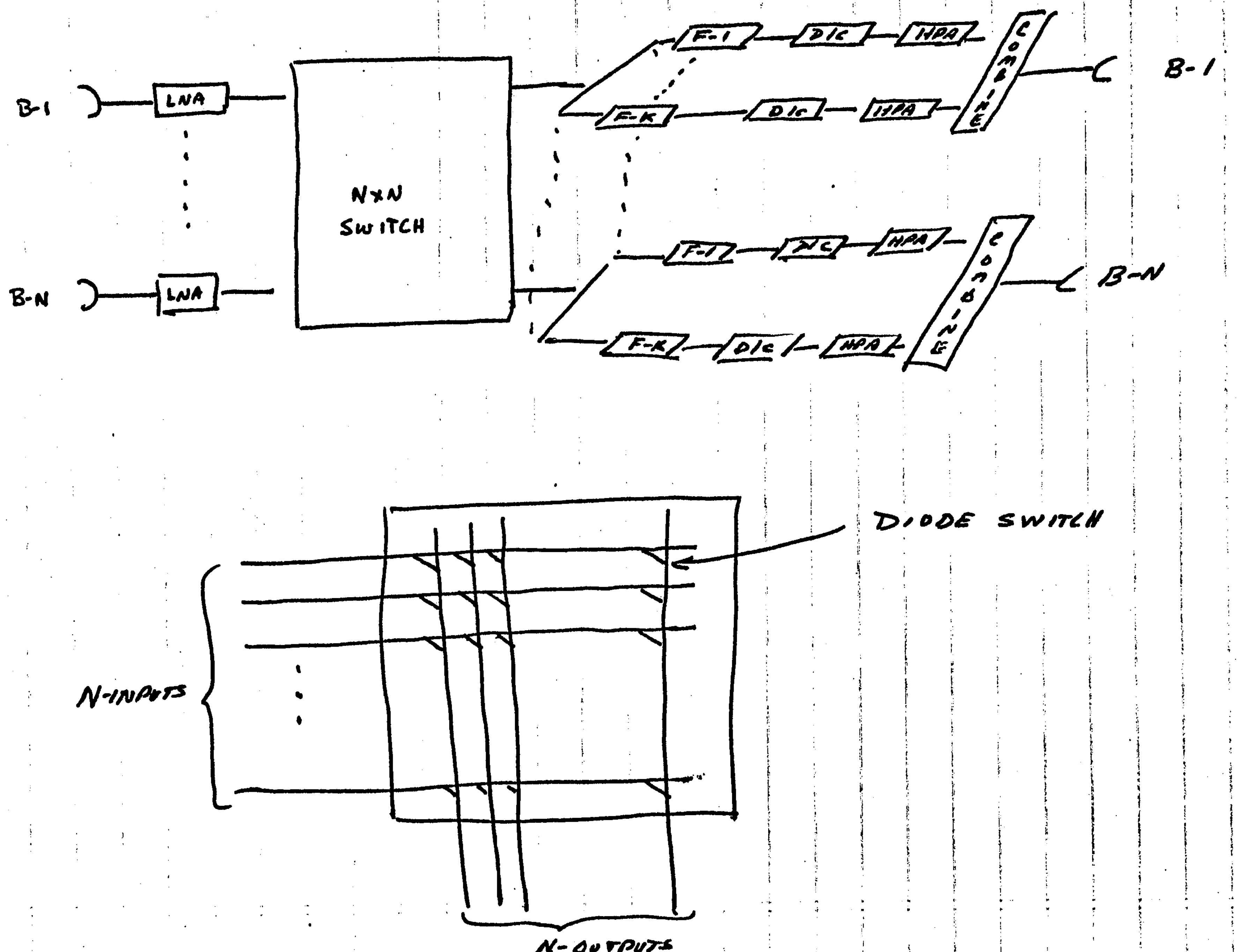
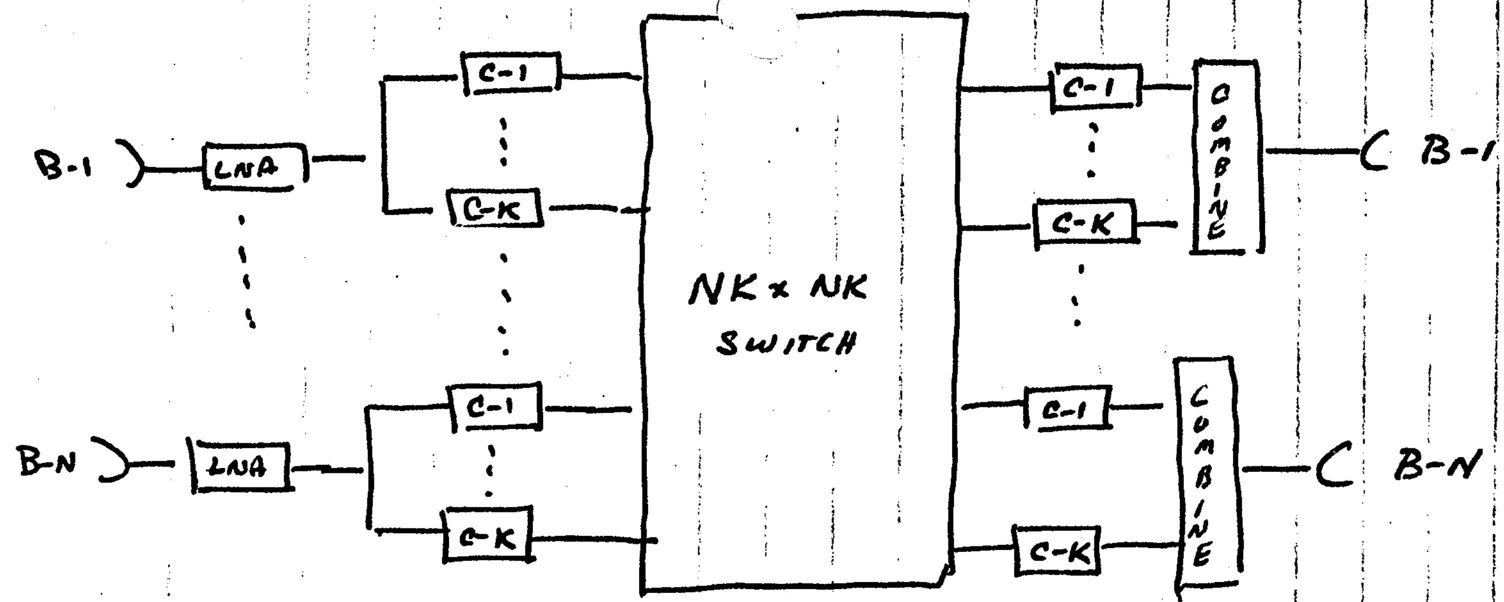



FIG 19 : RF SATELLITE SWITCH




Example: 12 Channels (K), 12 Beams (N) } 144x144 switch

FIG 20 : FULL RF SWITCH

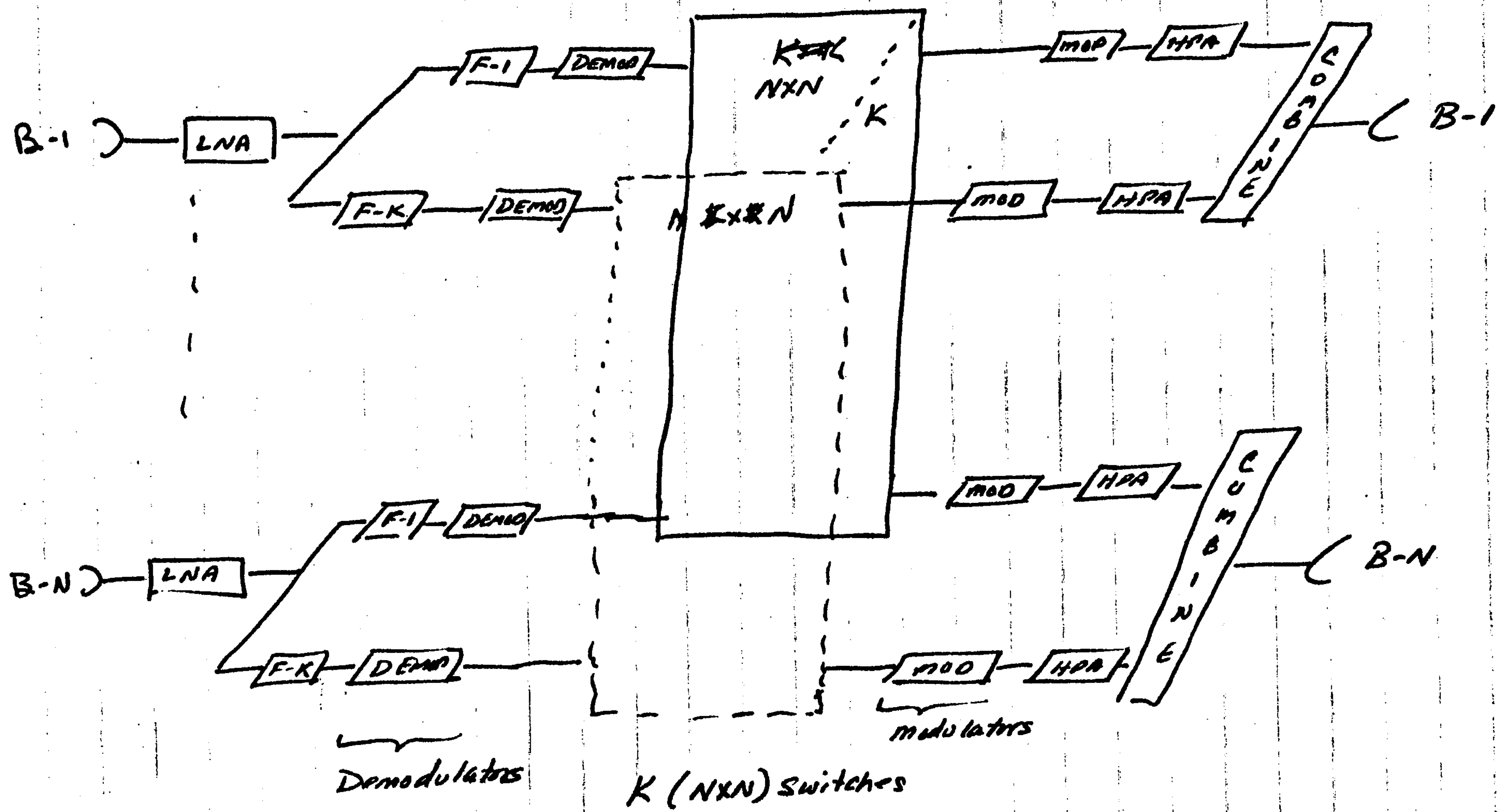



FIG 21 : Baseband Switch

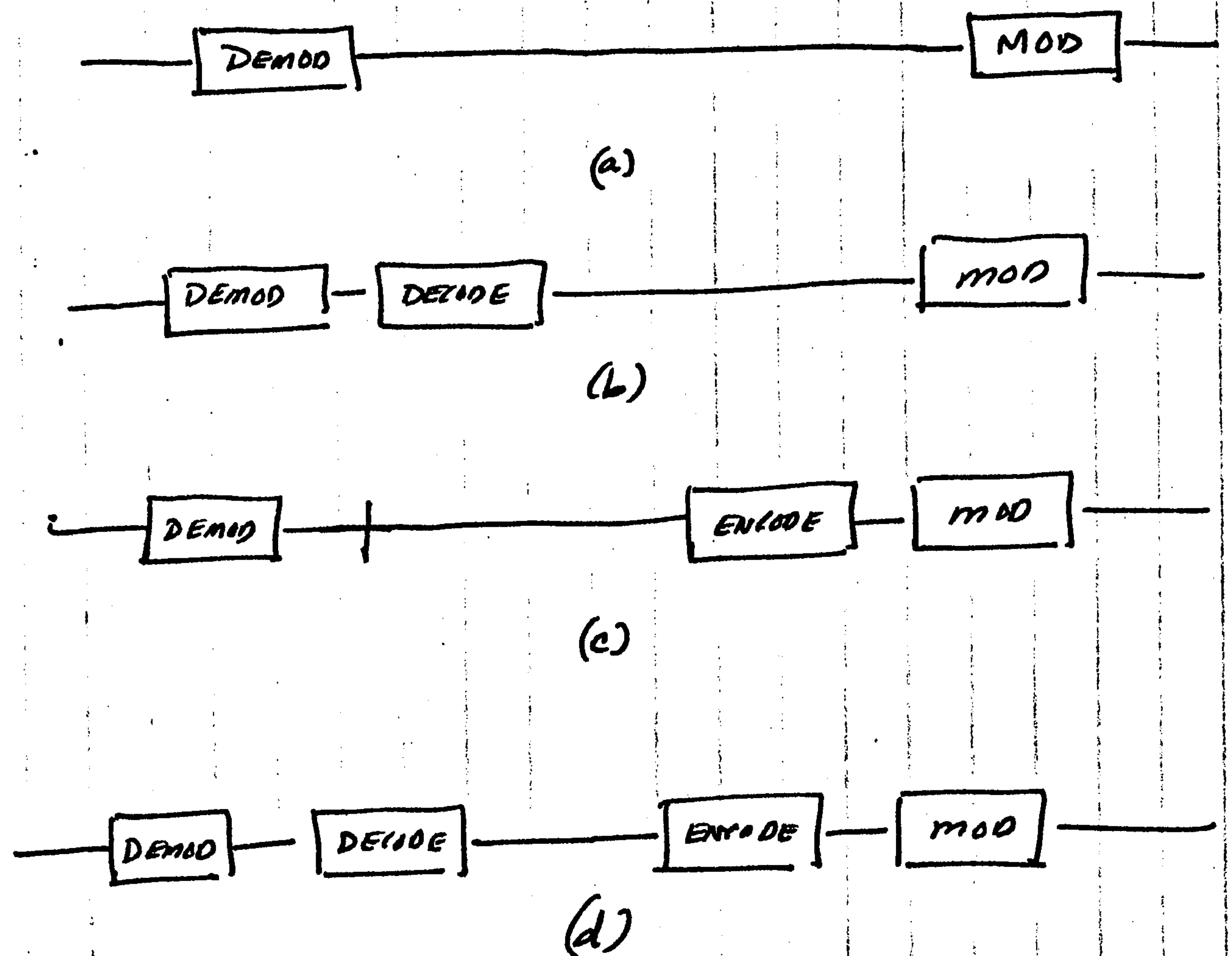



FIG 22 : Full Baseband

$$P[E] = Q\left(\sqrt{\frac{2E_b}{N_o}}\right)$$

where $E_b/N_o$ is the energy per bit to noise spectral density on the link.

Now let $(C/N_o)_u$ be the carrier power on the uplink to its noise spectral density. Then if R is the data rate in bits/sec;

$$\left(\frac{E_b}{N_o}\right)_u = \frac{1}{R}\left(\frac{C}{N_o}\right)_u$$

Similarly on the down link

$$\left(\frac{E_b}{N_o}\right)_d = \frac{1}{R}\left(\frac{C}{N_o}\right)_d$$

Now for a link that just amplifies we can show that

$$\left(\frac{C}{N_o}\right) = \left(\left(\frac{C}{N_o}\right)_n^{-1} + \left(\frac{C}{N_o}\right)_d^{-1}\right)^{-1}$$

u

Show in an appendix or give a reference!

or

$$\frac{E_b}{N_o} = \left(\left(\frac{E_b}{N_o}\right)_n^{-1} + \left(\frac{E_b}{N_o}\right)_d^{-1}\right)^{-1}$$

Now if we have two links in tandem the error probability is

$$P[E] = 1-(1-P_u[E])\cdot(1-P_d[E])$$

$$= P_u[E] + P_d[E] - P_u[E]P_d[E]$$

where $P_u(E)$ is the uplink BER and $P_d(E)$ is the downlink BER.

Thus we have P[E] for the nonprocessing satellite of

$$P_1(E) = Q\left(\sqrt{\frac{2E_b}{N_o}}\right)$$

and for the processing satellite.

$$P_2(E) = Q\left(\sqrt{2\left(\frac{E_b}{N_o}\right)_u}\right) + Q\left(\sqrt{2\left(\frac{E_b}{N_o}\right)_d}\right) - Q\left(\sqrt{2\left(\frac{E_b}{N_o}\right)_u}\right) Q\left(\sqrt{2\left(\frac{E_b}{N_o}\right)_d}\right)$$

lower the 2's! i.e. $\sqrt{2\left(\frac{E_b}{N_o}\right)_u}$

Let us now briefly consider two cases. First where both links are the same,

$$\left(\frac{E_b}{N_o}\right)_n = \left(\frac{E_b}{N_o}\right)_d = \alpha$$

Then;

$$P_1(E) = Q\left(\sqrt{\alpha}\right)$$

$$P_2(E) = 2Q(\sqrt{2\alpha})\,(2-Q(\sqrt{2\alpha}))$$

Thus if $\alpha \gg 1$ we see that the processing loop has almost a 3dB advantage. The factor of 2 becomes negligible. This is directly seen by noting that;

$$Q(\sqrt{2}) \cong \frac{1}{2} \exp\ (-\ \alpha/2)$$

Thus

$$\frac{P_2(E)}{P_1(E)} \simeq 2\ \exp(-\alpha + \frac{\alpha}{2})$$

$$= 2\ \exp(-\ \frac{\alpha}{2})$$

This 3dB gain can be signfiicant. The second case occurs where either link is dominant. That is;

$$\left(\frac{E_b}{N_o}\right)_u \gg \left(\frac{E_b}{N_o}\right)_d$$

or vice versa. In that case

$$\frac{E_b}{N_o} \approx \left(\frac{E_b}{N_o}\right)_d$$

so that

$$P_1(E) \approx P_2(E)$$

Thus if any one link dominates the gain is lost and performance is determined by the worst link.

These factors can be balanced by using the schemes in Fig. 22 b and 22c. Here if the uplink is poor we use coding and then decode and then transmit on the downlink. If the down link is poor we encode and decode as the down link to ensure overall performance.

The last configuration includes coding and decoding. Now it should also be mentioned that one may have an encoded uplink that is further encoded on the downlink. These techniques are used on military satellites and are used specifically for security and conti-jam purposes.

The final version of a full baseband processor is shown in Fig. 23 . Here we use both input and output buffers and there is one for each beam/transponder pair (NK in all). Data arrives, with routing information appended. The data in the buffers is already in a packetized form. The processor sorts through the buffers routing input to output and controlling the flow.

In this case, it is assumed that the data has been demodulated and packetized into data packets that contain data link control information as well as path control information. The processor must be capable at a minimum of scanning the routing data and schedule it accordingly.

The analysis of a processor of this type is considerably more complex than the others. The data into the queues is coming in at a constant rate but on the basis of the outgoing partitions it can be considered random. The service time at the processor could be considered constant. The model then could lead to an analysis based on queueing theory. Here performance is given in terms of delay and probability of buffer overflow. That analysis is beyond the scope of this text. A sample of this approach is covered in the paper by DeRosa.

This last approach also permits change of both modulation and multiple access on the down link. As we shall discuss, the uplink could be FDMN and the downlink TDMA. Many such combinations are possible.

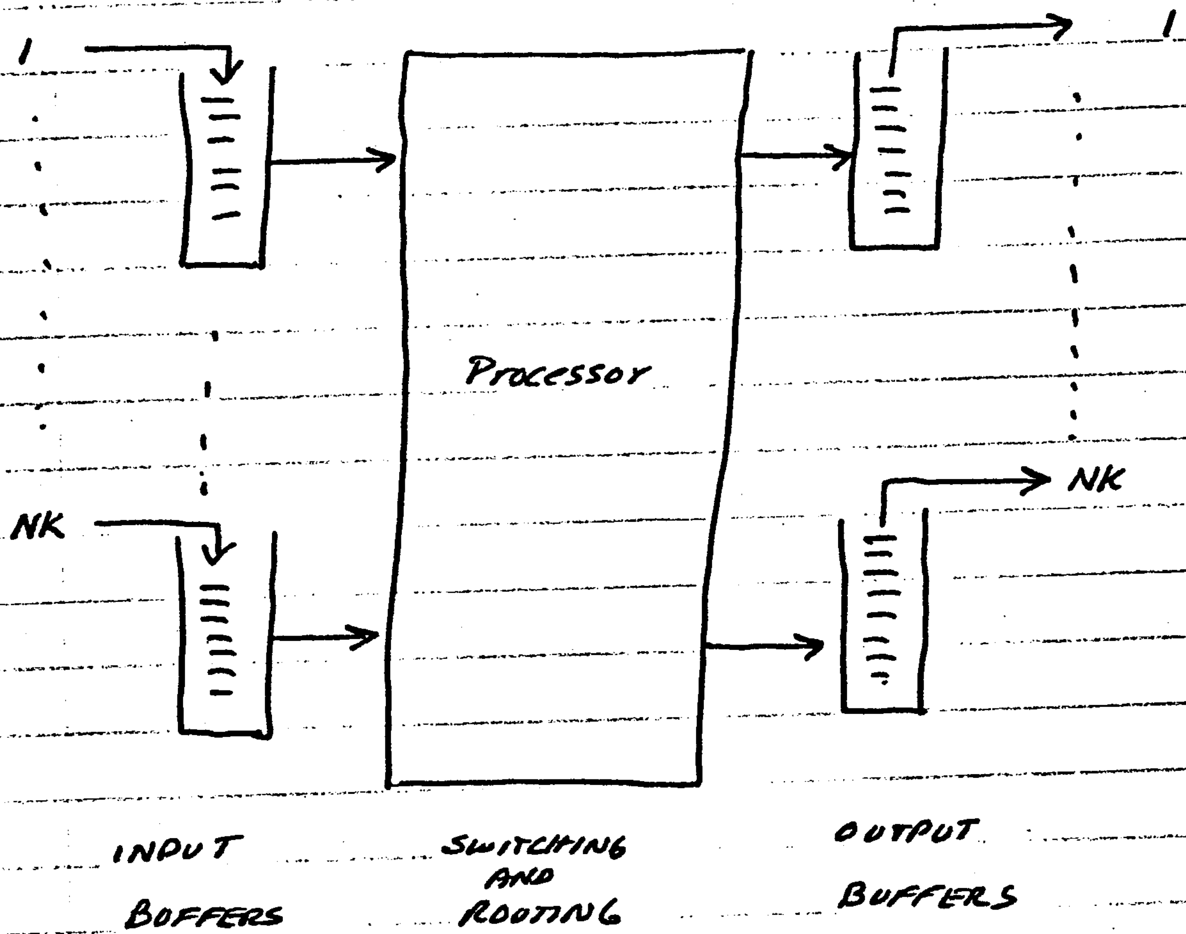



FIG 23: Generic Structure of Processor

## ~~4.5~~ 5.4 Intersatellite Links

An intersatellite link (ISL) is a two-way communication channel between two or more satellites. It provides a connection between users who access two different satellites in the same or different orbits. In so doing, it can significantly reduce the number of additional accesses needed, or provide connectivity between users who cannot see the same satellite.

The first reported intersatellite link was made in January 1975 between Oscar 6 and Oscar 7, the two VHF amateur radio satellites.[85] Later that year, a link was established between SPACELAB and ATS-6. In February 1976, the LES 8 and LES 9 (Lincoln Experimental Satellites; Massachusetts Institute of Technology, Lincoln Laboratory) established a link between two three-axis stabilized synchronous satellites at 36 GHz.~~[14]~~

An ISL capability can impact significantly upon system performance. The following five system design characteristics are most directly affected.

1. Connectivity: The ISL can be used to provide flexibility in satellite system design while still achieving a required level of interconnectivity within a community of users. Typically the total community of users will be composed of subgroups which form strong communities of interest; i.e., have a lot of internal traffic with a more modest, but non-zero, traffic requirement between these subgroups. Normally, if full interconnectivity is to be provided, then the total community must access a single large capacity satellite. An ISL capability allows subgroups of the total community to access separate lower capacity satellites, with the ISL accommodating the inter-community traffic so that the capability for full connectivity is retained.

2. Capacity: As described above, the provision of full connectivity in the usual fashion requires a large capacity satellite. Further, a part of this large capacity may be wasted because of the lower fill factors which normally

occur as the result of meeting the interconnectivity requirement. When an ISL is used, a higher fill factor or satellite use efficiency can be attained by choosing communities of interest on the separate satellites to maximize this factor.

3. Coverage: If a satellite system is to provide global coverage, then users in one region must be capable of talking to those in regions not covered by their set of satellites. This may be accomplished by terrestrial means or by "double hopping" the signal; that is, sending it up to one satellite and down and then repeating the process. This "double hopping" requirement can be eliminated by use of an ISL, with a resultant decrease in the transmission delay.

4. Orbit Utilization: By allowing more flexibility in orbit position without large connectivity and capacity penalties, the ISL can increase orbit utilization.

5. Cost: The cost of the earth segment can potentially be significantly reduced by the increased connectivity provided by the ISL. This results from a reduction in the number of earth stations required.

The ISL can also be used in a mobile communication system. One of the major problems with a mobile system is cost of the terminal equipment. A low powered small antenna mobile unit could easily transmit to a low orbiting satellite. That satellite could then use an ISL to uplink to a geosynchronous satellite and retransmit via a large gain antenna. The low orbit satellite could be used to reduce path losses and reduce costs of the earth segment. In turn, it can increase the capacity of the system.

An ISL system is composed of four generic subsystems--receiver, transmitter, acquisition and tracking, and an antenna/lens system. An optimized design is developed by tradeoffs among these subsystems. The subsystems perform the following functions.

1. Receiver: The receiver, in the most general case, detects the signal and demodulates and decodes the received information. It also provides the interface between the ISL and the normal mode of satellite transmissions, including any required format conversation. In some realizations, there will be no decoding and minimal format conversion.

2. Transmitter: In the general case, the transmitter selects those information signals that are to be sent across the link, encodes them and then provides modulation, frequency

translation, and amplification. In some systems, the transmitter may simply remodulate an appropriate high power source.

3. Tracking and Acquisitions: When two satellites are placed into synchronous orbit, the two systems must acquire each other and then track to an appropriate accuracy ($\approx$ 0.1 of the beamwidth) as the information is transmitted across the link. Such techniques as raster scan and monopulse can be used in this subsystem. It should be noted that it is not necessary to orient the entire spacecraft, to the required accuracies, but only a portion of the ISL subsystem; e.g., a lens, reflector, or antenna feed.

4. Antenna/Lens Subsystem: The aperture used to interface with free space depends upon the wavelengths used for transmission. For standard microwave or millimeter-wave transmissions, an antenna system such as a reflector or lens would be adequate. For optical wavelengths, an optical reflector or lens system would be required. The key differences are, of course, aperture sizes, surface tolerances and materials.

There are two frequency ranges that appear attractive for ISL operation--millimeter wave and optical. In both of these ranges there are sources available although there is a significant need for source technology improvement. In the millimeter band of 55 - 60 GHz, allocated to ISLs, there are TWTAs that can be used in a standard fashion. The tubes have in excess of 10 W peak output with about 10 - 15% efficiency. In order to provide low noise reception, a cooled paramp, which can yield a receive noise temperature of 750 K, is available. Present uncooled receiver noise temperatures for the 55 - 60 GHz range are less than 2500 K. However, using a passive cooler, this can be reduced. In the higher millimeter ISL bands, above 100 GHz, no strong technology base exists.

The optical frequency range is currently under development. For wavelengths in the range of .5 to 10.6 μm, increased antenna gains, increased path losses and changes in sources and detectors occur. The most promising sources in these frequency ranges are all lasers with coherent emissions. At the present time, there are four main candidates.

1. $CO_2$ (10.6 μm): The source here is a gas laser in the infrared region at 10.6 μm. This laser which has undergone considerable development, much of it under the support of NASA, is capable of delivering several watts of power and has high efficiency (~ 10%). Typical modulators use cadmium telluride crystals, often in an intracavity configuration to reduce the required driver power. Large bandwidths are available, on the order of 300 MHz;

and detection is performed by heterodyning with another laser and using a diode type detector, which must be cooled to approximately 110 K. Lifetimes of the $CO_2$ tubes are now in excess of 10,000 hours and 50,000 hours is expected by 1980. Bandwidths on the order of 5 to 10 GHz should be possible with improved modulator designs and advances in detector technology.

2. Nd:YAG (1.06 μm): The Neodinium Yttrium Aluminum Garnet (YAG = $Y_3Al_5O_{12}$) laser is similar to the $CO_2$ in that it has a high power output, but it is less efficient and is a solid rather than a gas. The Nd:YAG laser is presently being used in the U.S. Air Force 405B program.[86] Figure 18 shows an engineering model of the 405B transmitter. The Nd:YAG laser is externally pumped by a light source in an appropriate frequency band. Pump sources include the sun, low pressure discharge lamps, and solid state light-emitting diodes and lasers. Pump lifetime is a major problem when lamps are used. The use of sun-pumping puts some significant constraints on the satellite. Figure 19 shows a potential satellite configuration for a solar-pumped Nd:YAG ISL.

3. HeNe (.63 μm): This is the oldest source. It is a gas laser with very long lifetimes, but low output power (on the order of milliwatts) and low efficiency ($\approx 0.01\%$). It is useful for low data rate (< 100 Mbs) links that span reasonably short paths (< 10° of synchronous arc).

4. GaAs (0.9 μm): Gallium arsenide or GaAlAs is a solid state laser of high efficiency, potentially long life-time and modest output power, at least at present. At the present time the production yield is low, but this should improve. Because it can be directly modulated, tuned to various wavelengths by doping and is amenable to direct detection techniques, GaAs is an attractive source for those applications where it has sufficient power.

A direct comparison of optical and microwave receiver performance is not straightforward because millimeter receivers have standard kT noise; whereas, the fundamental noise for optical receivers is quantum in nature and depends on the detection technique.[87] Thus care must be taken in comparing various optical and microwave systems.

~~It is possible that th~~e frequency bands currently allocated for satellite use ~~will~~ became available, on a shared basis, for ISL use at the 1979 WARC. This ~~would~~ opens some new possibilities for ISL implementations that might be quite promising, particularly in the near term. It would, for example, allow the possibility of a 12/14 GHz ISL on a 12/14 GHz satellite, with an attendant increase in the uniformity of spacecraft implementation and potentially useful operational flexibility.

-- A decreased reliance on millimeter wave systems is foreseen as optical modulators and detectors become efficient and optical systems demonstrate their performance capabilities and reliability in space.

Given the above description of ISL technology, we now go through two examples of ISL systems to provide a measure of the performance that can be expected. In all cases we are assuming that the data rate is 1000 Mbps of 4Ø-PSK and for simplicity we also assume it to be un~~en~~coded. In terms of architecture, we also assume that the ISL channelization is equivalent to any other downlink beam on transmit or uplink beam on receive.

The system geometry is shown in Fig. 24. The ISL range depends on the angular separation and is

$$d = 2R \sin \frac{\theta}{2}$$

At a $10^{\circ}$ separation we have

$$d = 7363 \text{ km}$$

This range is fairly short compared to the synchronous orbit range. Greater separations are possible.

<u>Example</u>:

In this example we shall consider an RF link where we are using a 60 GHz band. The antenna is 1 foot in diameter on both transmit and receive. Thus the gain is (at 50% efficiency)

$$G = 48 \text{ dBi}$$

Define dBi

The 3dB beamwidth of this antenna is

$$BW = 0.95^{\circ}$$

The ISL antenna beam is composed of two sub beams, with their sum being used for communications. A difference beam is also generated. This beam is used for tracking. The tracking loop functions as follows. If the beam is centered, then the output is zero and it indicates a zero error voltage. If the loop is off center the voltages are positive or negative and this is used to drive the antenna back to zero. This is done in a monopulse tracking loop (Fig. 25).

This is a very cursory discussion. More details would be useful!

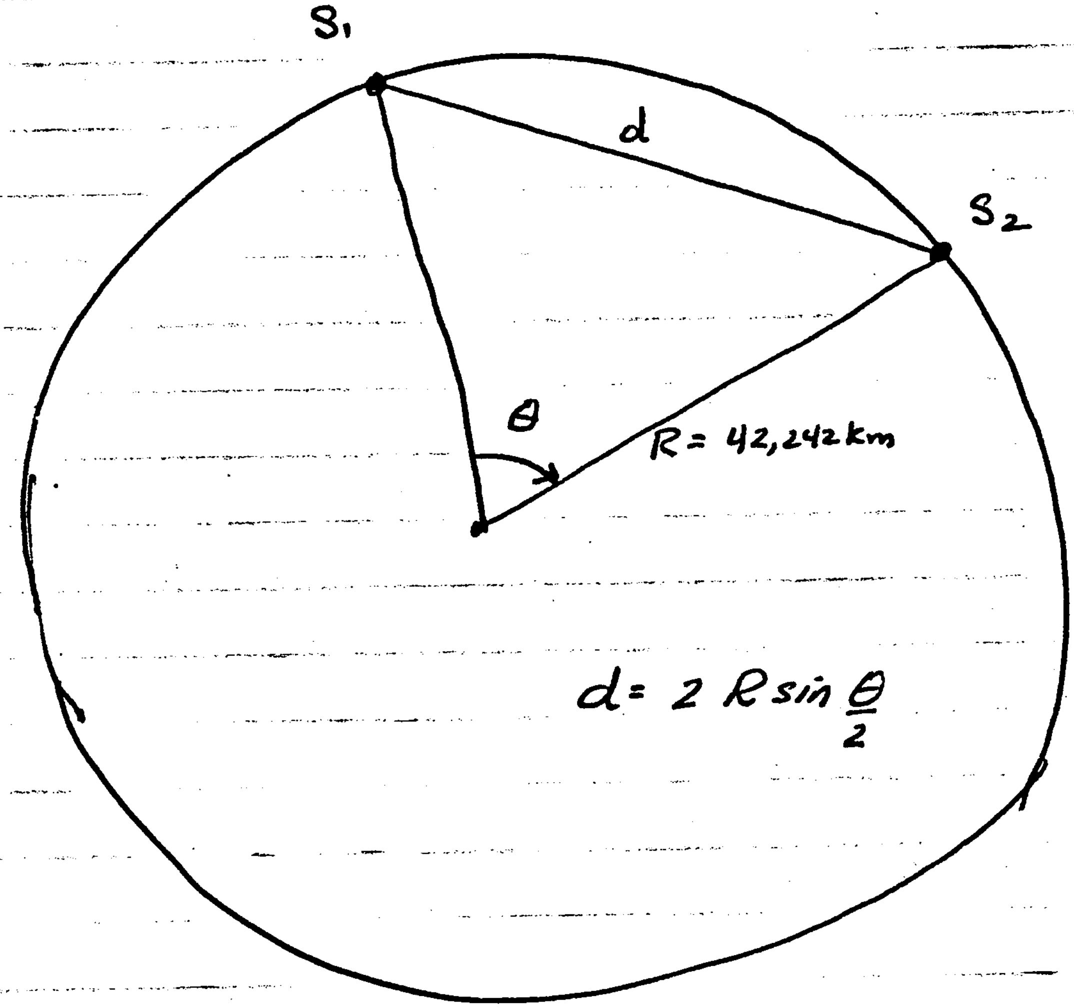



Fig 24 : ISL Geometry

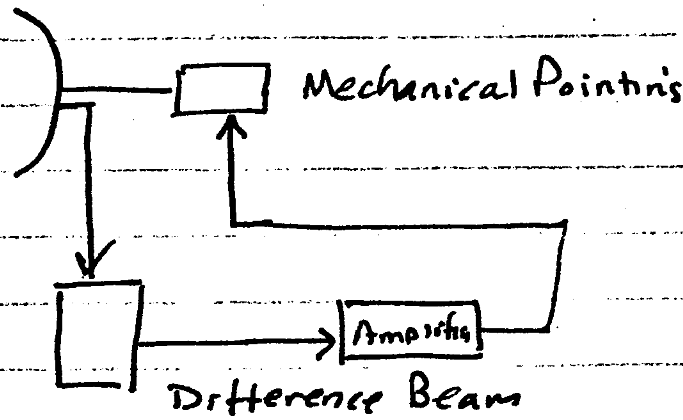
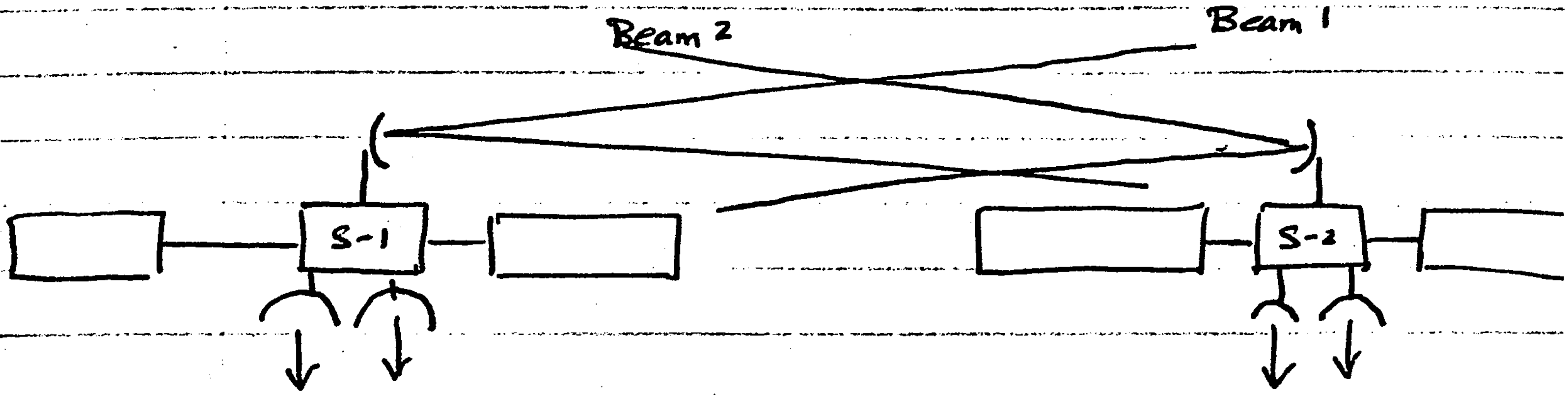



Fig 25 : Monopulse Tracking Loop




(a)

Satellite Allignment

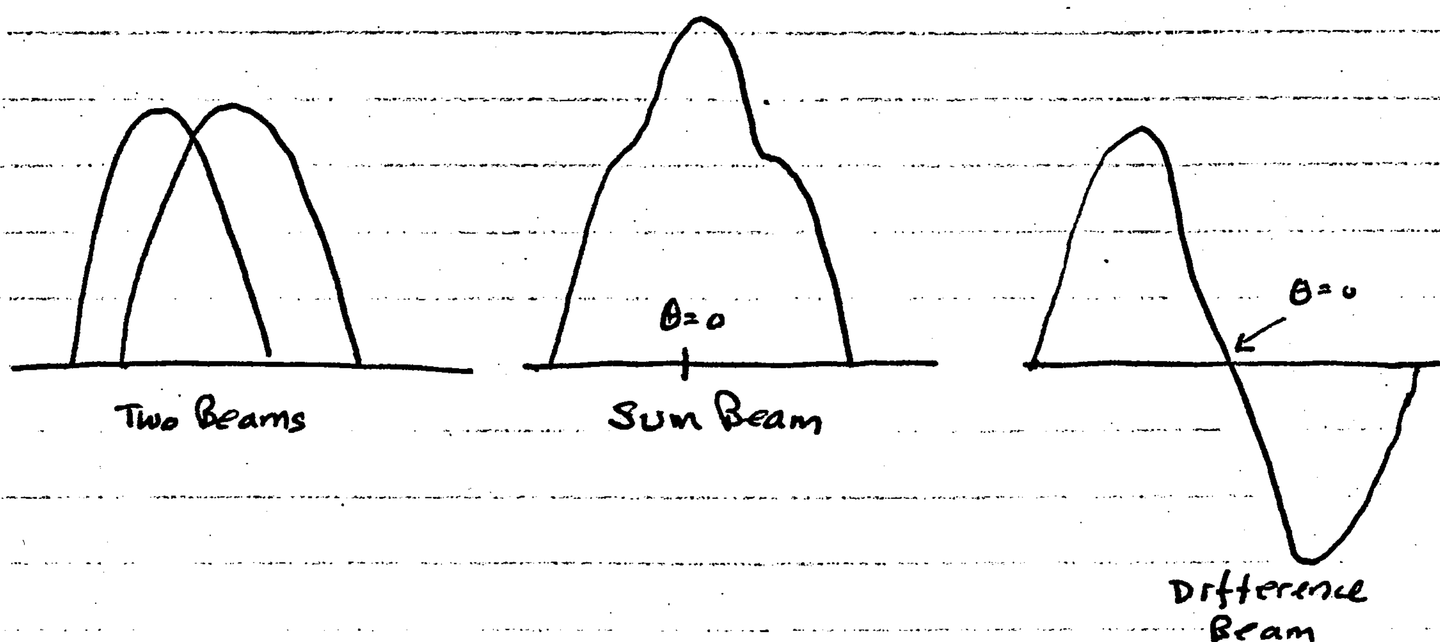



Fig 2b : ISL Pointing

The initial acquisition of these systems is also fairly straight forward. The beam of satellite 1 (S1) is pointed towards S2 and defocused to reduce gain and increase beamwidth. Then S2 antenna points toward S1 in a monopulse fashion to lock in. Then S1 focuses and also uses the tracking loop to look on. A roster scan search pattern may have to be employed depending on the initial uncertaintly.

Now returning to the data link we can determine the design parameter in a simple fashion. With this modulation scheme we need an $E_b/N_o$ to meet the BER. Thus we use this plus a margin of say 6dB. That is if $(E_b/N_o)^*$ is the desired value then

$$10\log_{10}(E_b/N_o) = 10\log_{10}(E_b/N_o)^* + 6\text{dB}$$

Req'd value is greater than theoretical (ideal) value by at least 2 dB! State this fact!

But $E_b/N_o$ can be calculated from the ISL parameters. Recall that

$$N_o = kT_e$$

where $T_e$ is the effective noise temperature of the ISL front end. At this frequency, $T_e$, will be in the range of $1000^{\circ}$K to $2000^{\circ}$K even with cooling.

$E_b$ is the energy per bit. R is the rate in bits per second or 1/R is the rate in seconds per bit. If P is the power received, then the energy per bit is $E_b = \frac{P}{R}$

The received power is calculated knowing the link. Let $P_s$ be the power of the ISL TWTA and G the antenna gain. Then the power emitted is

$$P_sG$$

Reference?

The free space attenuation can be shown to be

$$\Gamma = \left(\frac{\lambda}{4\pi d}\right)^2$$

so that the received power using a receive gain of G is

$$P_r = P_sG^2\left(\frac{\lambda}{4\pi d}\right)^2$$

Thus for the ISL link;

$$\frac{E_b}{N_o} = P_s(\text{dBw}) + 2G(\text{dBi})$$
$$- 20 \log_{10} \left(\frac{4\pi d}{\lambda}\right) + 228.6 \text{ (dbw)}$$
$$- T_e(\text{dB}^o\text{K}) - R \text{ (dB-Hz)}$$

From this we can obtain $P_s$ as required. This is called the link budget and we will develop it more fully later. For this example $P_s$ is ___?___ w. ?

Example:

In this example we consider an optical link. To perform this analysis we must introduce certain concepts about optical signals that relate to transmission and detection. We will consider four optical sources as discussed previously.

The optical system is presented in Fig. 27 where we have first a device that converts the modulated signal to an on-off signal used to modulate the laser transmitter. Thus a burst corresponds to a 1 and no burst to a 0. This then feeds a lens system with an aperture that radiates into space. The radiation and propagation follow the same principles as do RF signals. Thus a large aperture has a narrow beam and a high peak gain in a direction towards the receive end. We define the antenna gain as

typically off is 5% of on!

$$G = \left(\frac{\pi D}{\lambda}\right)^2$$

where D is the antenna diameter. As before the path loss is

$$\Gamma = \left(\frac{\lambda}{4\pi R}\right)^2$$

We now want to consider the on-off link of the ISL and determine error rates on that link. Again as a rule of thumb design goal if P[E] on this link is an order of magnitude below the overall required, we can then be assured that the ISL will not degrade performance.

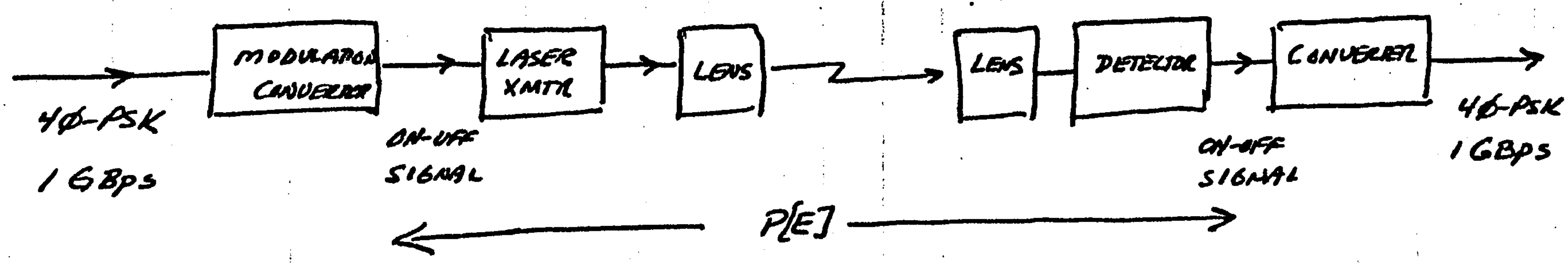



FIG 27: OPTICAL LINK ELEMENTS

Let us now more ~~clearly~~ look at the propagation of light in more detail. At low densities, what is received is really a photon stream. Each photon has energy of hf where h is Planck's constant ($6.63 \cdot 10^{-34}$ joule-sec) and f the frequency. Thus if the received power is P, the average number of photons per second, $\lambda$, is given by

$$\lambda = \frac{P}{hf} \quad \text{(photons/sec.)}$$

The detector is a device which, in effect, counts photons. Such an implementation is a photomultiplier, which converts a single photon, by a multiplier chain, into a single current burst. Thus the detector is a multiple event device. The demodulator part of the detector then has to determine whether the signal is a 0 or 1 in the bit interval. This decision process is different than that of signals in Gaussian noise.

In the optical case, if a signal has power P then it gives rise to photons at rate $\lambda$. It can be shown that N(T), the number of photons in interval (O,T) seconds is given by a Poisson distribution. Ref.?

$$P[N(T) = n] = \frac{(\lambda T)^n}{n!} \exp(-\lambda T)$$

This equation assumes the photomultiplicat (M) is unity I believe. If M>1 a different result applies.

What this says is that even with just signal present, the measurement is random. This contrasts with the RF case where the signal is deterministic With on-off signalling we have

$P = 0$ for 0

$P = P_s$ for 1

Thus either the signal is present or absent. If no other signals are present we use a count of $N(T) = 0$ as the threshhold. That is the demodulator says

If $N(T) = 0$ say 0

If $N(T) > 0$ say 1

What about thermal noise, dark current etc This seems quite unreasonable!

However we see that an error can occur if we sent 1 and do not detect a count. This is

$$P[E] = P[E/1]\, P[1]$$

$$= \frac{1}{2} P[N(T)=0/P_s]$$

Poisson

or using the Poisson result;

$$P[E] = \frac{1}{2} \exp(-\lambda T)$$

But

$$\lambda = \frac{P}{hf}$$

and T equals 1/R the rate. Thus

$$\lambda T = \frac{P}{R} \frac{1}{hf} = \frac{E_b}{hf}$$

where $E_b$ is the energy per bit. Therefore, in this case called the quantum limited case.

$$P[E] = \frac{1}{2} \exp\left(-\frac{E_b}{hf}\right)$$

This assumes that the device is perfectly efficient. If however it has an efficiency of ~~χ~~ η we need to reduce λ ~~to ηλ~~ by η. Thus

η eta not y!

$$P[E] = \frac{1}{2} \exp(-\eta E_b/hf)$$

is the final expression.

As an example, let us assume R of $10^9$ bits/sec and a range of 2275 n.mi. Let P[E] be selected as $10^{-6}$. Table 1 depicts four sources, their wavelengths, antenna size, efficiency and the received power required to attain the resultant P(E). Table 2 depicts the full link characteristics. What is important to note is the low level of power and the very narrow beamwidths.

It is also possible to have a set of noises due to either the devices itself, called shot noise, or due to background radiation. The background radiation has a power density that follows the blackbody radiation curve of a body at temperature $T_e$. The power density is

hf

$$S_n(f)\,(\mathrm{W/Hz}) = \frac{hf}{\exp(hf/kT_e)-1}$$

exp

If $hf/kT_e \ll 1$ then as before

$$S_n(f) \cong kT$$

so that the background power level in bandwidth B is

$$P_n = kTB$$

telescope diameter?

| Wavelength (micrometers) | Source | Modulation Format | Antenna Diameter | Receiver Structure | Detector Quantum Eff. η | Required Received Power (watts) |
|---|---|---|---|---|---|---|
| 10.6 | $CO_2$ | Binary on-off | 6" | heterodyne | 0.5 | -87 dBw |
| 1.06 | NdYAG | " | 6" | direct detection | 0.01 | -66 dBw |
| 0.53 | Doubled NdYAG | " | 6" | " | 0.3 | -77 dBw |
| 0.632 | HeNe | " | 6" | " | 0.08 | -73 dBw |

*Table 1: Optical Source and link Characteristics*

| | WAVELENGTH, λ | | | |
|---|---|---|---|---|
| | 0.53 μm | 0.6328 μm | 1.06 μm | 10.6 μm |
| Required Rec. Power (Table II) | -77 dBw | -73 dBw | -66 dBw | -87 dBw |
| Rec. Ant. Gain, $G_R = (\pi D/\lambda)^2$ | 119 dB | 118 dB | 113 dB | 93 dB |
| Xtr Ant Gain, $G_T = (\pi D/\lambda)^2$ | 119 dB | 118 dB | 113 dB | 93 dB |
| Spreading Loss S.L.= $(\lambda/4\pi R)^2$ | -280 dB | -278 dB | -274 dB | -254 dB |
| $(G_T G_R)/(S.L.)$ | -42 dB | -42 dB | -48 dB | -68 dB |
| Margin | 15 dB | 15 dB | 15 dB | 15 dB |
| Required Xtr Power | -20 dBw<br>0.01 watts | -16 dBw<br>0.025 watts | -3 dBw<br>0.5 watts | -4 dBw<br>0.4 watts |
| Noise Power Density per Hz | -176 dBw | -171 dBw | -164 dBw | -194 dBw |
| Xtr Power for SNR 17 dB in a 1 GHz bandwidth with 15 dB margin | -12 dBw | -7 dBw | +6 dBw | -4 dBw |
| 3 dB ant beamwidth | 3.5 μrad=0.0002° | 4 μrad=0.0002° | 7 μrad=0.0004° | 70 μrad=0.004° |
| c/N → | 61 db | 60 db | 48 db | 69 db |

Table 2: System Parameters and Levels for the Optical Link

When this noise exists, at the receiver we have

$$\text{if } 0: \quad P_r = P_n$$

$$\text{if } 1: \quad P_r = P_s + P_n$$

where again $\lambda$ is related to $P_r$ as before. Here we define

$$\lambda_o = \frac{P_n}{hf} = \lambda_n$$

$$\lambda_1 = \frac{P_s + P_n}{hf} = \lambda_s + \lambda_n$$

The detector now works as follows. A threshold is chosen as the count where

$$P[N(T) = k^*/\lambda o] = P[N(T) = k^*/\lambda_1]$$

Substituting;

$$\frac{(\lambda o T)^{k^*}}{k^*!} \exp(-\lambda_o T) = \frac{(\lambda 1 T)^{k^*}}{k^*!} \exp(-\lambda_1 T)$$

or,

$$\left(\frac{\lambda_1}{\lambda_o}\right)^{K^*} = \exp((-\lambda_o - \lambda_1)T)$$

or

$$k^* = \frac{(\lambda_1 - \lambda_o)T}{\ln\left(\frac{\lambda_1}{\lambda_o}\right)}$$

Then clearly P(E) ~~can be~~ is given by

$$P[E] = \frac{1}{2} P[E/\lambda_o] + \frac{1}{2} P[E/\lambda_1]$$

where:

$$P[E/\lambda_o] = \sum_{k>k^*} \frac{(\lambda_o T)^k}{k!} \exp(-\lambda_o T)$$

and

$$P[E/\lambda_1] = \sum_{k<k^*} \frac{(\lambda_1 T)^k}{k!} \exp(-\lambda_1 T)$$

Typically we are quantum limited in a space born ISL so the above analysis is not necessary. It should be emphasized, however, that the pointing and tracking problems are increased significantly due to the narrow beamwidth. This problem has been solved technically so it is not a real problem.

In conclusion, both of these technologies are available for ISL's and they both are well within the state of the art. It is expected that ISL's will find commercial use in the near future.

# CHAPTER 6

## SYSTEM CONSIDERATIONS

The previous chapters considered the elements that comprise the satellite communications system. It is the purpose of this chapter to bring them together in a total system context so that one can understand how all of the separate earth stations can communicate. We do this by first considering to total link and develop the link budget. This concept is key to the development or analysis of any satellite communications system. Expanding upon this we present the various anomalies that could occur in the link that could result in the degradation of overall performance.

The next development in this chapter is that of traffic definition and assignment. In any system we typically meet two conflicting goals; the first to carry a certain level of traffic and the second to meet the defined level of performance. The distribution of traffic must then be done to meet the performance constraint as determined by the link budget parameters. It is in this area that we see the trade off between satellites, multiple beam antennas, frequencies, modulation, and other system design elements.

The final issue discussed in this chapter is multiple access design. That is how do we actually partition the traffic; in time, frequency, space. We discuss the standard multiple access schemes and introduce the demand access schemes that are more assessable to very dynamic load conditions. It will be seen that this area will be significantly impacted upon as the costs of high speed digital processing are further reduced.

## 6.1 THE LINK BUDGET

As we have previously discussed, all the elements of the satellite communications link must be combined into a whole to evaluate the performance of the link in terms of its ability to carry traffic. We shall, in this section, concentrate upon the link budget as that means. We shall first begin with a single digital channel giving from one earth station to another and use several satellite models for that channel. Once this is completed, we can then in the next section consider link anomalies and evaluate their impact on the link performance.

Figure 1 depicts all of the elements of the link. It combines all of the items that we have previously discussed and characterized. The objective in any link budget analysis is to determine the design parameters of the system so that a given level of performance can be met. Specifically, we choose that measure to be the bit error rate or probability of error, P[E]. Recall that P[E] depends on $E_b/N_o$ which is shown at the output of the decoder. For a given modulation scheme and coding technique, the link $E_b/N_o$ then determines P[E].

The system designer may or may not have the choice of the major link parameters. This will typically depend upon the application. We can now go through the link and determine the power at each level.

Assume that the source is generating data at $R_o$ bits/sec. The encoder outputs $R_s$ symbols/sec and this implies a code rate of $R_o/R_s$ bits/symbol. The modulator converts the digital bit streams into some baseband modulation format as previously discussed. The upconverter places this signal in an up frequency of $f_u$ which is amplified by the HPA. The output power level of the HPA is $P_E$. This may be less than or equal to the saturation power level of this amplifier, $P_{E,SAT}$.

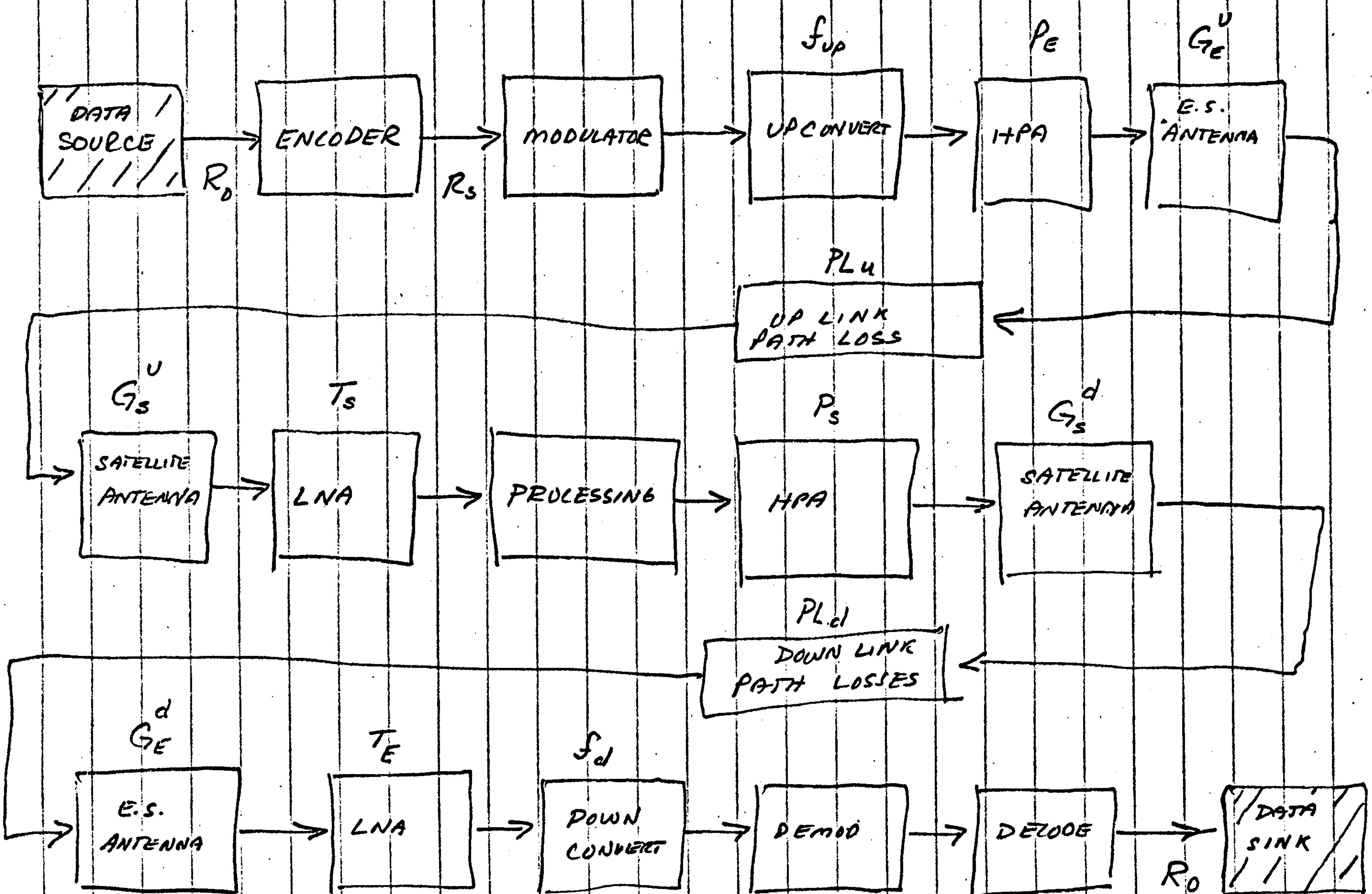



FIG 1 : BASIC ELEMENTS OF THE LINK BUDGET

The amplifier feeds the antenna along the waveguide. The earth station antenna has a peak gain of $G_E^u$ on the up link at frequency $f_u$. It will be important to note that $G_E^u$ is both the peak gain and frequency dependent. It also assumes that a certain efficiency has been used in the evaluation. The transmit eirp (equivalent isotropic radiated power) is defined,

$$\text{eirp} = P_E G_E^{\,u}$$

The radiation from the antenna is now propagated in free space at a peak power level of $P_E G_E^{\,u}$.

The power density (watts/m$^2$) decreases as distance such that, neglecting absorption, the total power remains constant. Thus at a distance r from the earth station the power density level at the peak is

$$G_E^{\,u} P_E \quad \frac{1}{4\pi r^2} \quad \text{watts/m}^2$$

Now if $G_E^{\,u}(\theta,\emptyset)$ is the total antenna pattern with

$$G_E^{\,u} = \max_{\theta,\emptyset} G_E^{\,u}(\theta,\emptyset)$$

then the integral over a sphere of radius r would conserve power.

Thus at the satellite the power density received is

$$G_E^{\,u} P_E \quad \frac{1}{4\pi R_u^{\,2}} \quad (\text{watts/m}^2)$$

where $R_u$ is the up-link range.

The satellite antenna has a collection area which is

$$\frac{\lambda_u^{\,2}}{4\pi} \quad G_S^{\,u}$$

where $G_S^{\,u}$ is the peak gain of the satellite antenna at the uplink frequency $f_u$, $\lambda_u$ is the uplink wavelength. Recall that $\frac{\lambda_u^{\,2}}{4\pi}$ is the cross section of the isotropic radiator

Thus the power level received at the satellite due to the signal at the earth station is

$$P_E^{\,u} = P_E G_E^{\,u} G_S^{\,u} \left(\frac{\lambda_u}{4\pi R_u}\right)^2 \qquad (\text{watts})$$

The LNA has a noise temperature which when combined with all other noise effects gives a front end noise temperature of $T_S$. This yields an equivalent front end noise which is white with spectral density $kT_S$ at the front end of the LNA. If the effective signal bandwidth is W then the front end noise power up, $N_u$, is

$$N_u = kT_S W \quad \text{(watts)}$$

Thus the satellite sees two signals $P_E^u$ and $N_u$.

The process of taking the input signals from the satellite to the output can involve one of several processes. The simplest is a pure gain, $G_{SAT}$, in power level across the satellite and a shift in center frequency from $f_u$ to $f_d$. This type of satellite is called a repeater or direct translation. It assumes no signal processing and assumes that the HPA acts in a linear mode. As was discussed previously the satellite may be in a processing mode where demodulation and decoding are performed or the amplifier may be saturated. We shall consider these cases later.

Thus the power coming out of the HPA due to the signal and noise respectively is:

$$G_{SAT} \; P_E^u$$

and;

$$G_{SAT} \; N_u$$

These powers are then fed to the antenna with gain $G_S^d$. Note that this may not be the same antenna but is most probably at a different frequency.

Following the same logic on the downlink we can determine the signal power at the front and $P_E^d$ as

$$P_E^d = G_{SAT} P_E^u \; G_S^u \; G_E^d \left( \frac{\lambda_d}{4\pi R_d} \right)^2 \qquad \text{(watts)}$$

and the noise power due to the satellite noise is

$$N_u^d = G_{SAT} N_u G_s^d G_E^d \left( \frac{\lambda_d}{4\,\pi R_d} \right)^2$$

The front end noise $N_d$ of the Earth station is

$$N_d = kT_E W \quad \text{(watts)}$$

where $T_E$ is the total effective noise of the receiving earth station front end. Thus the total noise is

$$N = N_d + N_u^d$$

and $P_E^d$ in the total signal power. The ratio $C/N$ is the ratio of signal power to noise power and it is

$$\frac{C}{N} = \frac{P_E^d}{N_d + N_u^d}$$

This can be reduced to a $C/N_o$ term, where $N_o$ is the effective noise spectral density, by dividing through the denominator by W. To simplify even further we can rewrite this as:

$$C/N_o = \frac{1}{\dfrac{N_d}{P_E^d/W} + \dfrac{N_u^d}{P_E^d/W}}$$

space

Now looking more closely at each expression we note:

space

$$\frac{P_E^d/W}{N_d} = \left(\frac{C}{N_o}\right)_d$$

which is the carrier power to noise spectral density of the down link only and

$$\frac{P_E^d\ W}{N_u^d} = \frac{P_E^u\ W}{N_u} = \left(\frac{C}{N_o}\right)_u$$

which is the carrier power to noise spectral density of the up link. Thus we can write

$$(C/N_o) = \left[(C/N_o)_u^{-1} + (C/N_o)_d^{-1}\right]^{-1}$$

Thus given $C/N_o$ we can readily calculate both $E_s/N_o$ and $E_b/N_o$. Since the power is constant we recall that energy is power times duration and the rates are inverse duration. Thus

$$\frac{E_b}{N_o} = \frac{1}{R_o}\frac{C}{N_o}$$

$$\frac{E_S}{N_o} = \frac{1}{R_S}\frac{C}{N_o}$$

Now recall that $E_b/N_o$ determines $P[E]$ and $E_S/N_o$ is useful in determining thresholds in PLL.

The calculation of $E_b/N_o$ now requires that of $(C/N_o)_u$, $(C/N_o)d$, $C/N_o$ and finally the use of $R_o$. This is all simplified by using logs, specifically $10 \log_{10}(\ )$ units. The following example demonstrates how this is done.

Example:

A small earth station with a 10' antenna is used to transmit and receive data over an INTELSAT satellite to a large Standard A INTELSAT earth station. The Standard-A is characterized by an antenna of about 30m. diameter with a nominal gain at 4GHz of 57 dB. It has a G/T of 40.7 dB /K minimum.

Two links are proposed; a data link from the small stations to the large at 4.8Kbps and a command link from the large to the samll at 96 bps. As we shall see, it is the return link into an earth station with a low G/T that causes the most difficulty. This is because more satellite power is needed on such a link.

The choice of modulation can be somewhat arbitrary but for simplicity 2Ø-PSK is chosen for the data and incoherent FSK for the command. The choice of the latter results, as we have discussed previously, because of the need to avoid phase noise distortions of the up and down converters.

To improve performance we also choose to use a rate 1/2 constraint length 7 convolutional code with a Viterbi decoder for this system. As we shall show later, this choice is optimal from a cost point of view.

The small earth station uses a GaAsFET with a $100^{o}$K noise temperature. As we discussed this is most cost effective and highly available. where was it discussed?

One last fact must be mentioned. For the INTELSAT Satellite the gain of the satellite is not given directly but implied by providing levels of saturation. What is provided is the incident flux density ($w/m^2$) that saturates the output HPA and the value of that saturated eirp. If we define $F_{SAT}$ as the saturated flux density then F the actual flux must be less than $F_{SAT}$. Similarly $(eirp)_{SAT}$ is the saturated eirp. Now the system works in almost a linear fashion. The input backoff, $BO_I$, is defined in dB as

$$BO_I(dB) = F_{SAT}\ (dB^{w/m^2}) - F\ (dB^{w/m^2})$$

The output eirp is then determined from the saturated value and the output backoff. Thus

$$\text{eirp (dBw)} = (\text{eirp})_{SAT}(\text{dBW}) - BO_o$$

For a given HPA (e.g, a TWTA), the input and output backoff are related by

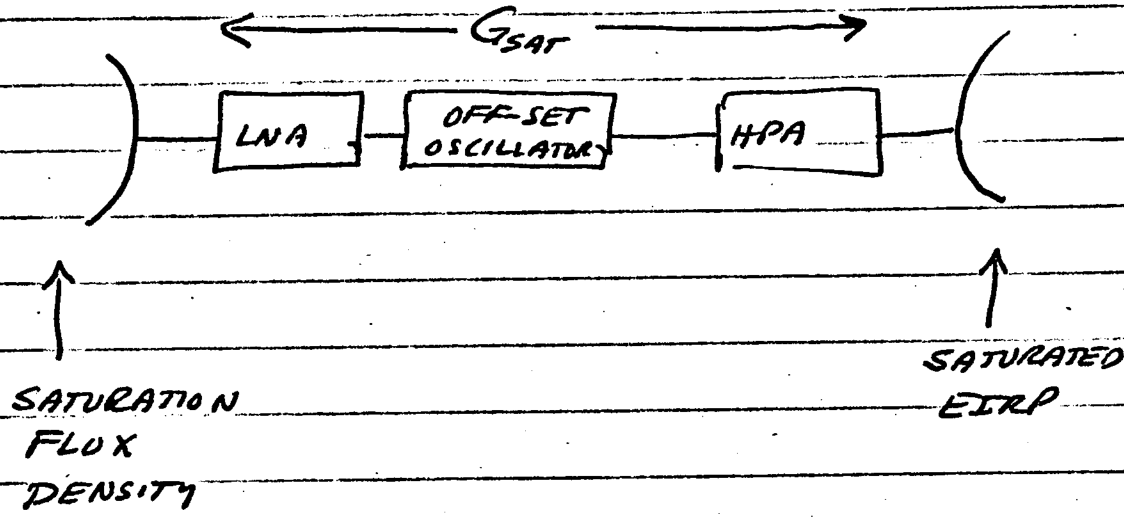
$$BO_I = BO_o + C \qquad (\text{dB})$$

where C is a dB offset which is typically positive. It results from the normal HPA level (see Fig. 2) and goes to zero if the input power level nears saturation. For INTELSAT this offset is typically 5 dB.

Tables 1 and 2 depict the link budgets for this system. Consider first the command link. The design process must be interative. First the $E_b/N_o$ needed for the BER is specified and the margin given. This, with the data rate determines the $C/N_o$. The only unknown is the transmit power at the INTELSAT station. To determine this an iterative process is used until the solution is obtained.

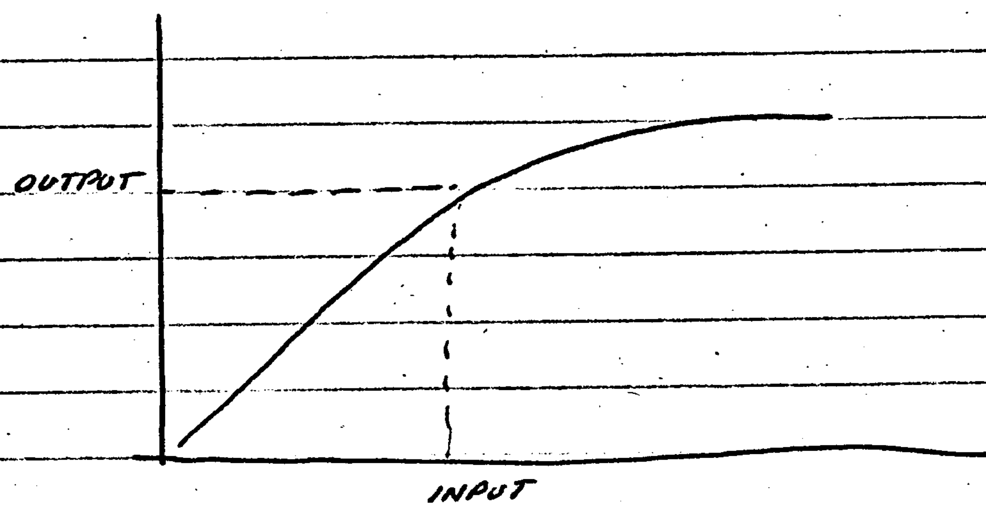
The data link functions much the same way except here we must include the effects of coding gain.

We may now expand the application to a transponder which does not act in a linear mode but in a saturated one. In this case, the satellite has a saturated power level of $P_{SAT}$. If the input consists of N signals of power levels $P_i^u$ (i=1,...N) then their corresponding output levels are




(a)

Block Diagram




(b)

TWTA Characteristic

Fig 2: Transponder Satellite with TWTA Characteristics.

Table 8: Link Budget for Data Link (continued)

| | |
|---|---|
| INTELSAT Earth Terminal Receive Power | -161.03 dBW |
| Receive Noise Temperature | 20.00 dB |
| Receive G/T | 40.70 dB/K |
| Boltzmann's Constant | -228.60 dBW/K-Hz |
| $(C/N_o)$ down | 51.27 dB-Hz |
| $(C/N_o)$ total | 50.43 dB-Hz |
| Data Rate (9600 bits/s) | 36.81 dB-Hz |
| $E_b/N_o$ | 13.62 dB |
| $E_b/N_o$ (BER @ $10^{-7}$) | 11.3 dB |
| Implementation Loss | 1.60 dB |
| Coding Gain | 6.0 dB |
| Margin | 7.02 dB |

Table 8: Link Budget for Data Link

| Up-Link (6327.3 MHz) | |
|---|---|
| NSS Power (HPA Flange) | 10.00 dBW |
| Feed Loss | 1.00 dB |
| NSS Antenna Gain | 42.50 dB |
| e.i.r.p. | 51.50 dBW |
| Radome Loss | 0.50 dB |
| Multipath Loss | 1.00 dB |
| Atmospheric Loss at 6 GHz | 1.50 dB |
| Pointing Loss (INTELSAT IVA) | 0.50 dB |
| Path Loss (nominal) | 200.00 dB |
| Power on Satellite | -152.00 dBW |
| Satellite G/T | -18.60 dB/K |
| Boltzmann's Constant | -228.60 dBW/K-Hz |
| $(C/N_o)_{up}$ | 58.00 dB-Hz |
| Satellite Saturation Flux | -67.30 dBW/m$^2$ |
| $4/\lambda^2$ (6327.3 MHz) | 37.47 dB/m$^2$ |
| Satellite Flux | -114.53 dBW/m$^2$ |
| Input Backoff | 47.23 dB |
| Output Backoff | 42.23 dB |
| Satellite Saturated e.i.r.p. | 22.00 dBW |
| **Down-Link (4102.3 MHz)** | |
| Satellite e.i.r.p. | -20.23 dBW |
| Atmospheric Loss at 4 GHz | 1.00 dB |
| Path Loss | 196.50 dB |
| Power at INTELSAT Earth Terminal | -217.73 dBW |
| INTELSAT Earth Terminal Antenna Gain | 57.00 dB |
| Feed Loss | 0.30 dB |

Table 7: Link Budget for Command Link

| Up-Link (6327.3 MHz) | |
|---|---|
| INTELSAT Earth Station Power at the OMT | 4.40 dBW |
| INTELSAT Earth Station Antenna Gain | 60.50 dB |
| e.i.r.p. | 64.90 dBW |
| Atmospheric Loss at 6 GHz | 1.50 dB |
| Path Loss | 200.00 dB |
| Power on Satellite | -136.60 dBW |
| Satellite G/T | 18.60 dB/K |
| Boltzmann's Constant | -228.60 dBW/K-Hz |
| $(C/N_o)_{up}$ | 73.40 dB-Hz |
| Satellite Saturation Flux | -67.30 dBW/m² |
| $4/\lambda^2$ | 37.47 dB/m² |
| Satellite Flux | -99.13 dBW/m² |
| Input Backoff | 31.83 dB |
| Output Backoff | 26.83 dB |
| Satellite Saturated e.i.r.p. | 22.00 dBW |
| **Down-Link (4102.3 MHz)** | |
| Satellite e.i.r.p. | -4.83 dBW |
| Radome Loss | 0.75 dB |
| Multipath Loss | 1.00 dB |
| Atmospheric Loss at 4 GHz | 1.00 dB |
| Pointing Loss (INTELSAT IV-A) | 0.50 dB |
| Path Loss | 196.50 dB |
| Polarization Loss | 0.10 dB |
| Power on Remote Earth Station | -204.68 dBW |
| Remote Earth Terminal Antenna Gain | 39.50 dB |

$$P_i^o = P_{SAT}\left(\frac{P_i^u}{\sum_{j=1}^{N} P_j^u}\right)$$

That is the power is divided ~~amongst~~ among all signals. Now let us assume that only one signal is present and that noise represents the only other power source. Then the output power of the signal component is

$$P_S^o = P_{SAT}\left[\frac{P_E^u}{P_E^u + N_u}\right]$$

and the output power of the noise is

$$P_N^o = P_{SAT}\left[\frac{N_u}{P_E^u + N_u}\right]$$

Now in our expressions we can replace the following

$$P_S^o \cong G_{SAT}\, P_E^u$$

$$P_N^o \cong G_{SAT} N_u$$

We can now rewrite the $C/N_o$ term as follows:

$$C/N_o = \frac{1}{\dfrac{N_d}{P_E^d W} + \dfrac{N_u^d}{P_E^d W}}$$

But now we have:

$$\frac{P_E^d W}{N_d} = P_{SAT}\frac{P_E^u}{P_E^u + N_u} \quad G_s^d\, G_E^d \left(\frac{\lambda d}{4\pi R_d}\right)^2 \quad \frac{1}{(N_d/W)}$$

6

and

$$\frac{P_E^d W}{N_u^e} = \frac{P_E^u\ W}{N_u} = \left(\frac{C}{N_o}\right)_u$$

This allows us to rewrite the first expression as:

$$\frac{P_E^d\, W}{N_u} = \left(\frac{C}{N_o}\right)_d \left[1 + W\left(\frac{C}{N_o}\right)_u^{-1}\right]^{-1}$$

Therefore;

$$(C/N_o) = \left[(C/N_o)_d^{-1} + (C/N_o)_u^{-1} + W(C/N_o)_d^{-1}\,(C/N_o)_u^{-1}\right]^{-1}$$

You are missing a minus sign

Thus for this type of channel there appears another term that is bandwidth dependent. It represents the "power robbing" factor on the link due to the use of a saturated amplifier.

## 6.2 LINK ANOMALIES

The link budget described in the previous section only accounts for the standard gains and losses. In this section we shall consider other link factors that tend to degrade the signal. Fig. 3 depicts a typical two way satellite link. In this case we have included a radome on one of the attennas. Radomes are typically employed in regions where icing is a concern but power is limited.

There are basically two types of additional link anomalies or losses; time invariant and time varying. The former may result in sets of static events such as additional absorption or constant system offsets. The latter losses result from such factors as satellite motion or transient absorption factors. Many cases of link performance require a close design in terms of available margin but at the same time require a level of performance to be met a given percent of the time. The approaches to the design of links with these anomalies fall into three categories.

1. Worst Case - In this case the losses are included at their maximum level so that the margin is ensured. The problem with this approach is either overdesign or failure to recognize real worst case values.

2. Statistical - In this case the losses are characterized individually with some statistical distribution both spatially and temporally. Then these variables are added to obtain the total statistical distribution. This then permits a probability density of the link budget to be obtained. The problem with this approach is that it requires assumptions on distributions and significant calculation to determine a total density. The advantage is that it permits the calculation of many statistical parameters and a real "look" at what the effect of the cumulative losses are.

3. <u>Simulated</u> - In this approach the loss mechanisms are modelled based either upon the actual phenomenology or statistical models. Then the link budget is calculated over time and statistics in its variance from levels determined. The disadvantages are similar to those of the statistical and the advantages are that "real" margains are obtained.

For both methods 2 and 3 such statistics as the following can be obtained;

- $F_{\Delta}^{\tau}$ = Prob [Fade $>\Delta$ for $\tau$ sec or more]
- $F_{\Delta}$ = Prob [Fade $>\Delta$]

In this section we shall examine losses due to several standard causes.

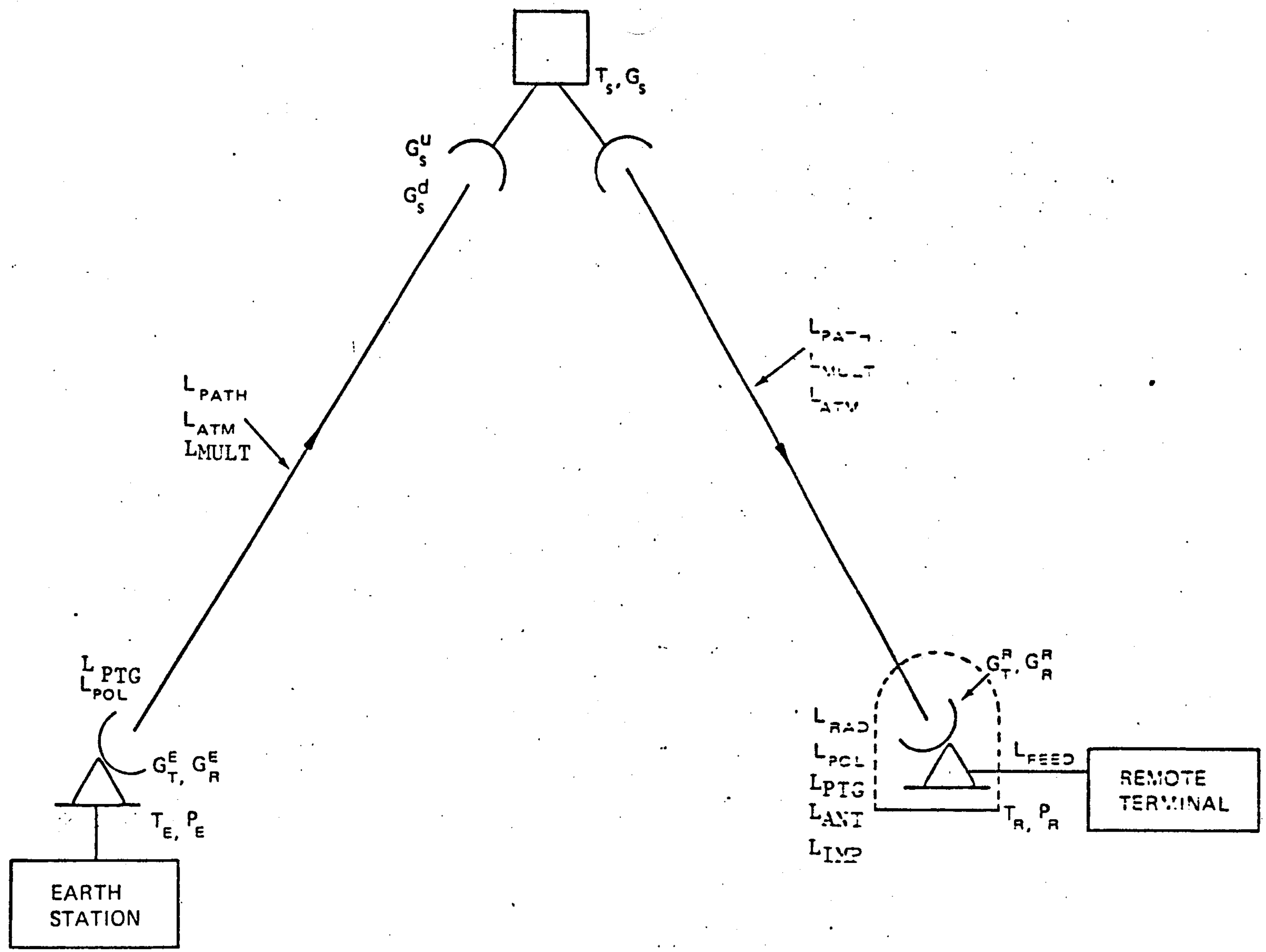



Figure 3. 2 Communications Links

Losses due to scintillation, pointing, atmosphere rain and radome are examined in depth to determine their cumulative effect over time. Since each occurs independently over time, and to varying degrees at each site, it is inappropriate to assume worst case losses for the link budget. Several additional losses are standard and worst case values can be used, since minimal site or time dependence has been exhibited in similar situations. The losses affecting the link budget are as follows:

i) Polarization Loss ($L_{POL}$) due to the misalignment of the receive and transmit polarization vectors.

ii) Atmospheric Loss ($L_{ATM}$) due to absorption in the atmosphere from water vapor and other factors. Included in this is also losses due to rain.

iii) Multipath Loss ($L_{MULT}$) due to the coherent adding of specular multipath signals at the antenna.

iv) Pointing Loss ($L_{PTG}$) due to the off boresight location of the satellite with respect to the antenna. This arises with a non-tracking antenna when the satellite moves, the antenna is initially misaligned or there is a shift in the pad on which the antenna is mounted. This loss can be analyzed in two ways. First by assigning the satellite fixed and assume errors in pointing the antenna. Second, assume the antenna fixed and let the satellite drift in a normal station keeping mode.

v) Radome Loss ($L_{RAD}$) due to absorption resulting from water on the radome cover.

vi) Antenna Noise Loss ($L_{RAD}$) due to the fact that the antenna is at a low pointing angle and sees the high ambient earth temperature with a portion of its beam.

vii) Feed Loss ($L_{FEED}$) due to waveguide losses on transmit and receive. Note that this is included in the losses due to the antenna and the resulting increased noise temperature on receive but must be included on transmit.

viii) Scintillation Losses ($L_{SCN}$) due to scintillation effects.

ix) Implementation losses ($L_{IMP}$) due to the implementation of demodulation, decoding, PLL synch and other hardware factors.

### 6.2.1 Pointing Loss

The antenna is assumed to be pointed at the satellite nominal orbit position. Slight variations in the satellite location produce antenna gain variations, known as pointing losses. The satellite is allowed to oscillate within a box centered on the nominal orbit position with dimensions 3 corresponding to the station keeping parameters. Table 3 gives these parameters for the INTELSAT System. Clearly, the worst case pointing losses occur for stations utilizing the Pacific INTELSAT IV Satellite.

Pointing loss is not a factor at the INTELSAT earth station since its antenna can be adjusted to handle slight positional changes of the satellite. It must be considered at the remote terminal on both uplink and downlink, however.

To characterize pointing loss as a function of time, it is first necessary to know the satellite's position at any given time. Satellite latitude motion has a period of 24 hours, and is bounded in magnitude by the inclination, I, in degrees. The inclination also changes with time according to a hyperbolic curve. Thus, a basic expression for satellite latitude, $\ell$, at any given time, t, in minutes would be (Ref?)

$$\ell = i \sin (nt)$$

where $n = \frac{2\pi}{1436}$ rad/min, i is the instantaneous

i,

inclinationa, and the asympotes for the inclination hyperbola give

3

Table 3 Station Keeping Parameters for the INTELSAT System

| SATELLITE | LONGITUDE (°E) WEST LIMIT | LONGITUDE (°E) EAST LIMIT | INCLINATION (DEGREES) |
|---|---|---|---|
| IV F-1 | 62.8 | 63.0 | .1 |
| IV F-2 | 355.0 | 357.0 | N/A |
| IV F-3 | 340.0 | 341.0 | .5 |
| IV F-4 | 178.5 | 179.5 | .5 |
| IV F-5 | 60.1 | 60.3 | .1 |
| IV F-7 | 358.5 | 359.5 | .1 (.5) |
| IV F-8 | 173.5 | 174.5 | .5 |
| IV-A F-1 | 335.4 | 335.6 | .1 |
| IV-A F-2 | 330.4 | 330.6 | .1 |
| IV-A F-4 | 325.4 | 325.6 | .1 |

$$i = -\frac{I}{30} T + I$$

$$i = \frac{I}{30} T - I$$

for T equal to the time in days.

The typical amount of time required between latitude maneuvers is 2 months. Figure 4 depicts these relationships for latitude motion. Thus actual latitude extent decreases as maximum longitude is reached.

Reference ?

The basic pattern of longitude motion can be given by

$$L = L_m + .03 \sin (nt + \phi)$$

where L = longitude and $\phi$ is a small phase shift angle. Then we also note that

$$L_m = \frac{1}{2} \frac{d^2L}{dT^2} T^2 = \frac{1}{2} \frac{d^2L}{dT^2} .0005 \text{ deg/day}^2$$ ?

T=t ? $\frac{d L_m}{dt} = 0$

Thus longitude motion follows a basic parabolic pattern with an additional oscillation term due to orbit eccentricity, as shown in Figure 5. Longitude maneuvers occur every 45 days, i.e., it takes 45 days for the satellite to move in longitude from one end of the box to the other.

The overall satellite motion resembles a slowly narrowing spiral as it moves from the nominal position to the maximum longitude. This emphasizes the fact that the satellite is typically not located in any of the 4 extreme corners of the box, but rather will have close to minimum extent in one direction when maximum extent is reached in the other. A sufficient worst case model

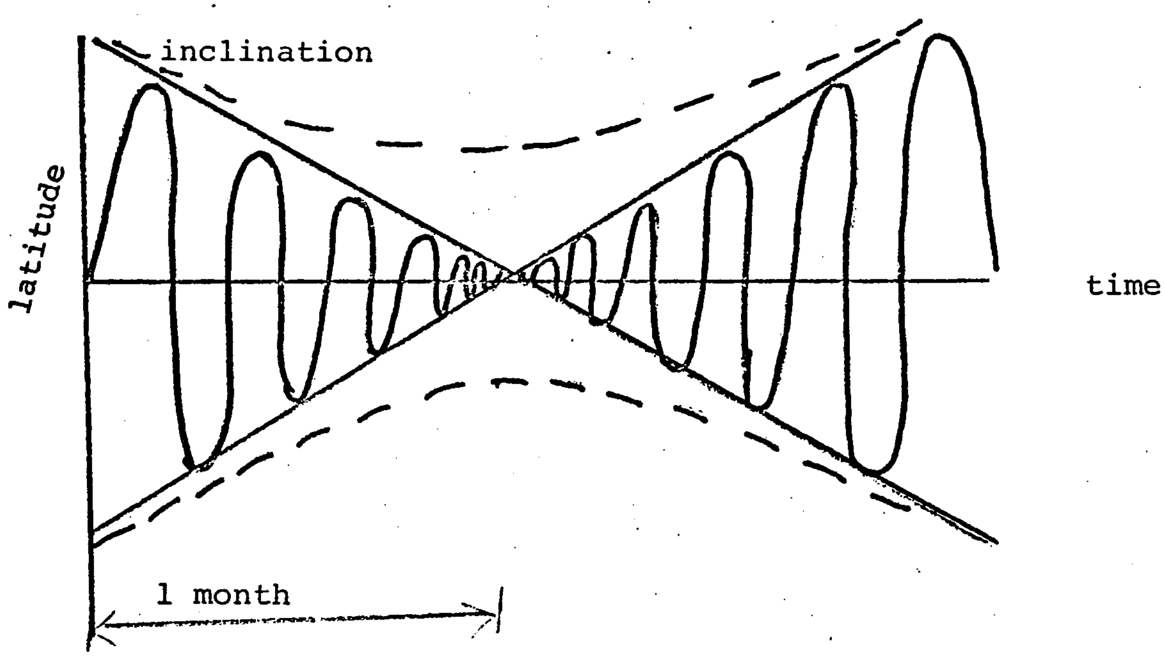



Figure 4 . Satellite Latitude Motion Vs. Time

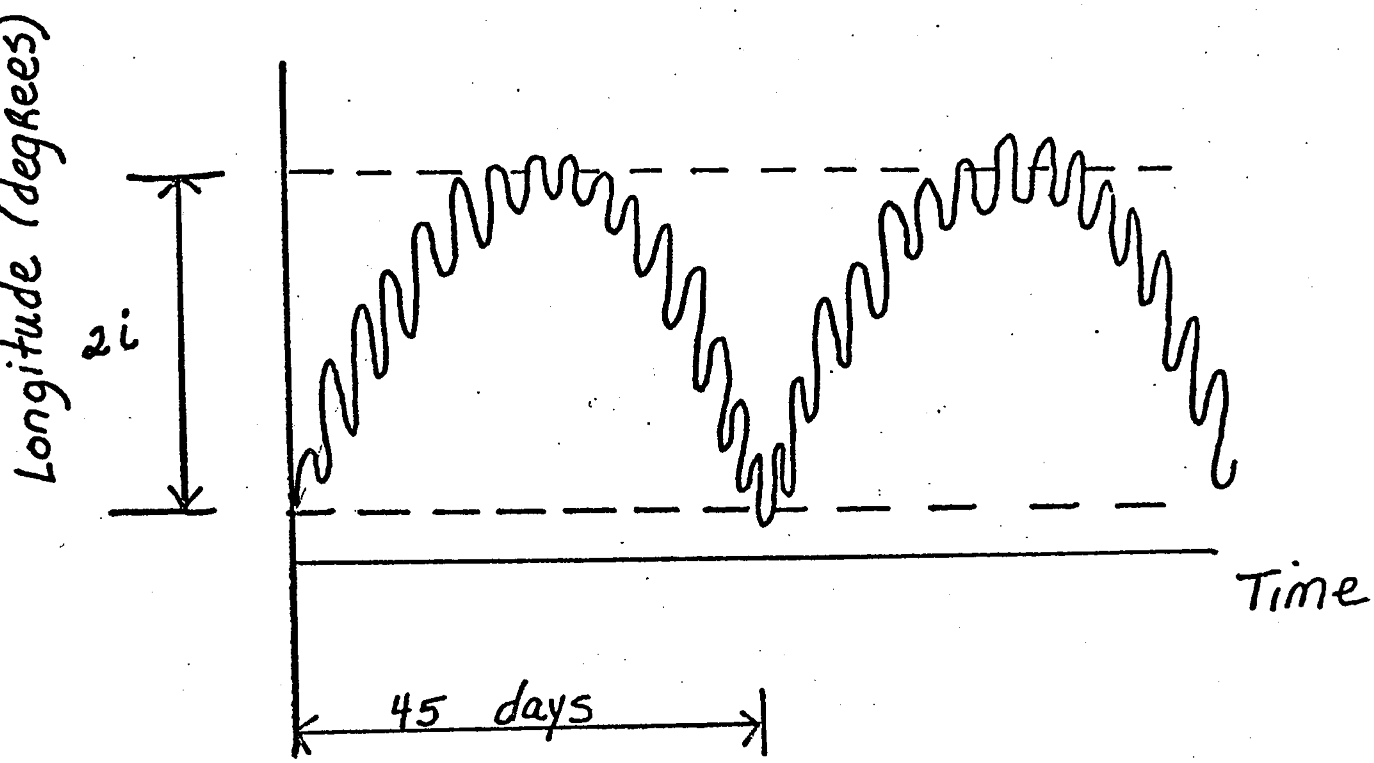



Figure 5 5. Longitude Motion Vs. Time

would be to take satellite latitude as zero, and longitude as maximum inclination.

earth station

Knowing satellite location, the offset angle, Θ, to any earth station can be calculated. Using the geometric relationships illustrated in Figure 6 , θ can be obtained from

$$\tan \Theta = \frac{|P \times P_c|}{P \cdot P_c}$$

where the vectors P and $D_c$ are defined and the dot and cross product used.

Pointing loss, or equivalently the antenna gain variation of the main lobe of a parabolic antenna, can now be characterized by the offset angle, Θ, according to the function $[J_1(x)]/x$ where $x = \pi D/\lambda \sin \Theta$. For small x, the following approximation is valid

$$\frac{J_1(x)}{x} = 1 - \frac{\pi^2 (D/\lambda)^2 \sin^2 \theta}{8}$$

Figure 7 plots pointing loss as a function of offset angle θ for a 10 ft antenna at 6 and 4 GHz. Worst case offset angle occurs for the Pacific satellite and earth station with the

earth &

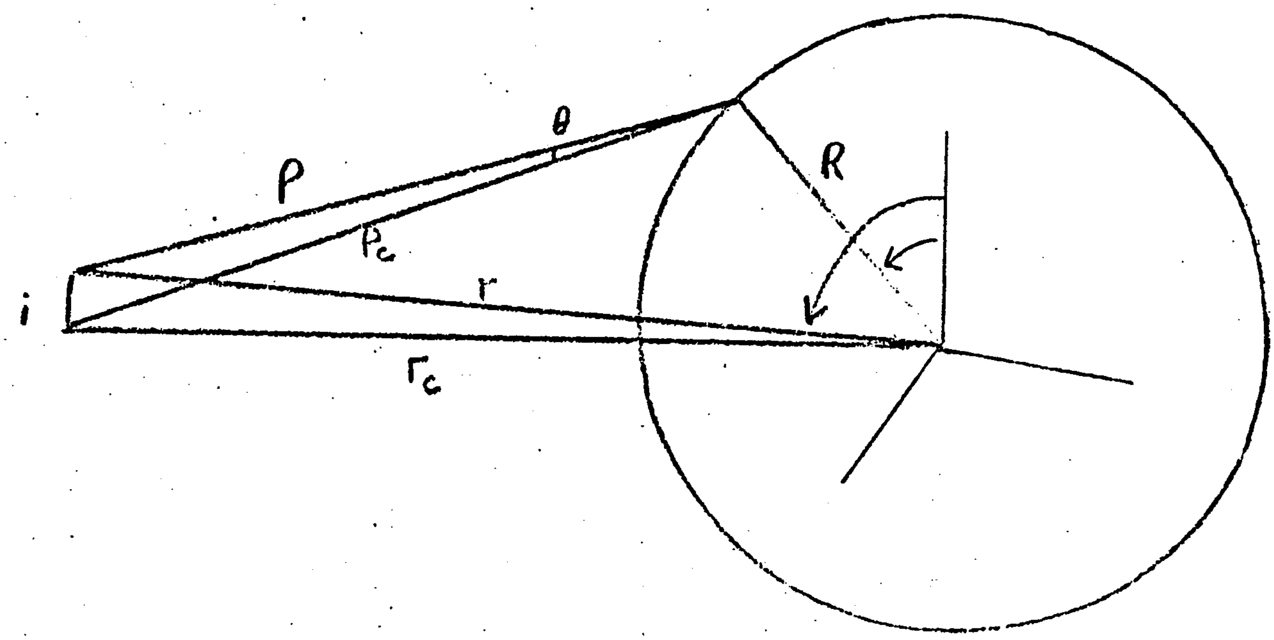
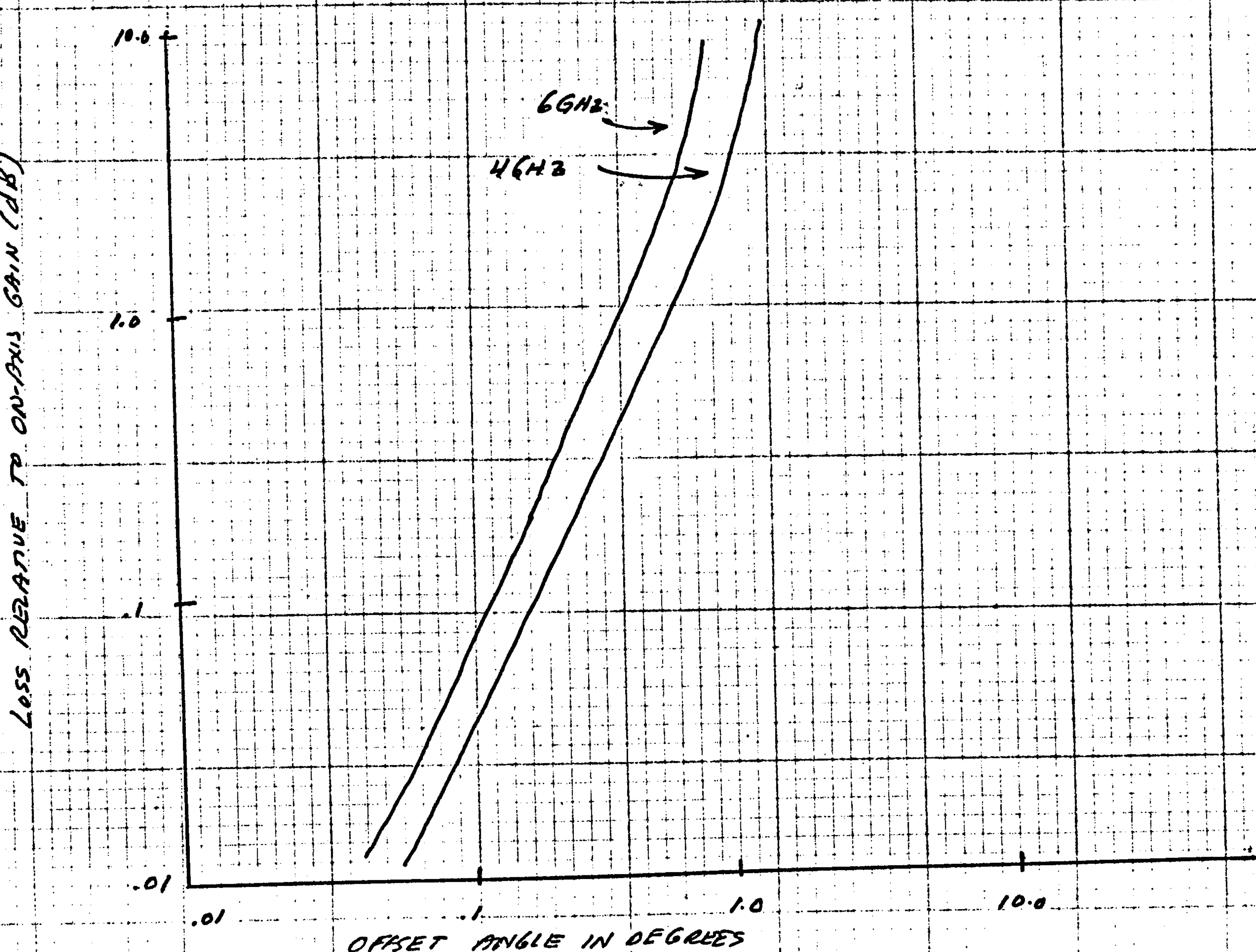



Figure 6 .. Satellite Position Geometry
6

FIG 7: POINTING LOSS F[illegible] A 3m. ANTENNA



largest elevation angle, with choice of satellite as the major factor. Figure 8 illustrates worst case pointing losses for each satellite and elevation angle.

Figure 9 illustrates pointing loss as a function of time over a ten day period for a typical station, at 6 GHz. The maximum pointing loss is sufficiently small to warrant a worst case value at both 6 and 4 GHz for all stations using the Indian Satellite.

Similarly, Figure 10 represents pointing loss over time using the Pacific satellite, at 4 GHz. At this frequency, worst case values can again be used for all stations utilizing the Pacific Satellite. Figure 11 is the corresponding curve for 6 GHz. It can be seen that the maximum pointing loss is high, but that value is reached for only a small percentage of the time. Further time distribution analysis may be necessary in this case, depending on the available link margin.

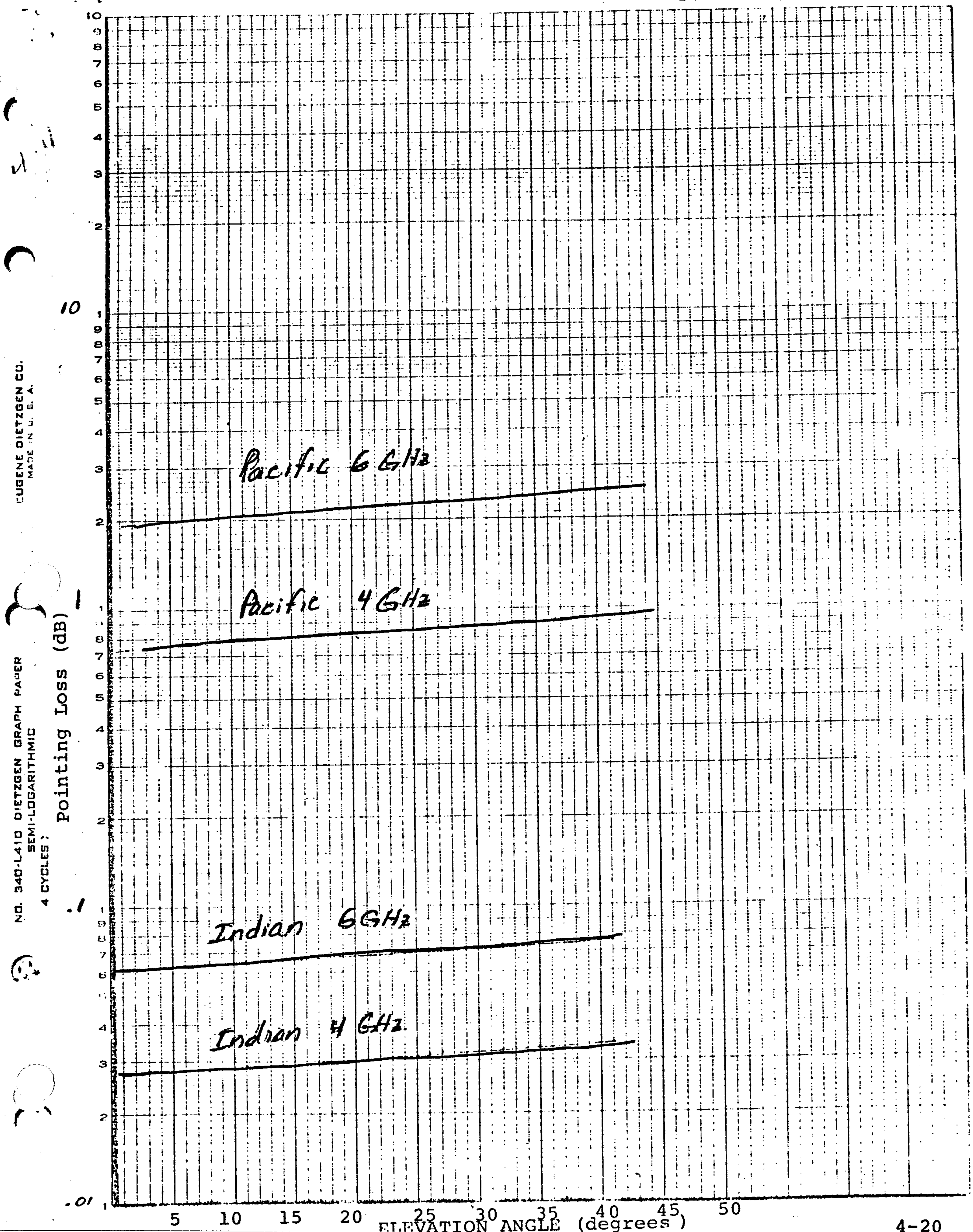
### 6.2.2 Scintillation Loss

Ionospheric scintillation is due to irregularities in the F-region of the ionosphere. In the 6/4 GHz frequency range, this type of scintillation has only been observed for earth stations near the geo-magnetic equator. It is safe to assume, for stations 25 degrees north or south of the geo-magnetic equator, that loss due to ionospheric scintillation is minimal.

Tropospheric scintillation is due to the refractive structure of the troposphere. For stations having an elevation angle greater than $9^{o}$, the loss due to tropospheric scintillation is low, and will not exceed 1 dB. Experimental data for low-elevation angle sites in Canada has been collected, but only one study represents high arctic data. Figure 10 gives the distribution of received signal strength as a function of elevation

 Corrected pg #'s

Figure 4.10. POINTING LOSS VS. ELEVATION ANGLE (WORST CASE SATELLITE POSITION)



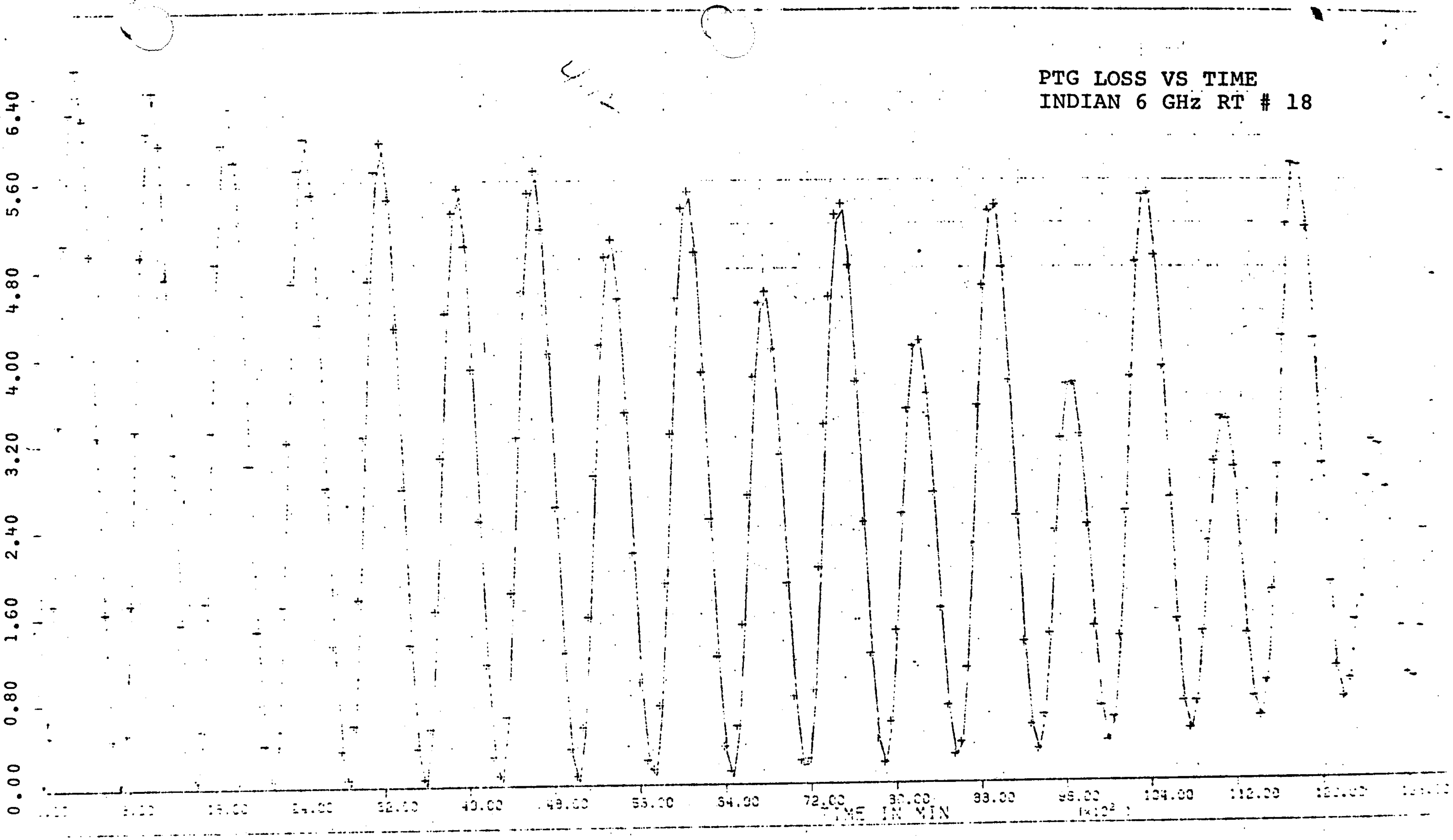

PTG LOSS VS TIME
INDIAN 6 GHz RT # 18
6.40
5.60
4.80
4.00
3.20
2.40
1.60
0.80
0.00
32.00
40.00
48.00
56.00
64.00
72.00
80.00
88.00
96.00
104.00
112.00
TIME IN MIN

Figure 4.11

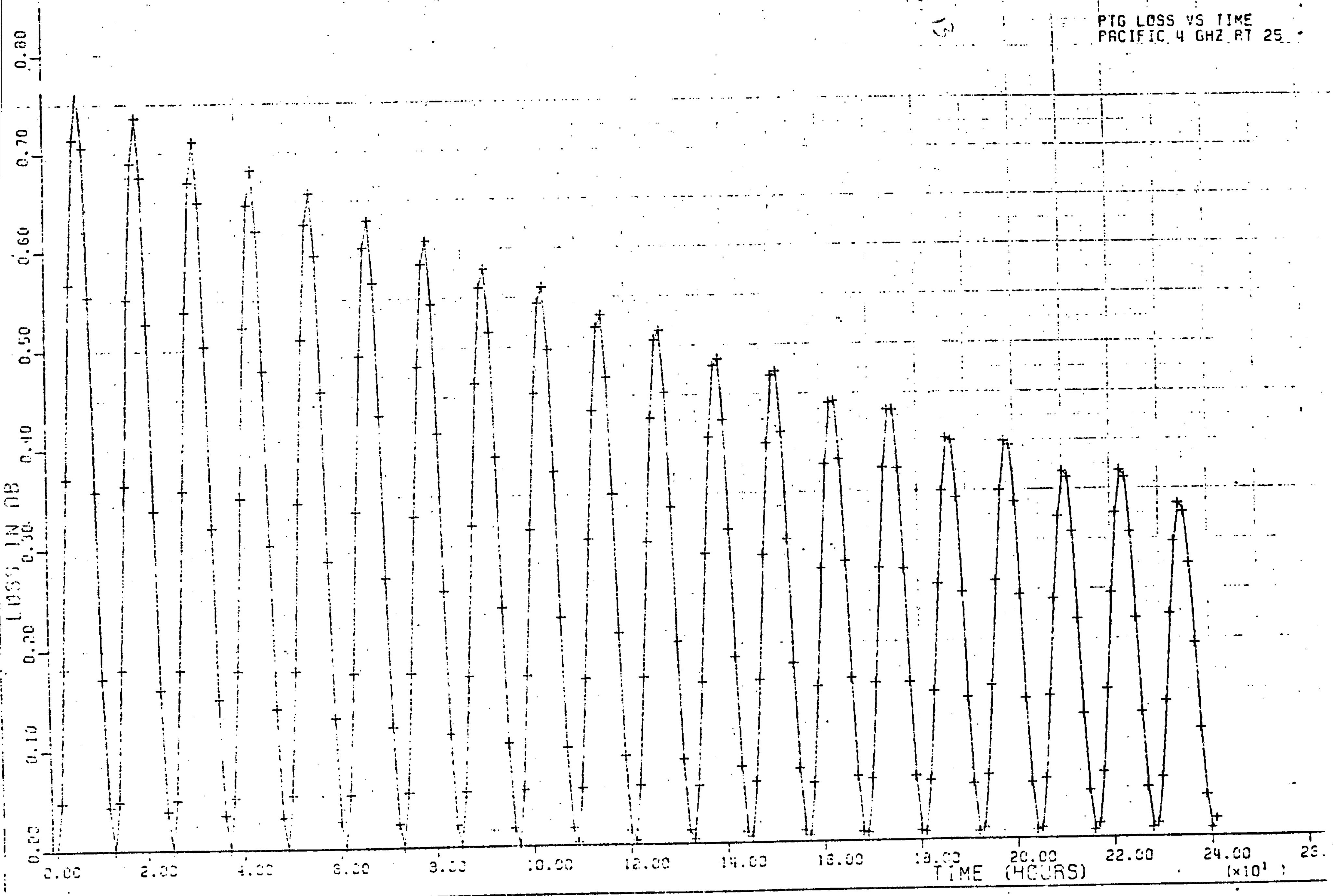
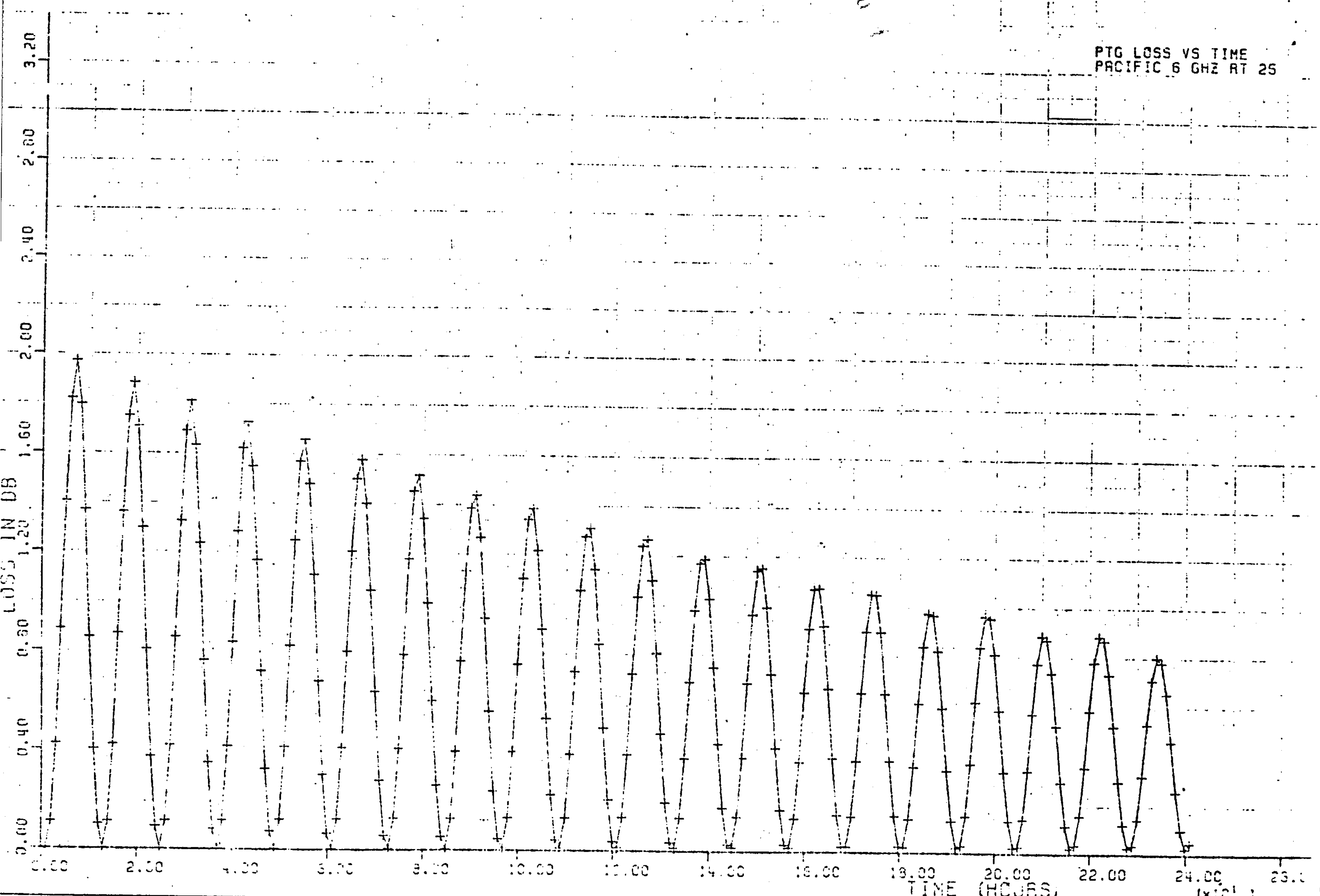



Figure 4.12



PTG LOSS VS TIME
PACIFIC 6 GHZ AT 25
LOSS IN DB
0.00
0.40
0.80
1.20
1.60
2.00
2.40
2.80
3.20
TIME (HOURS)
2.00
22.00
24.00
(x10^1 )


Figure 4.13

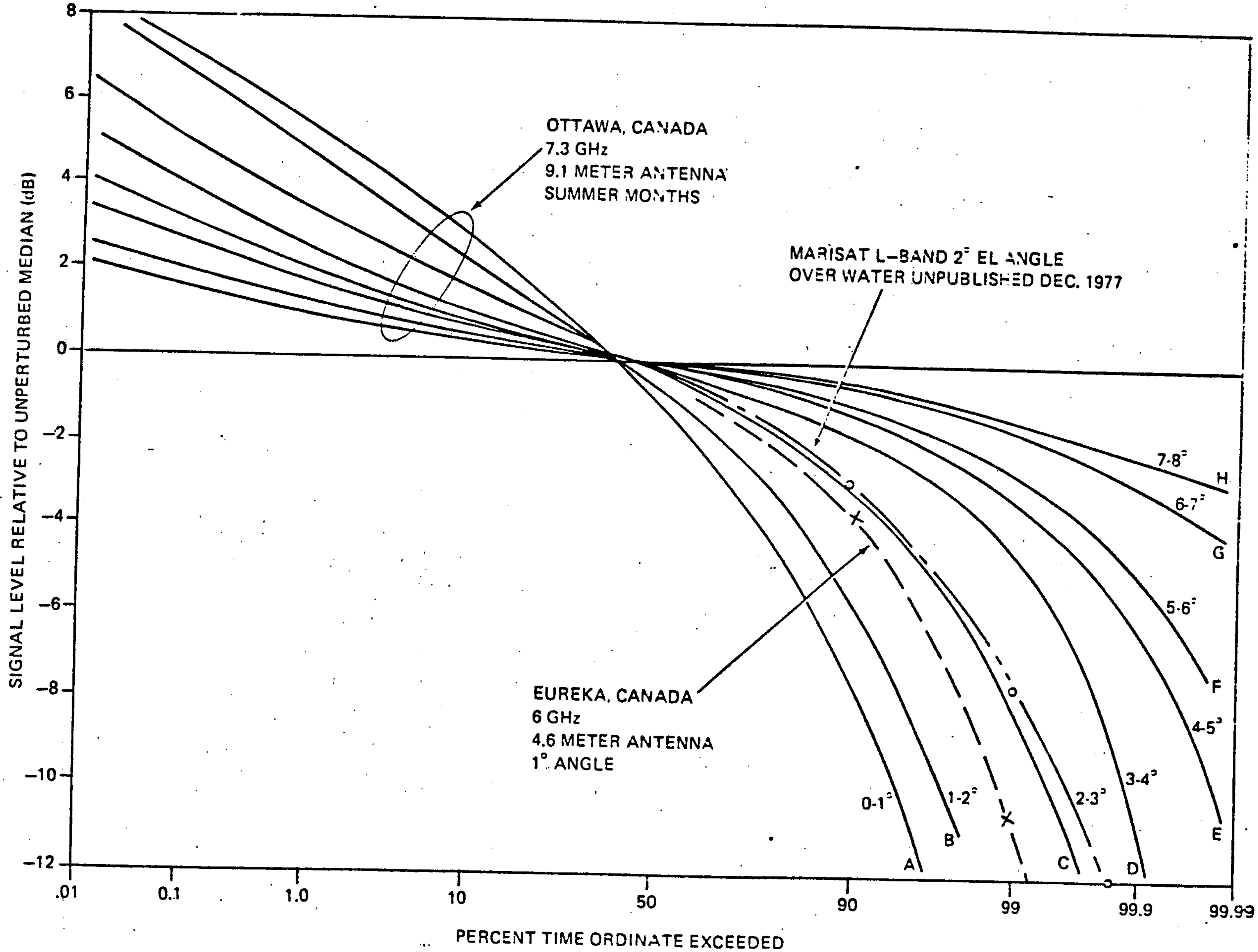



Figure 10 SCINTILLATION FADING

G–17389

angle. This data was collected in Ottawa, Canada using a 9.1 m antenna at 7.3 GHz. From that curve it can be seen that a margin of 8 dB would be required for 95% availability for such stations. For the station having an elevation angle of $3.2^{o}$, a margin of 8 dB yields 99.9% availability. This is overly conservative, due to the larger antenna, higher frequency and temperate climate for the data collection. A more likely estimate would be a 6 dB margin needed for 99.9% availability.

A theoretical Rayleigh distribution could be used to incorporate specific system factors, but this would also represent a worst case scenario, and would provide unrealistically high margins.

6.2.3 Radome Loss

Attenuation due to rainfall can be broken down into 2 categories: radome loss and raindrop loss.

Radome loss is dependent upon the surface-point rain rate, R, in mm/hr. The water thickness on the surface of the radome is given by: Reference?

$$t(\text{mm}) = .0349\ (\alpha\ R)^{\frac{1}{3}}$$

where α is the radome diameter in meters. Transmission loss for laminar flow can then be obtained as a function of water thickness according to Blevis. Ref ?

Rainfall distributions for earth station locations can be readily obtained. Combining this information with Figure 11 gives a cumulative distribution for radome loss, as seen in Figure 12.

This data indicates that loss due to water on the surface of a radome will obcur only 1% of the time in a given year for each ASO station. Typical dry radome loss has been quoted as .5 dB by manufacturers.

Again, the distribution curves are conservative but even for the worst case, a 5 dB radome loss occurs only .001% of the time in a given year.

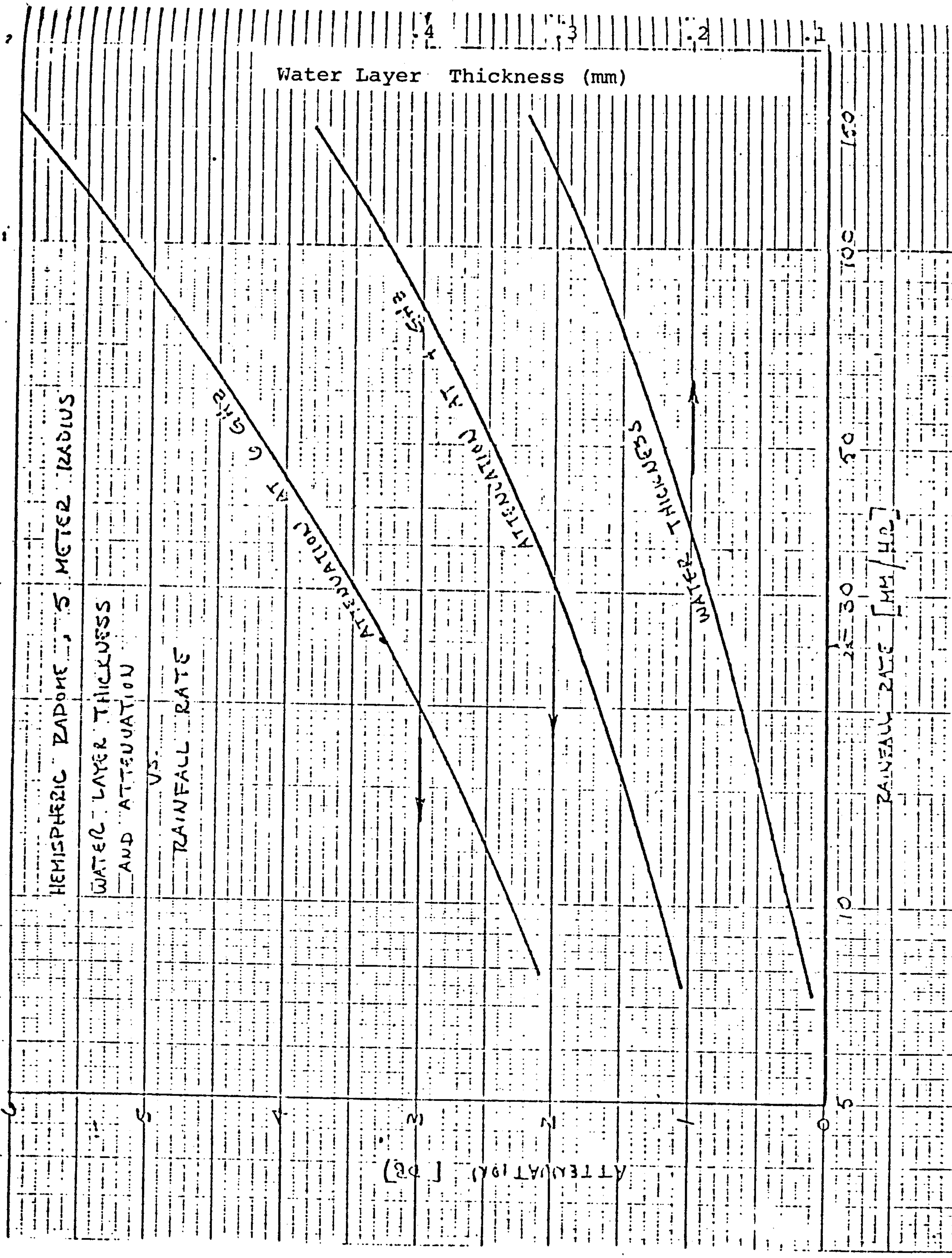
According to Hueter, radome loss dominates raindrop attenuation, and can generally be accounted for in the radome loss term.

6.2.4 Atmospheric Loss

Atmospheric loss is due to the presence of neutral oxygen and water vapor in the atmosphere. The amount of water vapor present is dependent on rainfall distribution and therefore varies with time. Benoit Ref calculates atmospheric attenuation as a function of frequency and elevation angle, avoiding the time distribution through the use of near-worst case attenuation parameters. Using the total vertical attenuation, $A_{gv}$, and the tropospheric equivalent depth, $r_v$ measured in km, the effective path length r, and atmospheric loss $A_g$ can be given by

$$r = [r_o^2 \sin^2 \varepsilon + r_v (2r_o + r_v)]^{\frac{1}{2}} - r_o \sin \varepsilon$$

$$A_g = \frac{A_{gv} r}{r_v}$$



Water Layer Thickness (mm)
.4
.3
.2
.1
HEMISPHERIC RADOME, 5 METER RADIUS
WATER LAYER THICKNESS AND ATTENUATION VS. RAINFALL RATE
ATTENUATION AT 6 GHz
ATTENUATION AT 4 GHz
WATER THICKNESS
RAINFALL RATE [MM/HR]
10
30
50
100
150
ATTENUATION [dB]


Figure 11

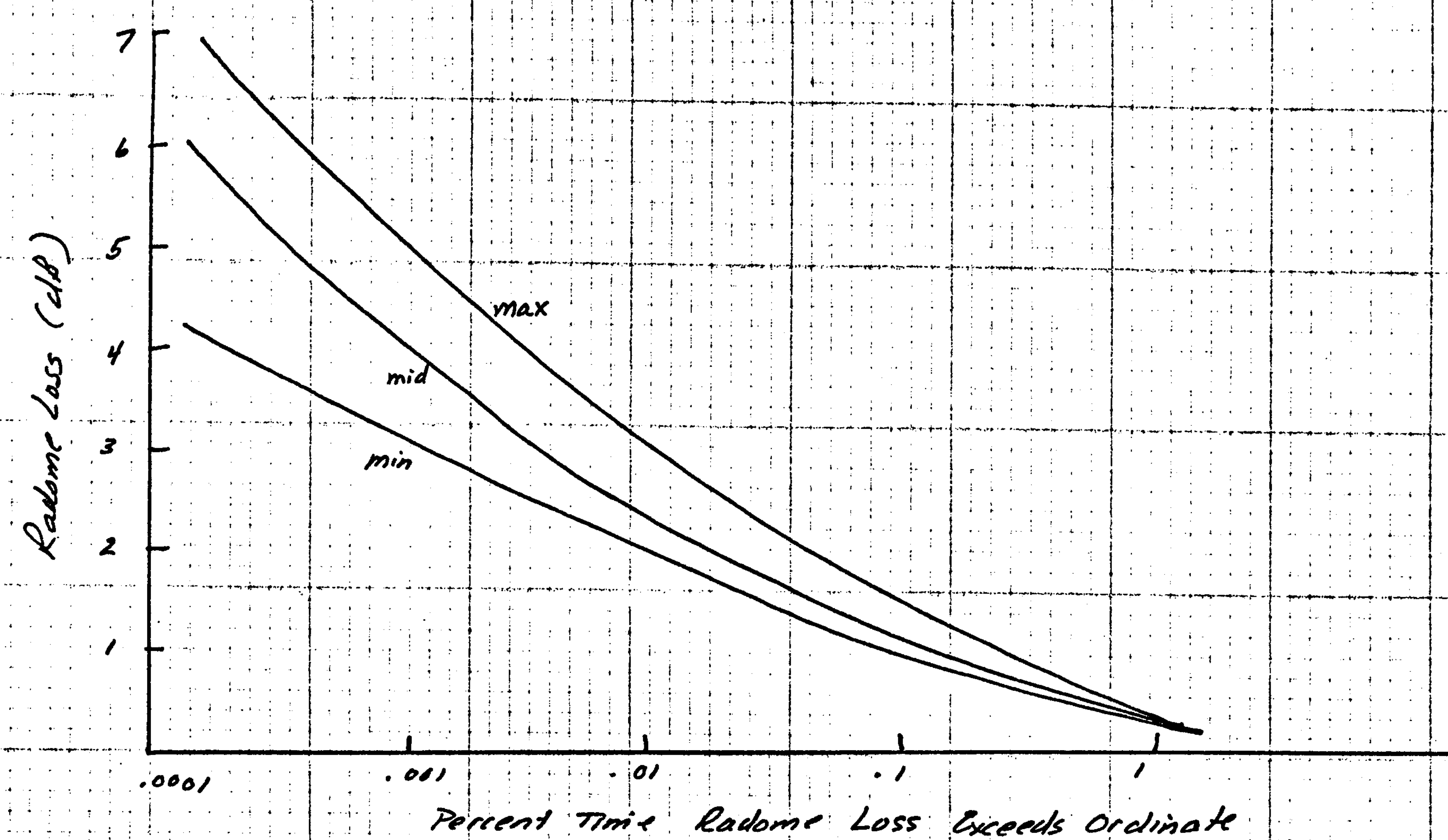



FIG 12 : Cumulative Radome Loss Distribution at 6GHz.

where $\varepsilon$ is the elevationangle and $r_o$ is the radius of the earth. Figure 13 plots atmospheric attenuation as a function of elevation angle and frequency.

From Figure 13 ,it is clear that allowing .5 dB atmospheric loss for both transmit and receive at the earth station will be more than adequate for any station chosen with an elevation angle greater than $8^o$.

### 6.2.5 Rain Losses

The loss of signal due to rain is a complex phenomenon. Power loss is due primarily to absorption due to molecular resonances and depolarizations due to scattering. Most models consider bulk loss characteristics and also assume single scattering and do not include multiple scattering. From a macroscopic point of view the loss of power over a distance $\Delta\ell$ is given by the following. Let P $(\ell + \Delta\ell)$ be the power at distance $\ell+\Delta\ell$ and P $(\ell)$ the power at distance $\ell$ $\ell$. Then

$$P(\ell+\Delta\ell) = P(\ell) - P(\ell)\int_0^\infty n(a)Q(a,\lambda)\,da \qquad (x)$$

where n(a) is the density of particles of radius a at distance $\ell$ and $Q(a,\lambda)$ is the extinction cross section for the particle diameter and wavelength $\lambda$. extinction

define the term!

### 6.2.6 Fixed Losses

The remaining loss are standard, and can be considered fixed over time. These include feed loss, implementation loss, polarization loss, antenna loss and multipath loss. Table 4 gives typical values for these losses as well as which link each affects.

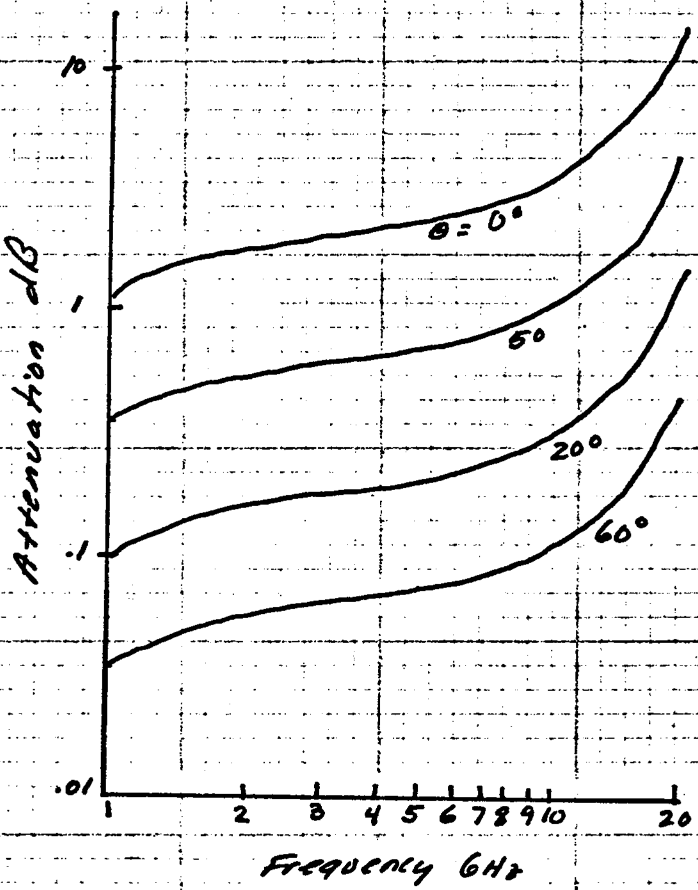



FIG 13 Atmospheric Attenuation due to Oxygen and Water at elevation angles θ. Ref?

earth station

Feed loss occurs only at the earth station on transmit.

Allowing 1 dB for feed loss is adequate. Implementation loss occurs on both links for receive only and assumes a value of 2 dB. Polarization loss is found at the remote terminal, on both transmit and receive, and is equal to .1 dB. Antenna loss occurs on receive at the remote terminal on transmit and receive for low elevation angle sites. It can be adequately covered by allowing 1 dB for multipath loss.

Returning to equation (X) as ~~As~~ $\Delta\ell$ goes to zero we obtain a differential equation

$$\frac{dP(\ell)}{dt} = -P(\ell)\,\alpha(\ell)$$

where

$$\alpha(\ell) = \int_0^\infty n(a)Q(a,\lambda)\,da$$

Solving this yields

$$P = P_0 \exp\left(-\int_\lambda^\ell \alpha(\ell)\,d\ell\right)$$

Now we must remember that n(a) depends also on time and is a statistical parameter. This only accounts for the absorption loss. The rain loss in dB is

$$10 \log_{10} \exp\left(-\int \alpha(\ell)\,d\ell\right)$$

The polarization loss is much more complex to analyze and is considered in the paper by Higg and Chu. Ref? Fig. 14 presents a cumulative distribution of losses for two frequency bands. It should be noted that rain losses are typically important at 6GHz and above. This figure shows the increase at the higher frequency. Fig. K depicts the losses as a function of vapor thickness as a function of frequency.

It should be pointed out that the rain losses are critical at the higher frequencies and they are combatted by increased power and diversity. Diversity means having several stations separated at distances large enough to decorellate the peak rain losses. To do this site surveys must be performed. In general the rain losses can be up to 20 dB and more! Thus care must be taken in earth station location ~~siting~~ and power control.

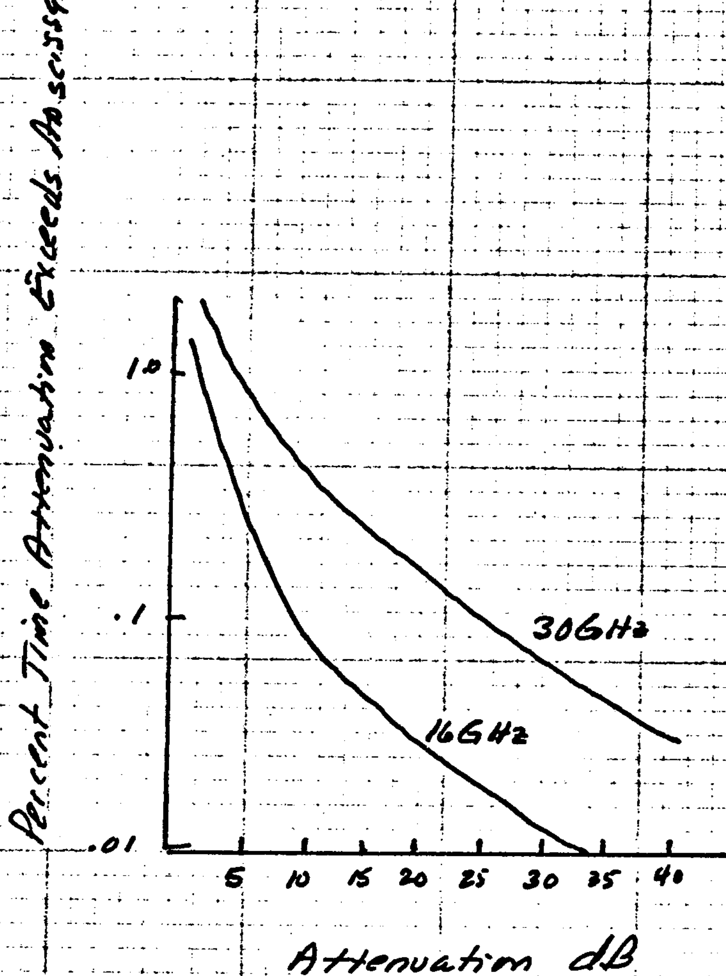
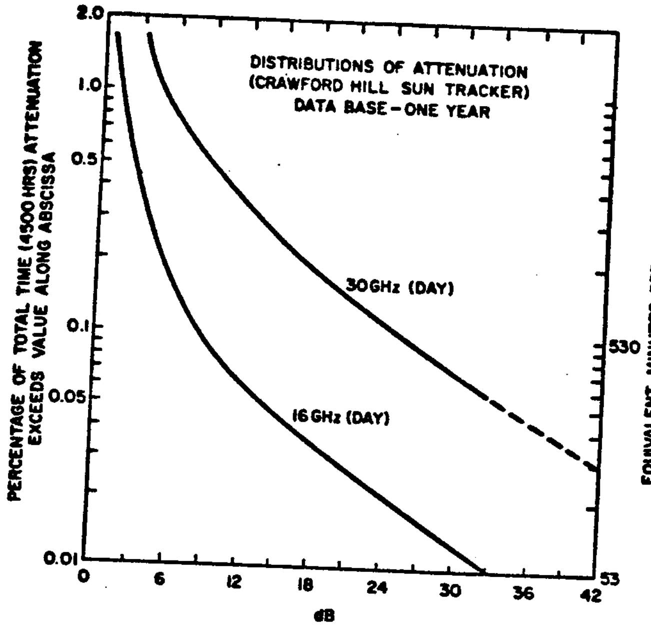



Fig 14 : Cumulative Rain Loss Distribution

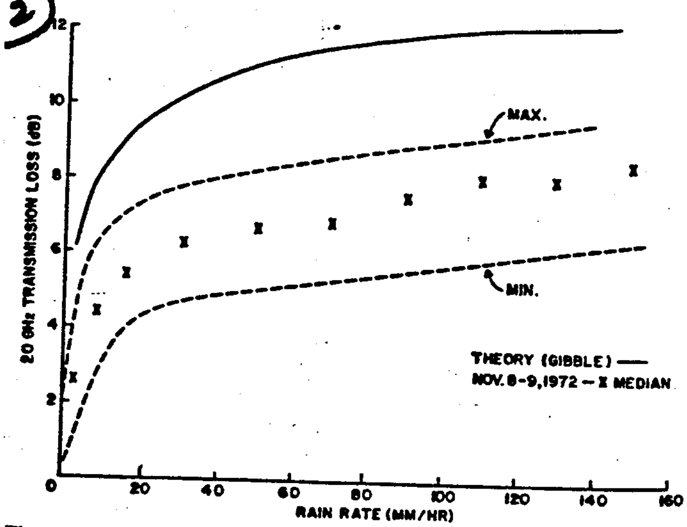



14 Fig. 21. Cumulative distributions of 16- and 30-GHz attenuation measured during 1968 in New Jersey using the suntracker of Fig. 20.

Ref ?

Fig. 45. Transmission through a layer of water at various common-carrier frequencies.

2




Change

15 Fig. 46. 20-GHz attenuation through a section of a large spherical radome versus measured rainrate.

Ref ?

| LINK | XMT FIXED | | RCV FIXED | |
|---|---|---|---|---|
| Command | $L_{MULT}$ | 0.0 | $L_{MULT}$ | 1.0 |
| | $L_{ANT}$ | 0.0 | $L_{ANT}$ | 0.0 (2.0) |
| | $L_{IMP}$ | - | $L_{IMP}$ | 2.0 |
| | $L_{POL}$ | 0.1 | $L_{POL}$ | 0.1 |
| | $L_{FEED}$ | 0.0 | | |
| TOTAL | | 0.1 | | 3.1 (5.1) |
| Data | $L_{MULT}$ | 1.0 | $L_{MULT}$ | 0.0 |
| | $L_{ANT}$ | 0.0 | $L_{ANT}$ | 0.0 |
| | $L_{IMP}$ | - | $L_{IMP}$ | 2.0 |
| | $L_{POL}$ | 0.1 | $L_{POL}$ | 0.0 |
| | $L_{FEED}$ | 1.0 | | |
| TOTAL | | 2.1 | | 2.0 |

TABLE 4.4 FIXED LINK LOSSES FOR COMMAND AND DATA

## 6.3 INTERCONNECTION

We have just described what is necessary to achieve performance on a single point to point link. That is for a given data rate and earth/space payment configurations, the resources in terms of eirp and G/T are specified. Now we want to consider a more complex problem, that of interconnecting users with various requirements.

Let us assume that we have N users that wish to communicate. Let us also assume that we have defined a traffic from user i to user j in terms of $T_{ij}$ which may be given in bits/sec. Then its NXN matrix $\underline{T}$,

$$\underline{T} = \begin{bmatrix} 0 & T_{12} & \cdots & T_{1N} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ T_{N1} & \cdots & T_{N,N-1} & 0 \end{bmatrix}$$

is called the traffic matrix. This matrix can be either time invariant or time varying. The INTELSAT system is an example of a relatively time invariant system. It is a trunking system that provides for interconnection. The SBS system, on the other hand, is more dynamic. That is $\underline{T}$ is changing rapidly with time and the satellite must reconfigure.

Now any satellite system is further partitioned into the following:

1. <u>Number of Satellites</u> - there may be one or more satellites located to cover a given region. In the INTELSAT case there are three satellites for the Atlantic Ocean region.

2. <u>Number of Beams</u> - each satellite has an antenna system that generates beams on both transmit and receive. These beams may be non-overlapping by use of spatial, frequency or polarization diversity. For example in INTELSAT there are beams at 6/4GHz that are spatially separated called hemi beams. There are 6/4GHz beams that are orthogonal in polarization called zone beams. There are 14/11GHz

beams called spot beams. Each of these beams can reuse the available frequency.

3. <u>Number of Transponders</u> - the satellite is channelized into transponders. This done for several reasons. Using limited bandwidth transponders (40-80 MHz) limits total data rate and thus peak power. Increased number of transponders increase availability by providing redundancy. The problem then is to route traffic through these.

The INTELSAT V coverage and transponder layout is shown in Fig. 16. Note that each transponder requires a separate up converter. Also note that the transmit plan, that is the assignment up, is exactly the same as the receive plan, assignment down, since this satellite just shifts all frequencies down by 2225 MHz.

The problem now is to take the traffic matrix, T, and partition it to satellites, beams, and transponders. That is, we want submatrices.

$$\underline{T}^{ijk}$$

such that $\underline{T}^{ijk}$ is the traffic in satellite i, beam j, transponder k. Note

$$\underline{T} = \sum_i \sum_j \sum_k \underline{T}^{ijk}$$

The constraint is that

$$\sum_l \sum_m T^{ijk}_{lm} < C_o$$

where $T^{ijk}_{lm}$ is the traffic from l to m on that satellite, in that beam and transponder and $C_o$ is the capacity of the transponder.

There are various techniques to solve this problem. Some of them use nonlinear integer programming techniques Ref? but most are just cut and try.

<u>Example:</u>

In the INTELSAT system the traffic for 1992 is estimated to be about 200,000 channels. The satellite beam configuration is shown in Fig. 17. It results in 17 different beams and the minimum transponder bandwidth is 40 MHz.

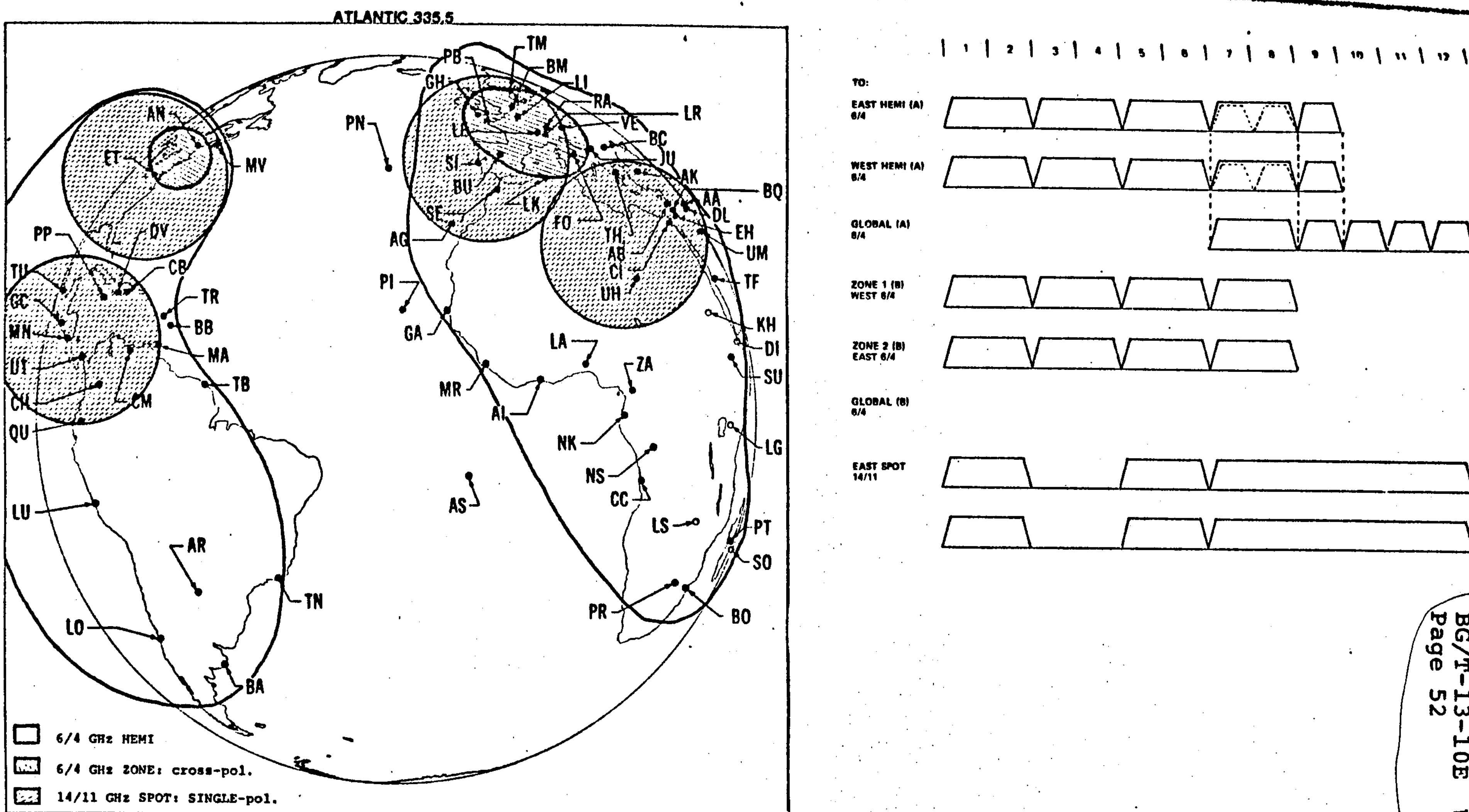



FIGURE 16 Coverages and Transponder Plan - Case B1

ATTACHMENT NO. 1 TO
BG/PC-5-5E W/10/75
BG/T-13-10E W/10/75
Page 35

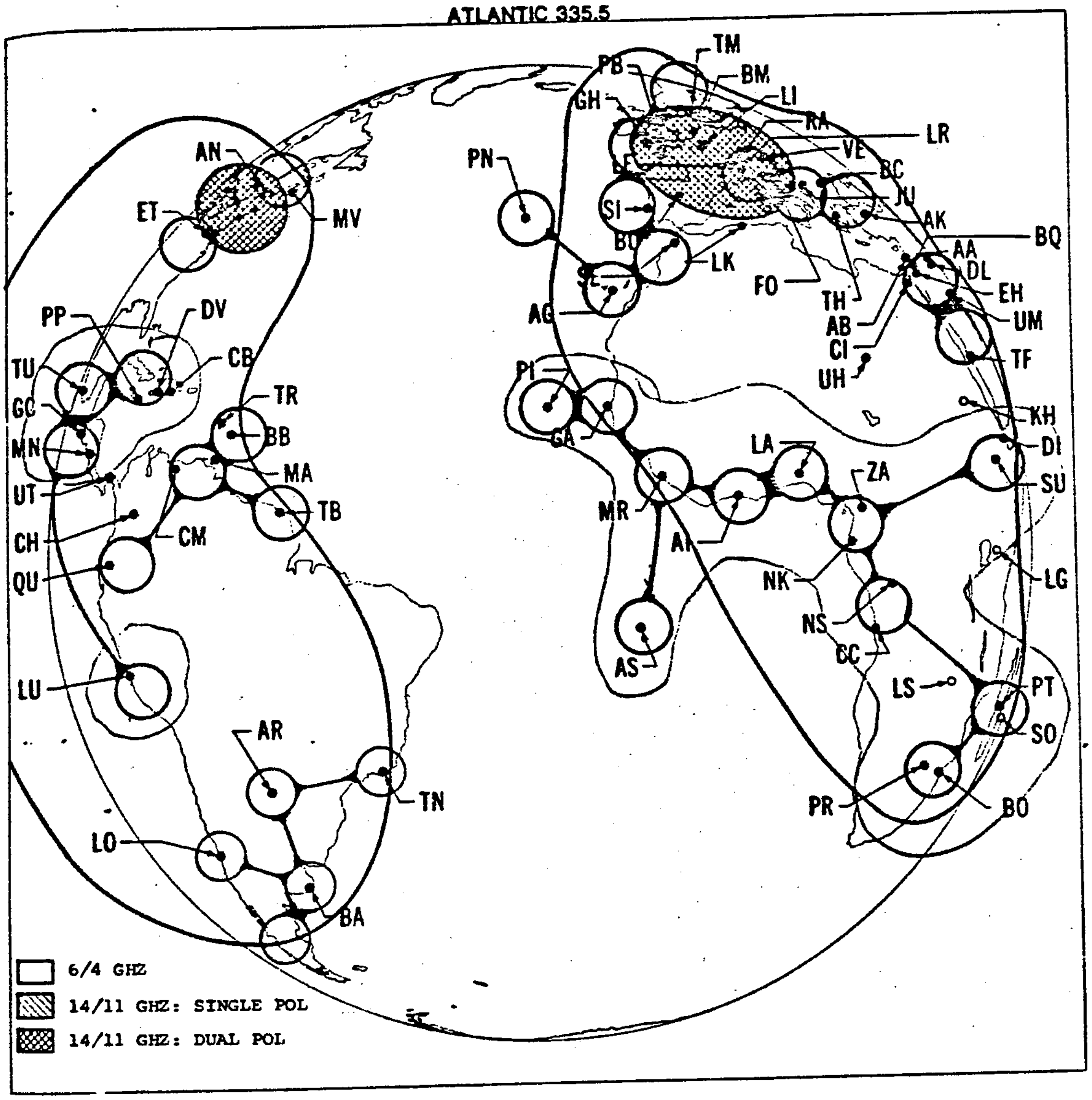



FIGURE 17 Atlantic Region Coverage - Concept A3

ATTACHMENT NO. 1 TO
BG/PC-5-5E W/10/75
BG/T-13-10E W/10/75
Page 37

| TO: | 1–2 | 3–4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|
| EAST HEMI (A) 6/4 | 1200 EAST HEMI | 1200 WEST HEMI | 600 EH | 600 WH | 600 WE | 502 AF | -- | -- | | |
| WEST HEMI (A) 6/4 | 600 W. SPOT (A) | 1200 EAST HEMI | | 600 EH | | 400 WH | -- | -- | | |
| GLOBAL (A) 6/4 | | | | | | | 271 G | SP | 271 G | TV |
| NORTHWEST–1(B) 6/4 | 1122 EAST EUROPE | 1109 WEST CARIB | 554 WC | 560 EE | 560 EE | 560 EE | 560 EE | 560 EE | | |
| NORTHWEST–2(B) 6/4 | | | | | | | | | | |
| EAST CARIB (B) 6/4 | 1026 E. SPOT (B) | 1116 W. SPOT (A) | | 503 WC | 362 EC | 40 WE | | 75 SA | | |
| WEST CARIB (B) 6/4 | 1071 E. SPOT (A) | 1109 NORTHWEST-1 | 554 NW-1 | 503 EC | 175 SA | 110 WC | | 25 ME | | |
| SOUTH AMER (B) 6/4 | 806 SOUTH AMER | | | | 175 WC | | | 75 EC | | |
| EAST EUROPE (B) 6/4 | 1122 NORTHWEST-1 | 769 E. SPOT (A) | 365 ME | 560 NW-1 | 560 NW-1 | 560 NW-1 | 560 NW-1 | 560 NW-1 | | |
| WEST EUROPE (B) 6/4 | 1082 AFRICA | 1086 W. SPOT (B) | 542 WS (B) | 190 ME | 600 EH | 40 EC | 540 AF | 504 WE | | |
| MID EAST (B) 6/4 | 956 W. SPOT (B) | 838 MID-EAST | 365 EE | 190 WE | 418 ME | 594 ES (B) | 594 ES (B) | 25 WC | | |
| AFRICA (B) 6/4 | 1082 WEST EUROPE | 1110 E. SPOT (B) | 555 ES (B) | 555 ES (B) | 555 ES (B) | 502 EH | 540 WE | 554 ES (B) | | |
| GLOBAL (B) 6/4 | | | | | | | -- | -- | 270 G | TV |
| EAST SPOT (A) 14/11 | 1071 WEST CARIB | 769 EAST EUROPE | 564 WS (A) | 564 WS (A) | 564 WS (A) | 564 WS (A) | 564 WS (A) | 564 WS (A) | 1126 W. SPOT (A) | |
| WEST SPOT (A) 14/11 | 600 WEST HEMI | 1116 EAST CARIB | 564 ES (A) | 564 ES (A) | 564 ES (A) | 564 ES (A) | 564 ES (A) | 564 ES (A) | 1126 E. SPOT (A) | |
| EAST SPOT (B) 14/11 | 1026 EAST CARIB | 1110 AFRICA | 555 AF | 555 AF | 555 AF | 594 ME | 594 ME | 554 AF | 1127 W. SPOT (B) | |
| WEST SPOT (B) 14/11 | 957 MID-EAST | 1086 WEST EUROPE | 542 WE | | | | | | 1127 E. SPOT (B) | |

FIGURE 18 Transponder Loadings for Atlantic Region Primary Satellite Concept A3

ATTACHMENT NO. 1 TO
BG/PC-5-5E W/10/75
BG/T-13-10E W/10/75
Page 38

| TO: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EAST HEMI (A) 6/4 | 1164 WEST HEMI | | 1163 WEST HEMI | | 479 NW-1 | 600 EH | 479 WS (B) | 479 NW-1 | | | | |
| WEST HEMI (A) 6/4 | 1164 EAST HEMI | | 1163 EAST HEMI | | 600 WH | 600 NW-1 | 600 NW-1 | 600 WH | | | | |
| GLOBAL (A) 6/4 | | | | | | | | | -- | -- | 466 G | 466 G |
| NORTHWEST-1 (B) 6/4 | 1152 WEST EUROPE | | 1192 E. SPOT (A) | | 479 EH | 600 WH | 600 WH | 479 EH | 576 WE | 575 WE | | |
| NORTHWEST-2 (B) 6/4 | 1092 EAST EUROPE | | 1091 EAST EUROPE | | 576 WE | | 597 ES (B) | | 315 ME | 557 SA | | |
| EAST CARIB (B) 6/4 | 1181 W. SPOT (B) | | 1162 E. SPOT (B) | | 581 ES (B) | 598 WC | 537 SA | 590 WS (B) | 590 WS (B) | 590 WS (B) | | |
| WEST CARIB (B) 6/4 | 1200 W. SPOT (A) | | 1200 W. SPOT (A) | | 600 WA (A) | 598 EC | | 470 ES (B) | 470 ES (B) | 471 ES (B) | | |
| SOUTH AMER (B) 6/4 | 1142 E. SPOT (A) | | 1114 W. SPOT (B) | | 572 ES (A) | 557 WS (B) | 537 EC | 534 SA | 534 SA | 557 NW | | |
| EAST EUROPE (B) 6/4 | 1092 NORTHWEST-2 | | 1091 NORTHWEST-2 | | 145 AF | 155 ME | | | | | | |
| WEST EUROPE (B) 6/4 | 1192 NORTHWEST-1 | | 464 AFRICA | | 576 NW-2 | | | | 576 NW-1 | 576 NW-1 | | |
| MID EAST (B) 6/4 | 784 E. SPOT (B) | | | | 315 WS (B) | 115 EE | | | 315 NW-2 | | | |
| AFRICA (B) 6/4 | | | 464 WEST EUROPE | | 145 EE | 563 ES (B) | | | | | | |
| GLOBAL (B) 6/4 | | | | | | | | | -- | -- | | |
| EAST SPOT (A) 14/11 | 1142 SOUTH AMER | | 1192 NORTHWEST-1 | | 572 SA | 596 WS (A) | 596 WS (A) | 597 WS (A) | 597 WS (A) | 597 WS (A) | 1194 W. SPOT (A) | |
| WEST SPOT (A) 14/11 | 1200 WEST CARIB | | 1200 WEST CARIB | | 600 WC | 596 ES (A) | 596 ES (A) | 597 ES (A) | 597 ES (A) | 597 ES (A) | 1194 E. SPOT (A) | |
| EAST SPOT (B) 14/11 | 784 MID-EAST | | 1162 EAST CARIB | | 581 EC | 563 AF | 597 NW-2 | 470 WC | 470 WC | 470 WC | 1194 W. SPOT (B) | |
| WEST SPOT (B) 14/11 | 1181 EAST CARIB | | 1114 SOUTH AMER | | 315 ME | 557 SA | 479 EH | 590 EC | 590 EC | 590 EC | 1194 E. SPOT (B) | |

FIGURE 19 Transponder Loadings for Atlantic Region Major Path 1 Satellite Concept A3

ATTACHMENT NO. 1 TO
BG/PC-5-5E W/10/75
BG/T-13-10E W/10/75
Page 39

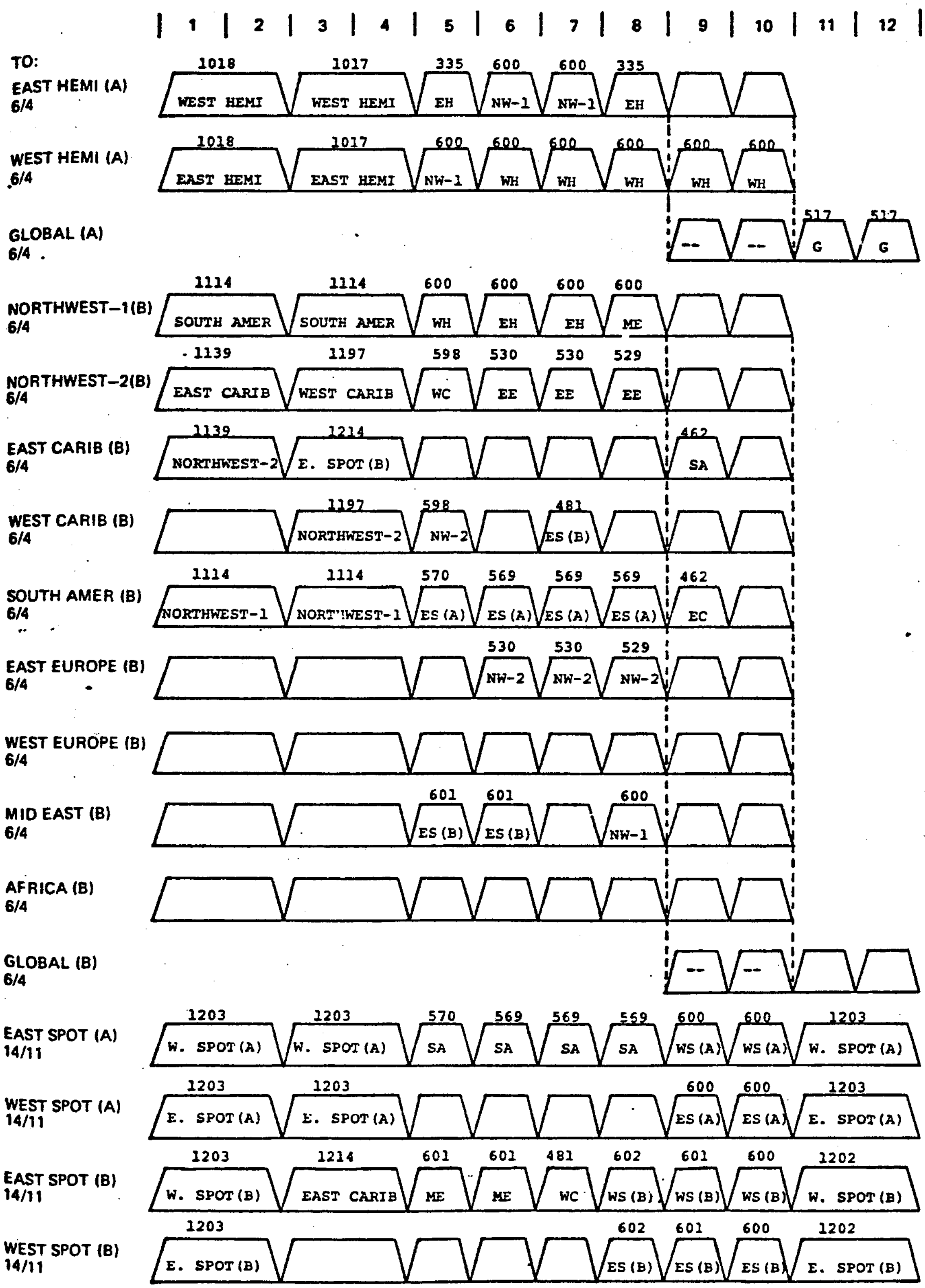



FIGURE 20 Transponder Loadings for Atlantic Region Major Path 2 Satellite (Concept A3)

ATTACHMENT NO. 1 TO
BG/PC-5-5E W/10/75
BG/T-13-10E W/10/75
Page 40

| ES. Name | No. | Primary Traffic Channels 6/4 | Primary Traffic Channels 14/11 | Major-1 Traffic Channels 6/4 | Major-1 Traffic Channels 14/11 | Major-2 Traffic Channels 6/4 | Major-2 Traffic Channels 14/11 |
|---|---|---|---|---|---|---|---|
| Algeria | 1 | 561 | | 0 | | 0 | |
| Angola | 2 | 544 | | 464 | | 0 | |
| Argentina | 3 | 125 | | 2389 | | 2294 | |
| Ascension Is. | 4 | 80 | | 0 | | 0 | |
| Barbados | 5 | 507 | | 507 | | 0 | |
| Belgium | 6 | 472 | | 906 | | 826 | |
| Brazil | 7 | 160 | | 3512 | | 3432 | |
| Cameroon | 8 | 784 | | 0 | | 0 | |
| Canada | 9 | 828 | 808 | 643 | 1870 | 1111 | 890 |
| Chile | 10 | 408 | | 1140 | | 1075 | |
| Colombia | 11 | 50 | | 1338 | | 1288 | |
| Dominican Rep. | 12 | 524 | | 524 | | 0 | |
| Ecuador | 13 | 683 | | 683 | | 0 | |
| Ethiopia | 14 | 381 | | 0 | | 0 | |
| France | 15 | 130 | 4473 | 1075 | 1796 | 0 | 1789 |
| French West Ind. | 16 | 805 | | 674 | | 0 | |
| Gabon | 17 | 433 | | 0 | | 0 | |
| Germany | 18 | 1637 | 643 | 60 | 2692 | 166 | 2708 |
| Greece | 19 | 255 | | 1084 | | 973 | |
| Guiana | 20 | 574 | | 0 | | 0 | |
| Iran | 21 | 1121 | | 0 | | 1071 | |
| Israel | 22 | 205 | | 1513 | | 1458 | |
| Italy | 23 | 1428 | 908 | 708 | 2156 | 40 | 2618 |
| Ivory Coast | 24 | 864 | | 473 | | 0 | |
| Jamaica | 25 | 267 | | 1653 | | 1386 | |
| Jordan | 26 | 571 | | 0 | | 0 | |
| Kuwait | 27 | 778 | | 311 | | 0 | |
| Malagasy | 28 | 271 | | 0 | | 0 | |
| Mexico | 29 | 1351 | | 1351 | | 0 | |
| Morocco | 30 | 503 | | 0 | | 0 | |
| Mozambique | 31 | 867 | | 0 | | 0 | |
| Netherlands | 32 | 261 | | 961 | | 841 | |
| Nicaragua | 33 | 932 | | 932 | | 0 | |
| Nigeria | 34 | 623 | | 0 | | 0 | |
| Nordic | 35 | 856 | | 856 | | 0 | |
| Panama | 36 | 200 | | 1348 | | 1218 | |
| Peru | 37 | 75 | | 1310 | | 1245 | |
| Portugal | 38 | 2087 | | 886 | | 422 | |
| Roumania | 39 | 321 | | 0 | | 0 | |
| Saudi Arabia | 40 | 591 | | 0 | | 0 | |
| Senegal | 41 | 714 | | 0 | | 0 | |
| South Africa | 42 | 706 | | 596 | | 0 | |
| Spain BU | 43 | 165 | 1357 | 407 | 1785 | 256 | 1399 |
| Spain AG | 44 | 291 | | 0 | | 0 | |
| Portugal VE | 45 | 201 | | 0 | | 0 | |
| Switzerland | 46 | 330 | | 1293 | | 1223 | |
| Trinidad | 47 | 508 | | 508 | | 0 | |
| Turkey | 48 | 524 | | 0 | | 0 | |
| U.K. | 49 | 873 | 5639 | 1097 | 4946 | 20 | 5677 |
| Iraq | 50 | 171 | | 0 | | 0 | |
| U.S. Mainland | 51 | 5995 | 9130 | 8927 | 12514 | 9879 | 8127 |
| Venezuela | 52 | 306 | | 3156 | | 2950 | |
| Yugoslavia | 53 | 567 | | 477 | | 0 | |
| Zaire | 54 | 834 | | 0 | | 0 | |
| Austria | 55 | 553 | | 0 | | 0 | |
| Egypt | 56 | 653 | | 0 | | 0 | |
| Guatemala | 57 | 554 | | 0 | | 0 | |
| Haiti | 58 | 533 | | 533 | | 0 | |
| Paraguay | 59 | 321 | | 0 | | 0 | |
| Liberia | 60 | 160 | | 0 | | 0 | |
| Portugal AZ | 61 | 635 | | 0 | | 0 | |
| Bolivia | 62 | 120 | | 0 | | 0 | |
| Lebanon | 63 | 371 | | 0 | | 0 | |
| U.S.S.R. | 64 | 161 | | 0 | | 0 | |
| Chile PA | 65 | 343 | | 0 | | 0 | |
| TOTAL | | 40698 | 22958 | 44259 | 27759 | 33174 | 23208 |
| Total per Sat. | | 63656 | | 72018 | | 56328 | |
| Total System | | 192056 | | | | | |

TABLE 5 Traffic Distribution by Earth Station and Frequency Band for System Concept A3

Using a cut and try approach the assignment is shown in Figs. 18 to 20. Here we have assigned units of traffic on a from/to basis. The source (beam) is shown in the transponder and the destination is shown to the left. For example, the East hemi transmits to the East hemi. The mid East group transmits to the West Spot.

This example shows a great deal of space unfilled. It represents an underutilized design. What is also important is that this system may start out at only 40,000 channels and be very much underutilized.

To date there is no well accepted optimal packing strategy that is time efficient. Significant work must still be done in this area.

The problem becomes more complex when the traffic is dynamic. In this case $\underline{T}$ (t) depends on time. We shall defer discussion of this until we get into the different multiple access schemes.

We can formalize the traffic assignment problem in the following manner. Let $\underline{T}$ be the traffic matrix and let there be N sources and N sinks. Let us assume that it is symmetric, although this is not a constraint. Let $T_{ij}$ be as before the traffic from i to j in suitable units. Then the total traffic is

$$T = \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij}$$

Now consider M beams that are non-overlapping. We shall begin with the assumption but we note that if we use two or more frequency bands or dual polarization that the non-overlapping assumption is not valid. We shall show how to ~~lift~~ remove this later. Let $B_i$ be the ith beam. Now the beams can either be defined a priori or we can define the beams as a posteriori. Either way, each user must be covered by one beam. We now define a binary variable, $V_{ij}$ which equals 1 if user i belongs to beam j and zero elsewhere. The matrix $\underline{U}$ is then called the association matrix, where $\underline{V}$ is NXM with elements $V_{ij}$. Ref?

$$\underline{U} = \left.\underbrace{\begin{bmatrix} U_{11} & & U_{1M} \\ \vdots & & \\ U_{N1} & & U_{NM} \end{bmatrix}}_{\text{Beam}}\right\} \text{User}$$

Recall that we can assign users to beams or beams to users. Also in the above matrix there are only N1 entries and

? N≤N zero entries. We further define beam vectors $V_i$ as

$$\underline{U} = \begin{bmatrix} | & & | \\ \underline{V}_1 & \cdots\cdots & \underline{V}_M \\ | & & | \end{bmatrix}$$

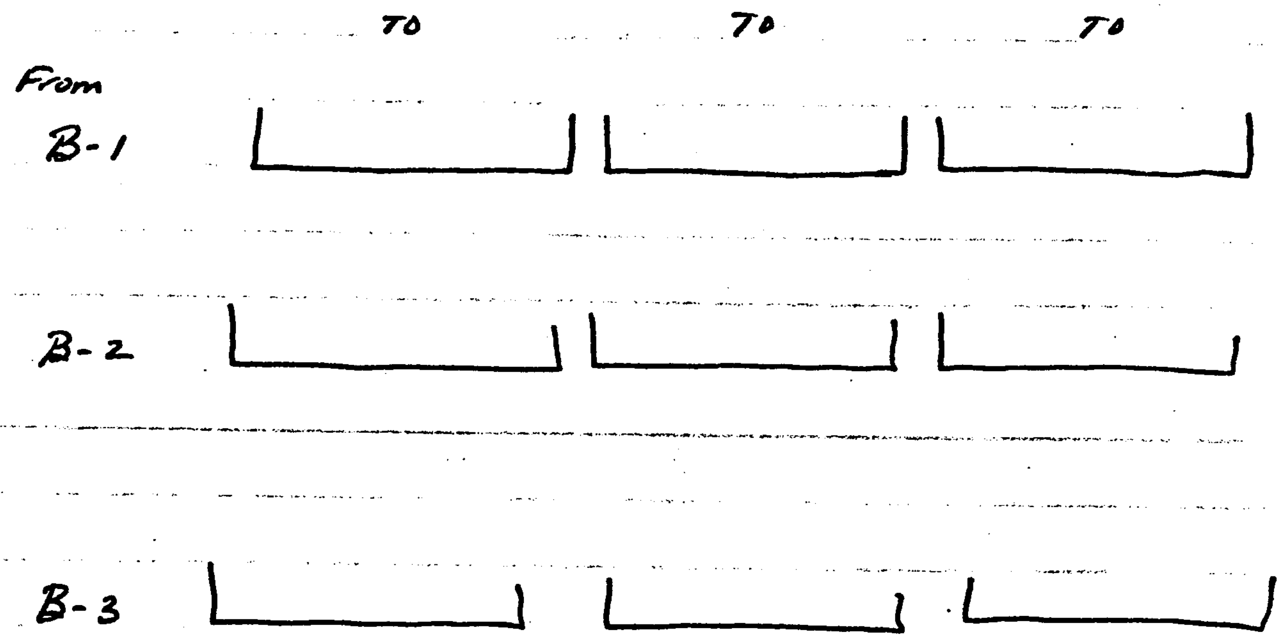



Fig 21 : Typical Beam to Beam Transponder Configuration

So show it!

where the elements are those of $U_{ij}$. It is easy to show that the total traffic in beam k is $S_k$.

$$S_k = \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} U_{ik}$$

$$= \underline{V}_k^T \underline{T} \underline{V}_k$$

where the second equality follows from the binary nature.

Fig. 21 depicts a typical transponder plan. In this case we have all circuits from B-1 going into three transponders with their down links to any other beam. We do the same for all other "from" beams. Thus a transponder is uniquely specified on a "from-to" basis. Note also that the transponders in the same column are in the same frequency so they cannot transmit to the same beam.

Each transponder has a capacity $C_o$. Let us assume that there are K transponders per beam. Thus the beam capacity is $KC_o$ and the total satellite capacity is $MKC_o$.

Given these for models we can now establish constraints between traffic, capacity, and allocation. First we must have

$$\sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} U_{ik} \leq KC_o$$

On the next level we note that the "from-to" transponders are quantized. Let $R_{mn}$ be the traffic from $B_m$ to $B_n$. ?
Then clearly

$$R_{mn} = \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} U_{im} U_{jn}$$

$$= \underline{V}_m^T \underline{T} \underline{V}_n$$

Define $N_{mn}$ as the number of transponders used to transmit $R_{mn}$. This is

$$N_{mn} = \left\lceil \frac{R_{mn}}{C_o} \right\rceil$$

where [x] is integer greater than or equal to x. Thus for each beam, m, we have the constant

$$\sum_{n=1}^{M} N_{mn} \leq K$$

The beam and transponder assignment problem can be stated as follows. Given a traffic matrix, T, how ~~do I~~ does one select the beams, that is the U matrix, so that all the constraints are satisfied. In particular how ~~do I do this~~ is this done with a minimum number of beams. Moreover, how ~~do I do it~~ is it done with beams that are physically realizable. We develop this concept in the following example.

<u>Example</u>:

Consider the traffic matrix shown in Table 6. We are assuming ten sources in a symmetric mode. Each of these is a separate earth station. Fig. 22 a shows the locations of these stations, in some geographical context. It may, for example, be global.

We now look for "natural" beams in terms of coverage and traffic constraints. We choose the following;

B1: 1,2,3
B2: 4,5,7
B3: 6,8,9,10

This choice is based on two factors. ~~First is~~ The first factor is the traffic constraint. Second B1 and B3 are "natural" beams. B-2 can be generated by two "natural" beams and the connection is only logical and not physical. With this partition we can generate a traffic matrix on a beam to beam basis. This is,

$$T = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{cc} \begin{matrix} B1 & B2 & B3 \end{matrix} & \\ \begin{bmatrix} 0 & 40 & 20 \\ 40 & 0 & 15 \\ 20 & 15 & 10 \end{bmatrix} & \begin{matrix} B1 \\ B2 \\ B3 \end{matrix} \end{array}$$

This satisfies our transponder constraints since we have assumed three transponders with a capacity of twenty each.

Fig. 23 depicts the loading on a beam to beam basis for this traffic matrix. Note that all transponders except 5, 7, and 9 are fully loaded. In Fig. 24 we present the complete loading of station pair i-j for beam 1. The triangular regions are used to represent loading only. One should be careful to ensure that intermodulation effects do not occur. This analysis, however, typically is performed after the loading.

As one final point in this example let us note that if the beam traffic matrix were;

$$\underline{T} = \begin{bmatrix} 5 & 35 & 20 \\ 35 & 5 & 15 \\ 20 & 15 & 10 \end{bmatrix}$$

why?

we could not satisfy it in this design. We would have to further channelize or add beams. However, as we shall note later, with an onboard switch, SS-TDMA could satisfy the traffic in the same configuration.

We can now ask how this traffic assignment can be automated. Typically in the INTELSAT case we start the traffic and advance to its beam assignment. However, it is important to factor in the issue of beam size into the development of the assignment. Thus besides a traffic matrix $\underline{T}$ we define a distance matrix $\underline{D}$, where $\underline{D}$ has elements $D_{ij}$ which represent the distance between i and j. Thus if $D_{ij}$ is small and we assign them to different beams this may require low sidelobes and thus a large aperture. Therefore $\underline{D}$ must enter into the beam assignment.

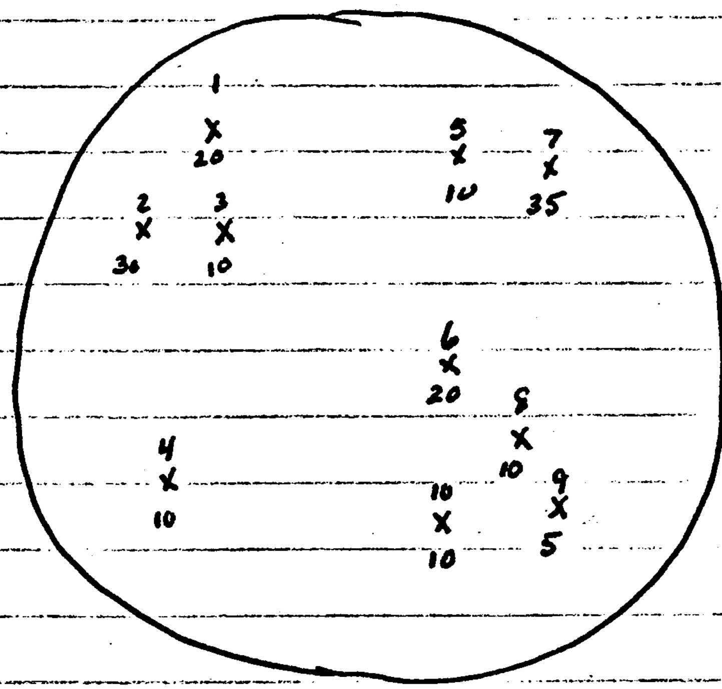
The following represents a typical assignment algorithm.

1. Order all stations in order of descending traffic.

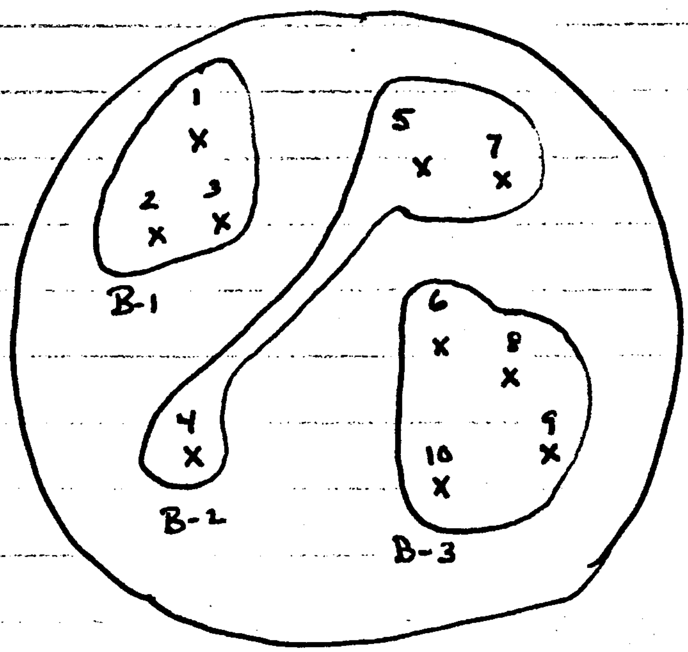
2. Start with 1 beam. If all the traffic fits stop.

$$
\begin{array}{c|c}
\text{From} \diagdown \text{TO} &
\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} &
\begin{bmatrix}
0 & 0 & 0 & 5 & 0 & 0 & 10 & 0 & 0 & 5 \\
0 & 0 & 0 & 5 & 5 & 10 & 10 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 5 & 0 \\
5 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \\
0 & 10 & 0 & 0 & 0 & 0 & 5 & 5 & 0 & 0 \\
10 & 10 & 5 & 0 & 0 & 5 & 0 & 5 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 5 & 5 & 0 & 0 & 0 \\
0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \\
 & \begin{array}{cccccccccc} 20 & 30 & 10 & 10 & 10 & 20 & 35 & 10 & 5 & 10 \end{array}
\end{array}
$$

Table 6 : Traffic Matrix Example




(a)
Locations




(b)
Beams

Fig 22: Locations and Beam Coverage

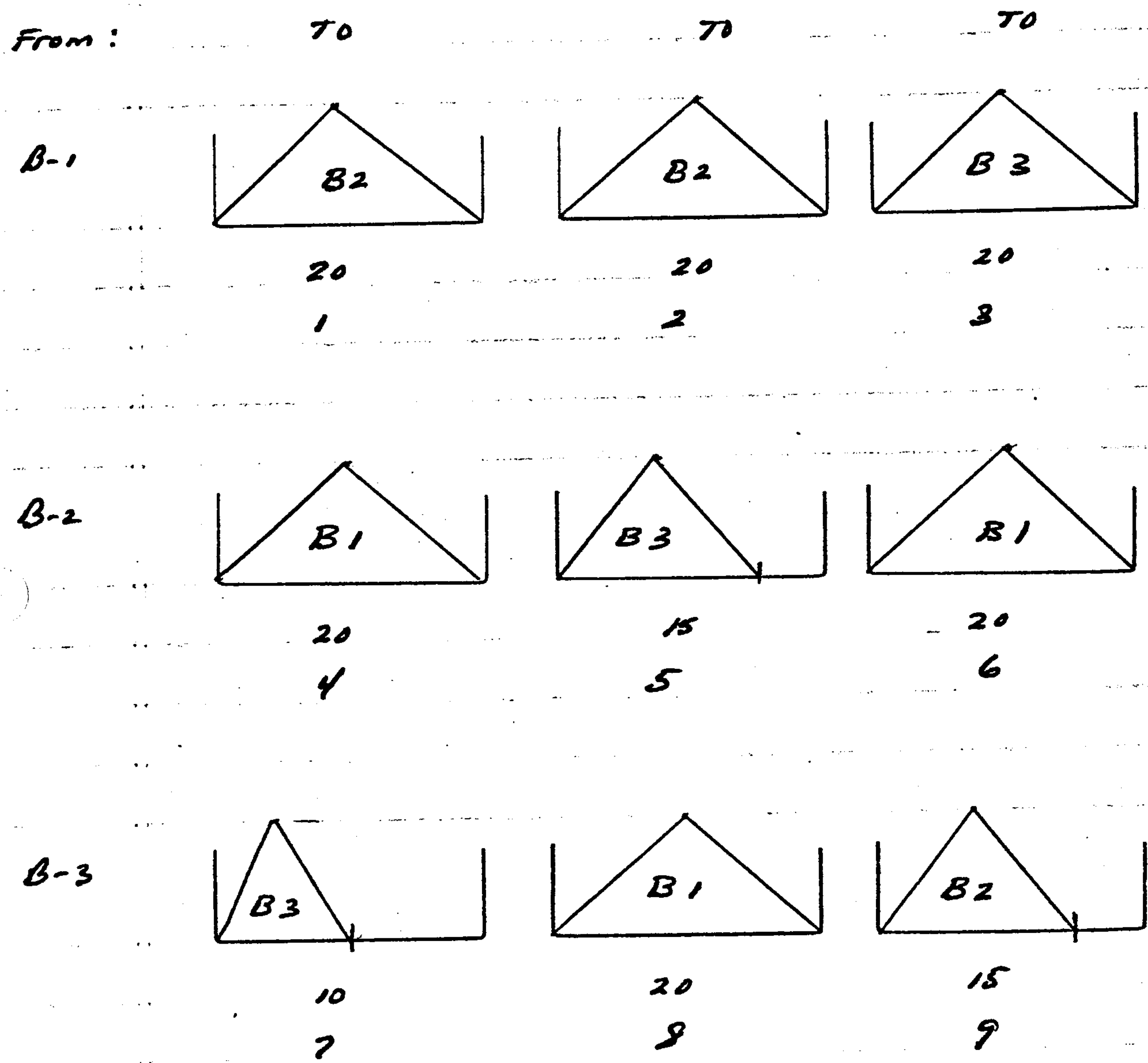



Fig 23 : Beam to Beam Transponder Loadings

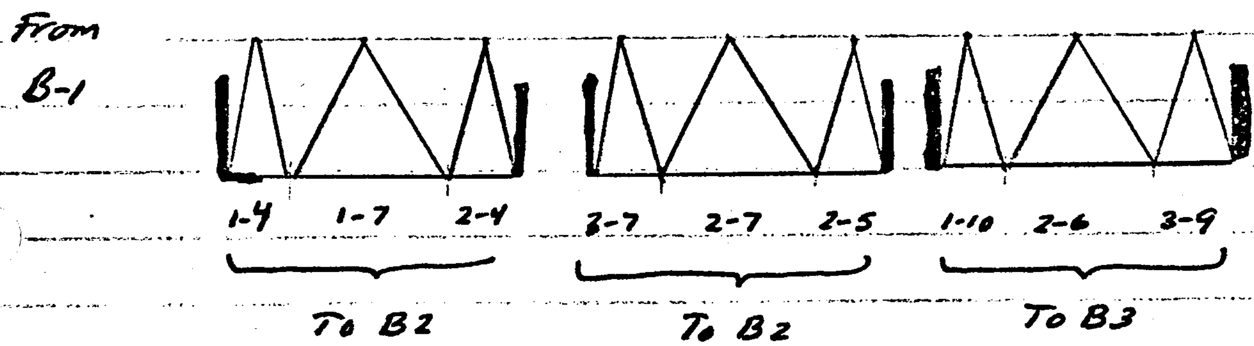



Fig 24: Beam 1 Transponder Loadings

3. Increase the number of beams by one. Choose the top N largest sources, where N now represents the number of beams. Examine $D_{ij}$ for those sources to see if any are too close. Remove the close ones that have least traffic and select more sources to increase it to N until satisfied. Call these N the beam kernels.

4. Add to the kernels from the bottom of the list up, checking to see if both traffic and beam constraints are mt.

5. If all are used up, stop and use the assignment. If there are some left that would result in a violation, go to (3) with an increase in beams.

This is a heuristic algorithm. It is possible to approach this also as an integer programming problem, but in general the complexity is too great. The above heuristic algorithm tends to adequately satisfy the requirements. It is a clustering algoirthm that is used in many pattern recognition schemes.

## 6.4 MULTIPLE ACCESS SCHEMES

As we discussed in the last section, with a given traffic matrix $\underline{T}$ we can, based upon general capacity constraints, partition $\underline{T}$ to a satellite, beam, transponder and in turn give the traffic in that slot from source i to destination j. This partitioning is done for a single instant of time so that if T changes drastically a repartitioning may be required. However, for most systems this global repartioning is not required, only a possible repartioning in a transponder. The purpose of this section then is two fold. Given a partition of traffic in a transponder, how do we physically provide for these multiple accesses. Then, if changes occur, how do we provide for fluctuations in demand.

The first question above is answered when we present the three multiple access schemes. The second question is answered by studying demand assignment schemes and some of the various packet techniques that are now employed. We can now begin our discussion of the three multiple access chemes; time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA).

### 6.4.1 TDMA

Let us assume that the partitioning process has given us a sub-traffic matrix with elements $T_{ij}$ indicating units of traffic from i to j. For simplicity let us also assume that all the earth stations in the network are the same and that the assigned transponder has an eirp that is known and constant. Now we can consider a scheme whereby each of the separate sources transmits independently of the other and that the transmission occurs at the peak rate allowed by the link budget.

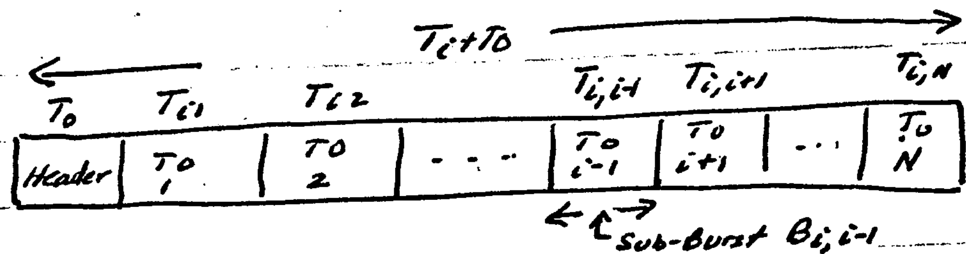
Let us consider a simple example of a 40 MHz transponder using 4Ø PSK (QPSK). We know that such a system could support a 50 mbps data stream. Now let us further assume that the traffic units are in bits/sec., e.g., $T_{ij}$ bps. Then clearly for a single transponder

$$\sum_i \sum_i T_{ij} < 50 \text{ Mbps}$$

Now let i=1 transmit its data first. Let it do this once a second so that it will transmit $T_i = \sum_j T_{ij}$ bits of data in one second. Then cycle through all the other users. We can view the transmission of any i as shown in Fig 25. Here we have a single Burst with $T_i + T_o$ bits where $T_o$ is a set of header bits which we shall return to shortly. This burst is called $B_i$. Now all stations then transmit on different times at the high burst rate so that in 1 sec we see all (N bursts). This called a frame of data as shown in Fig. 25. Note that we have left a guard band of $B_o$ bits between bursts to ensure that they do not overlap. This guard band is necessary because of timing errors that are present in each of the earth stations.

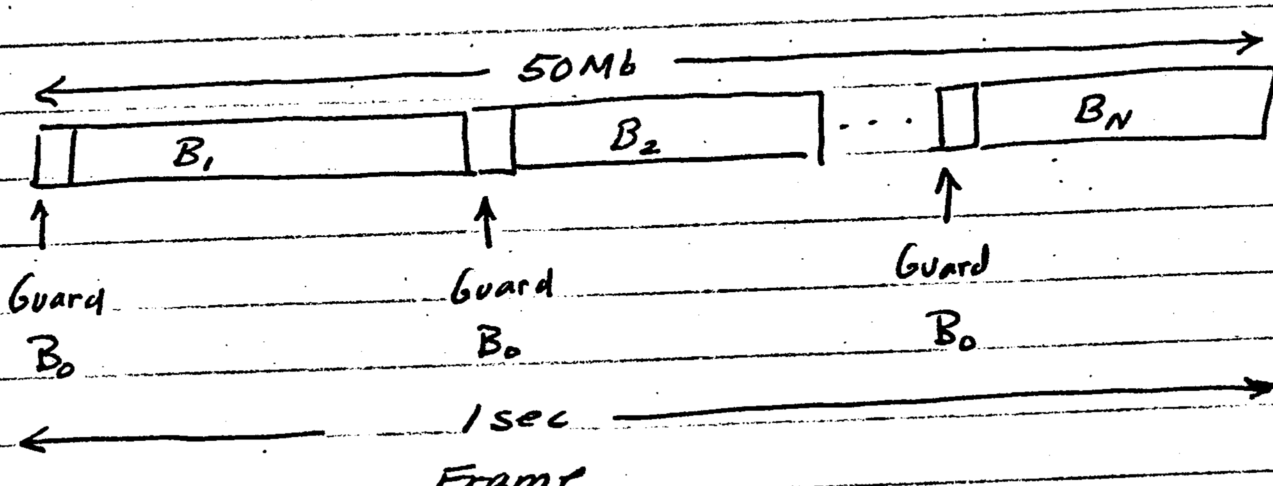
Let us return to this generic TDMA format and examine what the structure implies.

1. Transmit Time - If each of the earth stations in the net desire to communicate with this system then there must be some method of recognizing first when they should transmit and second where to put the sub-bursts. This information must therefore be available to all users in the local network. There are several ways in which this can be done. The first approach assumes that all stations have data maps. These maps serve several functions. First they tell each station when to transmit and how long that burst transmission can occur. Second they tell all receiving stations when they can expect to receive the burst from station i and are ready to look for the header. Third the maps must be detailed enough to define subburst boundaries and to explicitly allocate time slots in a burst to a sub-burst. This approach requires careful map up keep in order to maintain the timing as well as very accurate timing.




(a)

Data From ES $i$; Burst $i$.




(b)

Data Frame

Fig 25 : Generic TDMA Frame

A second approach is to add additional address information in both bursts and subbursts. This information may first be preceded by a unique identifier word and then the address. This of course increases the overhead but reduces the data storage. It further allows for increased flexibility in a burst by permitting sub-bursts to be increased on-line. (Fig. 26)

In this second case we can define an efficiency factor as the ratio of data to data plus overhead. This is:

$$\eta_{\text{self mapped}} = \frac{\sum_i \sum_j T_{ij}}{NT_o + NB_o + N^2A_o + \sum_i \sum_j T_{ij}}$$

Here we have N headers plus N guard bands plus N addresses per burst and N bursts per frame.

For the mapped case we have no addresses so that

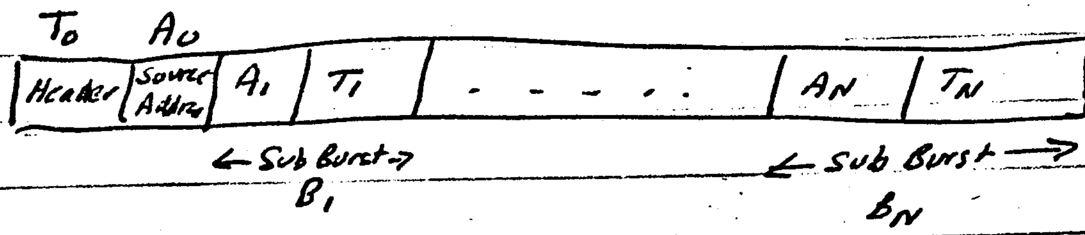
$$\eta_{MAP} = \frac{\sum_i \sum_j T_{ij}}{NT_o + NB_o + \sum_i \sum_j T_{ij}}$$

To have high efficiency we clearly want to keep the subbursts and bursts as long as possible since the addresses are fixed and so too are the headers. Yet two factors mitigate against this. First the larger they become the longer we wait between frames. This increases the guard band due to clock stability. The second reason is memory which we shall discuss below.

2. Memory - Consider the TDMA unit shown in Fig. 27. Here there are N sources of data destined for N other earth stations. They are entering continuously at rates $R_r$ bps. Let us assume that the burst duration for this station is $\Delta_i$ seconds and that the burst rate is $R_o$. The time between bursts is $\Delta$ sec where

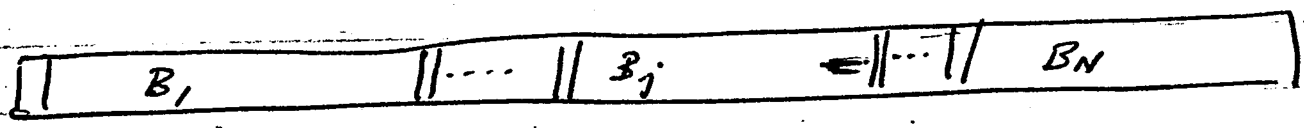
$$\Delta = \sum_i \Delta i + N\Delta_o$$

the sum of the individual burst rate plus N guard band times $\Delta_o$ where $\Delta_o$ equals $1/B_o$.




(a)

Burst ~~Frame~~; Addressed




(b)

Frame

Fig 26 : Self Mapped TDMA

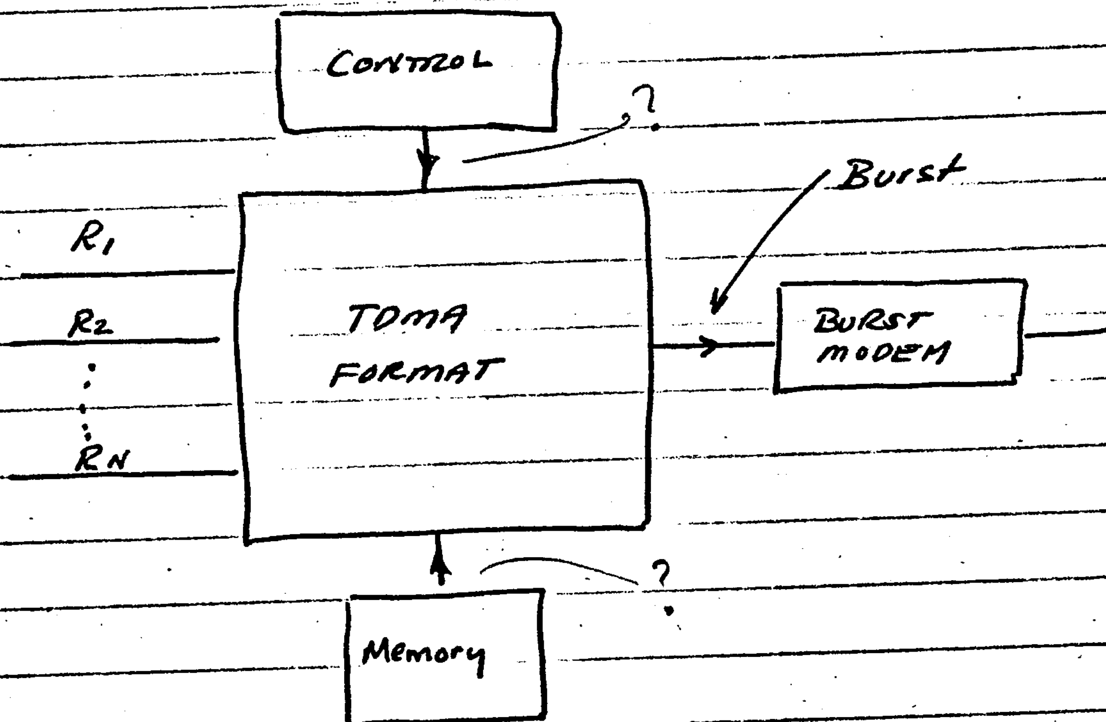



Fig 27 : Generic TDMA Architecture

?

Since the data is continuously flowing into the TDMA unit, it must be stored in memory. The data stored is the sum of the input rates times the duration. Let this be $D_i$, then

$$D_i = (\sum_k R_k)\,\Delta \text{ bits}$$

From this it is also clear that $D_i$ equals $T_i$ which is the number of bits in the burst. Let us assume that all sources have the same rate. Then let $T_i$ equal T and $D_i$ equal D. The efficiency equals

$$M = \frac{NT}{OH+NT} \qquad \eta = \frac{NT}{OH+NT}$$

where OH is the total overhead bits. We argued that we want T to be large to get M high. However large T means large D. For example with 10 users and a 50 Mbps channel, we assume that the overhead is 10,000 bits per frame. Let us assume that each earth station user generates a total of 4.8 Mbps. We can now trade off efficiency and memory as a function of the frame length.

The data storage is

$$D = 4.5 \text{ Mbps } \Delta \text{ (bits)}$$

and the efficiency is

$$\eta \quad M = \frac{45 \text{ Mbps } \Delta}{10{,}000 + 45 \text{ Mbps } \Delta}$$

In Fig. 28 we have plotted M and D versus Δ. At D of 10 msec the efficiency is 97% but the storage is 45Kb. To go to 99% we need 450Kb of storage. Thus in this design a 10 msec frame may be appropriate. Before answering that we must return to the system and see what that implies. This means that with 10 bursts per frame this is 500Kb per frame and 50Kb per burst. also this is about 5Kb per subburst. Now the question is does this type of partitioning make sense with the data source. If so, then we are complete.

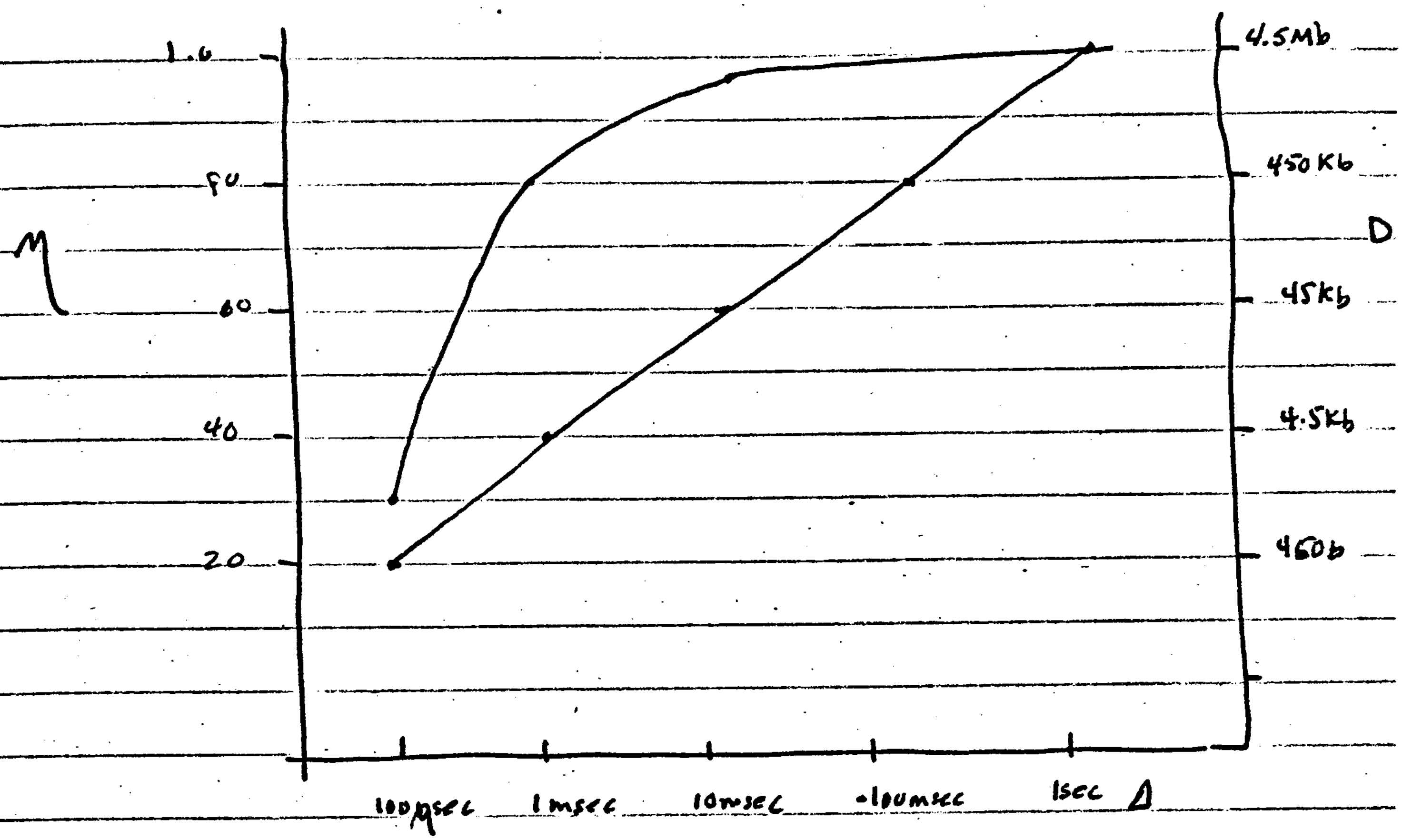



Fig 28: TDMA Efficiency and Storage versus Frame Time

When we have a multiple beam satellite system we can reduce the matrix to one that gives beam to beam traffic. If we define $T_{ij}$ as the traffic from beam i to beam j then we have

$$T = \begin{bmatrix} T_{11} & \cdots & T_{1N} \\ \vdots & & \vdots \\ T_{N1} & \cdots & T_{NN} \end{bmatrix}$$

If we use TDMA then we transmit from one beam to another through at least one transponder. For fixed TDMA we have a channelized system because the transponders are physically connected on a permanent basis from $B_i$ to $B_j$ (beam i to beam j). This tends to be an inefficient use of bandwidth at times because it requires many transponders of small bandwidth. A method to get around this is to use satellite switched TDMA, SS-TDMA, which works as follows. A switch is placed at the input of the satellite channels which allows $B_i$ to be connected by switch changes to any $B_k$. The only constraint is that at no time may two or more outputs be to the same beam. This switch must be centrally controlled and must be synchronized to all earth stations.

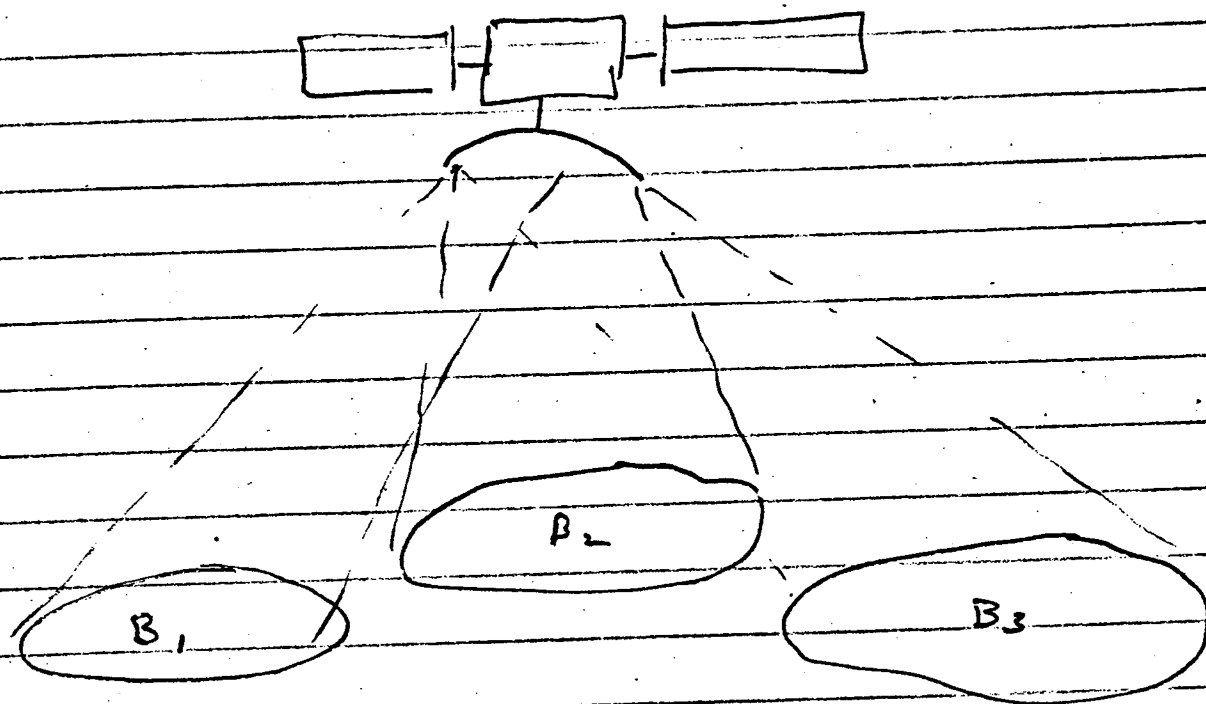
Now consider the system in Fig. 29. Here we have three beams and the traffic matrix as shown. In Fig 30 a we show how this would be done in fixed TDMA. We have nine transponders each of width 50. A total width of 150 is required. Now if we use SS-TDMA we can use only three transponders as shown in Fig. 30 b. Furthermore it is possible to arrange the bursts as shown so that the width needed is only 115. This is a savings of 35.

Muratani Ref ? has generalized this result as follows. Let $S_i$ be all the traffic leaving zone i,

$$S_i = \sum_j T_{ij}$$

and let $R_i$ be all the traffic coming into zone i

$$R_i = \sum_j T_{ji}$$




(a)

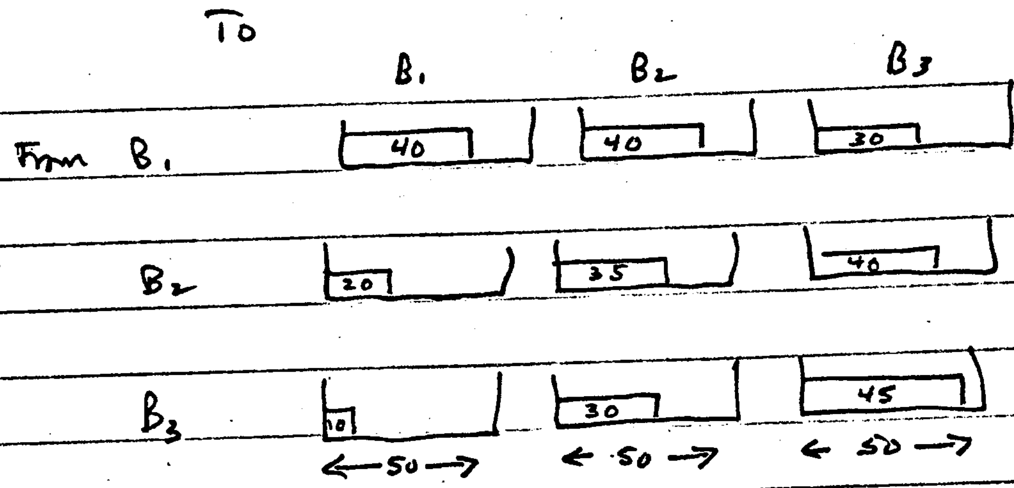
Beam Coverage

$$
T = \begin{array}{c|ccc}
\text{From} \backslash \text{To} & B_1 & B_2 & B_3 \\
\hline
B_1 & 40 & 40 & 30 \\
B_2 & 20 & 35 & 40 \\
B_3 & 10 & 30 & 45
\end{array}
$$

(b)

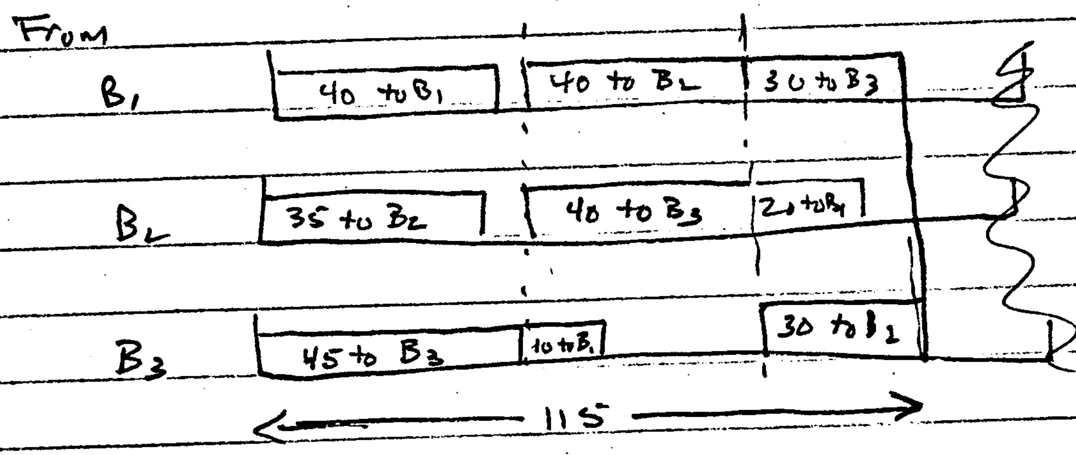
Traffic Matrix

Fig 29 : Multibeam Coverage and Traffic Matrix




(a)

Fixed TDMA




(b)

SSTDMA

Fig 30 : Comparison of TDMA and SSTDMA

Then let

$$T_R = \max\{S_i,\ R_i;\ i=1,\ldots,N\}$$

Then the maximum width to transmit all the traffic is $T_R$ and there exists a partition to do so. In our example $T_R$ is 115 and we showed the solutions. For more complex matrices this solution can be found by iterative procedures.

We can compare the channelized TDMA to SSTDMA by use of an efficiency factor ~~M~~ $\eta$. Let

eta?

$$\eta\ M = \frac{\text{Capacity Used}}{\text{Capacity Available}}$$

Clearly the capacity used ~~in~~ is ~~$\sum_i S_i$~~ $\sum S_i$. The capacity available is calculated as follows. For TDMA we need transponders of width $\max(T_{ij})$. We need $N^2$ for them. Thus,

$$\eta\ M_{TDMA} = \frac{\sum_i S_i}{N^2 \max(T_{ij})}$$

For SSTDMA we need a single transponder per beam of width $\max(S_i, R_i)$. Thus

$$\eta\ M_{SSTDMA} = \frac{\sum_i S_i}{N \max(S_i, R_i)}$$

For our example this yields;

$$\eta\ M_{TDMA} = 72\%$$

and

$$\eta\ M_{SSTDMA} = 84\%$$

This difference can represent a significant savings in satellite weight and cost.

### 6.4.2 FDMA

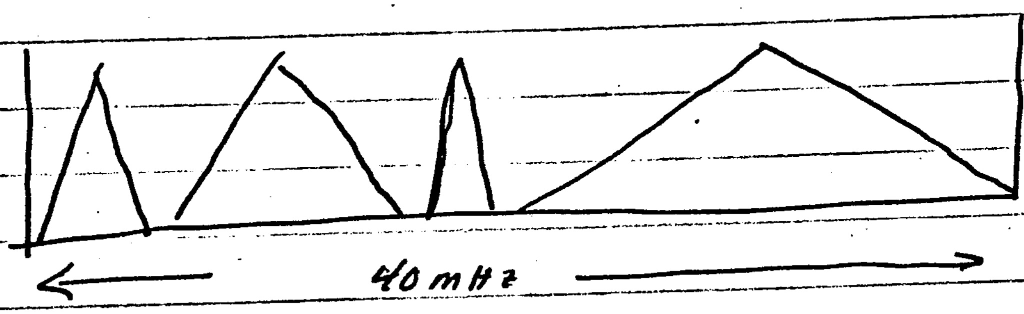
This form of multiple access is the oldest and is in most extensive operational use. In this case we begin as we did before by considering traffic on a satellite, in a beam and in a transponder. In that context let $T_{ij}$ be the traffic from i to j. Now $T_{ij}$ can be in units of bandwidth rather than bits per second. As we showed before this is a simple translation knowing the modulation scheme and the efficiency in bps/Hz.

In FDMA each traffic load occupies a different center frequency and a different bandwidth. This is shown in Fig. 31. Here there are four carriers up and respectively four down. Each has a different bandwidth. Now for each we must perform a link budget calculation.

The only problem with FDMA is the issue of interference due to third order harmonics. This results if the TWTA or HPA is driven too close to saturation where it behaves in a very nonlinear fashion. When this occurs the basic frequencies, $f_i$, mix and generate signals at frequencies of the form $f_i+f_j-f_k$. These are shown in Table 8. Note that there are eight of these. The task in FDMA is to choose the center frequencies $f_i$ so that the third order terms do not fall on any of the first order items. Again this is a cut and try process. Clearly as the number of carriers increase this problem becomes very complex.

The way that this problem is minimized is to keep the transponder well backed off; typically 6 to 10 dB from peak. The disadvantage of this is obvious. We cannot take full advantage of the full eirp capability of the system.

FDMA for quite a while was much less costly than TDMA since it merely required offset oscillators to center the different carriers. These are generally less expensive than a full TDMA system. In addition TDMA requires full capability right from




(a)

Transponder Loading

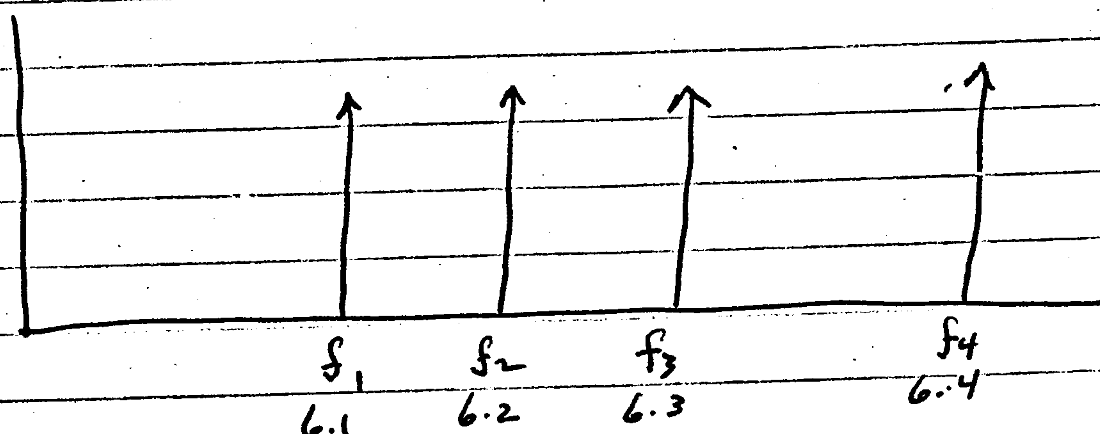



Fig 31 : FDMA Scheme

$$
\begin{array}{l}
f_1 \\
f_2 \\
f_3 \\
f_4
\end{array}
$$

$$
\left.\begin{array}{l}
f_1 + f_2 - f_3 \\
f_1 + f_3 - f_2 \\
f_1 + f_1 - f_2 \\
f_1 + f_1 - f_3 \\
f_2 + f_2 - f_1 \\
f_2 + f_2 - f_3 \\
f_3 + f_3 - f_1 \\
f_3 + f_3 - f_2
\end{array}\right\} \text{Third Order Harmonics}
$$

Table 8 : List of Fundamentals and Third order Harmonics.

the start. Thus the investment is large at the front end. In today's world of reduced cost digital circuits, this difference is disappearing for most large scale systems. For thin route systems however, FDMA is still the preferred choice.

### 6.4.3 CDMA

The same source characteristics apply to this case as to TDMA and FDMA. In TDMA we were time orthogonal and in FDMA we were frequency orthogonal. In CDMA we introduce the concept of code orthogonality. Let us assume that traffic $T_{ij}$ is as before but in bits per second. Let $W_S$ be the total bandwidth available. Clearly the bandwidth of $T_{ij}$ would be a very small fraction of $W_S$. Now we take this data stream and spread it over the available bandwidth by multiplying by a spreading waveform. For example if $s_i(t)$ is the data wave form from data source i then we take $s_i(t)$ and multiply it by a wave form $\tilde{s}_{ik}(t)$ which is the spread wave form. This latter wave form has bandwidth of $W_S$.

The spreading wave form may be of two types. The first is called frequency hopped. In this case we transmit $s_i(t)$ at its data bandwidth, $W_o$, but move the center or carrier frequency around all over the $W_S$ band. If we define n as $W_S/W_o$, then n is a measure of the number of hops. The pattern of hopping is also important. If we have many users in the system, say M, then if $n > M$ we can arrange it so they never hop into the same small frequency slot at the same time. The frequency of hopping is also a key factor as well as coherency. We develop these concepts in the problems.

The second form of CDMA uses a spreading waveform that is time modulated. In this case we let the data wave form be:

$$s_i(t) = \sum_{k=-\infty}^{\infty} w_{ik} u(T-kT)$$

where $w_{ik} = \pm \sqrt{E/T}$ and

$$u(t-kT) = \begin{cases} 1; kT \le t \le (k+1)T \\ 0; \text{elsewhere} \end{cases}$$

The spreading wave form however is:

$$\tilde{s}_{ij}(t) = \sum_{n=-\infty}^{\infty} v_{ijn}\, v(t-n\frac{T}{m})$$

where $v_{ijn} = \pm\sqrt{1/T}$ and

$$v(t) = \begin{cases} 1 & ;\ 0 \le t \le \frac{T}{m} \\ 0 & ;\ \text{eleswhere} \end{cases}$$

Here m is called the spreading factor. Since the data has bandwidth $W_o$ the spread bandwidth $W_S$ is m $W_o$. Note also in this model that we have synchronized the data and the spreading waveform. The data is determined by $w_{ik}$ which is random. The performance is determined by the spread sequence $v_{ijn}$ which we shall now concentrate on.

Figure 32 depicts the modulation and demodulation of the CDMA systems. Using this we can analyze its performance in a multiple access mode. At the demodulator we use $\tilde{S}_{ij}(t)$ first as the integrator correlator. Since we have synchronized the spread and data and we in addition assume synchronization of all other users (we may not require this in practice), we note that if

$$r(t) = s_i(t)\,\tilde{s}_{ij}(t) + n(t)$$

Then the output of the multipler is;

$$r(t)\, s_{ij}(t) = \tilde{s}_i(t)\,\tilde{s}_{ij}^{\,2}(t) + n(t)\, s_{ij}(t)$$
$$= s_i(t) + n(t)$$

not true, the spectral density is spread!

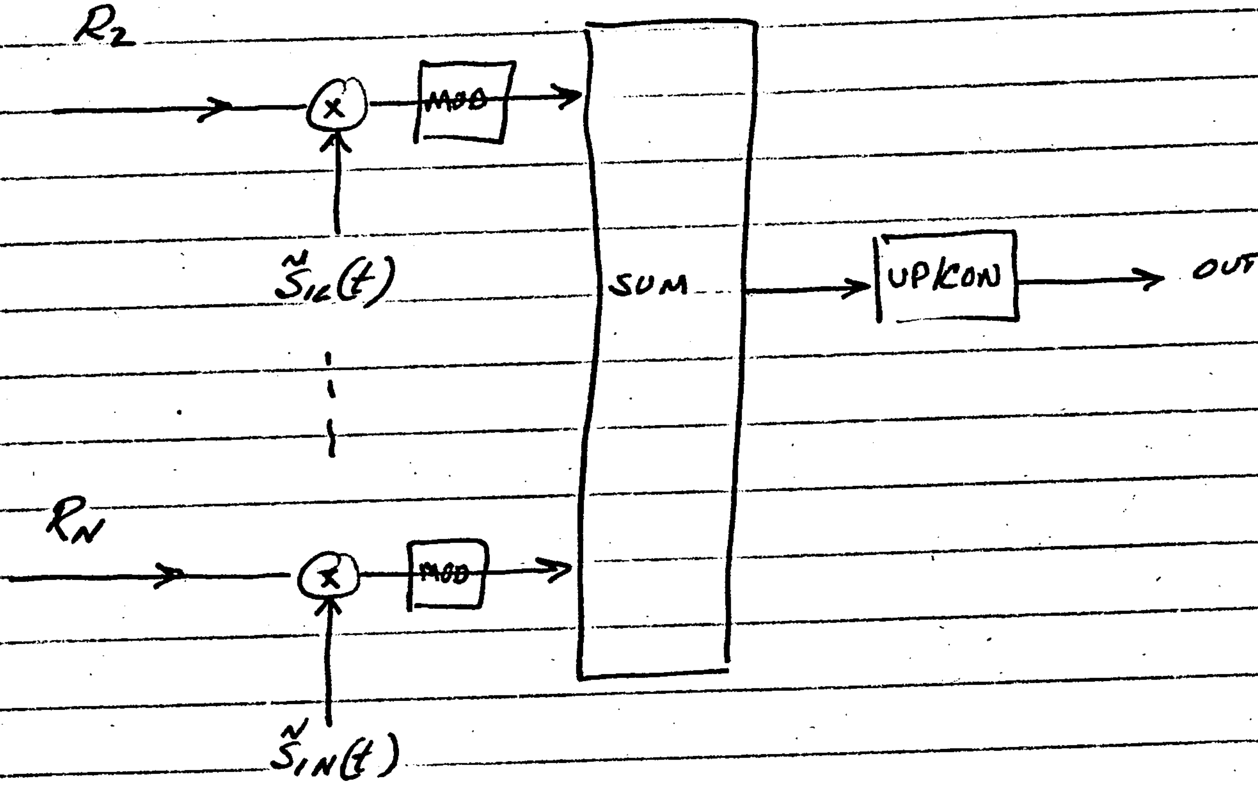
Since $\tilde{s}_{ij}^{\,2}(t)$ is 1 and the statistics of n(t) are not changed by the multiplication. Now if

$$r(t) = s_k(t)\,\tilde{s}_{kj}(t) + n(t)$$

then the multiplier yields

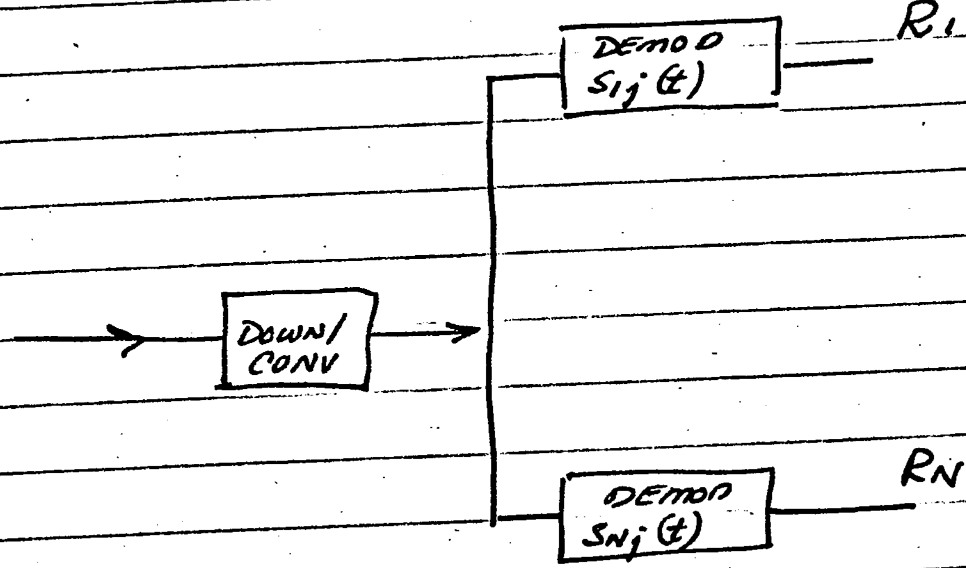
$$r(t)\tilde{s}_{i,j}(t) = s_k(t)\,\tilde{s}_{kj}(t)\,\tilde{s}_{ij}(t) + n(t)$$

Then if we integrate on kT to (k+1)T we obtain

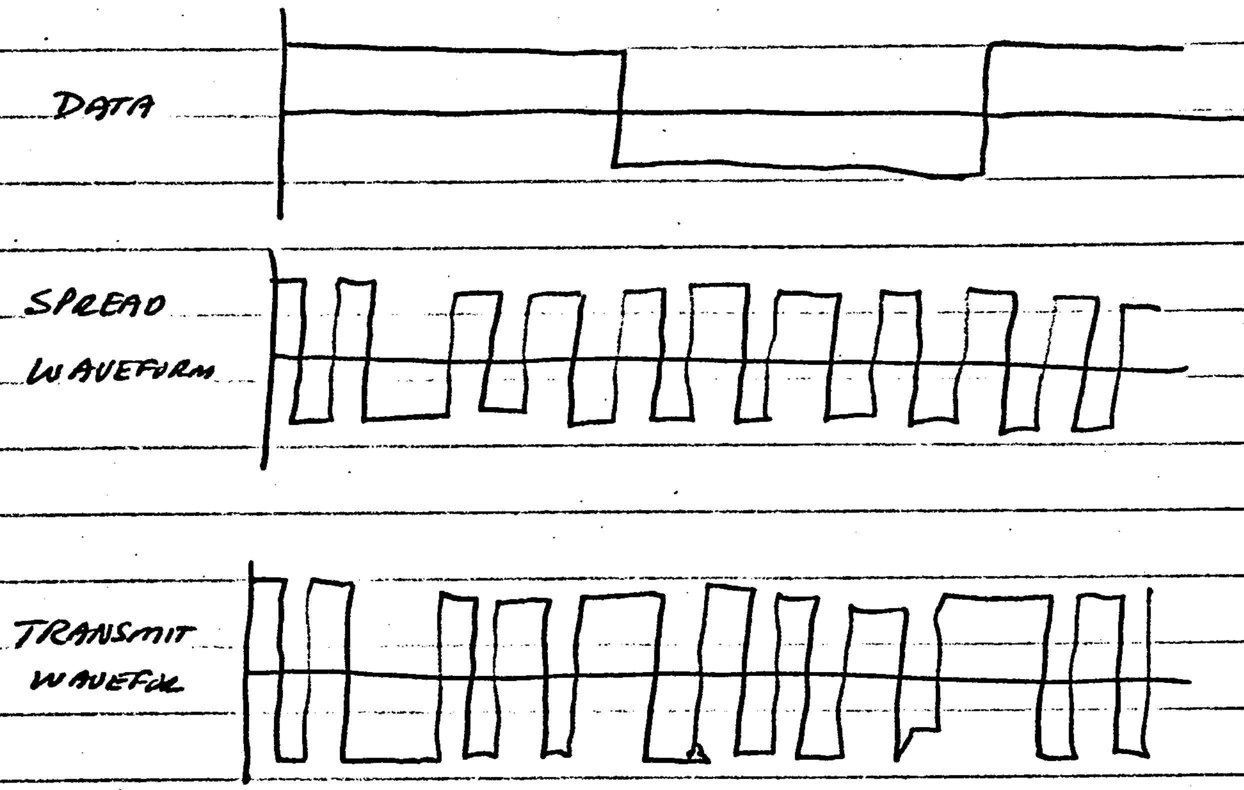



(a)

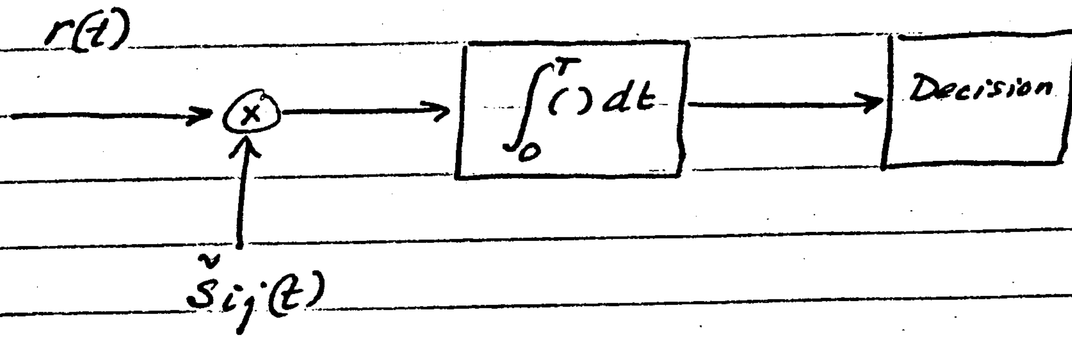
CDMA Transmit




(b)
CDMA DEMOD




(c)




(d)
Demodulator

Fig. 32 : CDMA Basic System Elements

$$r = \int_{kT}^{(k+1)T} r(t)\,\tilde{s}_{ij}(t)\,dt$$

$$= \pm \sqrt{E} \sum_{h=1}^{m} v_{kjn}\, v_{ijn} + n$$

If we choose m large enough, and we select the spreading sequences to be orthogonal, then;

$$\sum_{n=1}^{m} v_{kjn}\, v_{ijn} \simeq 0$$

and this means that all other signals in CDMA appear as noise only.

We can view this also by noting that when we correlate with the correct spreading sequence we reduce the output signal information bandwidth to $W_o$. However, if we correlate with another signal the output is still of bandwidth $W_S$. Thus another realization of the direct sequence spread receiver is shown in Fig. 33 where we follow the correlator with a filter, $W_E$ bandwidth equal to $W_o$.

We can now use the above concepts and return to the calculation of the link budget.

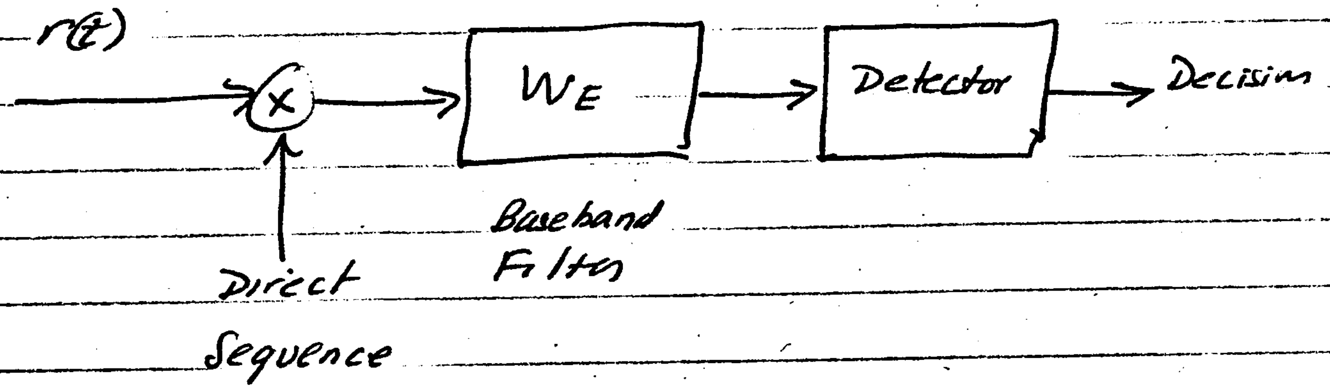



Fig 33 : Baseband CDMA Receiver

author:
You need a RF heterodyne
for a RF CDMA receiver!

A direct sequence spread spectrum system is depicted in Figure 34 . A data source generates $R_o$ bits per second. A direct sequence modulator produces a channel modulator of $R_c$ chips per second occupying a bandwidth of $W_s$ Hz. The earth segment transmits this with power $P_E$. The power received at the satellite on the uplink is

$$P_u^s = \Gamma_u P_E$$

Where $\Gamma_u$ is the uplink gain. The noise on the uplink is composed of receiver front end noise plus noise due to $\bar{n}$ other users. The total effective noise on the uplink is

$$N_u = kT_sW_s + \bar{n} P_u^s$$

where we have assumed that each user is transmitted $P_E$ over $\Gamma_u$ gain. We assume that the transponder is power limited so that the transmitted power due to the uplink signal is

$$P_d^s = P_s \frac{P_u^s}{P_u^s + N_u}$$

and the total down link noise is

$$N_d^s = P_s \frac{N_u}{P_u^s + N_u}$$

where $P_s$ is the peak satellite power.

Now at the receiver in bandwidth $W_s$, the total receiver power to noise power ratio is

$$\left(\frac{C}{N}\right)_{in} = \frac{P_d^E}{N_d^E + N_E}$$

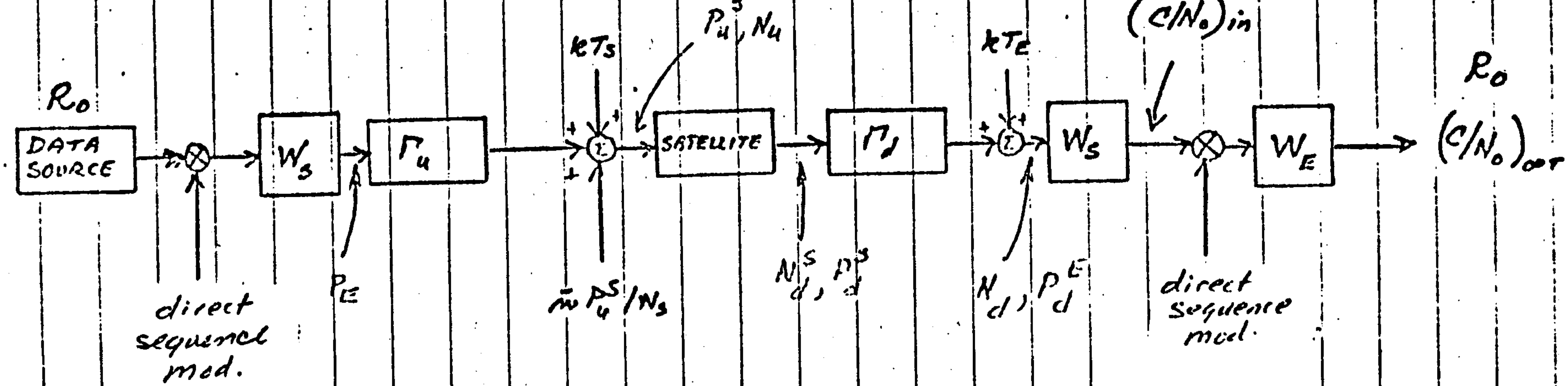



FIG 54 : Block diagram of Spread Spectrum System

where

$$P_d^E = \Gamma_d \, P_d^s$$

$$N_d^E = \Gamma_d \, N_d^s$$

and $N_E$ is the power due to front end noise;

$$N_E = kT_E W_s$$

By dividing we obtain

$$\left(\frac{C}{N}\right)_{in} = \frac{1}{\dfrac{N_d^E}{P_d^E} + \dfrac{N_E}{P_d^E}}$$

what relationship is the one?

Now using the relationship we can show that

$$\frac{P_d^E}{N_d^E} = \frac{\Gamma_d \, P_d^s}{\Gamma_d \, N_d^s}$$

$$= \frac{P_s \, (P_u^s / P_u^s + N_u^s)}{P_s \, (N_u / P_u^s + N_u)}$$

The dimensions are not correct!

$$\frac{(\text{Power})(\text{power}^2 + \text{power})}{\text{power}\,(\text{power}^2 + \text{power})}$$

$$= \frac{P_u^s}{kT_s W_s + \bar{n} \, P_u^s}$$ OK

$$= \frac{\Gamma_u \, P_E}{kT_s W_s + \bar{n} \, \Gamma_u P_E}$$ OK

Which is nothing more than the uplink carrier to noise (C/N) up.

Similarly

$$\frac{N_E}{P_d^E} = \frac{kT_E W_s}{\Gamma_d P_s (P_u^s \quad P_u^s + N_u)}$$ dimensionally not correct

$$= \frac{kT_E W_s}{P_s P_u^s} (P_u^s + N_u)$$ ok

$$= \frac{kT_E W_s}{\Gamma_d P_s} + \left(\frac{kT_E W_s}{\Gamma_d P_s}\right)\left(\frac{N_u}{P_u^s}\right)$$ ok

Now the first term in $(C/N)^{-1}_{down}$ and the second term is the product $(C/N)^{-1}_{up}$ $(C/N)^{-1}_{down}$. Thus;

$$(C/N)_{in} = \left[(C/N)^{-1}_{up} + (C/N)^{-1}_{down} + W_s (C/N)^{-1}_{up} \; (C/N)^{-1}_{up}\right]^{-1}$$

and this represents the carrier to noise ratio in bandwidth $W_s$.

Now when the signal is multiplied by the direct sequence again, the baseband signal appears in bandwidth $W_E$ (due to rate $R_o$) and all other sources of noise are spread over $W_s$ still. Thus by filtering to bandwidth $W_E$ we obtain a $(C/N)_{out}$ which is increased by this ratio. Specifically

$$(C/N)_{out} = \left(\frac{W_s}{W_E}\right) (C/N)_{in} \, .$$

where

$$\left(\frac{W_s}{W_E}\right) = G_p$$

is the processing gain of the spread spectrum system.

Now $(C/N_o)_{in}$ and $(C/N_o)_{out}$ can be defined as the carrier to noise spectral densities and these yield;

$$\left(\frac{C}{N_o}\right)_{out} = \left(\frac{C}{N_o}\right)_{in}$$

The corresponding $(E_b/N_o)$ is

$$\frac{E_b}{N_o} = \frac{1}{R_o}\left(\frac{C}{N_o}\right)_{in}$$

~~From (I-12)~~ we can obtain;

Indicate what equations are used!

$$(C/N_o)_{in} = \left[(C/N_o)^{-1}_{up} + (C/N_o)^{-1}_{down} + W_s\,(C/N_o)^{-1}_{up}\,(C/N_o)^{-1}_{down}\right]^{-1}$$

where

$$(C/N_o)_{up} = \frac{\Gamma_u\,P_E}{kT_s + \bar{n}\,\Gamma_u\,P_E/W_s}$$

$$(C/N_o)_{down} = \frac{\Gamma_d\,P_s}{kT_E}$$

Asymptotically we have;

$$(C/N_o)_{in} = \begin{cases} \dfrac{\Gamma_u\,P_E}{\bar{n}\,\Gamma_u\,P_E/W_s} = \dfrac{W_s}{\bar{n}} & ;\ W_s \to 0 \\[2ex] \dfrac{1}{W_s}\left(\dfrac{\Gamma_u\,P_E}{kT_s}\right)\left(\dfrac{\Gamma_d\,P_s}{kT_E}\right) & ;\ W_s \to \infty \end{cases}$$

and for both case $(C/N_o)_{in}$ goes to zero. Thus an optimum value exists. This can be seen by writing

$$(C/N)^{-1}_{in} = \frac{W_s kT_s + \bar{n}\,\Gamma_u\,P_E}{W_s\,\Gamma_u\,P_E} + \frac{kT_E}{\Gamma_d P_s}$$

$$+ \frac{kT_E}{\Gamma_d P_s}\left(\frac{W_s kT_s + \bar{n}\,\Gamma_u\,P_E}{\Gamma_u P_E}\right).$$

$\frac{\bar{n}}{W_s^2} =$

Differentiating yields;

$$-\frac{\bar{n}\,\Gamma_u\,P_E}{W_s^2\,\Gamma_u\,P_E} + \frac{kT_E}{\Gamma_d\,P_s}\left(\frac{kT_s}{\Gamma_u\,P_E}\right) = 0$$

and solving for the optimum bandwidth yields;

$$W_s^* = \left[\left(\frac{\Gamma_d\,P_s}{kT_E}\right)\left(\frac{\Gamma_d\,P_E}{kT_s}\right)\bar{n}\right]^{1/2}$$

as the optimum value. Equivalently this can be stated that $W_s^*$ equals the bandwidth that makes the signal to interference power ratio equal to the products of the uplink and downlink carrier to noise ratio independent of interference.

### 6.4.4 Preassigned vs. Demand Assigned

The schemes that we have presented describe methods to provide multiple users access to a satellite transponder or several transponders. There are two general cases that must be considered; pre-assigned or demand assigned. For the cash of pre-assigned traffic, the system is used as a trunking network and channels are requested, coordinated, and assigned. Most of the INTELSAT system operates on this fashion. Growth can be easily accommodated and channels are assigned to users. The disadvantage of some of these approaches is that the assigned channel may not be busy all the time. For example a TDMA time slot may not be carrying traffic. Since the traffic in the pre-assigned state is typically trunk traffic, the burden of efficiency is on the user.

? case?

If, however, there are many users, whose load is not constant, then it is not efficient to dedicate or pre-assign bandwidth, time slots or codewords. In this case we consider demand assigned systems. For these systems the user demands use in one of several ways.

1. Centralized. There is a central control station which receives all requests. The requests arrive on a separate channel called an order wire. The method of accessing the order wire is a separate issue that will follow more naturally later. Suffice it to say the central station could poll each station or the stations could contend for use of the order wire on a random basis. The central control when it receives requests, assigns capacity and them monitors the channel until its use is finished. There are delays in this system resulting from previous users and order wire access.

2. Distributed. In this approach each terminal is smart enough to share information of network state and listen to and respond to information that is broadcast. For this type of demand access to function, a terminal must be able to send up a request that is broadcast to all users. These users in turn must recognize the request and respond only if a conflict occurs.

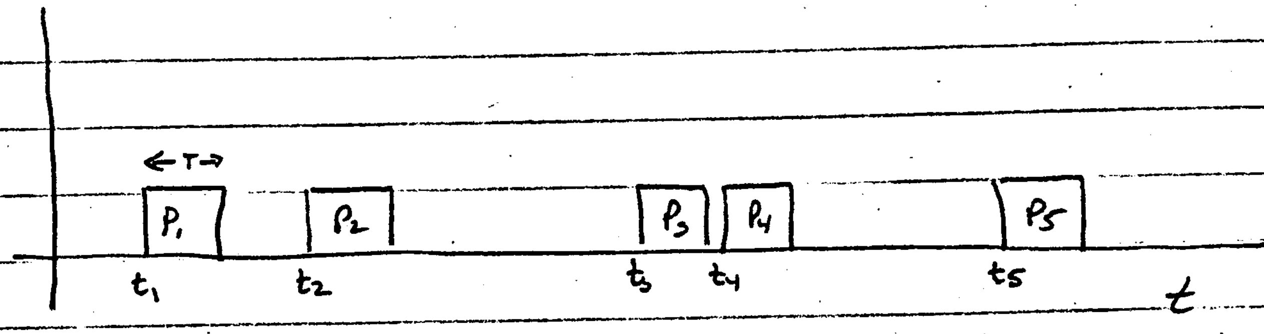
   For example in a distributed TDMA system each user would have a map of time assignments. This user now demands services and from the map selects an empty spot. He then broadcasts to all users his intent to use it. He can do so only if no other user beats him to the assignment. Once assigned, however, the slot is his and cannot be interferred with.

3. Random. In this scheme, no order wire is used. For each burst of data a contention scheme is used and assignments are random. We shall discuss this in more detail in the next subsection.

### 6.4.5 Random Multiple Access

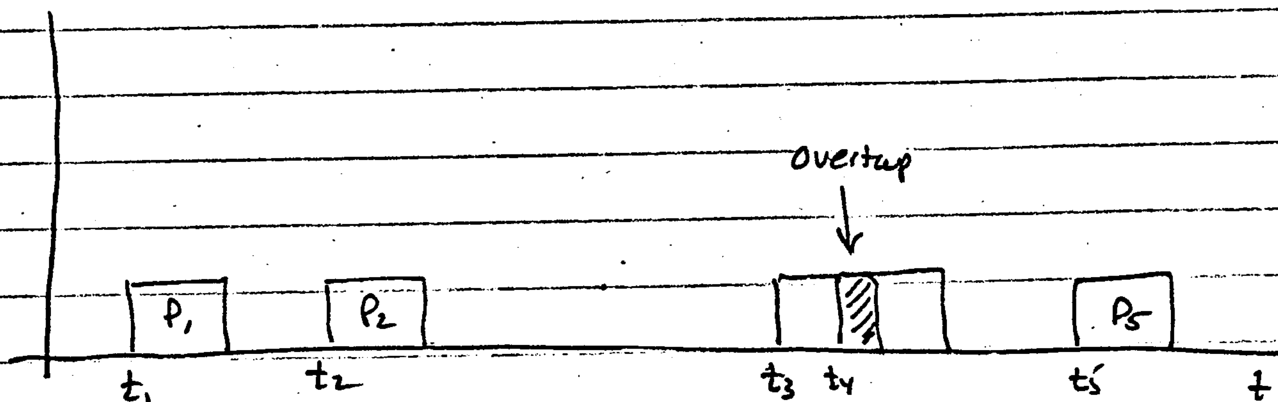
Random access schemes are used when the traffic is unpredictable and the user's time on the channel is short. Such a scheme can also be used as a subset of TDMA and others as a use for a pre-assigned channel. We shall consider here three types of RMA schemes; pure contention, called the Aloha scheme, a tree algorithm called the Capetenakis Algorithm and a hybrid RMA scheme using CDMA and contention.

Consider the scheme where we transmit a packet of data of length T sec. Depending on data rate, this packet can contain any number of bits. In Fig. 35 a we show 5 packets that are randomly sent. In this case, they do not overlap and are received properly. Consider now Fig. 35 b. Here $P_3$ and $P_4$ overlap. We then lose them because they are garbled. The packets are then retransmitted at random times. This retransmission increases the channel load without changing the real data flow.




(a)

Non Overlapping




(b)

Overlap

Fig 35 : RMA Packet

Let us now assume that there are n data sources in the network and each is generating packets with Poisson statistics at a rate of $\lambda$ packets/sec. Recall that the probability of k packets generated by a source in $T$ sec is $\tau$

$$P[k,\tau] = \frac{(\lambda\tau)^k}{K!}\exp(-\lambda\tau)$$ lower case k!

Now define S as the total average traffic offered the network. Clearly if there are k users and T sec we have;

$$S = k\lambda T$$

Now we let G be defined as the average channel traffic in T sec. This includes S plus all additional retransmissions due to collisions. If we assume that retransmissions are also Poisson then we can show that (see Schwartz) Ref?

$$S = Ge^{-2G}$$

Fig. 36 plots this result. Note that the channel throughput has a maximum of 1/2e. This is only 18%. What this says is that ~~I cannot use~~ this channel cannot be used for any more data. Thus if one ~~I have~~ has a 56Kbps channel, the best one ~~I~~ can do is get 10Kbps worth of data across it. Typically it will be 5Kbps! Ref?

Another approach has been developed by Capetenakis Ref? and uses a tree algorithm. This approach requires all users to be synchronized and be in slots. Aloha schemes similar to this were developed and are called slotted aloha. What makes the Capetenakis algorithm different is the way in which it resolves collisions. It does so in one of the most efficient ways possible. The tradeoff however is increased processing complexity.

The Capetenakis algorithm works in the following manner. Let each source be assigned a number. Now generate a binary tree such that it is deep enough to have end points that exceed or equal the number of sources. Recall that a binary tree has two branches at each node so that at depth n there are $2^n$ available points. Fig. 37 shows a binary tree for 16 sources.

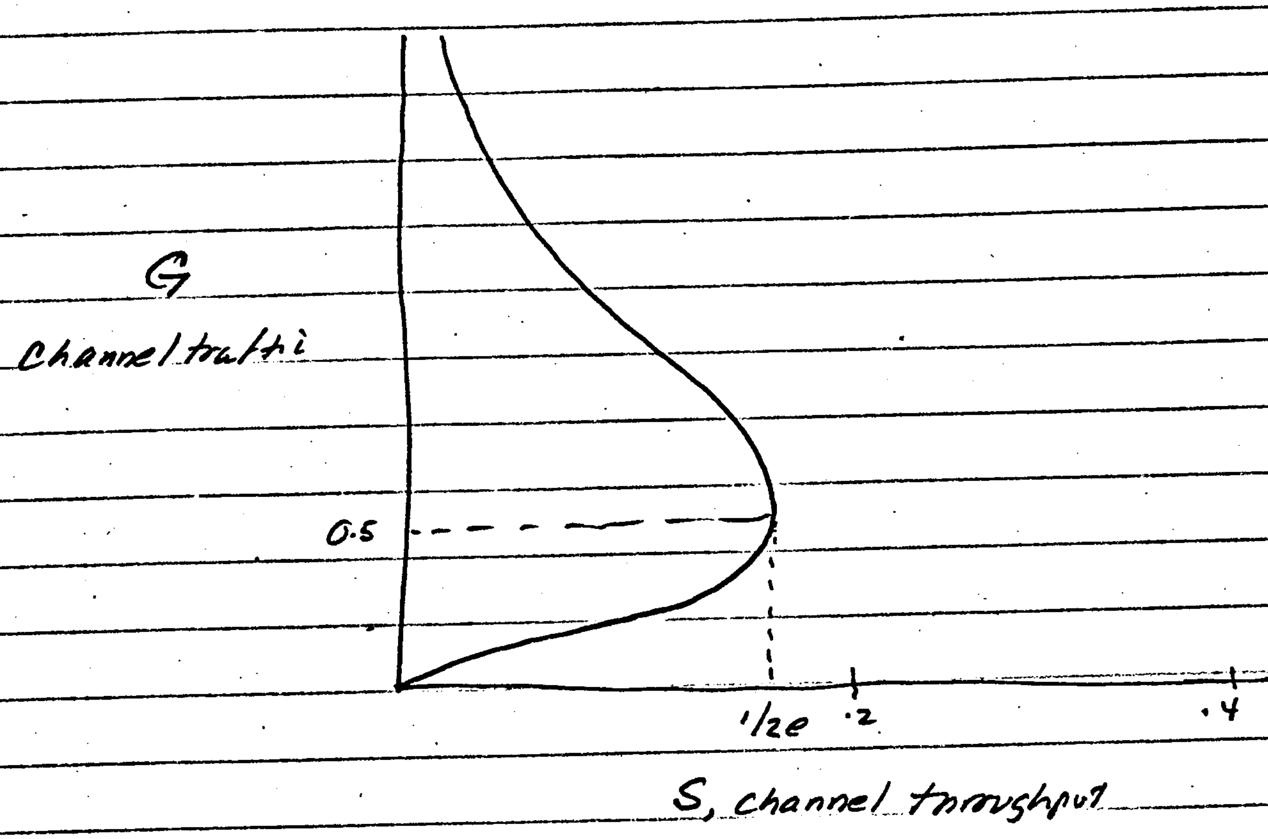



Fig 36 : RMA Throughput Performance

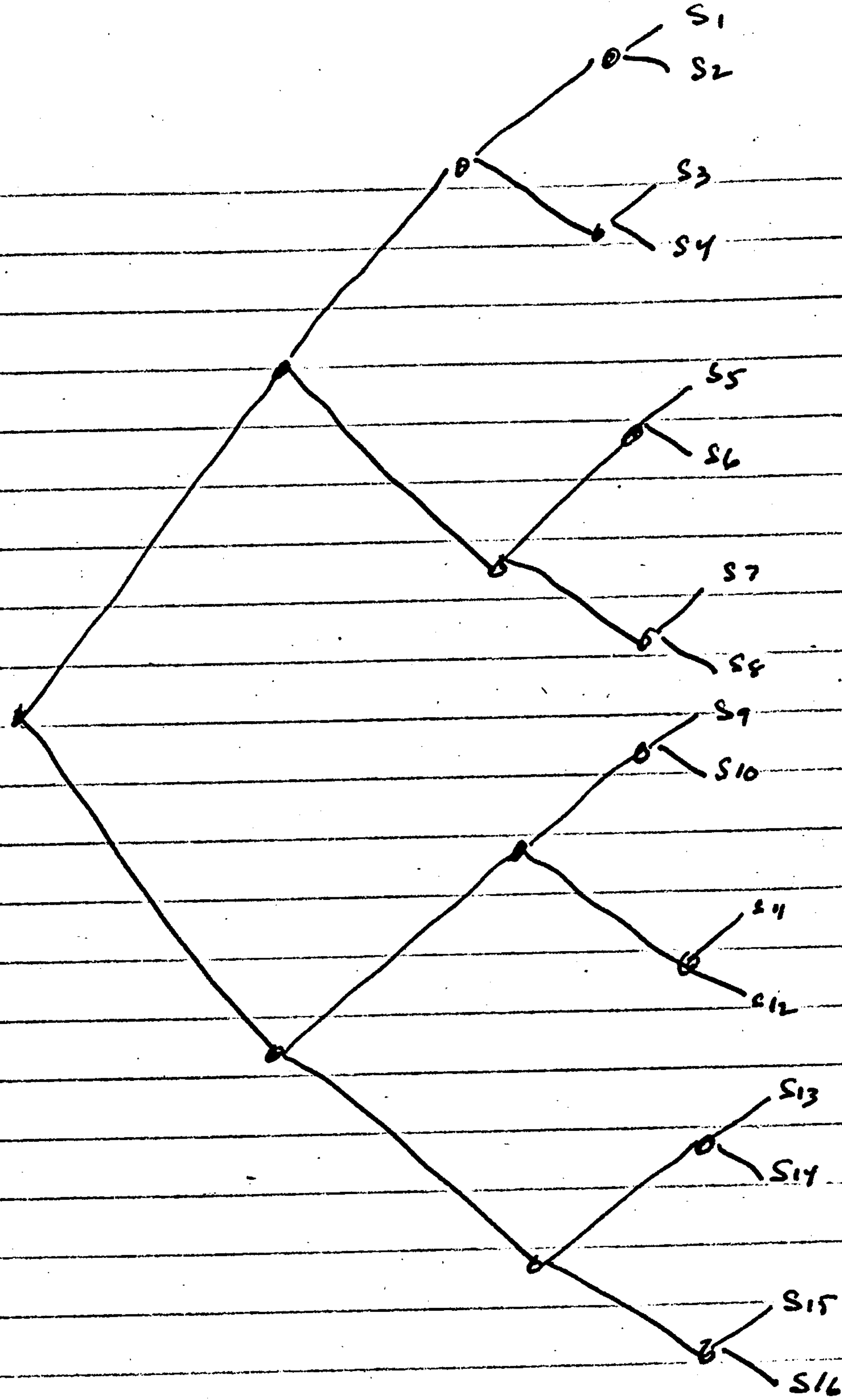


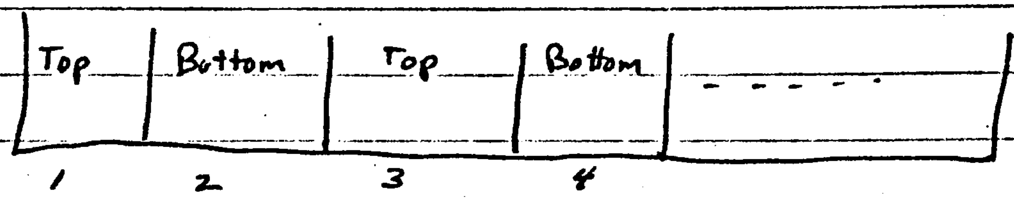

Fig 37 : Binary Tree

If there were thirteen we would have to increase the depth. Now the tree is first divided into two halves, top and bottom. Time slots are now assigned so that the top sources transmit in the odd slots and the bottom sources in the even (see Fig. 38 a). Now consider the case in Fig. 38 b where collisions occur. The algorithm resolves the collisions in the following manner. First start with the top and divide it into a top and bottom and have them transmit in odd and even respectively. When we do that we get S/4 through but still have a collision. So ~~do it~~ it is done again and ~~we are~~ it is successful. Now do the bottom and use the same algorithm.

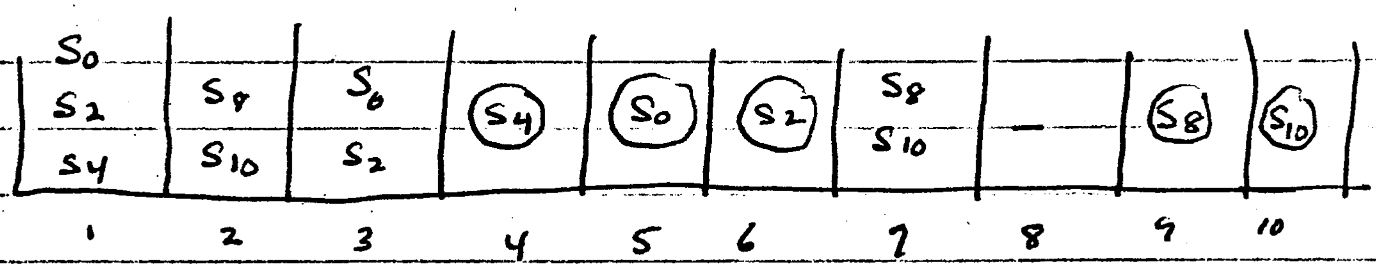
What happens is that eventually everything gets through but it is delayed. The delay depends on the total traffic offered. Fig. 39 plots the average delay versus the throughput. Note that it increases as the load increases.

We now want to consider a scheme where we use an Aloha system combined with CDMA. That is, bursts are sent at random but they are spread using orthogonal direct sequence codes.

We shall first assume a spread spectrum system such that a complete data word of M bits is transmitted once every T sec. in a free running manner. The channel has a bandwidth $W_S$ and for the proposed modulation scheme can support a rate of $R_C$ chips or symbols per sec. The source rate, $R_O$, is M/T bits per second. The number of bits per chip, $\alpha$, is given by $R_O/R_C$ and is typically much less than one. A code with a chip rate of $R_c$ chips/sec is to be selected so as to eliminate the system's self noise. A direct sequence spreading code will be used to spread the signal and we assume that the carrier powers are constant. The individual uplink transmitters are assumed to each saturate the satellite (e.g., the satellite does not operate in the linear region). The receiver carrier to noise ratio is:




(a)
Primary Slot




(b)
Algorithm Slotting

Fig 38: Capetanakis Algorithm.

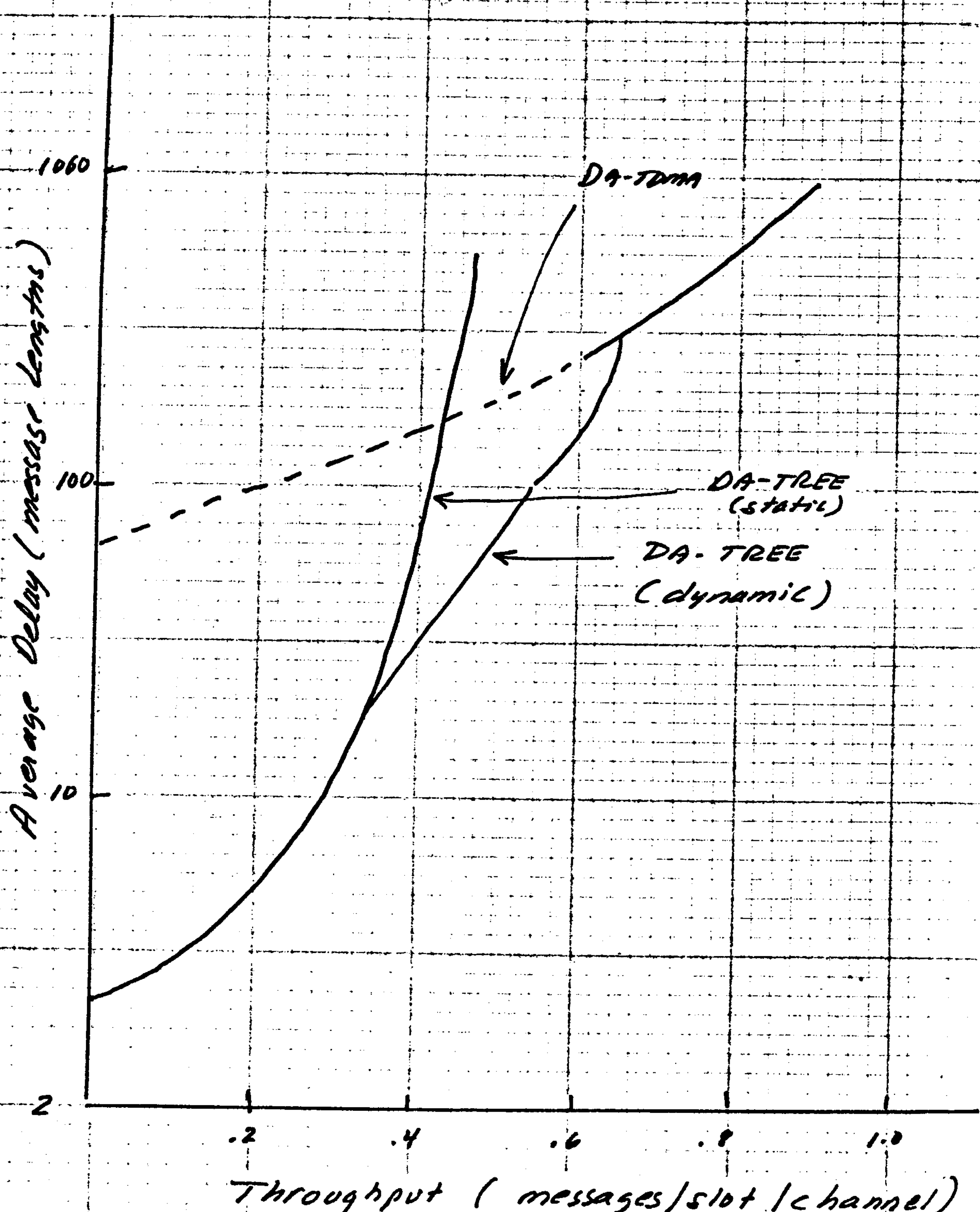



FIG 39: Performance Curve for TDMA Capetenakis Algorithm

Ref?

$$(C/N_o) = \left[ (C/N_o)_u^{-1} + (C/N_o)_d^{-1} + W_s (C/N_o)_u^{-1} \cdot (C/N_o)_d^{-1} \right]^{-1}$$

where the uplink carrier to noise ratio depends upon the satellite receiver noise temperature and the effects of all the other users, $\bar{n}$. Specifically,

$$\left(\frac{C}{N_o}\right)_u = \frac{\Gamma_u P_E}{kT_s + \bar{n}\,\Gamma_u P_E / W_s}$$

where $\Gamma_u$ is the uplink gain, $T_s$ the satellite receive noise temperature, $P_E$ is terminal uplink power, $W_s$ the spread bandwidth and $\bar{n}$ the average number of interferences. This model assumes that the other users appear as Gaussian white noise spread over the bandwidth W and of a total power equal to that of the sum total of users.

For the downlink;

$$\left(\frac{C}{N_o}\right)_d = \frac{\Gamma_d P_s}{kT_E}$$

with $P_S$ the satellite power, $\Gamma_d$ the downlink gain and $T_E$ the earth station temperature.

After dispreading the $E_b/N_o$, energy per bit to noise spectral density is given by:

$$\frac{E_b}{N_o} = \frac{1}{R_o}\left(\frac{C}{N_o}\right)$$

Note that $R_o$ is usually quite small compared to $W_S$.

$C/N_o$ as a function of the spread bandwidth shows that an optimum spread bandwidth exists. Typically, however, $W_S$ is given so that we must choose $P_E$ to provide the required $E_b/N_o$.

The second system is a pure random access system. Here there will be one bit per chip and each user bursts his message at the peak rate. The users use the channel at random so that collisions can occur. As is well known (e.g., the Aloha system, Kleinrock reference) as long as the fill factor, $\beta$, is less than $(2e)^{-1}$ we can maintain a low collision rate. Thus we design the system for a data rate, $R_o$. The rate at the receiver is $R_o$ since there is one bit per chip. Thus

$$\frac{E_b}{N_o} = \frac{1}{R_c}\left(\frac{C}{N_o}\right)$$

where $C/N_o$ is the carrier to noise spectral density of the channel in the absence of other users. It should be noted that in this approach $R_o$ is typically given, as well as $R_o$, but the average number of users, $\bar{n}$, is limited by the allowable fill factor, $\beta$.

It is possible to combine a random access with a spread spectrum system which we shall call a Code-division Random-Access Multiple Access (CRMA) system. That is, if the chip duration is held constant and the number of bits per chip varied, then a combined scheme may be developed.

We assume that we have a channel bandwidth of $W_s$ which is occupied if we use a chip transmitted at rate $R_c$ (chips/sec.). This chip rate is the fundamental constraint. In addition we have $\bar{n}$ users each of which must transmit M bits in T seconds for a source data rate, $R_o$, of M/T bits per second.

In the random access scheme each user is isolated in time so that $(C/N_o)$ is the greatest suffering from minimal mutual interference effects. However, the channel data rate, $R_D$, is the greatest and equals $R_c$. In contrast we have in the spread or code division scheme continual user overlap which maximizes mutual interference and thus reduces $(C/N_o)$. However, the data rate $R_D$ equals $R_o$ and is quite small. Thus the gain occurs by having the low data rate.

If we consider a hybrid scheme which slightly spreads the signal but in addition allows some overlap, we would like to obtain a combination of spread to maximize $(C/N_o)$ and minimize $R_D$ to effectively maximize $E_b/N_o$ for a fixed number of users or equivalently to maximize the number of users (e.g., the throughput) for a fixed $E_b/N_o$. Note that the bursts are random but they overlap. However, a code is used for protection in a correlation receiver. If we define a factor $\alpha$ as the number of data transmitted bits per chip, $\alpha = R_D/R_c$, then for the random access case, $\alpha = 1$, and for the code division use $\alpha = R_o/R_c$.

Now consider a channel that has bandwidth of $W_S$. Let the source transmit at a channel data rate $R_D$ and spread the signal but using rate $R_c$. If $R_D$ is less than $R_c$ we use the spread for protection against mutual interference. However, the average number of interferers, $\bar{n}_{eff}$, depends on $\alpha$. Let $\bar{n}$ be the maximum number of interferers as would be found in the spread case. If we now have a greater $\alpha$ the data pulse will be shorter, fewer chips, so that the effective number of interferers will decrease. These can be estimated in the following manner. There are a total of $R_cT$ cells that can be filled. If M bits or source information are to be transmitted then $M/\alpha$ cells are to be filled. Thus the probability that a single user fills a cell is $(M/\alpha R_C T)$, or $(R_o/R_c)\ (1/\alpha)$. If there are at most $\bar{n}$ users then

$$\bar{n}_{eff} = \bar{n}\,\frac{R_o}{R_c}\;\frac{1}{\alpha}$$

For the random access case $\bar{n}R_o/R_c$ is chosen to be 0.1 or less. Thus we typically choose a constant $\beta = \bar{n}(R_o/R_c)$, as a system design parameter. This implies $n_{eff} = \beta/\alpha$. The $E_b/N_o$ value is now determined by;

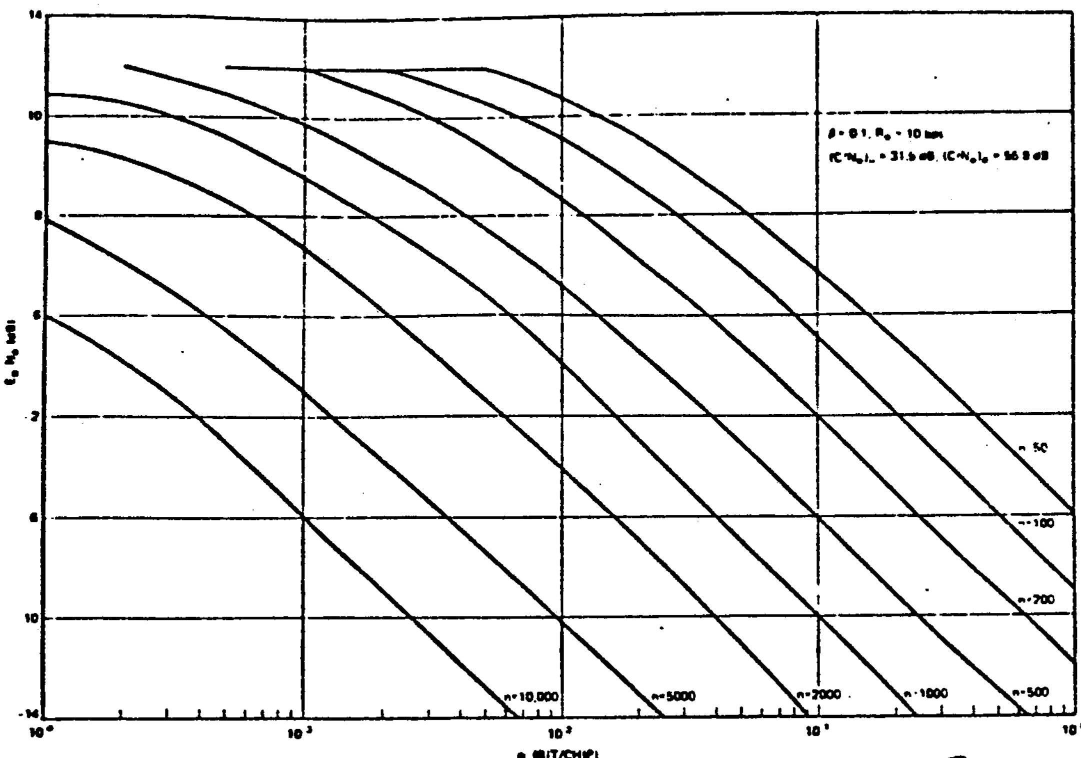
$$\frac{E_b}{N_o} = \frac{1}{R_D}\;\frac{C}{N_o}$$

Now $1/R_D$ equals $1/\alpha R_c$ and this factor decreases as $\alpha$ increases. Similarly $(C/N_o)$ decreases as $\bar{n}_{eff}$ increases or equivalently $(C/N_o)$ increases as $\alpha$ increases. Thus there exists an optimum value of $\alpha$ to maximize $E_b/N_o$.

We note that we may write $R_c$ as $\bar{n} R_o/\beta$ and let $W_s$ equal $2R_c$ we may obtain; Reference?

$$\frac{E_b}{N_o} = \frac{\beta}{\alpha \bar{n} R_o}\left[(C/N_o{}_{up}^{-1}(\alpha) + (C/N_o)_{down}^{-1} + \frac{2\bar{n} R_o}{\beta}(C/N_o)_{up}^{-1}(\alpha)\ (C/N_o)_{down}^{-1}\right]^{-1}$$

Now we can plot $\frac{E_b}{N_o}$ versus for various values of $\bar{n}$ as is done in Figure 35. Then by assigning a design of $(E_b/N_o)^{-1}$ we can obtain the throughput, $\bar{n}$, as a function of $\alpha$ and this is plotted in Figure 36. Note that in these curves we allow $R_c$ to vary. Note that $E_b/N_o$ and $\bar{n}$ both decrease from a peak value at the minimum $\alpha$ value $(R_o/R_c)$. Thus the CDMA system is the optional choice for these cases.

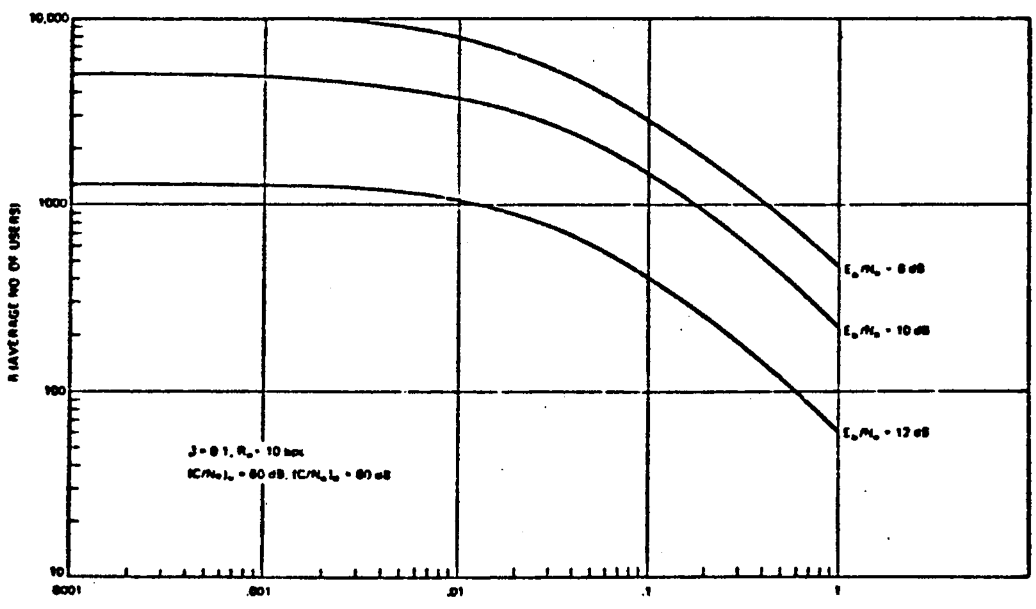



6-16407

FIGURE 35. PLOT OF MAXIMUM $E_b/N_o$ VERSUS α

Ref ?

35




6-16407

FIGURE 36. PLOT OF THROUGHPUT VERSUS α

Ref ?

36

# CHAPTER 7
# ECONOMIC ANALYSIS

This chapter focuses on the economic or cost factors of satellite system design. The reader should be cautioned that although numbers are used frequently, the actual costs are constantly changing with time so that this chapter represents an architecture in which to work.

We first begin with an overview of the basic cost factors and a brief discussion of capital budgeting issues. It is critical to include such factors in any future analysis of satellite design since cost/performance tradeoffs are the key issues. Then models for both earth station costing and satellite costing are presented.

Finally, we present several models for optimized designs. These models can be used in direct application by the user or modified for special applications. The approach of these three ~~approaches~~ Techniques is to use a case study method to develop concepts. The emphasis in on analytical model building to interrelate the system design factors and the financial performance. As satellite communications matures, the emphasis will clearly shift from the purely technical to the financial, and the techniques developed in this chapter will be key.

## 7.1 BASIC CONCEPTS

In this section we develop models for the basic cost elements and means for determining tradeoffs between differing capital investment strategies. In any economic enterprise its success is measured by revenue less expense, or profit. The revenue is generally determined in a free market context, yet this assumption often times has limited applicability to the regulated world of the communications. We shall avoid this issue by concentrating on the basic expense elements and make general assumptions concerning revenue.

Expenses can be categorized in various ways. We choose to characterize them first in terms of the basic system elements;

1. Earth Station
2. Satellite
   - leased space
   - total satellite
3. Terrestrial Interface
4. Local Loop Communications

Then each of these expenses can be broken down into nonrecurring and recurring elements. The nonrecurring parts are the one time charge in expense for installation of the terminal. For example, the installation of an earth station. The recurring expenses are those necessary to provide a continuity of service such as a tariff or operating and maintenance (O&M) costs.

The key issue of costs is its relationship to performance. As we note later, the design of any element of the system is such that if P indicates the level of performance, then there is a corresponding cost element C such that

$$C = f(P)$$

The objective of this chapter is to identify these relationships. For example, the gain of an antenna is related to the diameter. Its cost related to the area. Thus cost and gain can be interrelated.

Now for each of the basic system elements and for both the nonrecurring and recurring portions, we can further divide cost or expenses into elements that more closely relate to normal project elements.

Consider, for an example, the selection of an earth station. As we discussed before, the earth station has the following components.

1. LNA (Noise Temperature)
2. HPA (Power)
3. Antenna (Gain)
4. Up/Down Converters (Phase Noise)

Now we can write the total expenses of the earth stations in terms of the following elements.

1. Direct Labor - the cost of salaries of the people needed to perform the tasks. These tasks will be described below.

2. Indirect Labor - the overhead expenses for each direct employee.

3. Materials - these include the antenna, HPA, LNA, and all other elements of the design.

4. Other Direct Charges (ODC) - those expenses related to travel, computer time, etc.

5. General and Administrative (G&A) - those charges due to general non allocatable items.

Now each such element goes through at least four phases: design, development, deployment, and operations. The first three are nonrecurring and the last is recurring. For each phase, the set of expenses described apply; however, some of the elements such as materials may be zeroed out. (see Fig. 1)

Thus, for example, the per unit earth stations cost is broken down into nonrecurring, $C_{ES,N}$, and recurring, $C_{ES,R}$. These costs are then expressed in terms of the elements that are performance dependent;

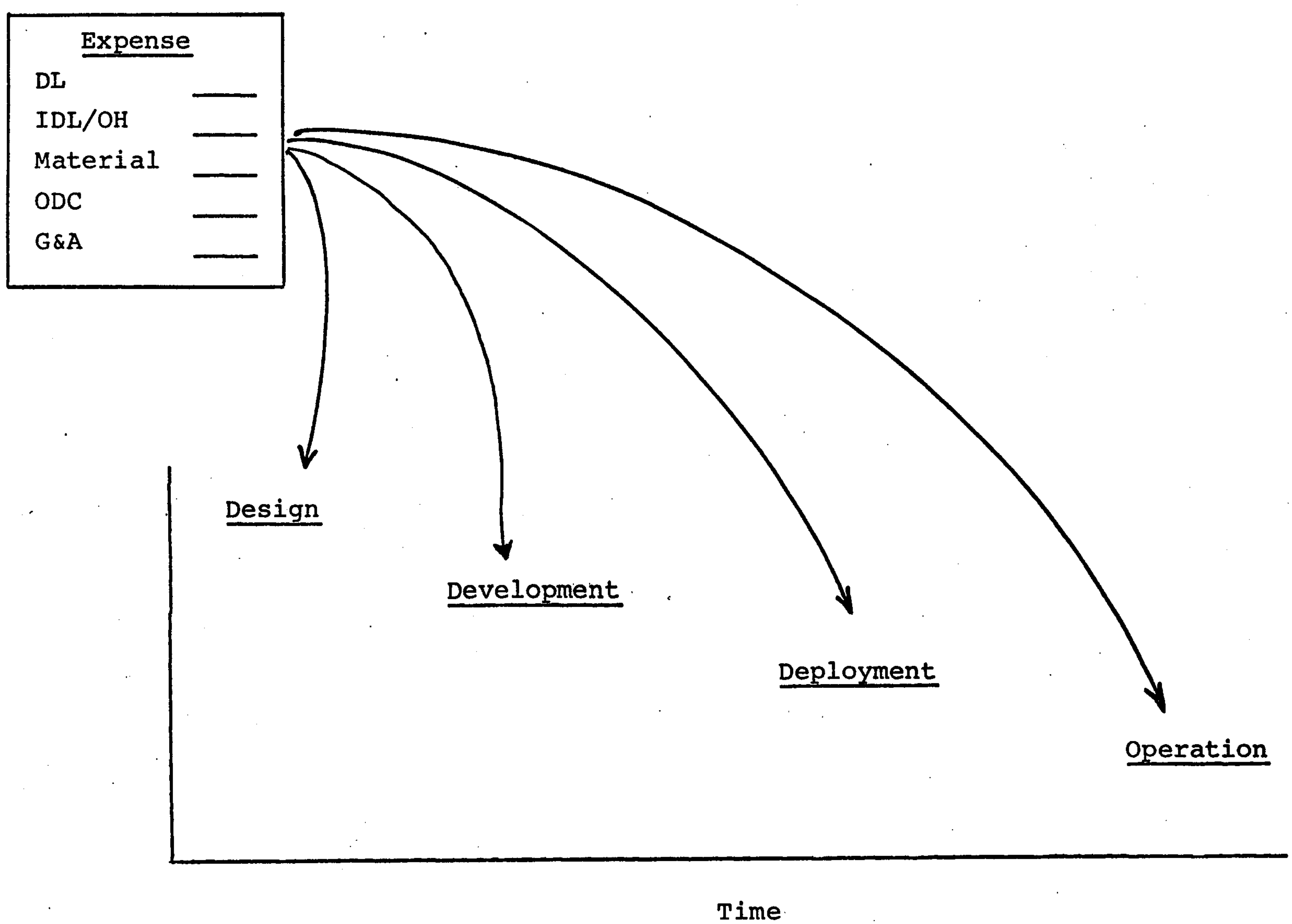



Fig 1: Cost Elements and Project Phasing

$$C_{ES,NR} = C_{ANT}(P_{ANT}) + C_{LNA}(P_{ANA}) + C_{HPA}(P_{HPA}) + C_{FIXED}$$

Here $C_i$ are the ith cost and $P_i$ the ith performance factor. $C_{FIXED}$ are those costs that are performance independent.

The recurring costs are typically stated on a monthly or annual basis.

The same type of breakdown applies to the other elements.

For those items that are directly purchased the emphasis would then be on material and less on labor, only that labor needed to specify, monitor, and test the procurement. As is noted, the cost typically increases as the performance level increases.

The costs can now be accumulated into expenses and these expenses related to the time elements of the project. Thus, if expenses are accumulated on a quarterly basis, they represent the sum of all cost elements that are either accrued in that quarter or put into service in that quarter. Thus if E(k) represents the expenses in quarter k, we have

$$E(k) = \sum_{k=1}^{n} C_{i,NR}(k) + \sum_{k=1}^{n} C_{i,R}(k)$$

Where i refers to the system element (earth station, satellite, etc.) and we have separated nonrecurring and recurring costs.

If we have a projection of revenue, R(k), per quarter then the cash flow (positive or negative) in quarter k is, CF(k),

$$CF(k) = R(k) - E(k)$$

It is this cash flow that defines the profitability of the project. A typical example is shown in Fig. 2.

Now careful attention may be paid to the interplay between revenues and expenses. Clearly to collect revenue we must have an adequate hardware base to support customer needs. Thus a large investment may be needed at the front end of the project.

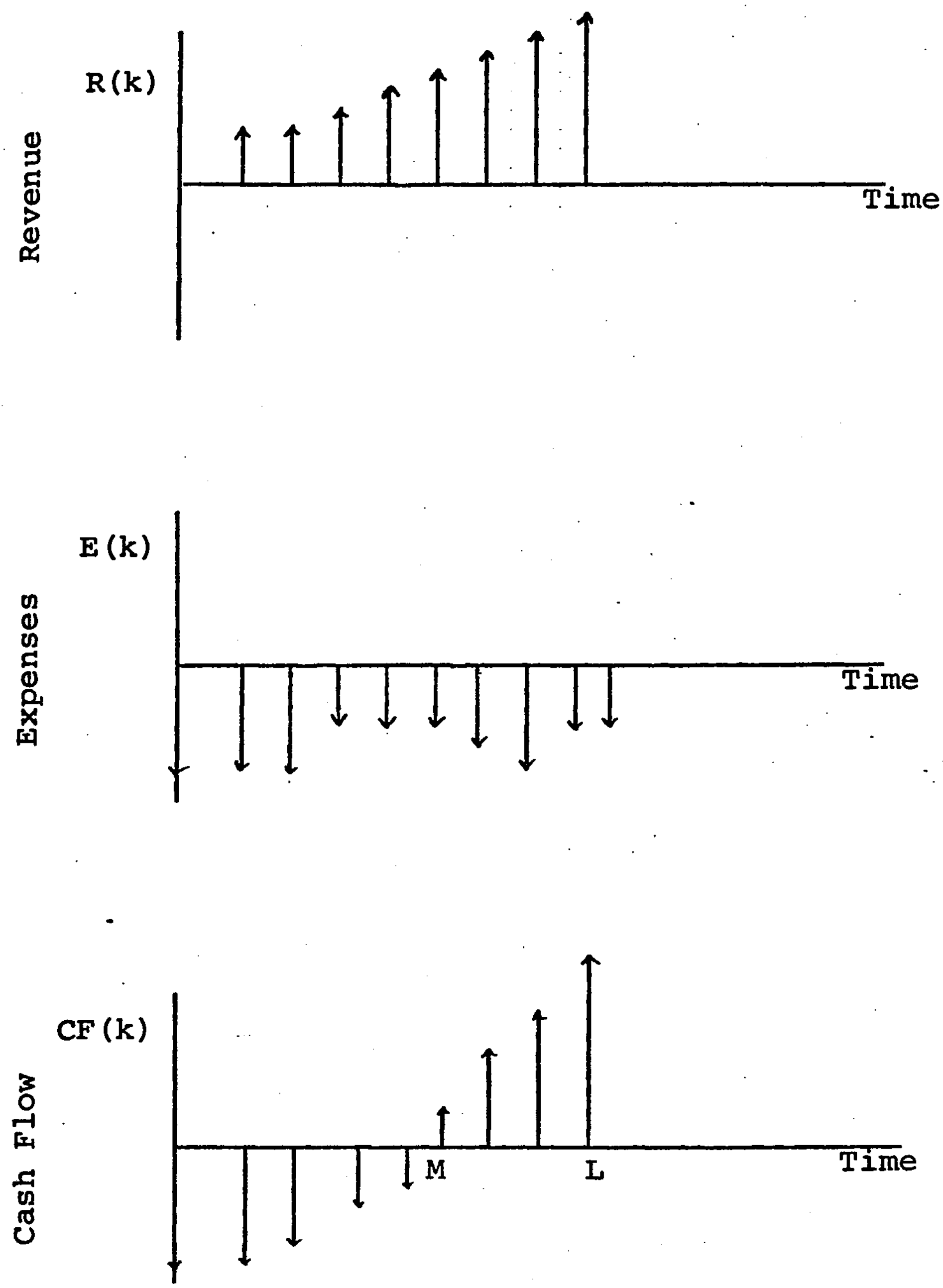



Fig 2: Sample Cash Flow Profile

Figure 2 depicts revenue, expense, and cash flow profiles on a quarterly basis for a typical project. Note how the revenue increases with time. The expenses, on the other hand, are negative at the beginning but decrease as we move from the initial investment stage to the recurring cost stage. The cash flow is the sum of these (note expenses are listed as negative or cash outflows). We can now define cumulative revenues, expenses and cash flows. To do so, we must define the time horizon of the project or its lifetime. This is the period over which we expect to have the major investment. For example in a satellite system, this could be the lifetime of the satellite, seven years. Let L be the lifetime in quarters. Then the cumulative numbers are;

$$R = \sum_{k=0}^{L} R(k)$$

$$E = \sum_{k=0}^{L} E(k)$$

$$CF = \sum_{k=0}^{L} CF(k)$$

One other item should be pointed out, that is the nature of the cash flow. In the example we noted that it became positive at quarter M. This is called the breakeven point of the investment. From a system point of view we want to make this as small as possible.

There are two other factors that we now want to consider, inflation and cost of money. Inflation will tend to drive up costs of implementing the system. A countervailing factor is that inflation may also increase revenue. The choice of including inflation in the analysis is typically left to the system designer, but it is frequently recommended.

The cost of money is the more important effect. This is the factor that results in the capital budgeting problem. To understand how to include this, let us first consider a simple example. Let us assume that we can receive $R_O$ dollars now or $R_M$ dollars M years from now where $R_M > R_O$. How do we decide on which is better.

If we could take the $R_O$ dollars and invest it at a return of p percent per year, then in M years we would have $(1+p)^M R_O$ dollars. Thus we would compare $R_M$ to $(1+p)^M R_O$. If they are equal, then they are equally useful. This implies that $R_M$ dollars in M years equals $R_M/(1+p)^M$ dollars today. This is called the present value of that revenue.

Using this concept we can return to our three cash flows and discount them by the percent interest per interval. This yields present values of

$$PVR = \sum_{k=o}^{L} \frac{R(k)}{(1+p)^k}$$

$$PVE = \sum_{k=o}^{L} \frac{E(k)}{(1+p)^i} \quad k?$$

$$PVCF = \sum_{k=o}^{L} \frac{CF(k)}{(1+p)^k}$$

Using this concept we can now look upon cash flows differently. We see that cash flows far out in time, even though large, are significantly discounted. Thus if we have two different system designs, one may have large CF values out in time, but large negative ones close and the other not as extreme. The latter may easily yield a greater PVCF due to this discounting. Clearly with the high cost of capital in today's market, this analysis is critical.

We can now apply these concepts to the system design. In any system we seek to obtain an overall performance of P. This performance depends upon the performances of all the constituent elements, $P_i$. We also have cost performance ratios between cash $P_i$. Thus if we have a constraint of

$$P = g(P_i, \ldots . P_N)$$

we then seek to choose the system design such that P is satisfied and some measure of system profit or cost is minimized.

Thus what we shall do in the remainder of this chapter is first develop those cost models and then develop and analyze approaches to this cost/performance tradeoff.

## 7.2 EARTH STATION COST MODELS

The earth station cost/performance model can be readily developed for a wide variety of earth station configurations. Figures 3 and 4 depict the FDMA configuration with M frequency chains and $N_i$ inputs per chain. We shall assume that all of the inputs have been digitized so that this process is not factored into the cost of the earth station. The TDMA station is shown in Figure 5. We shall in this section take each of the basic elements of the earth station and define the performance factors that relate to cost. We will do this for both nonrecurring and recurring elements of the system. The net result will be a cost model that yields cost, $C_i$, for each element in terms of performance factors $X_j$. Thus

i ?

$$C_i = f(X_1 \ldots\ldots X_N);\ \cancel{c}=1,\ldots\ldots M$$

The following are the basic earth station elements. We shall define the performance factors for each and concentrate on recurring costs first.

1. Baseband Interface

   This element provides for data packaging from or to many sources into a data stream that is then one carrier. In FDMA a small amount of memory may be required whereas in TDMA the memory may be great. The performance factors are:

   a. Data Rate (bps) - this is the rate that interfaces with the encryptions side.

   b. Source Number - the number of ports on the interface.

   c. Source Rate (bps) - the data rate per source.

   d. Storage Requirements (bits) - this is what is required to store data for packaging into from the transmit data steams. In FDMA if we have $N_i$ sources at a rate of $R_s$ each, and we transmit at rate $R_C$ where

   = ?

   $$R_C - N_i R_s$$

   then if we sample each line equally the memory depends on the frame length that we compress to. This is T sec,

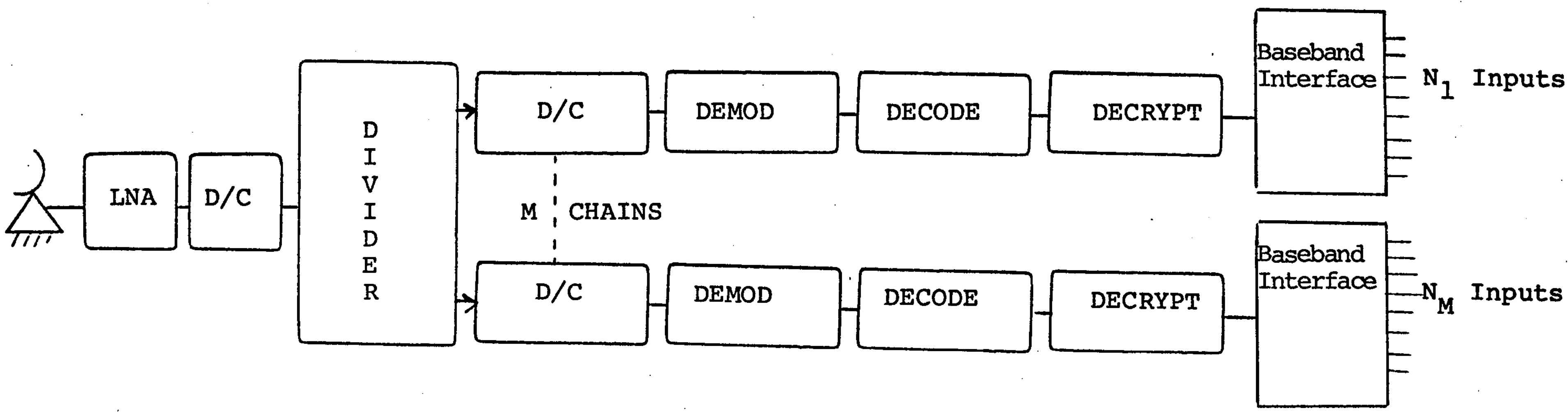



Figure 3 . FDMA RECEIVE STRUCTURE

arrowheads ?

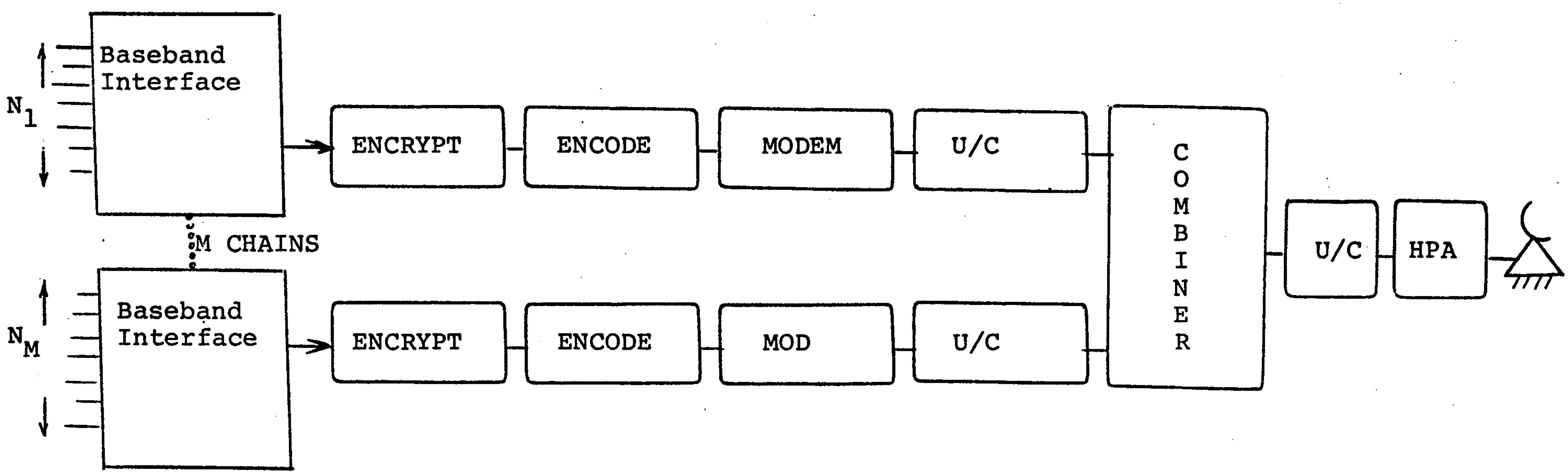
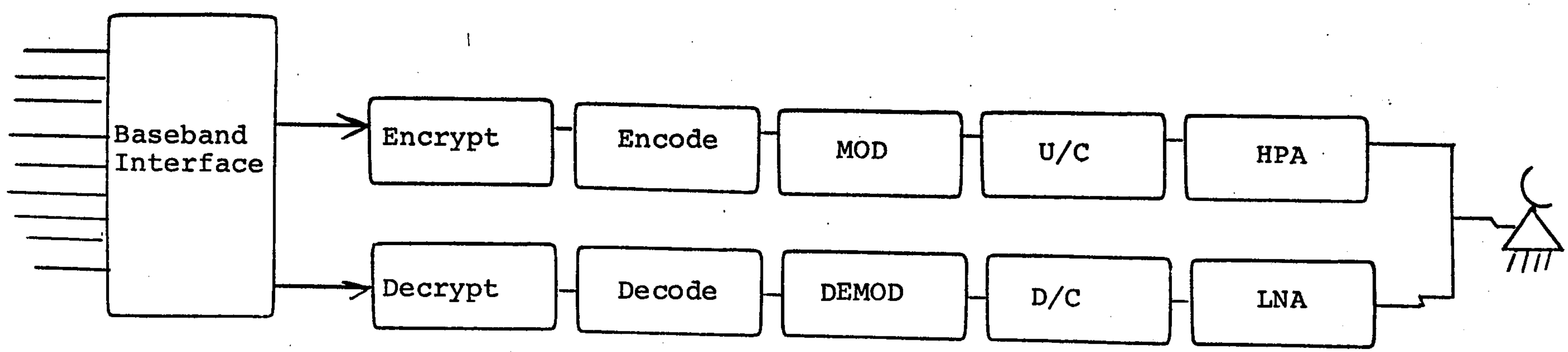



Figure 4. FDMA Transmit Structure

arrowheads!




arrowheads?

Figure 5. TDMA Transmit/Receive Structure

where once every T sec we transmit $R_S T$ bits in $(Rs/Rc)T$ sec. Thus, the memory is at the discretion of the user.

For TDMA on the other hand, we must compress or store data dependent on how many other users are in the system. As we discussed in TDMA, we have a burst time $T_b$ and a frame time $T_F$. If there are $N_u$ users in the system then:

$$T_F = N_u T_b$$

The storage requirement then is $R_S T_F$ bits. Note that if we have $R_S$ of 100 Kbps and $T_F$ of 100 MSEC then the storage is 10Kb.

2. Encryption

Here we have a choice of algorithms that can be used such as DES and Public Key. Both are complex polynomial algorithms that require significant compatibilities. The key factors are then,

a. Data rate

b. Key size - this gives the complexity of the polynomial and computation of complexity.

c. Memory which is similar to the baseband interface.

3. Coder/Decoder

This is similar to but less complex than the exception device. The key factors here are:

a. data rate
b. constraint length
c. number of soft decision levels
d. code rate
e. memory

4. Modem

There are wide variations of costs with modems. A key factor will be speed or data rate as well as whether it is continuous or burst. As speed increases, the type of devices used change in a step fashion, thus cost has step changes. In addition these boundaries as well as levels depend on time due to technical advances as well as the number of devices bought thus allowing large buy discounts. Although this latter factor applies to all elements, it is more sensitive to modems.

Thus with modems, the key cost factors are:

a. Modulation Type
b. Continuous or Burst
c. Speed or Data Rate
d. Processing loss - if we use a microprocessor matched filter modem, we get processing losses of 0.25 dB or less.

5. Up/Down Converters

The up and down converters depend on the following factors:

a. frequency
b. long term stability
c. phase noise
d. reliability

6. HPA

The HPAs can be of several types but costs tend to be somewhat independent of type (TWTA, GaAs FET, Impatt diodes, klystron). The performance factors are:

a. Gain or antenna diameter
b. Output Power
c. Efficiency
d. Bandwidth - here TWTA and GaAifet are useful.
e. Frequency

7. LNA

As the HPAs, the LNA depends less on type than on other factors. One typically chooses type to get to a certain range (e.g., cooled paramp for lower than 100$^{O}$K). The factors are:

paramp

a. noise temperature
b. gain
c. frequency
d. bandwidth

8. Antenna

Most earth station antennas are parabolic with only the feed changing. We shall not discuss torous antennas. We shall also not discuss pointing since that cost is somewhat fixed and one uses a monopulse technique.

Thus the key cost/factors are:

a. Gain (dbi)
b. Efficiency
c. Frequency
d. Beam width

Table 1 presents some cost model equations that have been found to be representative of the subsystem elements. The constants can be determined by taking manufacturer costs and performing a nonlinear regression fit. Figure 6 depicts such curves for LNAs and HPAs at three frequencies over a wide range of key factors.

Example: Consider now a TDMA and FDMA approach to a design that has an output rate of RMbps and a burst rate of 50 Mbps. Assume that each source generates 50 Kbps and that there are M chains. For TDMA we have a total cost of

$$C=C_{BI}+C_{ENC}+C_{CODEC}+C_{MOD}$$
$$+C_{U/D/C}+C_{HPA}+C_{LNA}+C_{ANT}$$

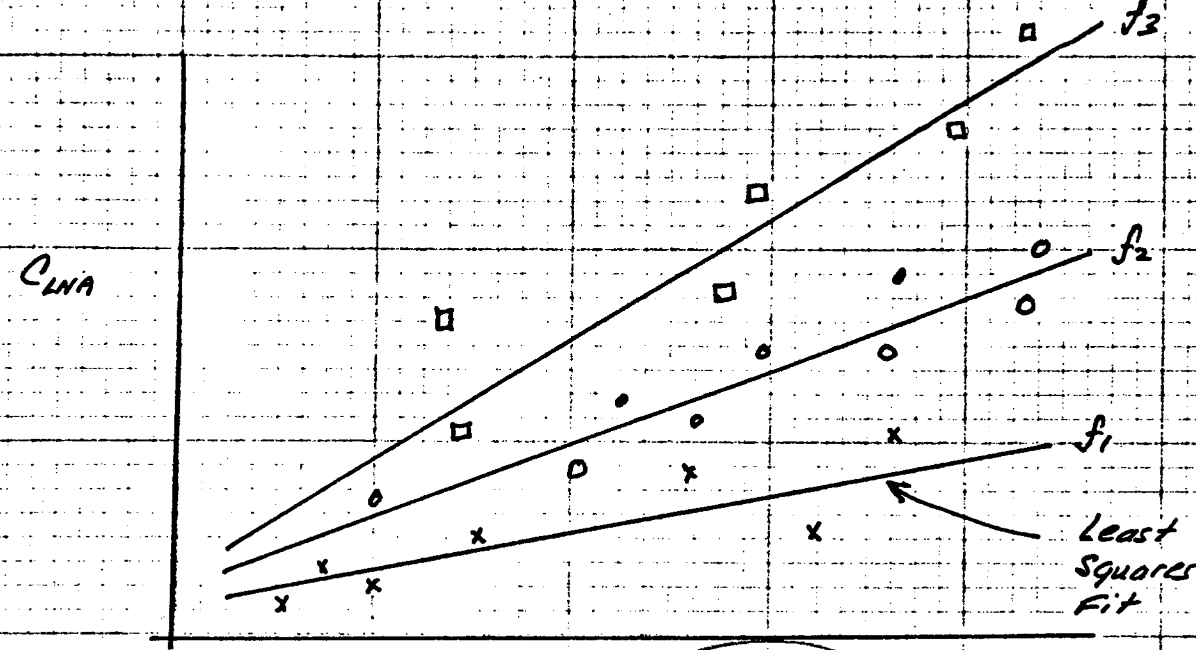
where the separate costs are obvious Fn FDMA;

$$C=M\ C_{BI}+M\ C_{ENC}+M\ C_{CODEC}+M\ C_{MOD}$$
$$+M\ C_{U/D/C}+C_{HPA}+C_{LNA}+C_{ANT}$$

We find that $C_{ANT}$ and $C_{LNA}$ are the same. In the FDMA case we must back off the HPA to avoid intermodulation, thus at a 10dB backoff, we find a factor of 10 cost difference. Qualitatively the other costs are dominated by data rate plus a constant. Thus if we plot the cost of both versus the number of chains, we note that a certain point TDMA becomes less expensive. Figures 7 and 8 show this cost tradeoff. What is more impressing is that if we plot the cost per source

| System | Cost Model | Factors |
|---|---|---|
| Baseband Interface | $C=C_o+C_1X_2X_1^{\alpha_1}X_3^{\alpha_2}$ $+C_2X_4$ | $X_1$= data rate<br>$X_2$= source number<br>$X_3$= source rate<br>$X_4$= storage |
| Encryption | $C=C_o+C_1X_1^{\alpha_1}X_2^{\alpha_2}$ | $X_1$= data rate<br>$X_2$= keysize<br>$0<\alpha_1\alpha_2<1$ |
| Codec | $C=C_o+C_1X_1^{\alpha_1}$ $+C_2X_2^{\alpha_2}$ ? | $X_1$= data rate<br>$X_2$= constraint length<br>$0<\alpha_2<\alpha_1<1$ |
| Modem | $C=f_1(X_1)+f_2(X_2)$ $+C_3X_3^{\alpha_3}$ | $f(\ )$ are discontinuous<br>$0<\alpha_3<1$ |
| Up/Down Converter | $C=C_o+C_1X_1^{\alpha_1}$ $+C_2X_2^{\alpha_2}$ $+C_2X_3^{\alpha_3}$ | $X_1$= frequency<br>$X_2$= long term stability<br>$X_3$= phase noise |
| HPA | $C=C_oX_2^{\alpha_2}$ $+C_1X_1^{\alpha_1}X_2^{\alpha_3}$ | $X_1$= power<br>$X_2$= frequency<br>$0<\alpha_2<\alpha_1<\alpha_3<1$ |
| LNA | $C=C_oX_2^{\alpha_2}$ $+\frac{C_2}{X_1^{\alpha_1}}X_2^{\alpha_3}$ | $X_1$= noise ? ← ?<br>$X_2$= frequency |
| Antenna | $C=C_o+C_1X_1$ $+C_2X_2^{\alpha_2}$ | $X_1$= gain<br>$X_2$= frequency<br>$0<\alpha_2<<1$ |

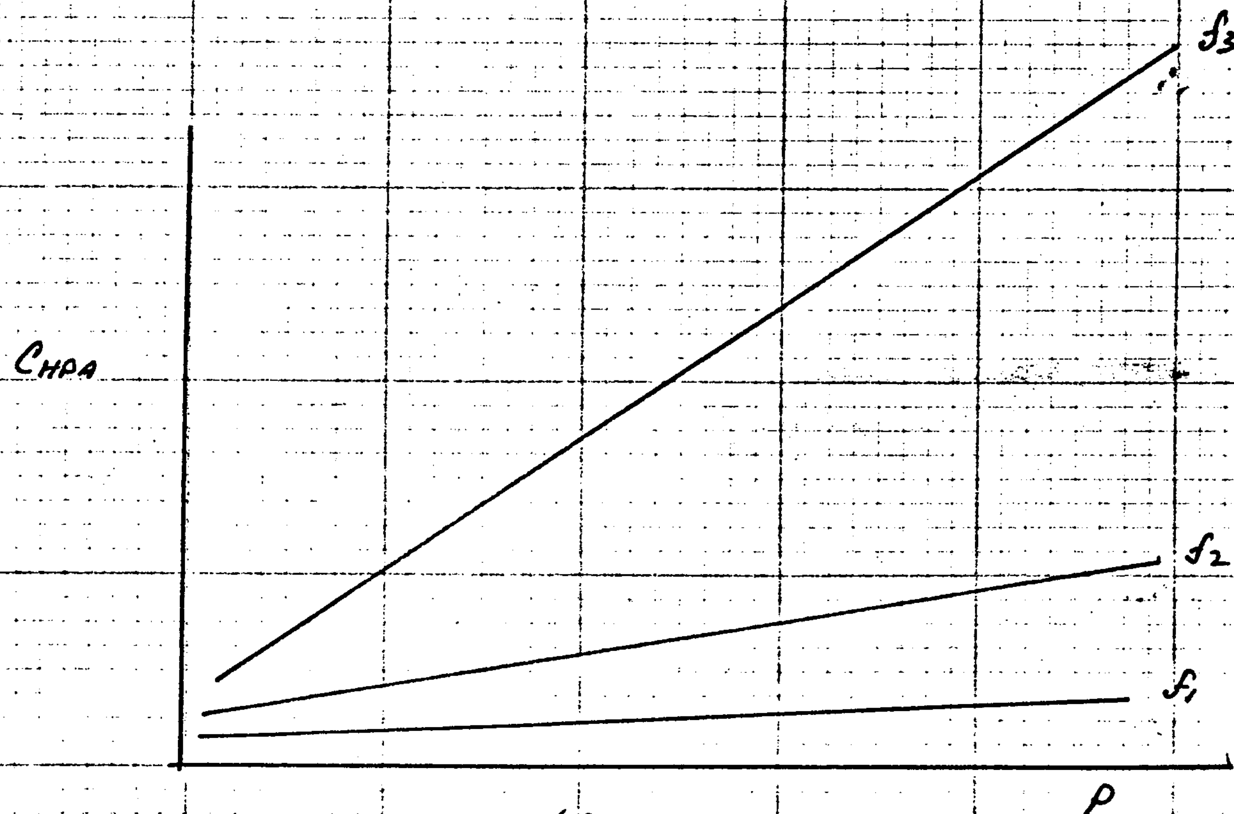
Table 1. Cost Models for Earth Station Subsystem




Scales ?

(a)

LNA COST MODEL WITH REGRESSION

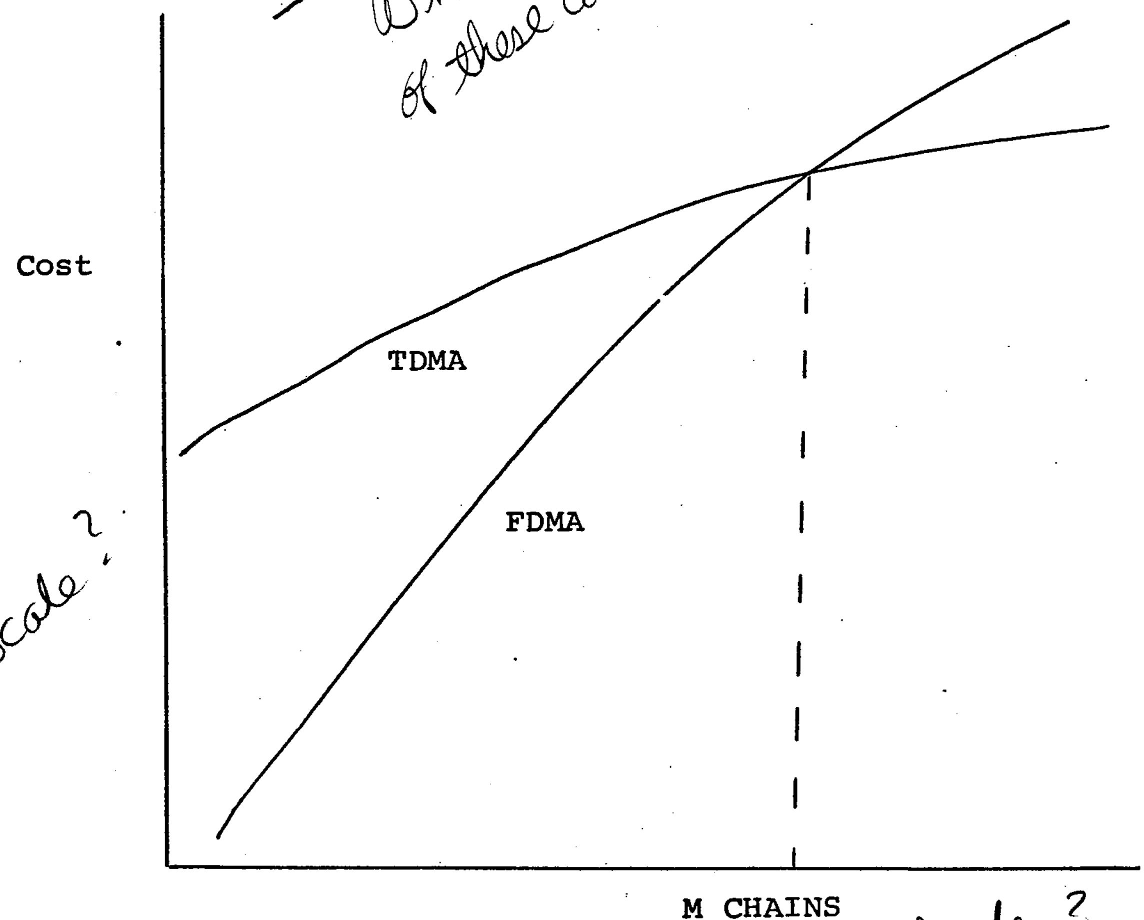



(b)

HPA MODEL

FIG 6: COST MODELS FOR LNA'S AND HPA'S

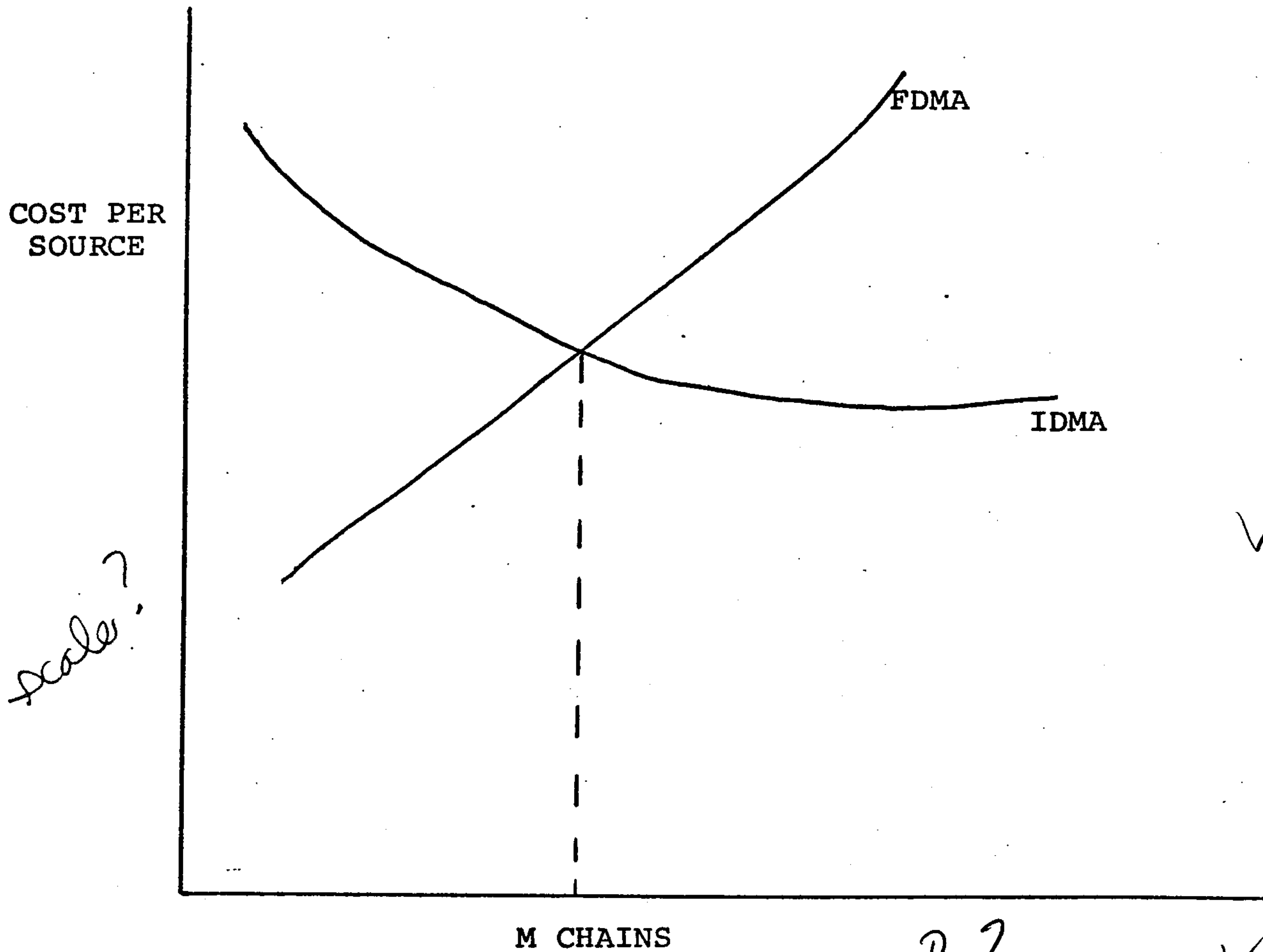
auth What is the basis of these curves?




Scale?

Scale?

Figure ~~8~~ 7. Cost Comparison of FDMA and TDMA au Basis of these curves?




Scale?

Figure ~~7~~ 8. Cost per Source Comparison we find FDMA slightly increasing whereas TDMA decreases! Thus there is an economy of scale factor in this design.

We shall expand these models in latter parts of this chapter. It should be emphasized, however, that the cost models should be revalidated due to the continuing change of technology.

## 7.3 SATELLITE COST MODELS

There have been several cost models developed for the space segment of a satellite communications network. They range from the simplest being so many dollars per pound to complex regressional techniques such as the SAMSO cost model. In this section we shall concentrate on the latter techniques. Ref ?

In developing a cost model for the space segment, we must recognize several factors. First, that launch costs fluctuate based on the vehicle. With the recent launch of the STS (Space Transportation System ) of NASA, the costs are expected to stabilize in time. Previous launch vehicles such as the Atlas-Centaur, Delta 2914 and 3914, and Titan, have costs that range from $20M to $45M per launch for a payload of about 1500 to 3500 lbs. It has been a previous rule of thumb that launch costs equal satellite costs.

The second factor is that there are recurring and nonrecurring costs. As we noted the nonrecurring costs include all fixed development costs that are expensed at the beginning of the project. The recurring costs are then the per satellite costs.

A third factor concerns the earth segment support necessary to operate a satellite system. This includes the TTC&M (Tracking, Telemetry, Command and Monitoring) stations as well as analysis stations. These generally are complex systems that provide for the terrestrial control node to the network. The cost model developed in this section will not address this issue. Typically the TTC&M costs are performance insensitive.

The model that we shall develop will relate cost to performance. It will include both recurring and nonrecurring and will be developed on a subsystem basis. The following is a list of the basic subsystem elements and their key performance factors.

1. Communications

   This includes the antenna, translation oscillators, LNA and HPA. Also, it would include any filters, switches, and any other RF elements. It does not include any base band processor equipment. We shall see that this is not included because the model uses historical data to generate costs and at present there are no such satellites (with the exception of LES 8/9).

The key performance factors for this are:

a. weight (lbs.) excluding structural support.

b. transmit power (w.)

c. transmit bandwidth (Hz)

d. information bandwidth (Hz)

e. wavelength (m.)

f. antenna gain (dBi); note that this must include both transmit and receive.

2. Structure

This includes all the mechanical structures in the satellite such as the antenna supports, the actual bus, and other elements. It is characterized by:

a. weight (lbs.)
b. volume ($ft^3$)
c. mass ($slug/ft^3$)

3. Propulsion

The propulsion systems is used to take the satellite from a parking orbit, through a transfer orbit and into the final circular geostationary orbit. These systems are called perigee kick motors or apogee kick motors. In the case of an Atlas Centaur or Delta launch only an AKM is required. It is an integral part of the spacecraft and is considered as separate from the launch vehicle. This system is characterized by:

a. weight (lbs.) of the motor and other elements

b. fuel weight (lbs.)

c. impulse capability of the motor (lb-sec)

d. burn time of motor (sec.)

It should be stressed that for geostationary orbits, the performance goal is to get to the final orbit position and this is relatively insensitive to the overall system performance. The larger the satellite, the larger the motor to transfer and finalize orbit. Advances in this area may include improved fuels or concepts such as the TUG for the STS.

TUG

4. TT&C

This portion of the spacecraft provides for communications to the earth TTC&M stations and is a monitoring element on the other spacecraft payloads. It primarily accumulates measurements, receives commands and transmits downlink. Depending on the capability of the satellite it may include processing, memory, and other functions.

The key factors are:

a. weight (lbs.)
b. output power (w)
c. bandwidth (hz)
d. wavelength (m.)
e. antenna gain (dBi); note that this may require a separate antenna
f. data rate (bps)
g. memory capacity (bits)
h. number of commands

5. Electrical Power Supply (EPS)

This portion of the spacecraft provides for power by means of solar arrays and batteries. The batteries are used when an eclipse of the sun occurs on a semi-annual basis. In addition it includes power control units that maintain several voltage levels, current levels and distributions. The key factors are:

a. weight (lbs.) of batteries, solar cells, and control units plus wiring.

b. battery weight (lbs.)

c. solar array weight (lbs.)

d. power (a)

e. solar array area ($m^2$)

f. design life (months)

6. Attitude Control System (ACS)

The ACS is used to keep the satellite positioned in orbit. The satellite moves about because of geopotential gradients, and it must be kept both in its orbit location and pointing toward the earth. If non-tracking earth stations with high gain, narrow beam antennas are used, this position must be fairly stable. For example with SBS this is $\pm$ 0.01°.

There basically are two types of satellite stabilization schemes; spin and three axis. The spin stabilized satellites are rotating constantly like a top and are kept stable in a direction perpendicular to the equatoral plane. Earth limb (edge) sensors are used to measure drift, and the ACS motors (jets) are used to position the satellite. Figures 9 and 10 show INTELSAT IV and IV-A which are spin stabilized. The antenna platform is decoupled so it can point through a rotary joint.

A three axis satellite is stabilized on all three axes by small rockets and a larger set of error sensors. Figure 11 shows INTELSAT V which is so stabilized.

The key factors are:

a. weight (lbs.)
b. fuel weight (lbs.)
c. impulse of jets (lb-sec)
d. lifetime (months)
e. number of axes stabilized
f. position accuracy (degrees)
g. number and type of sensors.

There are other minor elements of the spacecraft such as thermal control, dispenser and interstage which are not critical to the cost model.

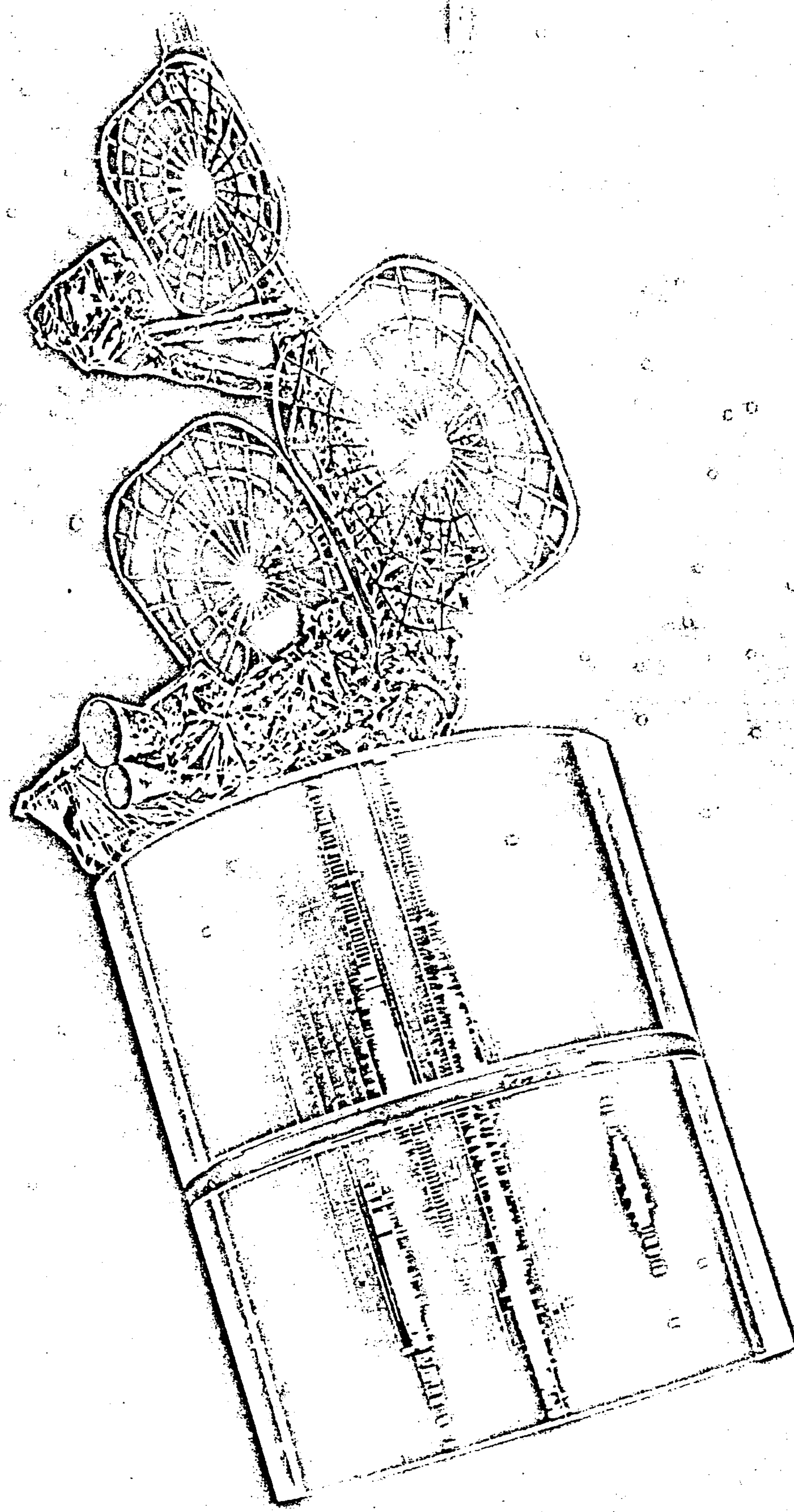


Figure 9 True!

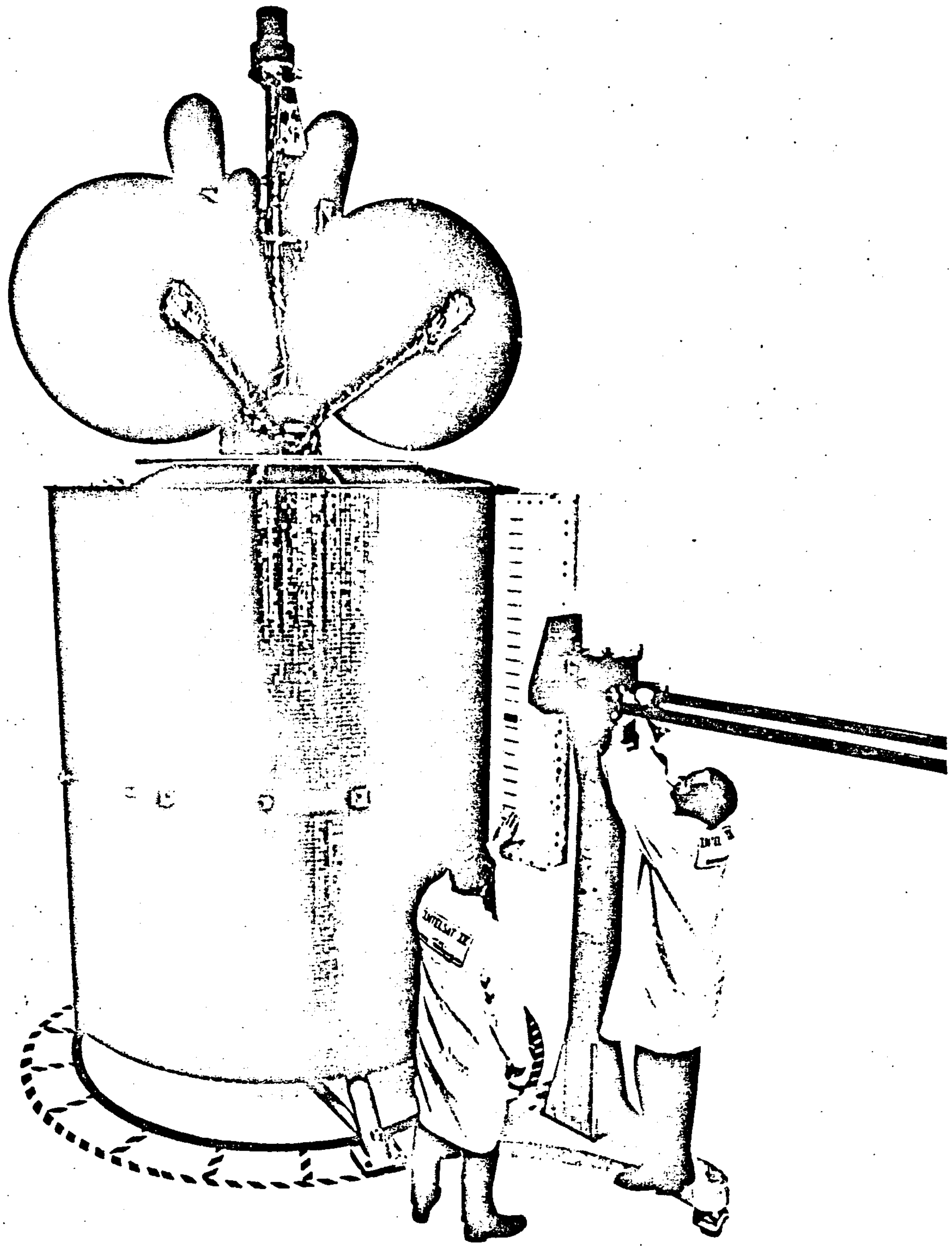


Figure 10.

Title!

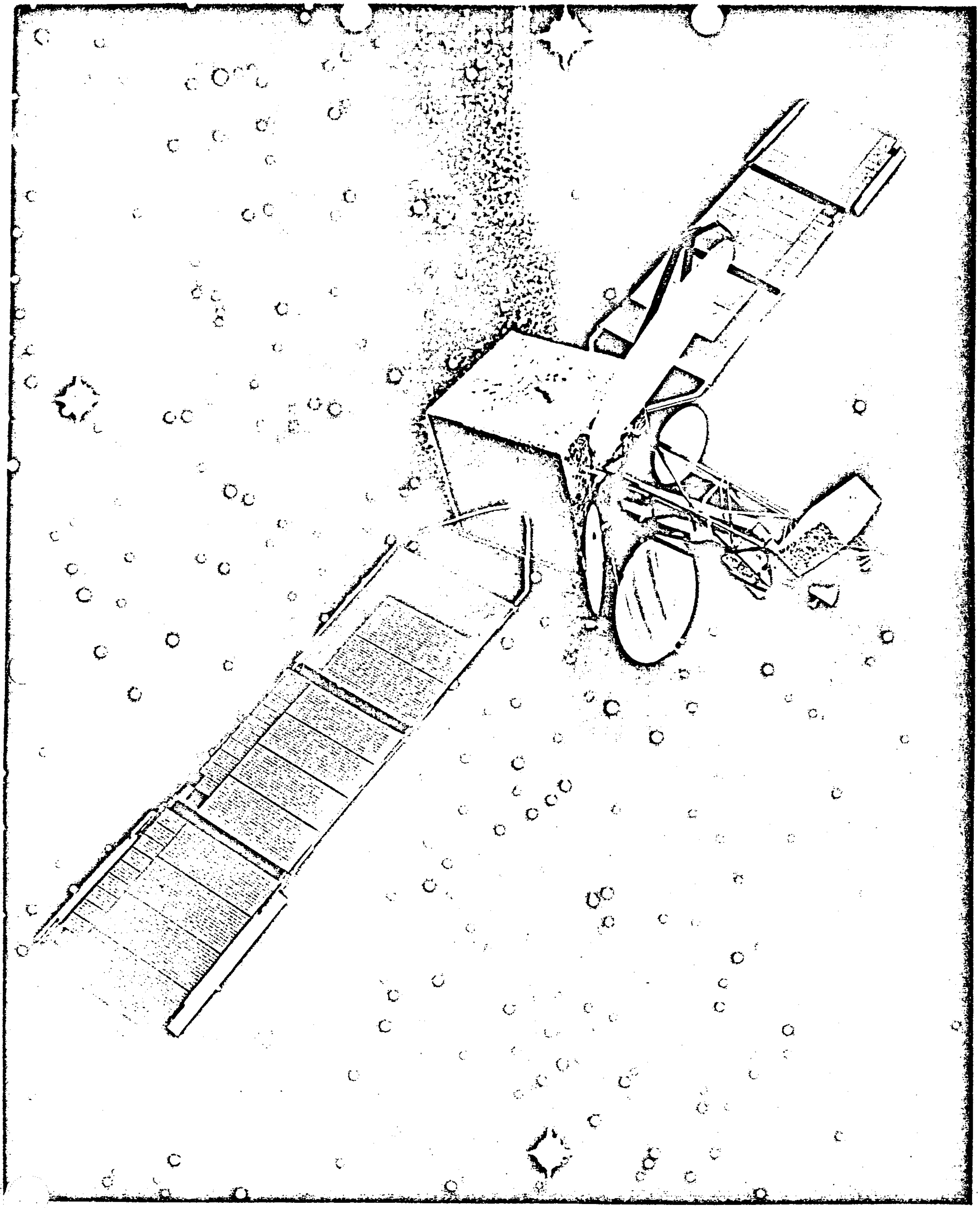


Figure 11 Title

The cost model approach is based on sampling various satellites that have been developed and determining the recurring and nonrecurring cost elements. These costs are parameterized on the key factors. Let us consider as an example the communications package. Let us also look at the key factors of weight, power, bandwidth, frequency and antenna gain. They define the following:

$C_{1,N}$ = nonrecurring cost of communications

$C_{1,R}$ = recurring cost of communications

$X_1$ = weight

$X_2$ = power

$X_3$ = bandwidth

$X_4$ = frequency

$X_5$ = antenna gain

Now we argue that both $C_{1,N}$ and $C_{1,R}$ depend on these five variables in the form

$$C_{1,R} = C_0 + C_1 X_1^{\alpha_1} + C_2 X_2^{\alpha_2} + \dots + C_5 X_5^{\alpha_5}$$
$$+ C_{12} X_1^{\beta_{11}} X_2^{\beta_{12}} + \dots + C_{45} X_4^{\beta_{45}} X_5^{\beta_{55}}$$
$$+ \dots \text{(other combinations)}$$

Is $\alpha_i$ an exponent or a multiplicative factor

The same holds for $C_{1,R}$. Now what we do is perform a regression analysis on all the data to determine the $C_i$, $\alpha_i$, $\beta_{ij}$ etc. to provide the best fit. For example if we find that $C_{1,R}$ depends only on power (X) then

$$C_{1,R} = C_0 + C_2 X_2^{\alpha_2}$$

Figure 12 depicts $C_1$,R versus power for various systems. The best fit curve is shown. Here $\alpha_2 > 1$ and $C_0 > 0$. We then do a standard regression to determine $C_0$, $C_2$, and $d_2$.

Table 2 summarizes the results for a wide variety of satellite systems (>30). What is surprising is that the subsystem costs depend on so few factors. Typical plots for the communications system, recurring and nonrecurring

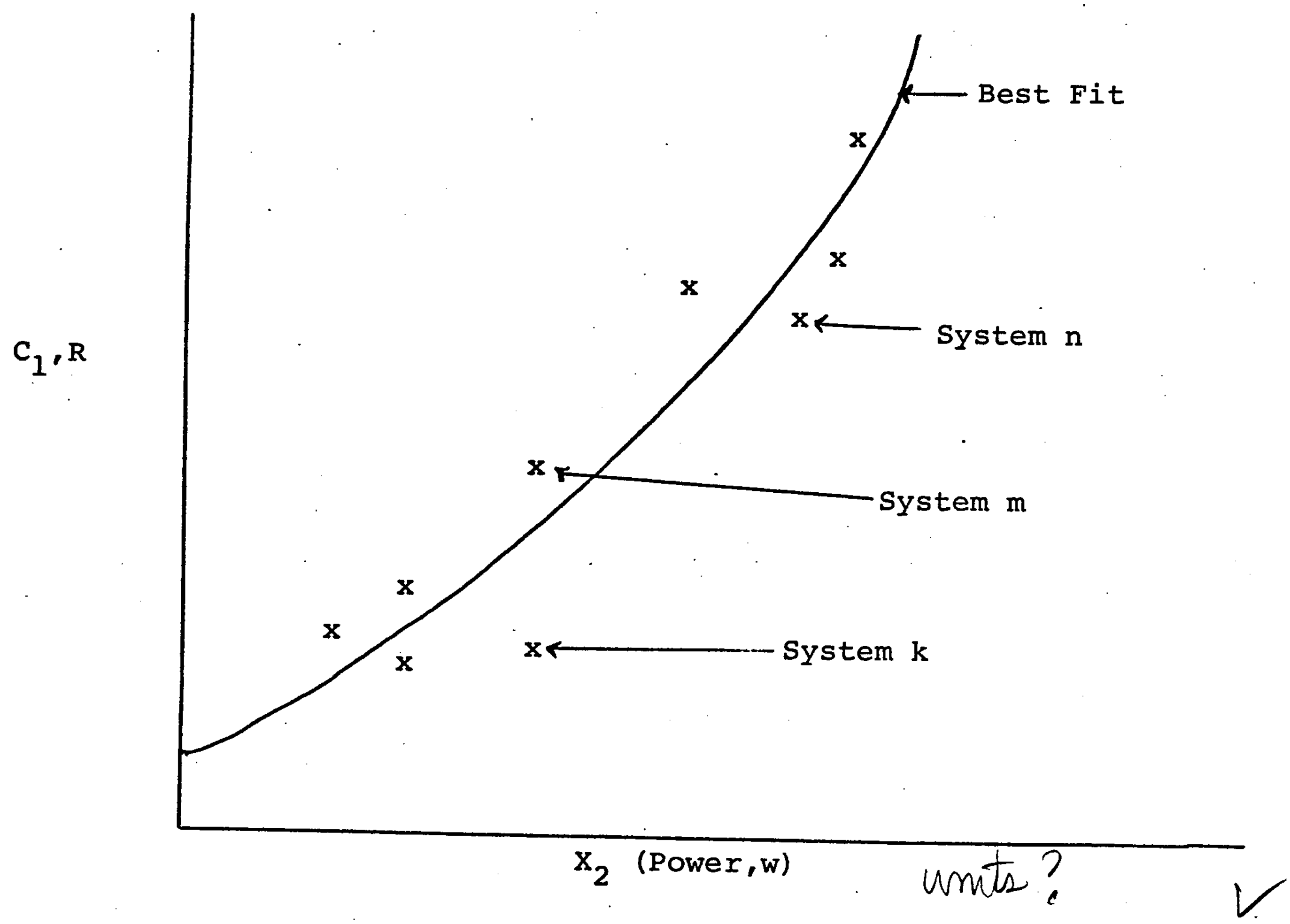



Fig 12: Least Squares Cost Fit

Table 2

Cost Relationships

Nonrecurring (NR)

| Subsystem | Cost ($\$10^3$) | Cost Drivers |
|---|---|---|
| Structure, TC & Interstage | $C = 550.3 + 106.5\ X^{.59}$ | Subsystem Weight |
| Electrical Power Supply | | |
| for Body Array | $C = 133.0\ X^{.5626}$ | Maximum Array Output |
| for Paddle Arrays | $C = 111.1 + 2.282\ X$ | Maximum Array Output |
| Attitude Control System | $C = 343.2\ X^{.5658}$ | Subsystem Dry Weight |
| Telemetry, Tracking & Cmd | $C = 593.3 + 34.27\ X$ | Subsystem Weight |
| Communications | $C = 979.7 + 95.66\ X^{.35}$ | Subsys Wt X Max Ar Output |
| Combined Comm & TT&C | $C = 1868.1 + 117.0\ X^{.68}$ | Maximum Array Output |

| Subsystem | Cost ($\$10^3$) | Cost Drivers |
|---|---|---|
| Structure, TC & Interstage | $C = -54.01 + 9.889\ X^{.74}$ | Subsystem Weight |
| Electrical Power Supply | $C = 8.061\ X^{.8352}$ | Maximum Array Output |
| Telemetry, Tracking & Cmd | $C = -11.22 + 16.78\ X$ | Subsystem Weight |
| Attitude Control System | $C = 14.72\ X^{.9031}$ | Subsystem Dry Weight |
| Communications | $C = 781.9 + .0143$ | Subsys Wt x Max Ar Output |
| Combined Comm & TT&C | $C = 1237.4 + .0154\ X$ | Subsys Wt x Max Ar Output |

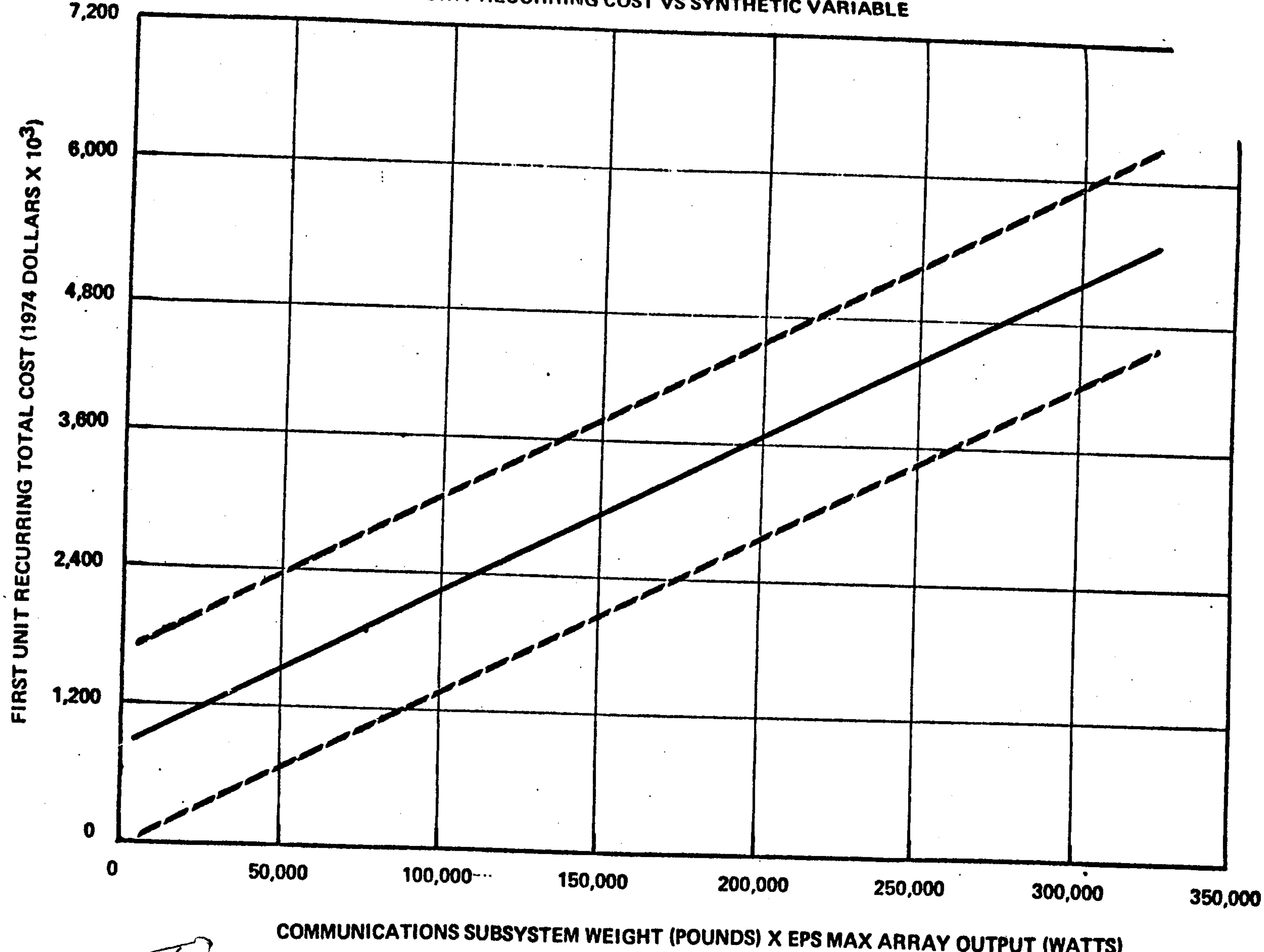
? ✓

are shown in Figures 13 and 14. These figures also show the 95 percent confidence regions.

With these models we can develop total cost by merely summing up the elements.

NORMALIZED

COMMUNICATIONS CER
FIRST UNIT RECURRING COST VS SYNTHETIC VARIABLE




Ref?

Figure [illegible]-14

## 3.4 SATELLITE SYSTEM DESIGN

In this section we apply the models that we have developed to a specific satellite system design. The objective, as we shall see, is to minimize overall costs. The analysis will demonstrate how the models developed can be used to determine the baseline for a system design.

The choice of the parameters of a satellite system design is generally made subject to performance and coverage constraints with the result that parameter tradeoffs result in changes in costs. In particular the per-unit or per-user costs represent a performance measure that can be used to obtain the RF parameters of the design which minimize such costs. A second method is developed which maximizes the energy per bit to noise specral density ratio ($E_b/N_o$) per beam.

There have been several previous attempts at design optimization. Schneider Ref proposed a model for a multiple beam satellite that determined a design for a mass constraint that maximized the satellite EIRP . The result that he obtained indicated that an equal split of mass between antennas and transmitters (including power sources) yielded this maximum. He however neglected the earth segment. In the work by Shaft and Roberts, Ref the residual satellite EIRP is maximized . This approach, however, does not provide a satellite design tool that yields the complete link parameters. The approach developed in this ~~paper~~ Book includes the effects of both earth segment and space segment.

The model used in this section assumes that the satellite has a multiple beam capability. In the satellite there is a tradeoff between the amount of mass used for antennas and that used for transmitters. As the mass of the satellite increases so does its cost (including launch). Thus larger mass allows for increased EIRP and G/T from the satellite. For the earth segment there is a tradeoff on cost between the antenna, the receiver, and the transmitter. A specific cost allocation or constraint to the earth terminal could be considered and the corresponding choices of gain, receiver noise termperature, and power obtained. However, when satellite and earth stations are combined in a system, the combination of costs attributed to satellite and earth stations are combined in a system, the combination of costs attributed to satellite and earth stations to minimize costs must be considered. This must be done subject to a performance constraint and a coverage constraint.

The approach developed in this section is not to be considered as a final design tool in the sense of providing a definitive system design. Rather, it provides a tradeoff model for the basic RF parameters that should permit the system designer to rapidly assess the impact on the system design of various changes in the basic design parameters. The model could be generalized to include the effects of factors such as modulation, access, source coding, channel coding, and onboard processing. However, such a large scale optimization is risky given the many uncertainties that would exist. The approach developed provides the basis for the development of optimized baseline designs that can be used in the development of more detailed systems. In addition, the technology impacts and tradeoffs as required can be quantitatively evaluated.

We shall consider a satellite system wherein a single satellite is used to provide communications between a set of users. The satellite uses a multiple beam antenna to provide services to the desired coverage area, and it has the same number of beams on the uplink and downlink. The multiple beams may or may not be used to provide frequency reuse.

Each beam of the antenna will be of identical beamwidth and each of the system users will have identical transmitters and receivers. It will also be assumed that the frequencies for transmit and receiver are reasonably close to that so no gross changes in the antenna pattern will occur.

A general model of this type is applicable to a mobile communications system or a regional system composed of many equally loaded sources and sinks of information.

In addition, we shall assume that the system is a purely digital one operating in a burst mode (e.g., TDMA). No multicarreir systems will be considered, although the model can be generalized to include such designs. It will also be assumed that full beam-to-beam interconnectivity is provided by either satellite switching or earth station network control.

Each uplink beam will be equipped with a separate receiver and each downlink beam will be driven by a single amplifier. This is a simplification that can be altered in both the actual design and the anaytical model.

No form of modulation, coding, source coding, or multiple access scheme will be assumed. The primary design emphasis will be in obtaining the RF parameters of the earth terminal and the satellite. This system is typified by the earth station parameters of receiver temperature $T_E$, transmit power $P_E$, and antenna gain $G_E$. Similarly the satellite is characterized by $T_S$, $P_S$, and $G_S$ as well as the total number of beams N. We shall concentrate on the selection of these parameters in an optimal fashion subject to the simplifications stated above.

There are many measures of performance that can be used to evaluate a system design, that can be used as design objectives, or that can be used as criteria for system optimization. Two of specific interest are the energy per bit to noise spectral density per beam $E_b/N_o$ and the per-user cost $C_n$. The former is a measure of the system quality for a given set of coding, access, and modulation schemes. The latter is useful in determining the cost effectiveness of this system design. In the system assumed, each beam transmits at a rate of R bits per second. Thus the total system capacity is NR bits per second. The carrier power to noise spectral density ratio $C/N_o$ depends upon the system parameters that were previously defined. In turn, we have the relationship for each beam:

$$E_b/N_o = R^{-1} C/N_o$$

For the costs of the system we have both earth segment costs $C_E$ and space segment costs $C_S$. Typically $C_E$ depends upon $T_E$, $P_E$, and $G_E$ plus amounts that are functions of coding, modulation, and access techniques. For our purposes we shall consider these costs as fixed and concentrate on the RF portion. Similarly for the satellite, $C_S$ depends upon $T_S$, $P_S$, and $G_S$ as well as N, the number of beams. If we assume that there are n users, each with a receiver/transmitter, then the cost per user is defined as

$$C_n = C_E + C_S/n.$$

It should also noted that the rate R depends upon the number of users and that this dependence is a function of user type and distribution.

We can now state the two optimization criteria in terms of the above definitions. The first would maximize $E_b/N_o$ subject to a fixed $C_n$ (or $C_E$ and $C_S$) and the second would minimize $C_n$ subject to a fixed $E_b/N_o$.

There are three constraint equations that will be useful in developing the model further. The first is that constraint that deals with coverage. We shall assume that the system must provide coverage to a given region. Thus as we increase the satellite antenna gain we decrease the width of the beam. This therefore requires more beams, an increased N, so that the desired region is continually covered.

We pose the coverage constraint of the following form:

1) a) The region is of arbitrary shape but of maximum extent $\phi_1$ in east-west (E-W) and $\phi_2$ in north-south (N-S).

   b) The region is to be covered by N circular beams of beamwidth $\phi_b$ in such a way that minimizes N.

2) No two adjacent beams are to use the same frequency band.

3) Users not assigned directly to a beam are assigned to the closest beam in terms of distance to beam center.

It will be convenient to perform the analysis assuming that the satellite antenna is a parabolic reflector. For this case, assuming an aperture efficiency of 50 percent, we can relate antenna gain and beamwidth $\phi_b$ via the following form (4):

$$G_S = (70\pi)^2/\phi^2_b$$

Now is can be shown that for both rectangular and circular coverage constraints we can relate $\phi_b$ and N as follows:

$$\phi_b = \beta/N^{\frac{1}{2}}$$

where $\beta$ is a constant that depends upon the geometry of the constraint. Thus combining relationships, the coverage constraint indicates that $G_S$ is linearly related to N, the number of beams.

The second constraint equation is for the earth station cost. The cost $C_E$ is the sum of antenna, receiver, and transmitter costs plus other fixed costs. Thus

$$C_E = C_A + C_R + C_T + C_{EO}$$

where

$$C_A = A_A G_E^{\alpha/2}$$

$$C_T = A_T P_E$$

$$C_R = A_R/T_E + A_{R0}$$

Where the quantities $A_A$, $A_T$, $A_R$, and $A_{R0}$ are constants to be obtained from price information ( ? ), $\alpha$ is a positive constant (typically 2), and $C_{E0}$ is the cost of all non-RF equipment ( ?.

The third constraint is that for the satellite cost. It will be assumed that a cost per unit mass can be ascribed to the satellite and that the major mass components are the antennas and transmitters. To simplify the design we then let $T_s$ be specified since it will have no significant effect on satellite mass. This results in a reduction of the variables to six. The satellite mass $M_S$ is thus given by

$$M_S = \gamma(M_A + M_T)$$

where $\gamma$ is a constant (greater than 1) that relates communication system mass to total satellite mass. Here it can be shown that for a parabola with the coverage constraint we have

$$M_A = N\left[(1.64\rho/\beta)\ (70/\pi)^2 + M_F\right]$$

where $\rho$ is the antenna density in kilograms per meter squared and $M_F$ is the mass per beam feed. Also,

$$M_T = NA_S P_S$$

where $A_S$ is a constant that for the satellite power choice relates power to mass. These units have been previously analyzed by Schneider ( ). Thus if $C_p$ is the cost per pound of the satellite we find that Ref?

$$C_S = C_p M_S$$

which is in terms of N and $P_S$ only, having eliminated $G_S$ using the coverage constraint.

Combining these results provides a total per-user cost in terms of the basic set of RF parameters.

The total link carrier to noise spectral density is given in terms of the individual carrier to noise densities for both the uplinks and downlinks. These separate link parameters are given by

$$(C/N_o)_i = \begin{cases} G_E G_S P_E / kT_S PL_u, & i=u \\ G_E G_S P_S / kT_E PL_d, & i=d \end{cases}$$

where u is uplink, and d is downlink, the $PL_{u,d}$ are the link path losses. For a satellite with a linear repeater we have the standard form for the total link $C/N_o$:

$$C/N_o = [(C/N_o)_u^{-1} + (C/N_o)_d^{-1}]^{-1}$$

where we interpret $P_S$ as the repeater gain times the uplink received power. For the more common case of a power limited repeater we have

$$C/N_o = [(C/N_o)_u^{-1} + (C/N_o)_d^{-1} + W(C/N_o)_u^{-1}(C/N_o)_d^{-1}]^{-1}$$

where W is the bandwidth of the channel. Typically we can neglect the third term in parentheses above when each $(C/N_o)_{u,d}$ is greater than unity and the uplink carrier to noise power ratio is also large. In this case the noise does no significant power robbing of the uplink and the expression reduces to that of the linear repeater. We shall deal with this case throughout and consider the more general case only in passing. It should also be noted that we neglect the possibility of an onboard processor. Models including these effects are considered by Schneider Ref?

As discussed in the previous section, the first optimization problem to be considered entails a maximization of of $E_b/N_o$ per beam given a constraint on $C_E$ and $C_S$. The second problem is to minimize $C_n$ given a constraint on $E_b/N_o$. Before proceeding, recall that R, the rate per beam is considered to be the same for each beam. The system is designed to serve n users with a rate of $R_o$ per user (on the average). Thus $R=(n/N)R_o$. This implies that both problems inherently depend upon n, the number of users.

## $E_b/N_o$ Maximization

From the previous analysis, for both the linear and power limited transponder (if we assume no uplink power robbing),

$$E_b/N_o=(n/NR_o)P_SP_EG_SG_E/(PL_ukT_SP_S + PL_dkT_EP_E).$$

Now we typically choose $T_S$ because it does not impact the satellite cost significantly. Similarly from the coverage constraint $G_S$ is given in terms of N. We can then use the $C_S$ constraint to express $P_S$ in terms of N. Also we can eliminate $T_E$ using the constraint on $C_E$. This still leaves the expression to be maximized in terms of three variables, and a closed form analytical result is difficult.

Two special cases are worth considering: the uplink and downlink limited cases. Each permits considerable simplification, yet provides insight into the general case. In particular, there are many designs where one of these links is the dominant factor, such as a ship-to-shore or shore-to-ship link in maritime communications.

Case I. Uplink Limited: In this case

$$E_b/N_o = R^{-1}G_SG_EP_E/kT_SPL_u$$

Based upon the model developed , $T_S$ is a fixed parameter in that it does not enter into the mass constraint. Thus, since $P_S$ is not contained in the experience for $E_b/N_o$, we shall assume that it is fixed and thus determines $G_S$. Thus the maximization is on the product of $G_E$ and $P_E$ which relate to the cost constraint. In addition, we assume that $T_E$ is fixed so that the cost constraint becomes

$$C_E = A_A G_E^{\alpha/2} + A_T P_E + C_{EO}$$

Using this in the $E_b/N_o$ expression and differentiating and solving for $G_E$ yields

$$C_E = (C-C_F)/[(1+ \alpha/2) A_A]^{2/\alpha}$$

as the optimum choice. Note that if $\alpha$=2, then the cost of the antenna would equal the cost of the transmitter.

Case II. Downlink Limited: For this case,

$$E_b/N_o = R^{-1}\, G_S G_E P_S/kT_E PL_d$$

As with the uplink case, we must fix certain paramters. Here we fix $P_E$ and this then implies that the cost constraint becomes

$$C_E = A_A G_E^{\alpha/2} + A_R/T_E + C_F$$

where $C_F$ represents all constant fixed costs. Similarly, using the mass and coverage constraint R, $G_S$ and $P_S$ can be written as functions of N, the number of beams. Thus we also note

$$\max_{N,G_E} (R^{-1}\, G_S G_E P_S/kT_E PL_d) = (kPL_d)^{-1}$$

put on same line!

$$\cdot \max_{N} (G_S P_S/R) \max_{G_E} (G_E/T_E)$$

That is, the maximization separates to a maximization of the normalized EIRP and of the receive G/T. This implies that the satellite and earth station can be optimized separately.

As a specific example, we can consider the case of a dedicated channel with circular coverage. The EIRP becomes

$$G_S P_S/R = (N/nR_o)\ (70\pi^2)/(\beta^2/N)$$
$$\cdot \{[M_C - NM_F - (C_o/\beta^2)N\ \ /N(C_p + C_W)]\}$$

This then yields for an optimum number of beams N*

$$N^* = M_C/2(M_F + C_o/\beta^2).$$

Note that N* increases linearly with communication systems mass. The two factors in the denominator are $M_F$, the feed mass, and $C_o/\beta^2$, a ratio of the wavelength normalized antenna mass to coverage constant.

Similarly the maximization on $G_E$ yields

$$G^*_E = (C_E - C_{EO})/2A_A$$

where $C_{EO}$ represents all the constant costs.

$C_n$ Minimization

The dual problem can be stated as follows. Define an M X 1 vector x whose components are $N, P_S, G_E, P_E, T_E$ (where m=5) and define a scalar function h(x) as

$$h(x) = E_b/N_o.$$

From a system design viewpoint we want h(x) to equal a value $h_o$, which is the design value (e.g., 15 dB). Now we want to choose that vector x which minimizes $C_n$ subject to the constraint $h_o$ and in addition subject to the constraints that each $x_i$ (the components of x) is positive. That is, define $C^*_n$ as

$$C^*_n = \min_{h(x)=h_o} C_n, \; x_i \geq 0, \; i-1,\ldots,M.$$

A direct analytical solution to this problem has not been found. Although one could solve it using the Kuhn-Tucker theorem, Ref? we propose an alternative approach. We solve the problem first of maximizing $E_b/N_o$ given $C_E$ and $C_S$. This is done for an array of $C_E$ and $C_S$ values and the set of $C_E$, $C_S$ points yielding a maximum $E_b/N_o$ equal to $h_o$ is obtained. Then over this set the pair minimizing $C_n$ is obtained. That is, we solve the cost minimization by imbedding the $E_b/N_o$ maximization in it directly. Although this is not the formal dual problem, a conceptual duality still exists.

The evaluation of the minimum $C_n$ result has important implications in the area of telecommunications economics. Specifically, if we define C as the total system cost, then C equals $nC_n$. $C_n$ is typically called the average cost. The marginal cost $C_n$ is defined as

$$C_n = \partial C/\partial n$$

and it is the ratio of marginal to average costs that is called the scale factor . The analytical result thus developed, although restricted by the assumptions, provides a basis for demonstrating the economies of scale that exist in a satellite communications system. Specifically, if the average cost per user decreases as the number of users increases, then the system is said to have scale economies. As we shall show, this holds for the example given in the next section.

The model and optimization procedure developed in this section can be used to obtain a set of baseline design parameters as well as to determine the possible tradeoffs that can be performed among the different technological alternatives.

In this section we analyze a particular example to demonstrate a set of possible tradeoffs. We consider a domestic communications system that must provide coverage to the continental United States (CONUS). The choice of frequency was in the L band region (1.5 GHz) for both uplink and downlink. A possible example of such a design would be a domestic mobile satellite communications system that requires a mobile-to-mobile link.

There is an assumed CONUS data rate that must be satisfied based upon user characteristics, system performance goals, and source coding techniques. We define this total rate as $R_T$. It will also be assumed that the traffic will be distributed uniformly over each beam so that the required rate per beam R is $R_T/N$. It should be clear that this constraint can be modified to accommodate systems with differing rates; however, from an operational point of view such fine tuning often reduces the flexibility of the total system design.

Table 3 lists the set of parameters that are used for this particular model. We have simplified the analysis by assuming that the mobile terminal has a fixed gain. Specifically, we assumed that a 3-dB dipole antenna was used. The cost data for the earth terminal was based upon a survey of the present available prices for L band front-end equipment. These data are plotted in Fig. 15 and were used in the model. The satellite cost per pound, $C_p$ in Table 3, includes shuttle (STS) launch costs.

We first maximize $E_b/N_o$ for the system defined in Table 3. To do this we fix $C_E$. Then analytically the value of $E_b/N_o$ is maximized and parameterized on $C_E$, $C_S$, and N. In Fig.16, $E_b/N_o$ is plotted as a function of N, the number of beams. The maximum point is clearly evident in each curve. This demonstrates that for all cases, the multiple beam satellite is important in the design process. However, an excess of beams reduces performance for a fixed satellite cost, that is, mass.

Once the maximum $E_b/N_o$ is obtained, called $(E_b/N_o)^*$, the effect of varying $C_E$ or $C_S$ can be considered. These results are presented in Figs.17 and 18 for the system constraints given in Table 3. These curves show the sensitivity of peak $E_b/N_o$ to costs, and both curves demonstrate a saturation effect with increasing cost, demonstrating a decreasing cost effectiveness. These curves can also be used to determine the minimum cost design. Specifically, if we choose a design value for $(E_b/N_o)^*$, including all margins, and then consider the set of $C_E, C_S$ values that yield this design value, $C_n$ can be calculated and the minimum point obtained.

This process of design optimization is performed for two additional cases. Table 4 presents the results for Case I, which is that in Table 3, and Cases II and III which have more users and increasing $R_T$.

The evaluation of $C, C_n$, and $c_n$ as a function of optimization is presented in Fig. 19. It should be noted that $C_n$ and other costs neglect the fixed costs of the individual earth terminals due to coding, modulation, and access components. The economy of scale effect is evident in the curves with their decreasing costs. Each

TABLE 3

**Basic Design Constants for Example Optimization**

| | |
|---|---|
| $R_T$ | 2.7 Mbit/s |
| $n$ | 20 000 |
| $f$ | 1.5 GHz |
| $C_p$ | \$20 000/lb |
| $\gamma$ | 0.3 |
| $M_F$ | 0.4 kg |
| $A_S$ | 0.2 kg/W |
| $\rho$ | 1.38 kg/m$^2$ |
| $T_S$ | 400 K |
| $G_E$ | 3 dB |
| $A_R$ | \$100/deg K |
| $A_{R_0}$ | \$50 |
| $A_T$ | \$100/W |

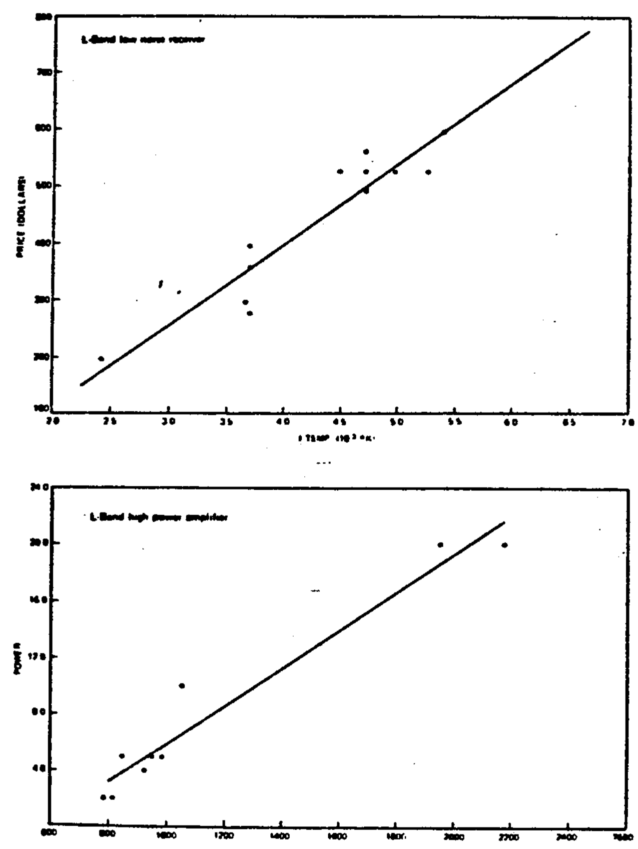



Fig. 1. Cost data for earth station receiver and transmitter in L band.

15

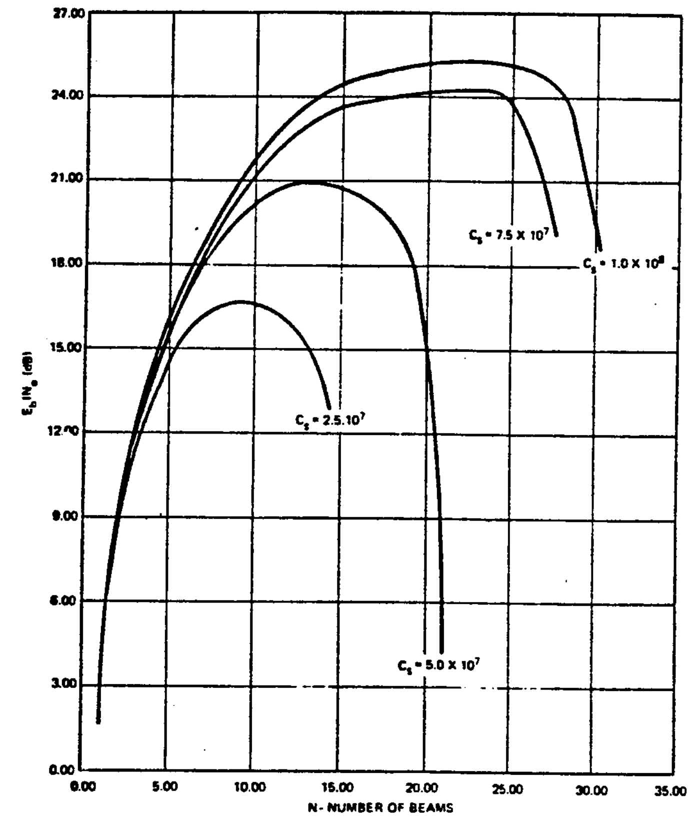



Fig. 2. Plot of $E_b/N_0$ versus $N$ showing maximization.

16

Ref's?

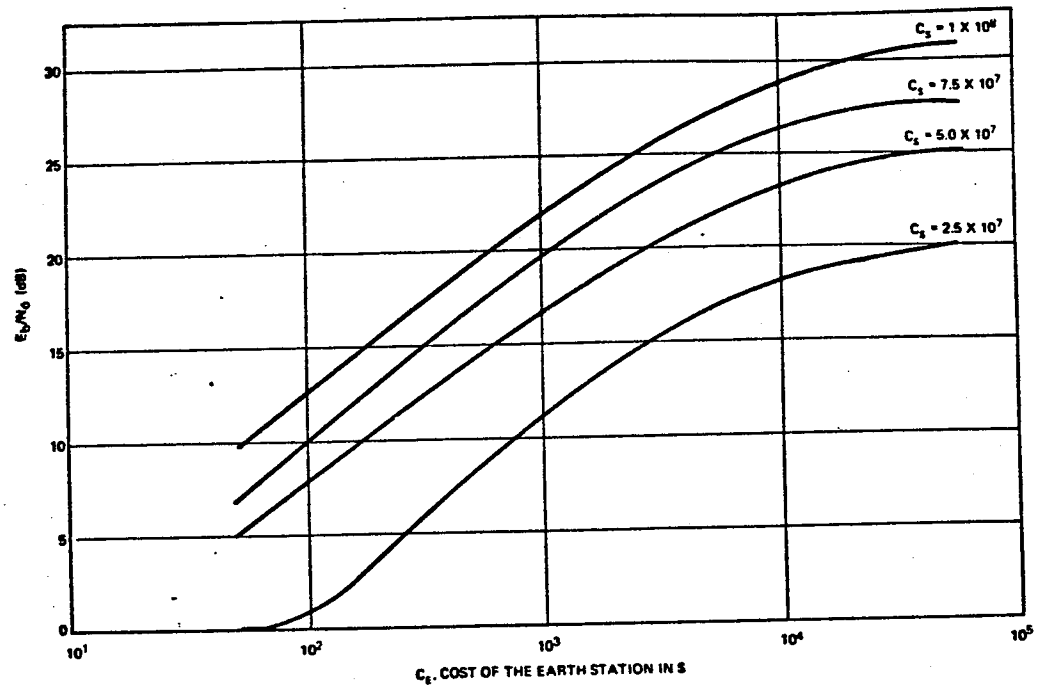



Fig. ~~3~~. Optimum $E_b/N_0$ versus $C_E$.
17

Ref 1 p. 9

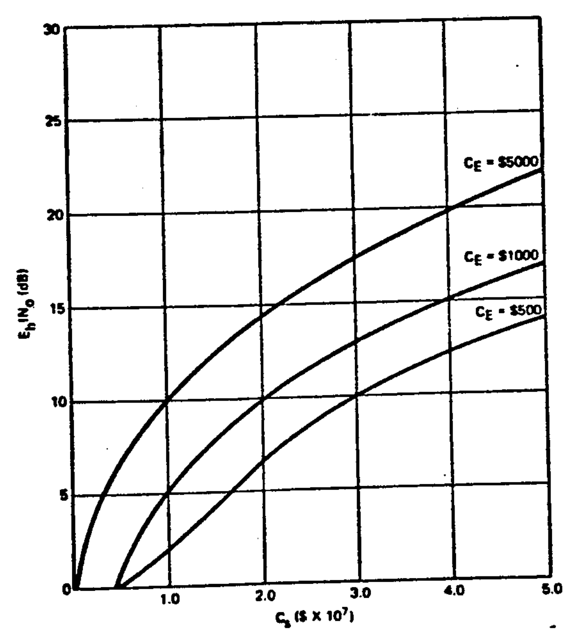



Fig. ~~4~~. Optimum $E_b/N_0$ versus $C_S$.
18

**TABLE 4**

Satellite Parameters for Various Optimized Designs

| | Case I | Case II | Case III |
|---|---|---|---|
| Number of users | 20 000 | 100 000 | 200 000 |
| Bit rate per beam (Mbit/s) | 0.27 | 0.38 | 1.0 |
| Beams | 10 | 22 | 28 |
| Beamwidth (deg) | 1.68 | 1.27 | 0.92 |
| Diameter (m) | 7.5 | 10.5 | 13.5 |
| $G/T$ (dB/deg K) | 16.3 | 18.7 | 21.5 |
| EIRP | 55.5 | 55.7 | 61.6 |
| Satellite mass (kg) | 682 | 1136 | 2273 |
| Satellite power (W) | 200 | 250 | 600 |

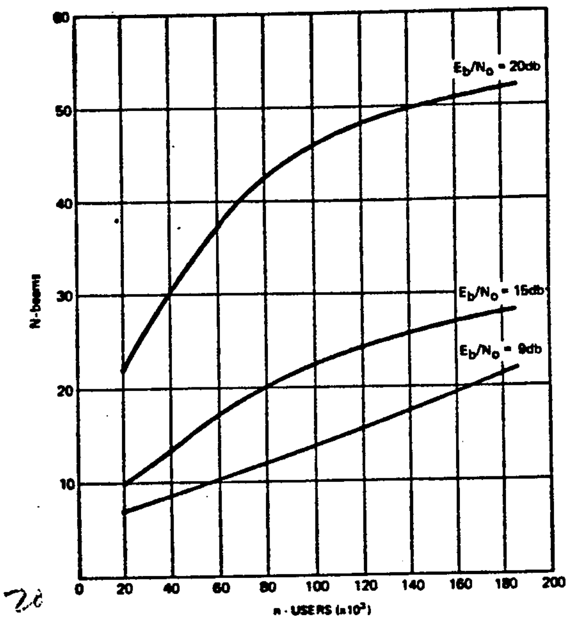



Fig. 6. Variation of optimum number of beams as a function of number of users for fixed performance.

Ref?

✓

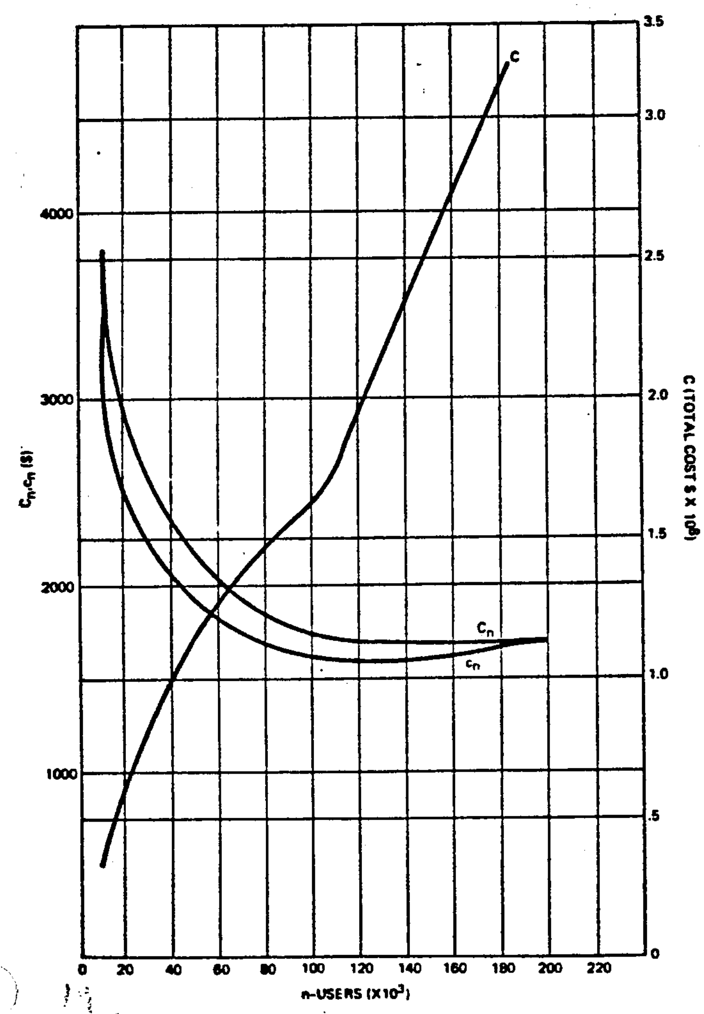



Fig. 5. Evaluation of economic factors demonstrating economy of of scale factor for the example.

point on the $C_n$ curves was an optimized design obtained by plotting the number of users and the corresponding total data rate $R_T$. Cases I to III of Table 4 are only three points of this curve.

A final result is the dependence of the number of beams N on the number of users. This has been plotted in Fig. 20 for varying values of $E_b/N_o$ constraints. 20

The model developed provides a method of trading off several important design factors in a satellite system design to achieve specific performance objectives in terms of cost or capacity. The model allows the system designer to rapidly develop baseline configurations that can be used in more detailed studies.

As a result of the example considered, it has been observed that a multiple beam design provides considerable improvement in system performance and can, in turn, lower the per-user costs. It should be noted, however, that this assumes that suitable interconnectivity can be provided on the satellite. This factor, combined with the ability to physically fit a large antenna in a launch vehicle, is a limiting factor on the applicability of multiple beam antenna use.

It was observed that with the more complete model, the 50/50 split between antenna mass and transmitter/prime power mass on the spacecraft still held, as was observed by Schneider Ref. 5. A similar split in cost can be observed for the earth segment.

The case considered showed that the optimum design point was broad. Yet, outside of this broad range, the slope of the curves were steep. This indicates that design insensitivity to beam size and number exists over a useful range, but that caution should be exerted in staying within that range. This optimization procedure can be extended to satellite designs where the link frequencies are different, e.g, UHF/SHF, by applying two constraints on $E_b/N_o$ and by expanding the number of variables to be optimized. Although straightforward to formulate, the solution is quite difficult.

One of the most useful advantages of this model is that it provides a cohesive method of evaluating the impact of technologies on system performance and cost. The use of reduced cost transmitters and receivers or lightweight antennas can be consistently considered. In addition, by using the satellite mass constraint and relating it to costs, optimized designs for future shuttle based launches could evolve.

Finally, the development of an analytical model which determines the economy of scale factor is an important result not only from a system design point of view but also in terms of the nature of the business investment that must be made.

## 3.5 EARTH STATION OPTIMIZATION

We now want to consider a different form of optimization. Here we assume that the satellite is given as would be the case in a leased operation (e.g., WESTAR, RCA SATCOM, SBS, etc.). Thus satellite costs are defined by tariffs that depend on bandwidth or power use. The satellite is leased so that it becomes a continuing expense. We do however wish to install a set of earth stations to use that leased capacity. The problem then is to determine a least cost design using our concepts of discounted cash flow.

Consider a simple communication network consisting of two identical earth stations and a leased satellite segment. It is assumed that:

a. the cost of satellite usage is proportional to the extent of transponder usage.

b. the satellite's amplifier has linear gain,

c. multiple channels are used,

d. path losses are independent of earth station designs,

e. coding gain is a function of coding rate and bit error rate but not modulation,

f. a broad range of high power amplifiers, antennas, and low noise amplifiers may be obtained,

g. component costs which are not known may be approximated from known prices with the least squares method,

h. no back-up systems are used, and

i. maintenance and operational costs for earth stations do not depend on the devices implemented in the stations.

It is necessary to keep performance criteria in mind when considering possible earth station designs. A constraint equation to do this may be obtained by focusing attention, as before, on the system energy per bit per noise spectral density,

$\left(\frac{E_b}{N_o}\right)_{available}$. This is given by

$$\left(\frac{E_b}{N_o}\right)_{av} = \left(\frac{E_b}{N_o}\right)_{REQ} \bullet \frac{\text{coding gain}}{\text{processing loss}} \bullet \text{actual margin}$$

4r

$\left(\frac{E_b}{N_o}\right)_{REQ}$ is the minimum energy per bit per noise spectral density which yields bit error rates at a desired level.

Coding gain is the factor dependent on coding rate, which is a measure of the amount of redundancy put in a signal. In order to maintain the same bit error rate, $\frac{E_b}{N_o}{}_{av}$, the energy per bit of information per noise spectral density, must be increased. Such an increase is reflected by the coding gain factor.

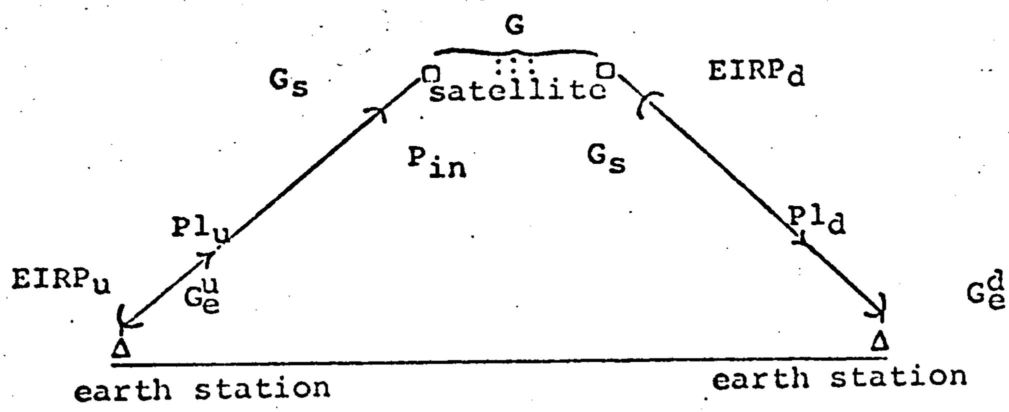
Processing loss refers to the losses encountered in generating the signal and actual margin refers to the power beyond the required amount which is put into the signal for safety. A factor of 3 dB is commonly used for the actual margin.

The link analysis relates transmitter power, antenna diameter, and receive noise temperature to the carrier to noise spectral density $\left(\frac{C}{N_o}\right)_T$, which is related to

$\left(\frac{E_b}{N_o}\right)_{av}$ by:

$$\left(\frac{C}{N_o}\right)_T = \left(NR_o\right)\left(\frac{E_b}{N_o}\right)_{av}$$

where N is the number of channels used and $R_o$ is the rate of transmission for each channel. From this it is possible to generate earth station designs in accordance with the desired system characteristics by computing either power, diameter or temperature as a function of the other two and $\frac{C}{N_o}{}_T$. Fig. 21 and the definitions will be used in doing this.




Cut
F. 21

EIRP = effective isotropic radiated power

$G_e$ = gain of the antenna at the earth station

$P_e$ = earth station power

$PL_u$ = losses on the uplink

$PL_d$ = losses on the downlink

$G_s$ = gain of the antenna of the satellite

G = gain of the satellite itself

$P_{in}$ = power coming into the satellite

R = bit rate (bits/sec)

N = number of channels or users

$R_o$ = rate per user

$\frac{C}{N_o}$ = carrier to noise spéctral density

$T_e$ = receive noise temperature

f = wavelength = c/f

Figure 21.

The total carrier to noise spectral density, $\left(\frac{C}{N_o}\right)_T$, is given in terms of $\left(\frac{C}{N_o}\right)_u$ and $\left(\frac{C}{N_o}\right)_d$, the carrier to noise spectral densities for the uplink and the downlink respectively. As before we write: State where this result was cited in your book!

$$\left(\frac{C}{N_o}\right)_T = \left[\left(\frac{C}{N_o}\right)_u^{-1} + \left(\frac{C}{N_o}\right)_d^{-1}\right]^{-1}$$

But from the link analysis $\left(\frac{C}{N_o}\right)_u$ and $\left(\frac{C}{N_o}\right)_d$ are found to be:

$$\left(\frac{C}{N_o}\right)_u = \left(\frac{P_e\, G_e^u}{PL_u}\right)\left(\frac{G_s^u}{T_s k}\right);$$

and

$$\left(\frac{C}{N_o}\right)_d = P_{in}\left(GG_s^u\right)\left(\frac{1}{PL_d}\right)\left(\frac{G_e^d}{T_e k}\right); \text{ where } P_{in} = P_e G_e^u \left(\frac{1}{PL_u}\right).$$

and

$$GG_s^u = \frac{4\pi}{\lambda^2} \bullet \frac{\text{satellite saturation EIRP}}{\text{satellite saturation flux}} \bullet (\text{input backoff})$$

Substituting these expressions for $\left(\frac{C}{N_o}\right)_u$ and $\left(\frac{C}{N_o}\right)_d$ in $\left(\frac{C}{N_o}\right)_T$ and solving for the earth station power yeilds the desired constraint equation for $P_e$ in terms of antenna diameter, receive noise temperature, and total carrier to noise spectral density. This is:

$$P_e = \left(NR_o\left(\frac{E_b}{N_o}\right)_{av}\right)\left[\frac{k\ (T_s\ PL_u\ G\ \ G_e^d + T_e\ PL_u\ PL_d)}{GG_s^u G_e^u G_e^d}\right].$$

A communication system such as the one being considered requires an initial investment for construction of the earth stations and an annual expense for lease of the satellite. The cost of an earth station is comprised of the costs of its components and of a series of fixed costs. The cost of an earth station, $(C_E)$, is computed as follows:

$$C_E = C_T + C_A + C_R + C_{EO}$$

where

$C_T$ = transmitter cost,

$C_A$ = antenna cost,

$C_R$ = receiver cost,

and

$C_{EO}$ = all design invariant costs

Note that here are concentrating only on the RF parameters and assuming all others as given.

The lease cost of the satellite per year is $C_L$, and is given by:

$$C_L = C_S \bullet \mathrm{MAX}\left(\frac{BW}{40\ \mathrm{MHZ}}, \frac{\text{P out of satellite}}{\text{satellite saturation EIRP}}\right)$$

where $C_S$ equals the cost of full usage of the satellite transponder and we assume it to be \$1.2 million per year. Also

$$\text{P out of satellite} = P_E \circ G_e^{u} \bullet \frac{1}{P_{lu}} \circ (GG_S^{w}),$$

$BW$ = bandwidth = $NR_O \bullet (MEF) \bullet (BWEF)$

MEF = modulation effeciency factor

BWEF = bandwidth efficiency factor = 1/ coding rate.

For an N year contract subject to a constant inflation rate and nominally constant leases the present value is given by: auth: note that $\sum_{n=1}^{N}\left(\frac{1}{1+i}\right)^n = \frac{1-\left(\frac{1}{1+i}\right)^N}{i}$

$$\text{Present Value} = C_E + C_L + \sum_{n=1}^{N} \frac{C_{Lease}}{(1+i)^n}$$

The annual cost of the system is that annuity which would result in a present value of investment equal to the system's present value. Annual cost is of interest because it effectively spreads the initial investment over the length of the contract and it is given by:

au: $\sum_{n=0}^{N}(1+i)^n = \frac{(1+i)^{N+1}-1}{i}$

$$\text{Annual Cost} = C_{Lease} + \sum_{n=0}^{N} \frac{C_E}{\left(\frac{1}{1+i}\right)^n}$$

Using the previous model we can analytically derive a minimal cost design, but the large number of parameters involved and the nature of the lease cost make such an analytical optimization difficult to obtain. If only general relationships are sought, though, it is possible to simplify the analysis by grouping satellite parameters, path losses, and market information into single expressions. Also, the difficulty posed by the dependence of lease cost on either power or bandwidth may be alleviated by considering the two cases individually. The analysis which follows will obtain expressions for the present value of the system for both such cases, and the implications of these expressions will be considered.

From before,

$$P_E = NR_o\left(\frac{E_b}{N_o}\right)_{av}\left[\frac{k\,(T_s P L_u G G_e^d + T_e P L_u P L_d)}{G G_s^u G_e^u G_e^d}\right]$$

since $G_e = \left(\frac{\pi}{\lambda} D\right)^2$ this may be simplified to be

$$P_e = NR_o\left(\frac{E_b}{N_o}\right)_{av}\left[\frac{C_1}{D^2} + \frac{C_2 T_e}{D^4}\right]$$

where $C_1$ and $C_2$ are defined appropriately.

When power dominates, lease cost is proportional to the power coming out of the satellite. Representing this by $C_{LP}$

$$C_L \propto P_e G_e^u = NR_o \left(\frac{E_b}{N_o}\right)_{av} \left[C_3 + \frac{C_4 T_e}{D^2}\right]$$

Omitting the cost of a low noise amplifier, which is negligibly small, the expressions for the present values of the system are given by:

$PV_P$ = present value when power dominates

= fixed costs + antenna cost + transmitter cost + lease cost

$$PV_P = (C_5) + (C_6 D) + NR_o \left(\frac{E_b}{N_o}\right)_{av} \left[\left(\frac{C_7}{D^2} + \frac{C_8 T}{D^4}\right) + \left(C_3 + \frac{C_4 T}{D^2}\right)\right]$$

Now in a similar fashion when bandwidth dominates.

$PV_{BW}$ = present value when bandwidth dominates

= fixed cost + antenna cost + transmitter cost + lease cost

$$PV_{BW} = (C_5) + C_6 D + NR_o \left(\frac{E_b}{N_o}\right)_{av} \left[\left(\frac{C_7}{D^2} + \frac{C_8 T}{D^4}\right)\right] + C_{LBW}$$

where $C_{LBW}$ equals lease cost in this case. Recall that it depends on bandwidth.

Solving for the antenna diameter at the intersection yields

$$D_{int} = \sqrt{\frac{C_4 T \left[NR_o \left(\frac{E_b}{N_o}\right)_{av}\right]}{C_{LBW} - C_3 \left[NR_o \left(\frac{E_b}{N_o}\right)_{av}\right]}}$$

The minimums of these are $D_p$ and $D_{BW}$, and equal

$$A \quad \cancel{D_p} = \left[NR_o\left(\frac{E_b}{N_o}\right)_{av}\right]\left[\frac{C_7+C_4T_e}{D_p} + \frac{3C_8T_e}{D_p^3}\right]$$

and

$$2 \quad \cancel{D_{BW}} = \left[NR_o\left(\frac{E_b}{N_o}\right)_{av}\right]\left[\frac{C_7}{D_{BW}} + \frac{3C_8T_e}{D_{BW}^3}\right]$$

Comparison of equations shows that the power limited minimum is less than the bandwidth limited minimum.

Assuming that, the diameter of intersect, $D_{int}$, exists, the following possibilities exist for $PV_p$ and $PV_{BW}$ as shown in Figs. 22 and 23.

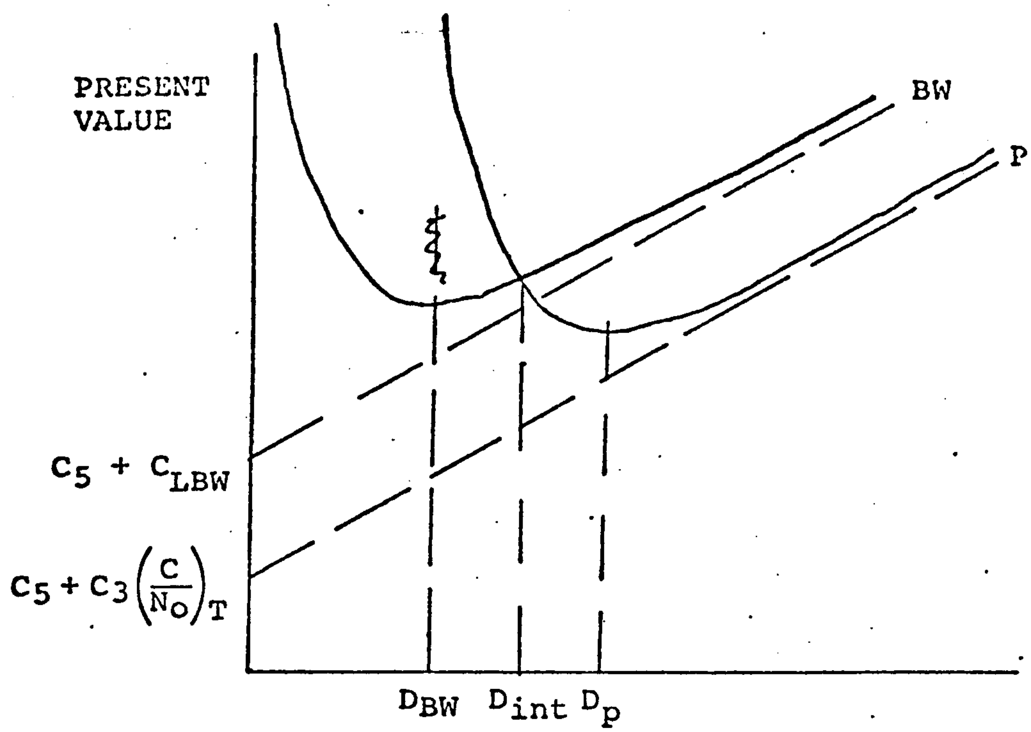



Figure 1

Possible present value curves for the cases of bandwidth and power dominance.

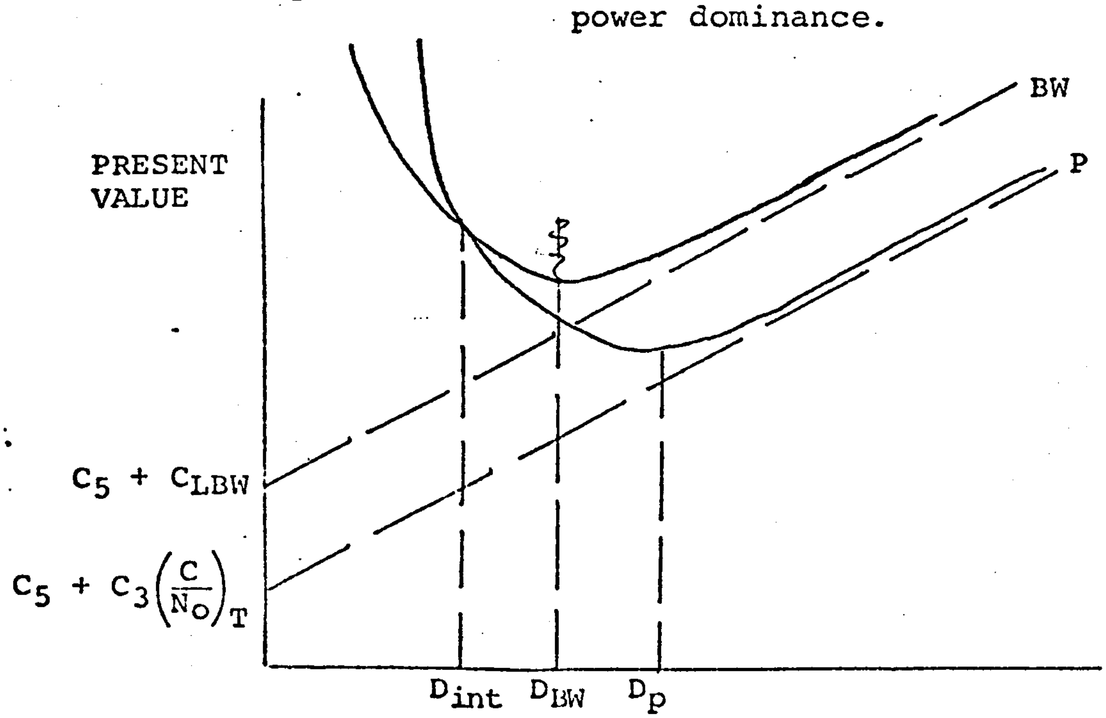



Figure 2

Present value curves for bandwidth and power dominance. Bandwidth dominates at optimal point.

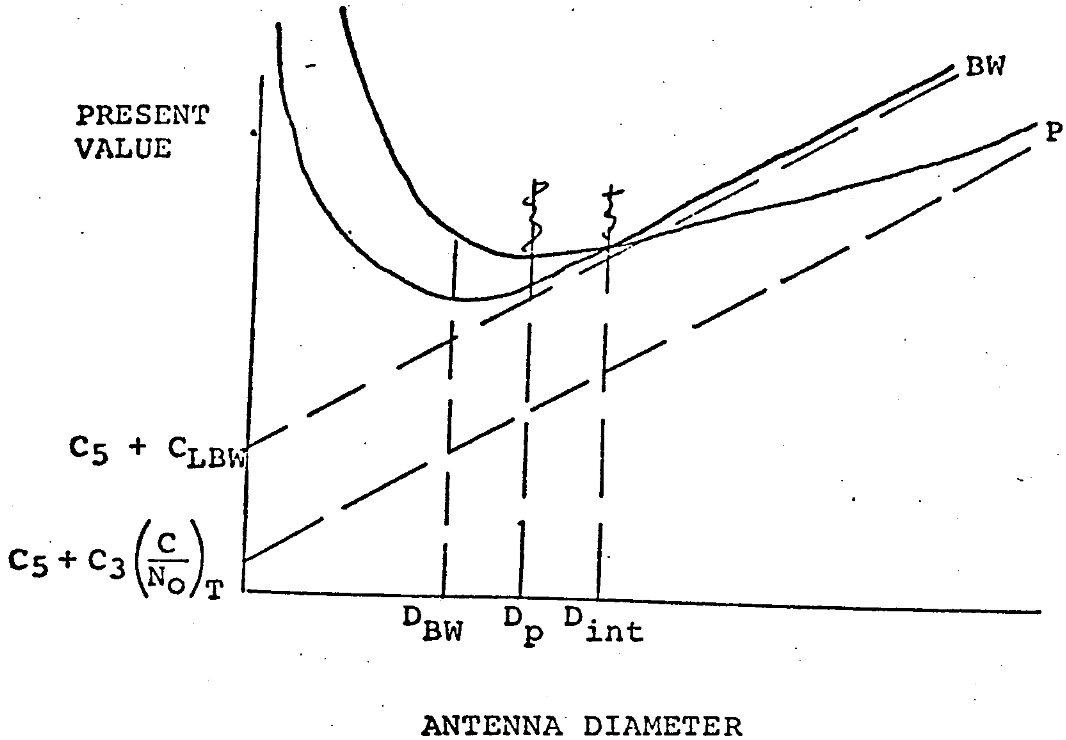



Figure 24 24

Present value curves for bandwidth and power dominance. Power dominates at optimal point.

The heavy sections of Figs. 3 and 4 mark the present value of the station. These diagrams show that shifting the point of inflection or the minimums may result in the use of different operating points. If the curves do not intersect, power must dominate throughout, and the problem of minimizing cost is equivalent to that of minimizing $NR_o\left(\frac{E_b}{N_o}\right)_{av}$.

Since for a fixed modem and rate, $NR_o\left(\frac{E_b}{N_o}\right)_{av}$ and $C_{LBW}$ are constant, an increase in temperature produces increases in present values and in the diameters at which their minimums occur. Thus, for a fixed modem and rate,

the graph in Fig. 25.

25

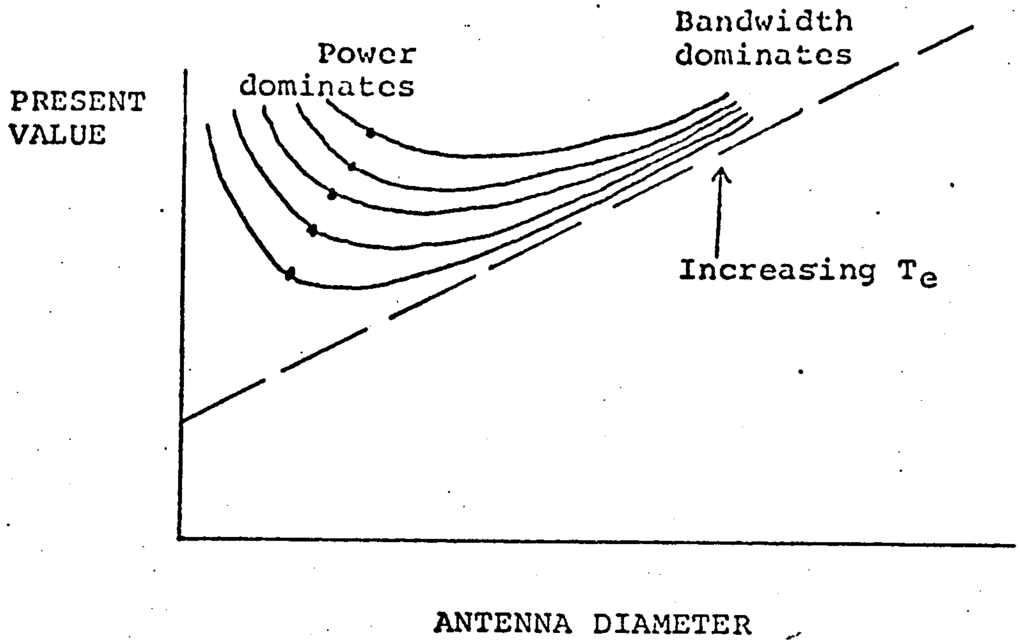


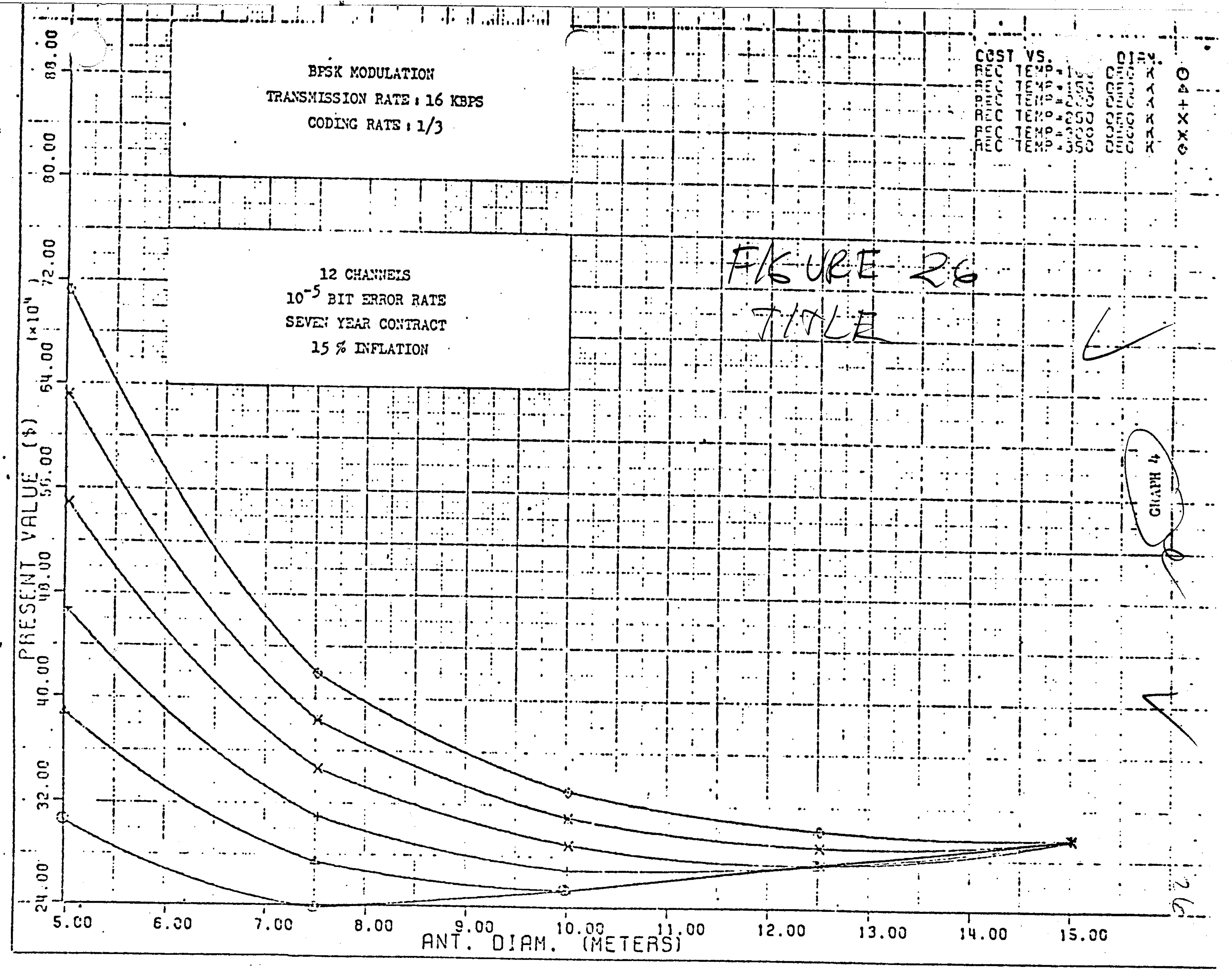

Figure 25 25

Present Value for a Fixed Modem and Rate

As diameter increases for a fixed temperature, required power decreases while bandwidth remains constant. This accounts for the shift from power to bandwidth represented by the heavy dots in Figure .. Fig Fig.26 shows the computer's version of this graph for a particular modem. The effect of higher costs for increasing temperatures is due to the linear dependence of PV on $T_e$. For a fixed modem and rate low temperatures are most economical.

The curves in Figure 26 are affected when either the modem or the transmission rate is changed. The effects on these curves of changing one factor at a time will be analyzed next.

Variation of the number of channels or of the transmission rate produces the same effect since both factors



BPSK MODULATION
TRANSMISSION RATE : 16 KBPS
CODING RATE : 1/3
12 CHANNELS
$10^{-5}$ BIT ERROR RATE
SEVEN YEAR CONTRACT
15 % INFLATION
COST VS. DIAM.
REC TEMP=250 DEG K
REC TEMP=350 DEG K
FIGURE 26
TITLE
GRAPH 4
26
PRESENT VALUE ($) ($\times 10^4$)
88.00
80.00
72.00
64.00
56.00
48.00
40.00
32.00
24.00
ANT. DIAM. (METERS)
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00

always appear together in the equations. If either is increased, the following events occur:

$D_{int}$ remains the same

$D_p$ and $D_{BW}$ both increase

$PV_p$ and $PV_{BW}$ increase for all diameters

Thus, higher transmission rates produce higher costs. Also, there is a tendency to shift from power dominance towards bandwidth dominance.

If modulation efficiency is decreased, then

- Bandwidth increases,
- $D_{int}$ increases,
- $D_p$ and $D_{BW}$ stay the same,
- $PV_p$ stays the same, and
- $PV_{BW}$ decreases.

The shift in $D_{int}$ that accompanies a decrease in modulation efficiency favors power over bandwidth in the lease cost . Decreasing the efficiency while power dominates will therefore not alter the cost, but doing so while bandwidth dominates lowers the cost until power dominance is reached. The similarity of BPSK and QPSK at low transmission rates, where power dominates, supports this, as does the shift from bandwidth dominance in BPSK to power dominance in QPSK when the coding rate is 1/3 and the transmission rate is 56kbps. Graphically, a decrease in efficiency corresponds to lowering the bandwidth present value curve while maintaining the power curve stationary.

Effects due to changes in required $\frac{E_b}{N_o}$ are not nearly as clear as those due to variations in rate or modulation efficiency. An increase in $\frac{E_b}{N_o}$ produces increases in $PV_p$, $PV_{BW}$, $D_p$, $D_{BW}$, and $D_{int}$. In other words, the curves in Figure are shifted upward and to the right. Accordingly, BFSK is always more expensive than BPSK since the only difference between the two lies in their required $\frac{E_b}{N_o}$'s, and QPSK is slightly more expensive than BPSK.

An increase in coding rate is accompanied by an increase in $\left(\frac{E_b}{N_o}\right)_{av}$ and a decrease in BW. Present value increases if power dominates. Such a statement may not be made for the case of bandwidth dominance.

Example: We are to consider designing an earth station that is a multichannel TDM system that communicates with an identical terminal. Tables 5 - 7 consist of a list of values used for satellite parameters, signal losses, gains.

## TABLE 5

### Assumptions

Satellite = INTELSAT IV-A

$T_s$ = 34.6 dB - $^oK$

$G_s$ = 16 dB

G = 140 dB

Satellite G/T = -18.6 dB/K

Satellite saturation flux - -80 dBW/$M^2$

Satellite saturation EIRP - 22 dBW

Input back off at satellite saturated EIRP - 5 dB W/$M^2$

Cost of full transponder - $1.2 million

Number of channels - 12 @ 56 Kbps per channel

Losses (in dB):

Processing - 2 dB

| | Uplink | Downlink |
|---|---|---|
| Feed | 1 | .5 |
| Radome | 0 | 0 |
| Multipath | .1 | .1 |
| Atmosphere | .7 | .7 |
| Scintillation | 0 | 0 |
| Pointing | 0 | .2 |
| Path | 200 | 196.5 |
| Polarization | - | .1 |
| Antenna | - | .5 |

6

TABLE ~~II~~

Required $\frac{E_b}{N_o}$

| BER | Modulation BPSK | QPSK | BFSK |
|---|---|---|---|
| $10^{-3}$ | 6.7 | 7.0 | 9.7 |
| $10^{-5}$ | 9.6 | 9.9 | 12.6 |
| $10^{-7}$ | 11.3 | 11.6 | 14.3 |

Why is $\left(\frac{E_b}{N_o}\right)_{QPSK} > \left(\frac{E_b}{N_o}\right)_{BPSK}$ ?

TABLE ~~III~~ 7

Coding Gain

Convolutional codes – Viterbi decoding ?

Coding Rate

| BER | 1/3 | 1/2 | 3/4 | 7/8 |
|---|---|---|---|---|
| $10^{-3}$ | 4.2 | 3.8 | 3.1 | 1.4 |
| $10^{-5}$ | 5.7 | 5.1 | 4.6 | 3.0 |
| $10^{-7}$ | 6.2 | 5.8 | 5.2 | 3.6 |

(1) What codes are used?
(2) What constraint length?

Figs. 27 to 31 depict various tradeoffs with modulation and a summary design.

The results from the previous section along with an analysis of the equations that generated them reveal the following set of conclusions with respect to the design of efficient earth stations:

(1) the lowest available temperature should always be used

(2) BFSK is always more expensive than BPSK or QPSK

(3) the lowest acceptable transmission rate should be used

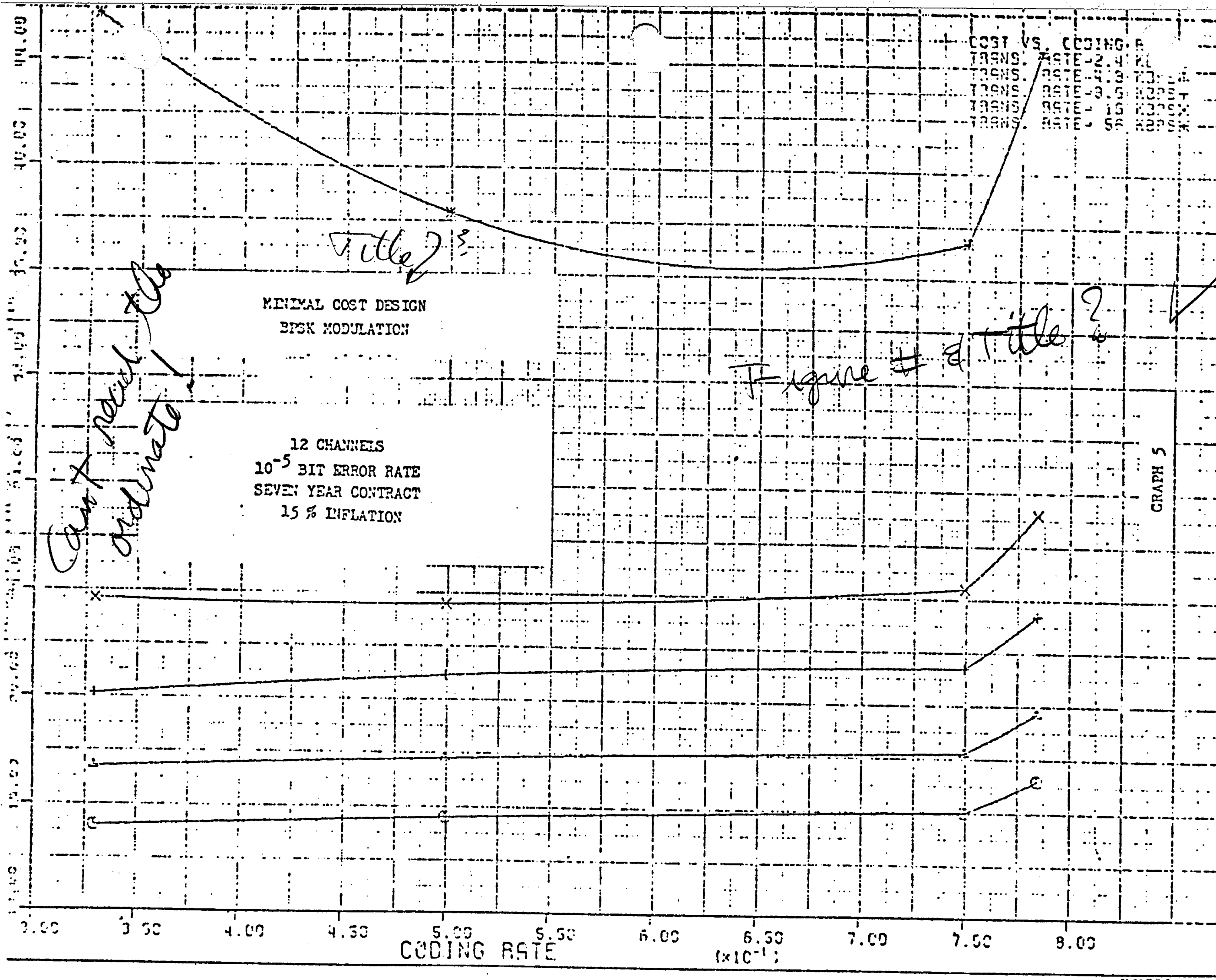
(4) low coding rates should be used while power dominates in the lease cost

(5) high coding rates should be used while bandwidth dominates in the lease cost

(6) BPSK should be used over QPSK if the lease costs for both modulation schemes are entirely dominated by either power or bandwidth over the range for coding rate under consideration. If a transition occurs, an individual case study must be conducted.

(7) Design is not significantly affected by inflation.

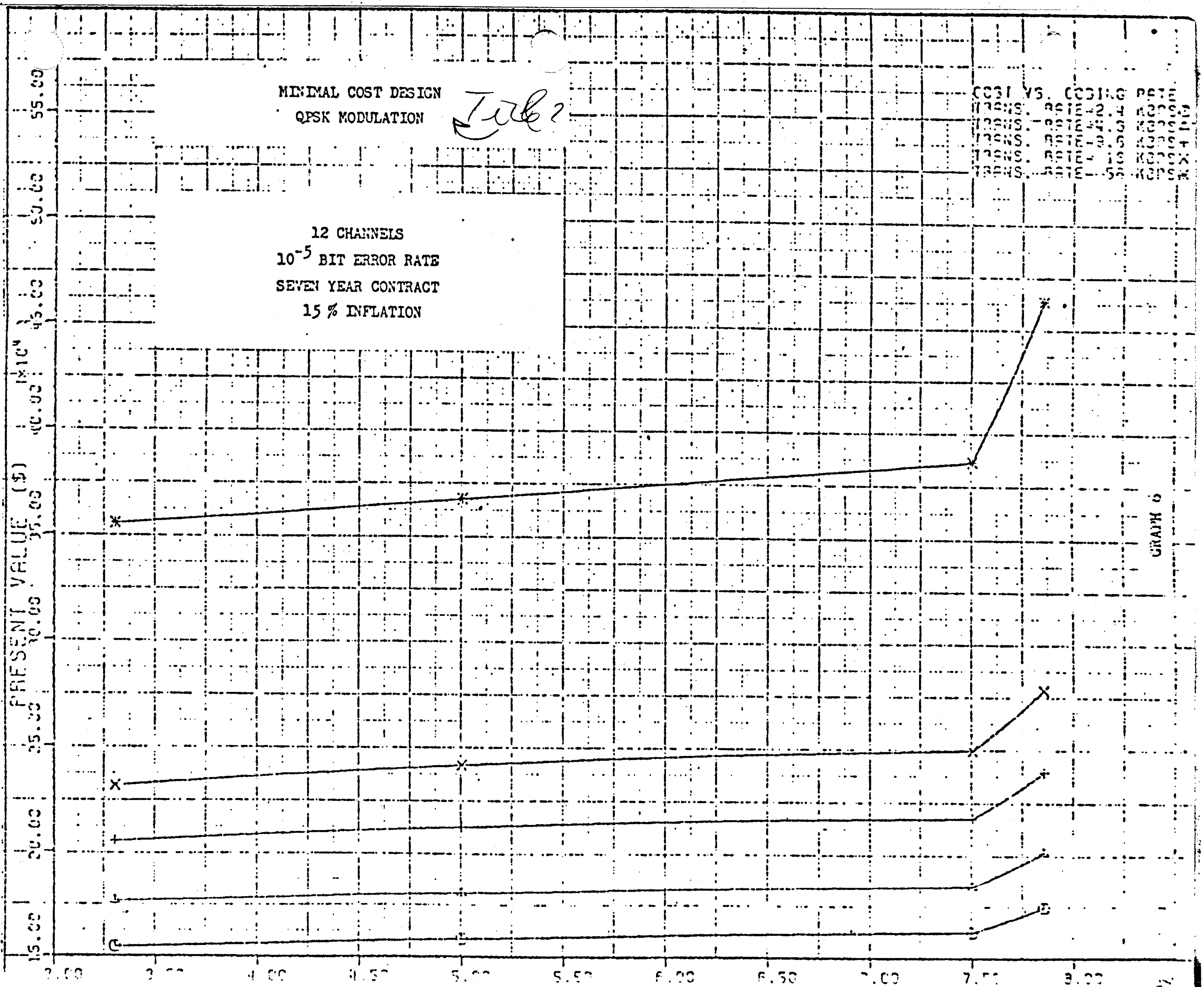
The analysis is subject to certain limitations. From an engineering viewpoint, it is unrealistic to design stations with no back-up systems. The introduction of this factor into the model would certainly raise cost estimates.



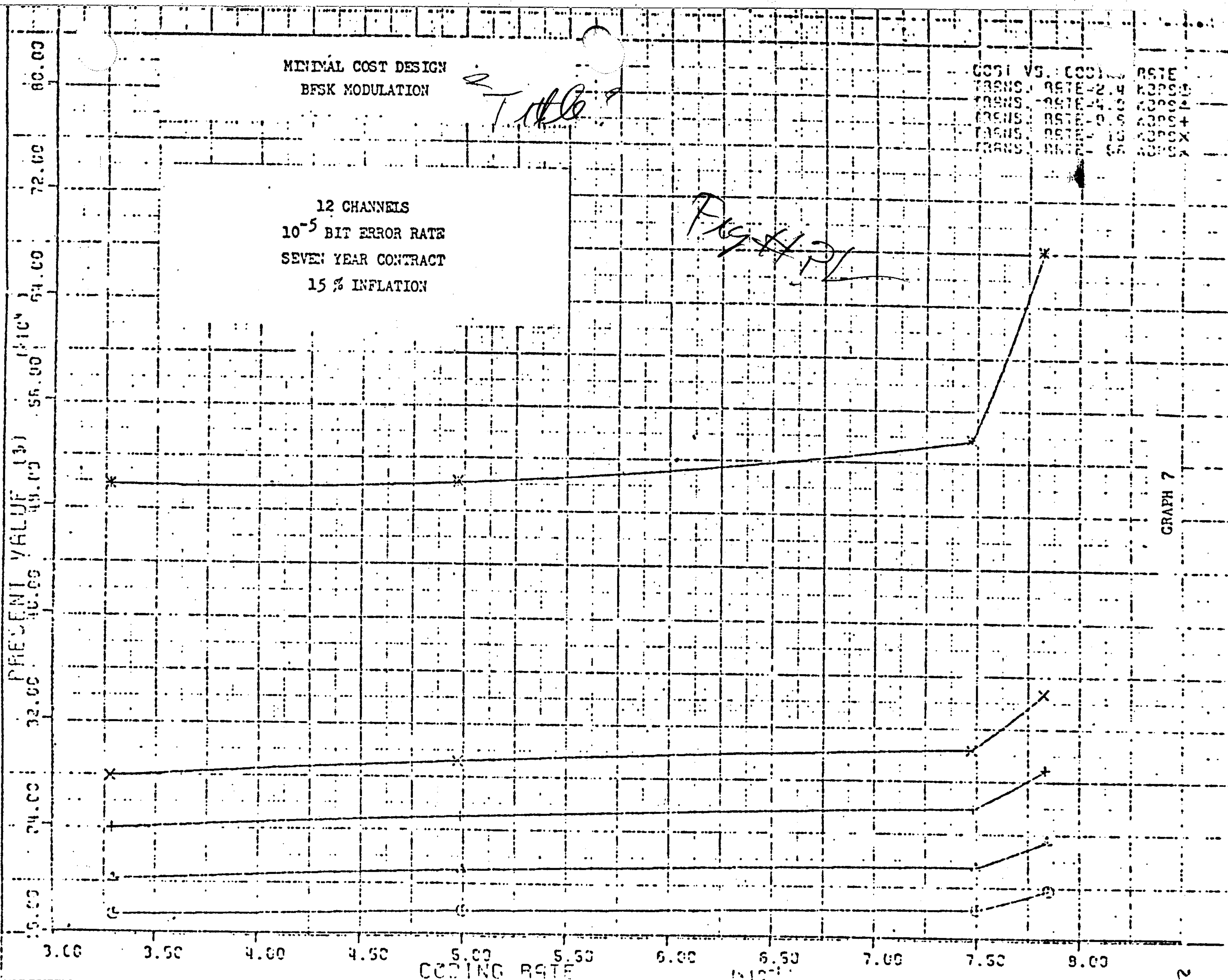

Title?

Figure # & Title?

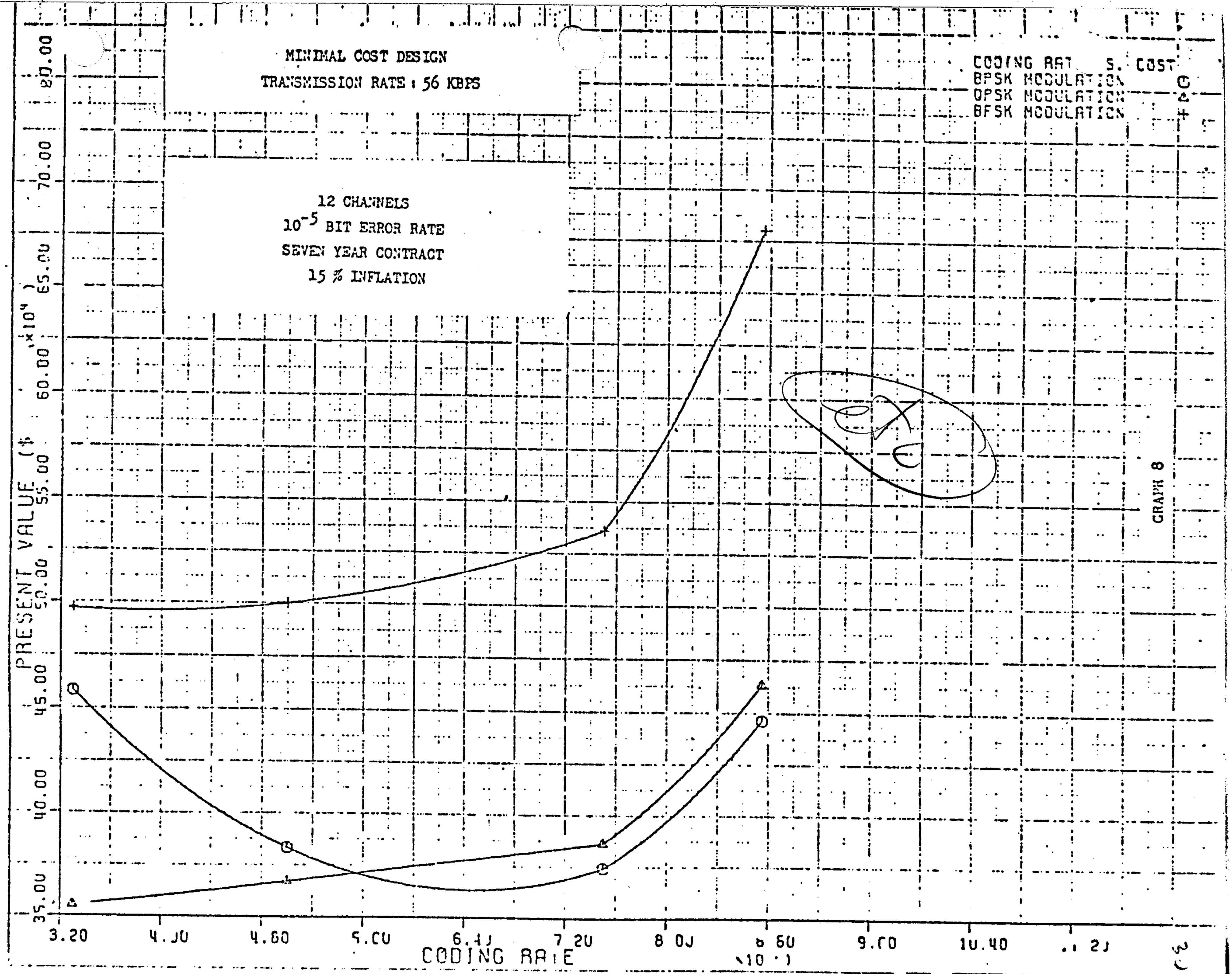
Can't read the ordinate!

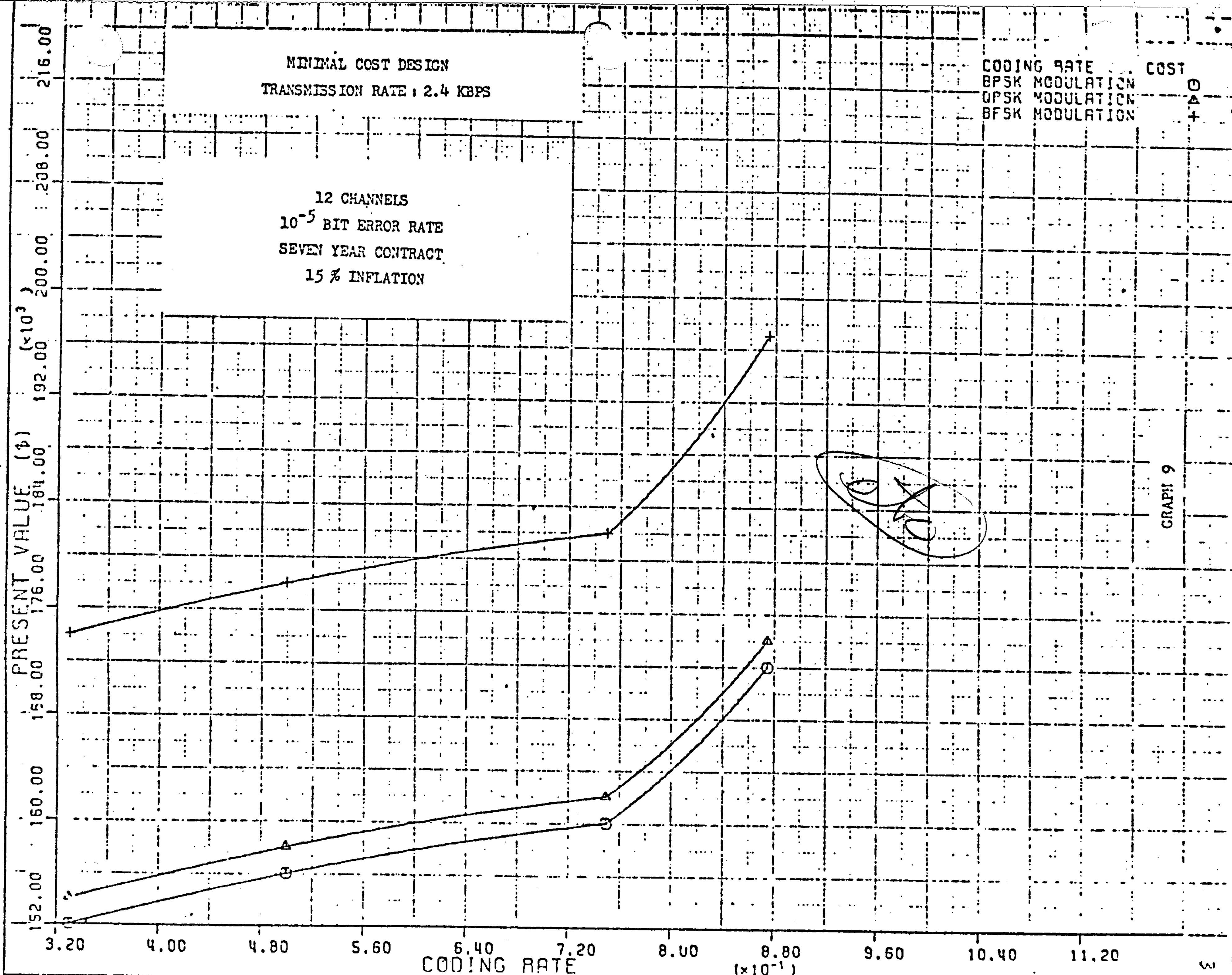


MINIMAL COST DESIGN
QPSK MODULATION
12 CHANNELS
$10^{-5}$ BIT ERROR RATE
SEVEN YEAR CONTRACT
15 % INFLATION
COST VS. CODING RATE
GRAPH 6
PRESENT VALUE ($)
15.00
20.00
25.00
30.00
35.00
40.00
45.00
50.00
55.00
5.00
6.00
6.50
7.00
8.00


MINIMAL COST DESIGN
BFSK MODULATION
12 CHANNELS
$10^{-5}$ BIT ERROR RATE
SEVEN YEAR CONTRACT
15 % INFLATION
COST VS. CODING RATE
PRESENT VALUE ($)
CODING RATE
3.00
3.50
4.00
4.50
5.00
5.50
6.00
6.50
7.00
7.50
8.00
16.00
24.00
32.00
40.00
48.00
56.00
64.00
72.00
80.00
Title
Fig #21

GRAPH 7


MINIMAL COST DESIGN
TRANSMISSION RATE : 56 KBPS
12 CHANNELS
$10^{-5}$ BIT ERROR RATE
SEVEN YEAR CONTRACT
15 % INFLATION
CODING RAT S. COST
BPSK MODULATION
QPSK MODULATION
BFSK MODULATION
GRAPH 8
PRESENT VALUE ($) (x10⁴)
35.00
40.00
45.00
50.00
55.00
60.00
65.00
70.00
80.00
CODING RATE (x10⁻¹)
3.20
4.00
4.60
5.00
6.40
7.20
8.00
8.60
9.00
10.40


MINIMAL COST DESIGN
TRANSMISSION RATE : 2.4 KBPS
12 CHANNELS
$10^{-5}$ BIT ERROR RATE
SEVEN YEAR CONTRACT
15 % INFLATION
CODING RATE . . COST
BPSK MODULATION
QPSK MODULATION
BFSK MODULATION
GRAPH 9
PRESENT VALUE ($) ($\times 10^3$)
152.00
160.00
168.00
176.00
184.00
192.00
200.00
208.00
216.00
CODING RATE ($\times 10^{-1}$)
3.20
4.00
4.80
5.60
6.40
7.20
8.00
8.80
9.60
10.40
11.20

3/

### 3.6 INVESTMENT ANALYSIS

In the previous two sections we considered techniques to design such systems or just earth segment portions. In both cases we concentrated on the cost of the service and not on the profitability. In this section we develop the structure for the analysis of this profitability. We must introduce the concept of a market and revenue flow and combine that with the expense flow to get a total cash flow profile. It is through an analysis of this type that any new systems can and must be judged. The approach taken in this section is to perform the analysis via an example.

We are considering the design of a satellite system to provide service to a community of mobile users. The sets of users have been divided into six sections.

1. Land Mobile - Trucking
2. Land Mobile - Police
3. Land Mobile - Telephony
4. Maritime - Inland and Coastal
5. Commercial Aviation
6. General Av~~a~~iation

The communication needs of each group are closed, meaning they do not wish to communicate with other groups.

The basic components of the system are the earth segment, consisting of mobile terminals and base stations; and the space segment, consisting of one operational satellite and a spare. There will be one mobile terminal for each user with the capability of accessing any fixed base station in the system via the satellite. The number of base stations will be determined according to the number of users. The service provided will be interconnection for digitized voice traffic (@ 32 Kbps per voice channel) between mobiles and base stations, but no interconnection between mobiles. The system will operate over CONU~~C~~S (Continental U.S.), providing a basic coverage constraint.

Three scenarios are examined;

i. High
ii. Medium
iii. Low

In any analysis we try to define what we expect (medium), what is best (high), and what is worst (low). This gives a sensitivity analysis to the overall result. They are defined by the number of users and total traffic capacity.

The lifetime of the system is 7 years. The launch year is chosen as 1984, and focus of this study runs from 1982 to 1991.

### 3.6.1 Mathematical Model

This model has two major subdivisions which shall presently be discussed. The model herein used has been developed to systematically analyze cashflows (including costs, revenues, and taxes) relative to an envisioned mobile satellite system.

#### Market Model

The market model calculates the size of the market in year K for each scenario in terms of number of units in use in each year. These values are the basis for the positive cash flow or revenue flow calculations. Six markets are considered for each of 3 scenarios, as defined previously. For this model we let K=0 to 7, corresponding to the years 1984-1991 and let i-1 to N, where N represents the Markets. We now need the following quantities to quantify the market size.

$N_i(k)$ = total number of units in market i in year k

$E_i(k)$ = total erlang load in market i in year k

$e_i(k)$ = erlang load per user in market i in year k

$g_i$ = annual growth rate of market i.

$T_i$ = turnover time of equipment in market i.

$P_i$ = percent market penetration for market i.

Note that in the above the erlang load represents the fraction of time a single user desires to use the channel for voice traffic.

Assume that the initial number of units, $N_1(0)$, to be given for baseline 1984. Then $N_i(k)$ can be found by

$$N_i(k) = N_i(0)\ (1+g_i)^k$$

This represents the total market potential in available units. The number of units obtained as customer base for capture is given by $N^1{}_i(k) = P_iN_i(k)$ ? Why must we?

Now the potential market number, $N_i^1(k)$, can be captured eventually in total; however, we must try to model a slow penetration to simulate market lag. This lag results from such factors as existing service to some customers, lack of marketing force, unavailability of hardware. Thus the actual number of users in year k is $X_i(k)$ for market i. Now we ask how do we reflect what is available, $N_i^1$ (k), of what we get. This is given by saying that in year k we get some of what is available in k, some of what is in (k-1), and some all the way back to 0.
We can represent this concept by: ok?

Let $X_i(k)$ be the number of users in market i in year ?. Then

$$x_i(k) = \sum_{j=o}^{k} h(k-j)N_i^1(j)$$

where h(k-j) is a filter selected to provide market lag. Three models will be considered:

1. $h_1(k) = \begin{cases} 1/T_i & 0 \leq k \leq (T_i-1) \\ 0 & \text{elsewhere} \end{cases}$

2. $h_2(k) = \begin{cases} \alpha a_i^k & 0 \leq k \leq \infty \\ 0 & \text{elsewhere} \end{cases}$

3. $h_3(k) = \begin{cases} \beta k b_i^k & 0 \leq k \leq \infty \\ 0 & \text{elsewhere} \end{cases}$

where

$$a_i = \exp(-\gamma/T_i) \Rightarrow b_i \text{ ?}$$

$$b_i = \exp(-\gamma/T_i)$$

and $\gamma/T_i$ is the growth constant for market i. The particular market lag for each market is chosen according to the growth type, L, where L=1,2,3.

An additional constraint on market lag must be considered.

$$\sum_{j=0}^{\infty} h(j)=1$$

which indicates that we eventually get all of $N_i(k)$.

A cumulative total of users in all 6 markets in year k is given by

$$X(k) = \sum_{i=1}^{N} x_i(k)$$

These totals are used to calculate positive cash flow in the financial analysis by multiplying by the usage and tariff values.

The market model has the potential to calculate total erlang load for each market in year k according to the formula

? Ref & definition ?

$$E_i(k) = e_i(k) x_i(k)$$

These values are utilized in the mobile satellite system design of Section 3.4 rather than in the economic analysis presented in this section.

The necessary constants for each market are presented in Table 8, and the data for each scenario is given in Table 9.

| MARKET | $N_i(0)$ NUMBER OF UNITS IN 1984 | $e_i$ (0) ERLANG LOAD PER USER 1984 | $g_i$ GROWTH RATE | $T_i$ TURNOVER TIME (Years) |
|---|---|---|---|---|
| POLICE | 106,000 | .00142 | 10% | 10 |
| TRUCK | 42,100 | .00416 | 8% | 7 |
| TELEPHONE | 3,660,000 | .0205 | 10% | 10 |
| MARITIME | 366,000 | .0050 | 10% | 10 |
| GEN. AV. | 38,000 | .00526 | 6% | 10 |
| COM. AV. | 3,680 | .1111 | 3% | 10 |

TABLE 8 8 : MARKET CHARACTERISTICS

| SCENARIO / MARKET | HIGH | | | MEDIUM | | | LOW | | |
|---|---|---|---|---|---|---|---|---|---|
| | $P_i$ MARKET CAPACITY | L GROWTH TYPE | $\gamma/T_i$ GROWTH CONSTANT | $P_i$ MARKET CAPACITY | L GROWTH TYPE | $\gamma/T_i$ GROWTH CONSTANT | $P_i$ MARKET CAPACITY | L GROWTH TYPE | $\gamma/T_i$ GROWTH CONSTANT |
| POLICE | 50% | 3 | 2/T | 50% | 3 | 2/T | 25% | 3 | 1/T |
| TRUCK | 100% | 2 | 2/T | 70% | 2 | 2/T | 30% | 2 | 1/T |
| TELEPHONE | 1% | 2 | 2/T | 0.5% | 3 | 1/T | 0% | -- | -- |
| MARITIME | 40% | 3 | 1/T | 25% | 3 | 1/T | 10% | 3 | .5/T |
| GEN. AV. | 50% | 2 | 2/T | 25% | 3 | 1/T | 10% | 3 | .5/T |
| COM. AV. | 75% | 2 | 2/T | 50% | 2 | 1/T | 25% | 3 | 1/T |

TABLE 9. 9 : SCENARIO GROWTH DATA

## Financial Model

The financial model is based on a 7-year system lifetime, beginning in 1984 and terminating in 1991. The launch year is designated as 1984. There will be a two-year lead time for all payments on equipment, so the analysis will run from 1982 to 1991. All calculations are done on a yearly basis, although quarterly computation is an acceptable alternative.

In order to standardize notation, let k=0...9 for years 1982-1991 for the financial model.

Then

$$x'(k) = \begin{cases} 0 & \text{for } k=0,1 \\ x(k-2) & \text{for } k=2...9 \end{cases}$$

Thus, $x'(k)$ represents the number of users in year k, expanded to include 1982 and 1983.

The satellite cost for each scenario has been determined based on the minimum cost optimization discussed previously, and the results are as follows:

i. High - $100M

ii. Medium - $50M

iii. Low - $30M

These values include launch costs. The first satellite will be launched in early 1984, and a spare launched 6 months later. There is a 2-year lead time for payment. Let

$C_s$ = cost of the satellite

$S_j(k)$ = cash paid for satellite j in year k, j=1,2

$CF_s(k)$ = total cash flow due to satellites in year k

Then $CF_s(k) = \sum_{j=1}^{2} S_j(k)$

where

$$S_1(k) = \begin{cases} C_s/2 & k=0,1 \\ 0 & \text{elsewhere} \end{cases}$$

$$S_2(k) = \begin{cases} C_s/4 & k=0,2 \\ C_s/2 & k=1 \\ 0 & \text{elsewhere} \end{cases}$$

Let us now consider the cost of a large base station. We again use the optimum technique. The optimum base station cost obtained from previous calculations is \$50,000. There will be one base station for every 200 users. The 2-year lead time on equipment can be interpreted to mean that equipment becoming operational in year k will be paid for in years k-2 and k-1, or equivalently, that half the cost of new equipment in year k+1, and half the cost for year k+2, will be paid in year k.

Let

$C_B$ = cost of base station

$U_B(k)$ = new base stations becoming operational in year k

$CF_B(k)$ = total cash flow for base stations in year k

Then

$$U_B(k) = \begin{cases} 0 & k=0,1,\ k>9 \\ \left\lceil \dfrac{x'(k)-x'(k-1)}{200} \right\rceil & 2 \le k \le 9 \end{cases}$$

where [x] is least integer greater than x.

Then

PRIME!

Where does this come from?

$$CF_B(k) = C_B \frac{U_B(K+2)}{2} + C_B \frac{U_B(k+1)}{2}$$

The optimum mobile terminal cost, as determined previously, is \$3,000. There will be one mobile terminal for each user, with a 2-year lead payment time. Let

$C_M$ = cost of mobile terminal

$U_M(k)$ = new mobile units becoming operational in year k

$CF_M(k)$ = cash flow for mobile terminals in year k

Then

$$U_M(k) = \begin{cases} 0 & k=0,1,k>9 \\ x'(k)-x'(k-1) & 2 \leq k \leq 9 \end{cases}$$

and

$$CF_M(k) = \frac{C_M U_M(k+2)}{2} + \frac{C_M U_M(k+1)}{2}$$

?

Operations and maintenance costs are calculated for the total number of mobile units and base stations in service in year k. The cost is estimated at 5% of the total value of equipment. Let

$E_q(k)$ = value of equipment entering the system in year k

$$E_q(k) = U_B(k) + U_M(k)$$

Then if

$$P_{om} = \text{percent cost for operations and maintenance}$$

$$O_m(k) = \begin{cases} 0 & k=0,1 \\ P_{om}E_q(k) + O_mU(k-1) & k>2 \end{cases}$$

where $O_m(k)$ is the total amount spent on operations and maintenance for all working equipment in year k.

Engineering is considered to be a fixed nonrecurring cost, paid in full the first year, 1982. The value of this cost is \$30,000,000, and it will be denoted by $E_n(k)$ where

$$E_n(k) = \begin{cases} 30,000,000 & \text{if } k=0 \\ 0 & \text{elsewhere} \end{cases}$$

Marketing costs are computed according to number of people required to market the product. During the years before the system becomes operational, the marketing costs should be 20% of the total required for 1991. After the launch, these costs should level off, according to the number of users. There should be one marketing representative for every 1000 users.

Let

$R_m$ = marketing rate, one person for every 1000 users (.001)

$C_{MK}$ = cost per year for one marketing representative

where $C_{MK}$ = \$75,000

Let $P_o$ = initial percentage of total 1991 market determining the number of market representatives in 1982, 1983

where $P_o$ = 20%

Define $CF_{MK}(k)$ = cash flow in year k due to marketing costs. Then

$$CF_{MK} = \begin{cases} P_O C_{MK}[x'(9)\ R_M] & k=0,1 \\ P_O C_{MK}[x'(9)\ R_M] + C_{MK}\ [x'(k) R_M] & k=2..9 \end{cases}$$

where [x] is the smallest integer greater than x.

As we install equipment in the system, it is considered all to be capitalized and to be placed as an asset. Capital assets are not expensed when they are incurred because it is recognized that they have a longer lifetime than one year. What we then do is to extend the cost over several years and expense an amount equal to a loss of usefulness. This amount is called the depreciation of the asset. Thus on a cash flow basis the asset is only expensed on a fractional basis. There are many techniques to calculate depreciation. One such method is called straight line depreciation. In this case the expense for a year is the total cost divided by the number of useful years. The total depreciation in any year equals depreciation of equipment in that year plus all past depreciations that are allowed in that year.

Depreciation is calculated for the equipment in service in year k. This equipment includes the satellites, base stations, and mobile terminals. Let

$E_q^{\ell}(k)$ = value of all equipment, including satellites, entering the system

Then

$$E_q^{\ell}(k) = \begin{cases} 0 & k=0,1 \\ 2C_s + E_q(k) & k=2 \\ E_q(k) & k=3...9 \end{cases}$$

Ref ?

Depreciation in year k is calculated by

$$D(k) = .2E_q^{\ell}(k) + .18E_q^{\ell}(k-1) + .16E_q^{\ell}(k-2) \ldots$$

or equivalently $D(k) = \sum_{j=0}^{k} (.2-.02j) E_q^{\ell}(k-j)$

An investment tax credit is given for the new equipment becoming operational in year k. It has no direct relationship to the cash flow analysis presented here, but is important to the corporate tax picture as a whole.

Let

$$I_{tc}(k) = \text{investment tax credit in year } k$$

The percentage credit differs depending on the value of new equipment in a given year. The first $25,000 spent receives 10% credit, and anything beyond that amount receives 5% tax credit.

$$I_{tc}(k) = \begin{cases} .1E_q^{\ell}(k) & ; E_q^{\ell}(k) \leq 25{,}000 \\ 2500 + .05(E_q^{\ell}(k) - 25{,}000) & ; E_q^{\ell}(k) > 25{,}000 \end{cases}$$

The inflation rate is chosen to be a constant rate of 6%. Although this analysis begins in 1982, both positive and negative cash flow must be inflated from the current year

Let

$\Delta$ = number of years from current year to start year

Define $P_{in}$ = percent inflation rate = .06

$I(k)$ = inflation factor in year k

Then $I(k) = (1+P_{in})^{(k+\Delta)}$

The discount rate, or cost of money, is a constant value throughout the lifetime of the system. Several values in the range of 14%-18% will be considered. As before, money must be discounted from the current year. Let be as defined above, and

$$P_{cm} \quad = \quad \text{percent discount}$$

Then the cost of money factor, m(k), can be given by:

$$M(k) \quad = \quad (1+P_{cm})^{-(k+\Delta)} \qquad (1+P_{cm})^{-(k+\Delta)}$$

We are now in a position to calculate positive cash flow and negative cash flow.

Positive cash flow is based on the charge per mobile minute and the number of users in the system. The optimum charge per mobile minute has been previously determined as $3. Assuming that each user has 6 minutes of transmission per day, 75 days per quarter, a yearly charge per user can be calculated. Let

$$C_u \quad = \quad \text{charge per user per year} = \$5{,}400$$

and

$$P_{CF}(k) \quad = \quad \text{discounted positive cash flow in year k}$$

Then the discounted positive cash flow $P_{CF}(k)$ can be given by:

$$P_{CF}(k) \quad = \quad C_u\, x'(k)\ m(k)I(k)$$

where M(k) is the cost of money factor, I(k) is the inflation factor, and $x'(k)$ is the number of users in year k.

Negative cash flow is the cumulative sum of all costs in a given year. It is affected by inflation and cost of money.

Let

$$N_{CF}(k) \quad = \text{discounted negative cash flow in year k}$$

Then

$$N_{CF}(k) \quad = \quad [CF_S(k)+CF_B(k)+CF_M(k)+O_m(k)+E_n(k) + CF_{MK}(k)]\ m(k)I(k)$$

where $CF_s(k)$ is the cash flow due to the space segment, $CF_B(k)$ is cash flow due to base stations, $CF_M(k)$ the mobile terminals, $O_m(k)$ the operating costs, $E_n(k)$ the engineering costs, and $CF_{MK}(k)$ the marketing costs, all in a given year k. The positive cash flow can equivalently be called pre-tax revenue. Then define taxable revenue to be

$$R_T(k) = P_{CF}(k) - D(k)$$

where D(k) is the discounted depreciation of equipment in year k. The tax on this amount is given by

$$T(k) = \begin{cases} 0 & \text{if } R_T(k) \leq 0 \\ .2R_T(k) & \text{if } 0 < R_T(k) \leq 25{,}000 \\ 5{,}000 + .22\,(R_T(k) - 25{,}000) & \text{if } 25{,}000 < R_T(k) \leq 50{,}000 \\ 5{,}000 + 5{,}500 + .48\,(R_T(k) - 50{,}000) & \text{if } R_T(k) > 50{,}000 \end{cases}$$

where the first $25,000 of taxable revenue is taxed 20%, the second $25,000 is taxed 22%, and the remainder is taxed 48%.

Revenue is the total positive cash flow after taxes have been paid. Define

$$R(k) = \text{post tax revenue in year } k$$

Then

$$R(k) = P_{CF}(k) - T(k)$$

where $P_{CF}(k)$ is the positive cash flow, and T(k) is the tax in year k.

The total cash flow is this revenue less the previously computed negative cash flow.

$$T_{CF}(k) = R(k) - N_{CF}(k)$$

where $T_{CF}(k)$ is the total cash flow, the desired end result to be used in the financial analysis.

### 3.6.2 Financial Analysis

An understanding of the cost of money concept is essential for the analysis of the results. Concerning project finance, the cost of money concept has direct relevance in two areas:

- the cost of money (which could be conceptualized as 'cost of funding') used to finance the project should be viewed as a weighted average of the various types of funds a corporation uses (i.e., debt, stock and equity). There can be one figure used by management to represent this cost of money, or a range.
- the committal of these funds means, obviously, that they will be unavailable for other projects. Foregone opportunities represent an implicit cost[4] to the corporation in the project finance process.

In this regard, by incorporating the cost of money concept into the mathematical model, it can be treated as a cost, just like any other, to the corporation. The use of this concept in this manner contributes toward a more realistic picture of the project.

Hypothetically, if the funding costs are 14%, and the return on the project (after-tax profits) is 14%, the management would not pursue the investment. All other things being equal in this example the wider the gap between 14% and the envisioned rate of return on an investment option (e.g., 14%), the more desirable that option will be.

---

[4] Explicit costs relative to the envisioned project are defined as the satellite segment cost, base station costs, marketing costs, etc.

As noted previously, inflation was included in the model at a rate of 6% p.a. While this figure does not necessarily reflect the current rate of inflation, from a historical point of view it would be judged as relatively high. Consequently, there is a somewhat negative bias regarding the positive cash flows since the costs associated with the envisioned mobile system have been inflated at a higher than average rate. This is consistent, however, with the model's philosophy which is conservative. In other words, by raising component costs in this matter, the date and extent of the breakeven point is delayed. The unpredictability of inflation coupled with its erosive tendancies regarding a project's profitability warrants such a valuation of inflation since it increases the elements of risk which is inherent in any project.

In order to discuss depreciation as it relates to the model, it must be remarked that any depreciation charge has the effect of reducing the total cashflow (e.g., revenues subject to taxation) by a certain amount; in this matter the value of a corporation's tax liability is reduced. Thus, the pre-tax revenue of the project less depreciation yields taxable revenue. Further, this amount less tax yields the final revenue from which the "return on investment" graphs are computed.

There are various types of depreciation methods to adopt. The method (sum-of-years' digits), which was chosen herein, allows for a variable depreciation rate, e.g., a larger initial rate and deductions in the early years of use, and smaller later. For equipment with a life of three years or greater, this chosen method provides the largest deduction in the initial years. Such a depreciation method was adopted to mitigate, albeit partially, the relatively high start-up costs associated with the system.

With regard to tax, its determination was based on standard IRS procedure. The principle behind the inclusion of the "tax" concept in the mathematical model was to view the project as revenue earned independent of other corporate endeavors. Since tax on earnings is charged on the total revenues of a corporation (less any appropriate deductions), the assumption herein is that this project's revenues will contribute to the overall liability of the Corporation in an amount equal to the percentage tax rates as stipulated by IRS. The tax is derived from IRS corporate tax schedules.

Lastly, the investment tax credit will be discussed. This credit is allowed for investment generally in machinery and equipment. The credit results in the abatement of all or part of a corporation's income tax liability for the year. The credit is approximately 10% of the value of qualified property placed in service during the relevant year. Qualified property is defined as having a life (in terms of usage) of 7 years or more. Relevant to the analysis of the envisioned mobile system, the space segment, the mobile terminal and the base stations (equipment, not building) are interpreted to be qualified property. The inclusion of the investment tax credit is designed to indicate to the reader the amount by which this specific project will reduce the corporation's tax liability. While this in no matter relates to the cashflow *per se*, it is felt that its positive impact relative to the total corporate picture is of a large enough degree to warrant mentioning.

While various results could be generated, they are of a specific nature and relate only to the particular cost of money and scenario involved. Consequently, attention will herein be directed only toward those general trends which appear in the model. In this regard the following should be noted:

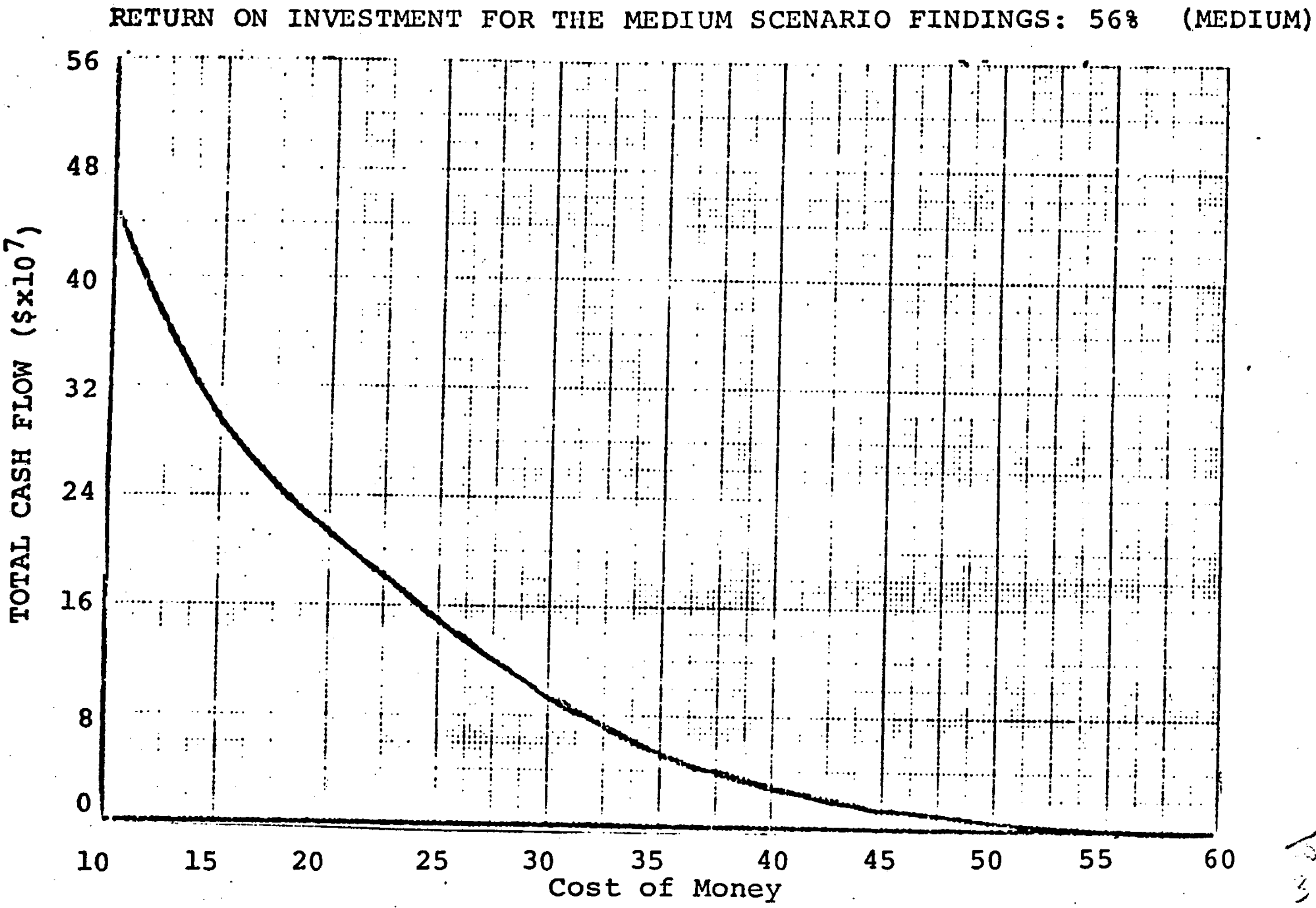
1) Irrespective of the cost of money, the low, medium, and high scenarios break-even in 1990, 1988, and 1987 respectively.

2) The degree of profitability of the system varied significantly as a result of the interaction of the cost of money and the usage (as defined in terms of the three scenarios) variables. Therefore, proper long range planning must seriously take account of these variables.

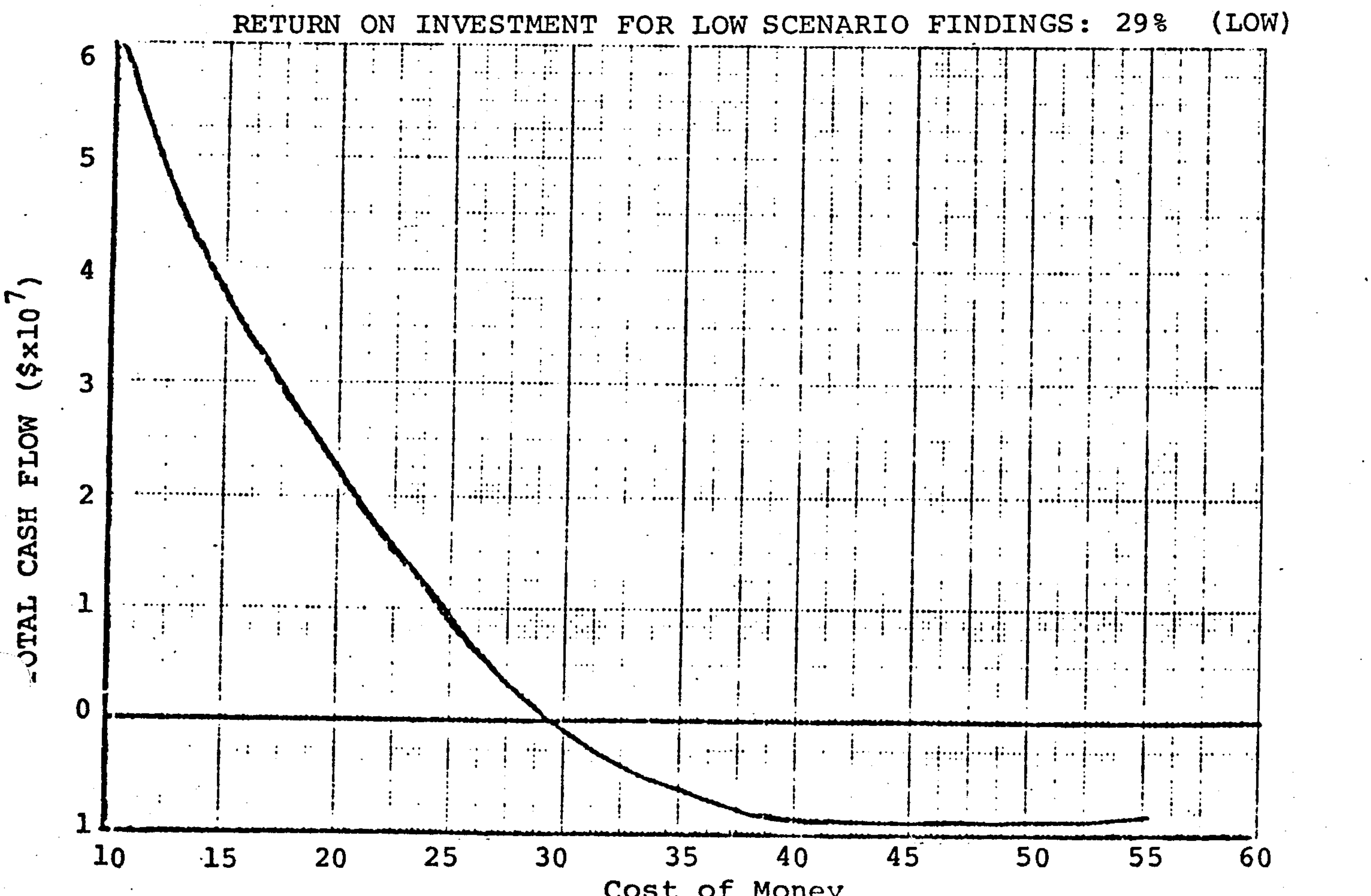
3) As Graphs 32 to 34 indicate, the return on investment (ROI) ranges from 29% to 63% depending upon the scenario. The specific 14%, 16% and 18% costs of money were included in the ROI calculations. Given the degree of risk inherent in all high technology projects, the ROI is reasonable. Further, the ROI can be adjusted by varying the charge per minute.

4) The cash position of the Corporation will be more quickly enhanced by the return inherent in the high scenario. Any decision made, however, relative to the contribution of this envisioned project to the total Corporate picture must be made within the context of other Corporate investments.

5) As is indicated in Graphs 35 to 37, in order to realize the high scenario, the largest outlay of Corporate funds (negative peak cash flow) is required. However, the more immediate and larger return which is characteristic of high scenario somewhat mitigates this large initial outlay. This remark is also germane to the medium scenario though to a lesser degree.

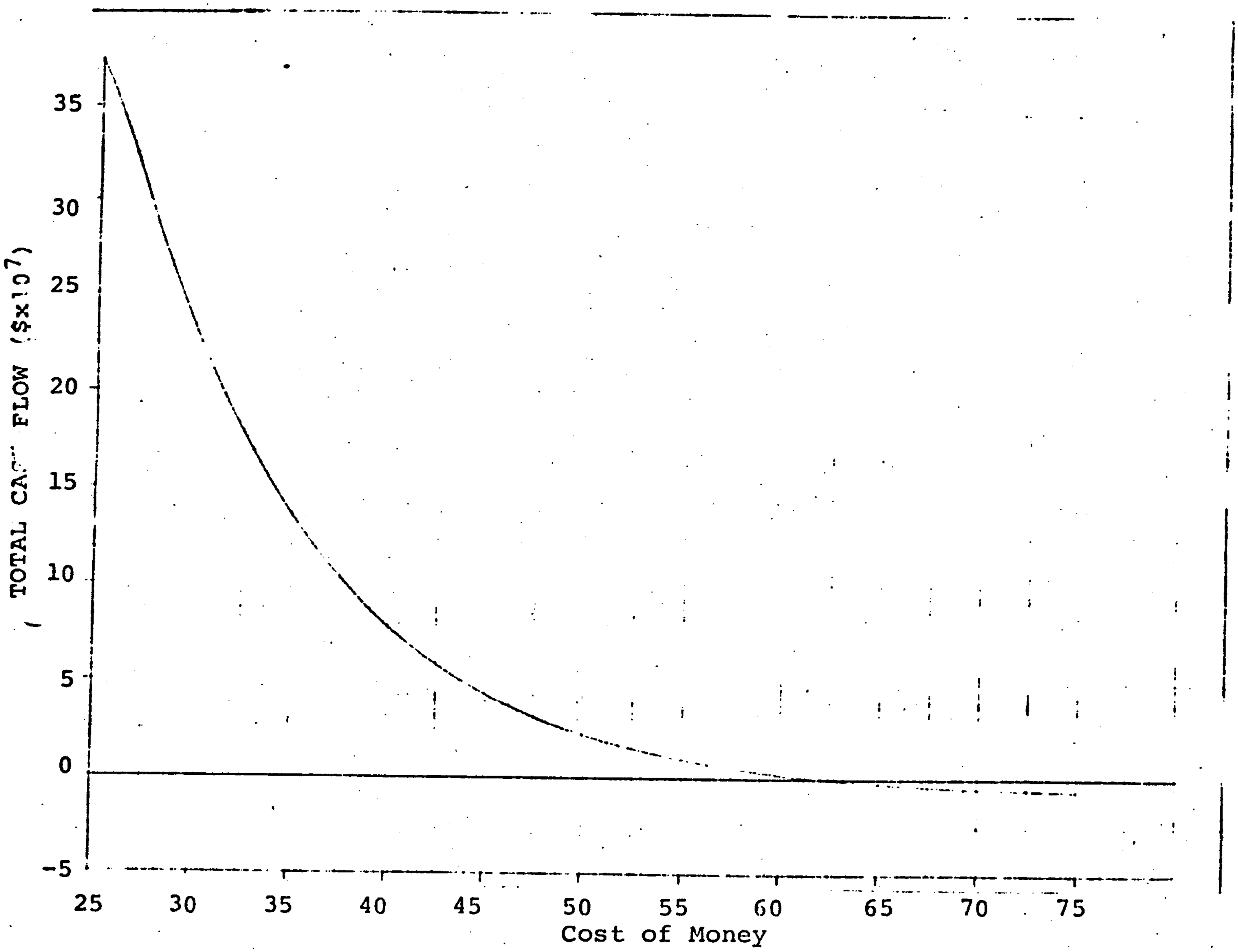


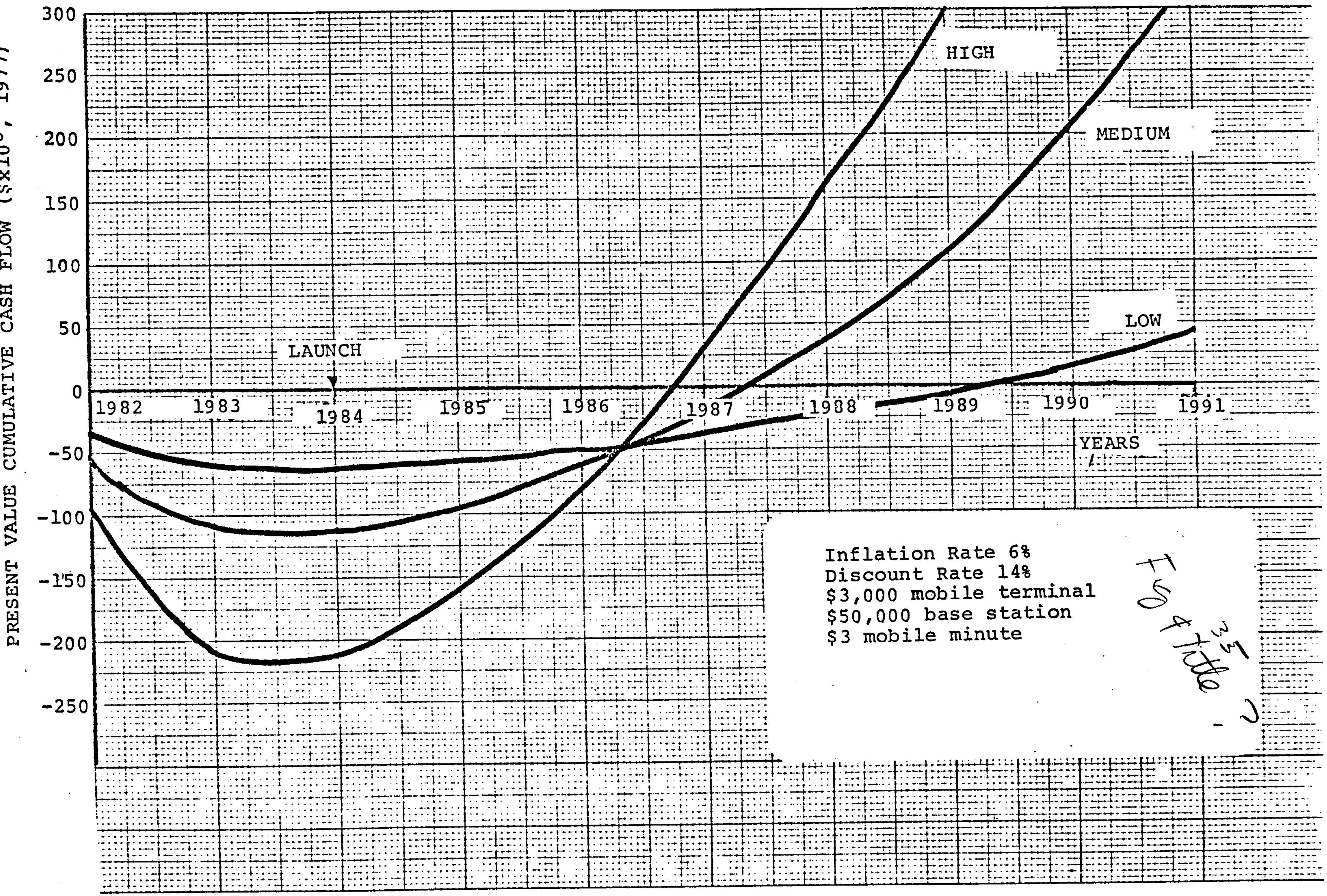
RETURN ON INVESTMENT FOR THE MEDIUM SCENARIO FINDINGS: 56% (MEDIUM)
TOTAL CASH FLOW ($x10^7)
56
48
40
32
24
16
8
0
10 15 20 25 30 35 40 45 50 55 60
Cost of Money

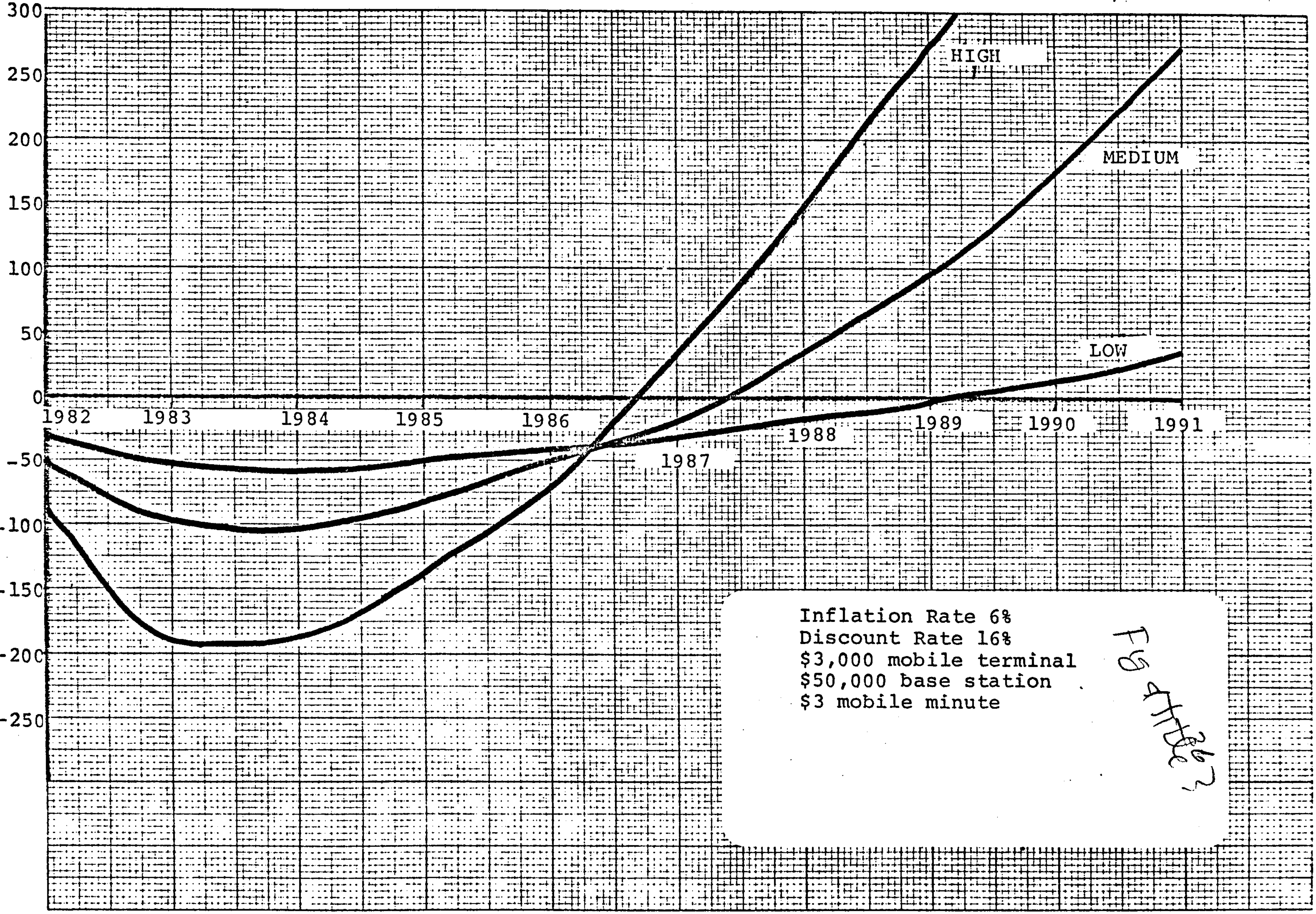



RETURN ON INVESTMENT FOR LOW SCENARIO FINDINGS: 29% (LOW)
TOTAL CASH FLOW ($x10^7)
6
5
4
3
2
1
0
1
10 15 20 25 30 35 40 45 50 55 60
Cost of Money

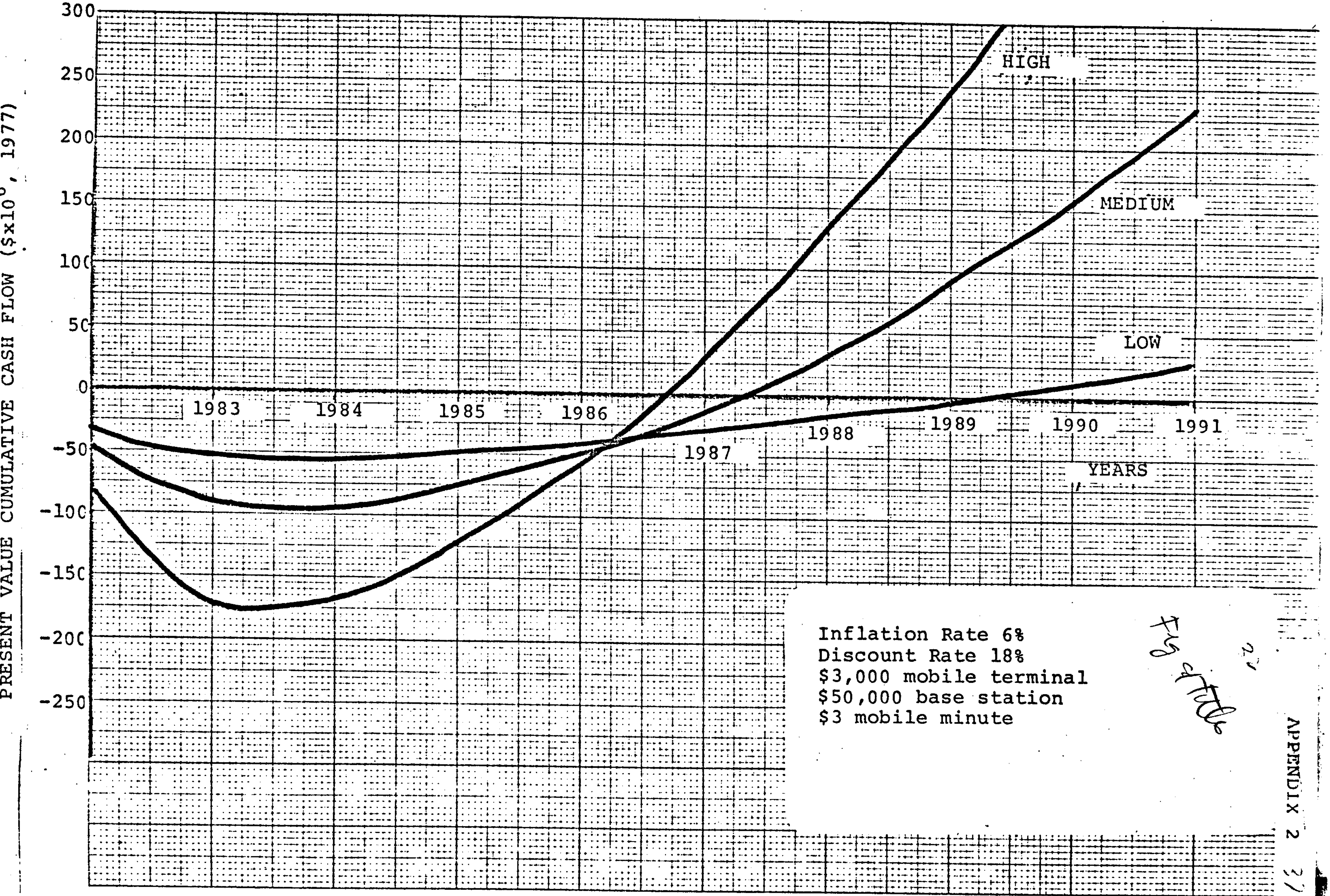
RETURN ON INVESTMENT FOR THE HIGH SCENARIO FINDINGS: 63% (HIGH)




PRESENT VALUE CUMULATIVE CASH FLOW ($x10^6, 1977)
300
250
200
150
100
50
0
-50
-100
-150
-200
-250
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
YEARS
LAUNCH
HIGH
MEDIUM
LOW
Inflation Rate 6%
Discount Rate 14%
$3,000 mobile terminal
$50,000 base station
$3 mobile minute
Fig 4 35 Title ?


300
250
200
150
100
50
0
-50
-100
-150
-200
-250
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
HIGH
MEDIUM
LOW
Inflation Rate 6%
Discount Rate 16%
$3,000 mobile terminal
$50,000 base station
$3 mobile minute
Fig 4 Title?

## BREAKEVEN GRAPHS



BIBLIOGRAPHY

Aarons, J., Whitney, H.E. and Allen, R.S.

(1) "Global Morphology of Ionospheric Scintillations," Proc. IEEE, Vol.59, No.2, (Feb 1971), 159-172.

Abutaleb, G.E.A.

(1) "The COMSTAR Satellite System," COMSAT Tech. Rev., Vol.7, No.1, (Spring 1980), 35-84.

Ananasso, F.

(1) "Coping with Rain Above 11 GHz," Microwave Systems News, Vol.10, No.3, (March 1980), 58-72.

Arthurs, E. and Dym, H.

(1) "On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise - A Geometric Interpretation and a Study of Three Basic Data Transmission Systems," IRE Trans. Comm. Sys., Vol.CS-10, No.4, (Dec 1962), 336-372.

Aulin, T. and Sundberg, C.-E.W.

(1) "Continuous Phase Modulation - Part I: Full Response Signaling," IEEE Trans. Comm., Vol.COM-29, No.3, (March 1981), 196-209.

Baghdady, E.J., Lincoln, R.N. and Nelin, B.D.

(1) "Short-Term Frequency Stability: Characterization, Theory, and Measurement," Proc. IEEE, Vol.53, No.7, (July 1965), 704-722.

Bantin, C.C. and Lyons, R.G.

(1) "The Evaluation of Satellite Link Availability," IEEE Trans. Comm., Vol.COM-26, No.6, (June 1978), 847-853.

Barry, J.D., et al.

(1) "1000 Mbits/s Intersatellite Laser Communication System Technoloby," IEEE Trans. Comm., Vol.COM-24, No.4, (April 1976), 470-478.

Barton, D.K. and Ward, H.R.

(1) Handbook of Radar Measurement, Prentice-Hall, New Jersey, 1969.

Bekey, I. and Mayer, H.

(1) "1980-2000: Raising Our Sights for Advanced Space Systems," AIAA Astron. and Aeron., (July/Aug 1976), 34-63.

Benoit, A.

(1) "Signal Attenuation Due to Neutral Oxygen and Water Vapor, Rain, and Clouds," Microwave Jnl., (Nov 1968), 73-80.

Blevis, B.C.

(1) "Losses Due to Rain on Radomes and Reflecting Surfaces," IEEE Trans. Ant. and Prop., Vol.AP-13, No. , (Jan 1965), 175-176.

Bleviss, Z.O.

(1) "Expendable Launch Vehicles for Synchronous Communications Satellites," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-274.

Bond, F.E. and Curry, Jr., W.H.

(1) "The Evolution of Military Satellite Communications Systems," Signal, Vol.30, No.6, (March 1976), 39-44.

Brian, D.J. and Mark, J.R.

(1) "The Disc Antennas," IEEE International Conference on Satellite Systems for Mobile Communications and Surveillance, London, UK, (March 1973), 14-17.

Burton, H.O. and Sullivan, D.D.

(1) "Errors and Error Control," Proc.IEEE, Vol.60, No.11, (Nov 1972), .

Bux, W., Kummerle, K. and Truong, H.L.

(1) "Balanced HDLC Procedures: A Performance Analysis," IEEE Trans. Comm. Vol.COM-28, No.11, (Nov 1980), 1889-1898.

Cacciamani, E.R. and Kim, K.S.

(1) "Circumventing the Problem of Propagation Delay on Satellite Data Channels," Data Comm., Vol.4, No.7-8, (July/Aug 1975), 19-24.

Calvit, T.O.

(1) " Marisat - Prelude to Global Maritime Satellite Communications System," IEEE Western Electronic Show and Convention (WESCON 75), San Francisco, California, (Sept 1975), .

Campanella, S.J., et al.

(1) "The INTELSAT TDMA Field Trial," COMSAT Tech. Rev., Vol.9, No.2, (Fall 1979), 293-340.

Capetanakis, J.I.

(1) "Coding for a Satellite Channel,"

(2) "The Multiple Access Broadcast Channel: Protocol and Capacity Considerations," Ph.D. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1977), 224 p.

(3) "Generalized TDMA: The Multi-Accessing Tree Protocol," IEEE Trans. Comm., Vol.COM-27, No.10, (Oct 1979), 1476-1483.

Carter, C.R.

(1) "Survey of Synchronization Techniques for a TDMA Satellite-Switched System," IEEE Trans. Comm. Vol.COM-28, No.8, (Aug 1980), 1291-1301.

Cook, J.S., et al.

(1) "The Open Cassegrain Antenna: Part I," Bell System Technical Jnl,, Vol.44, No.7, (Sept 1965), 1255-1300.

Craft,Jr., H.D. and Westerland, L.H.

(1) "Scintillations at 4 and 6 GHz Caused by the Ionosphere," AIAA 10th Aerospace Sciences Meeting, San Diego, CA, (Jan 1972),

Crane, R.K.

(1) "Ionospheric Scintillation," Proc. IEEE, Vol.65, No.2, (Feb 1977), 180-199.

Cuccia, C.L.

(1) "Bandwidth Conservation is Essential," Microwave Systems News, Vol.8, No.10, (Oct 1978), 67-72, 161.

Curry,Jr., W. H.

(1) "The Military Satellite Communications Systems Architecture," AIAA Satelite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-268.

Cutler, L.S. and Searle, C.L.

(1) "Some Aspects of the Theory and Measurement of Frequency Flucturations in Frequency Standards," Proc. IEEE, Vol.54, No.2, (Feb 1966), 136-154.

Das, S.

(1) "Multibeam Antennas Improve Satellite Communication," Microwave Systems News, Voo.7, No.12, (Dec 1977), 42-56.

Dayton, A.D. and Jain, P.C.

(1) "MILSATCOM Architecture," IEEE Trans. Comm., Vol.COM-28, No.9, (Sept 1980), 1456-1459.

DeRosa, J.K., Ozarow, L.H. and Weiner, L.N.

(1) "Efficient Packet Satellite Communications," IEEE Trans. Comm., Vol.COM-27, No.10, (Oct 1979), 1416-1422.

Diffie, W. and Hellman, M.E.

(1) "New Directions in Cryptography," IEEE Trans. on Info. Theory, Vol.IT_22, No.2, (Nov 1976), 644-654.

(2) "Privacy and Authentication: An Introduction to Cryptography," Proc. IEEE, Vol.67, No.3, (March 1979), 397-427.

Dill, G.D., et al.

(1) "Application of SS-TDMA in a Channelized Satellite," Proc. IEEE International Conference on Communications, Philadelphia, PA, (June 1976), 51-1 to 51-5.

Dixon, R.C.

(1) Spread Spectrum Systems, Wiley, New York, 1976.

Doll, D.R.

(1) "Calculating Throughput on Full-Duplex Data-Link Controls," Data Comm., Vol.5, No.1-2, (Jan/Feb 1976), 25-28.

Donnan, R.A. and Kersey, J.R.

(1) "Synchronous Data Link Control: A Perspective," IBM Systems Jnl., Vol.13, No.2, ( 1974), 140-162.

Dorian, C.

(1) "The 1979 World Administrative Radio Conference and Satellite Communications," COMSAT Tech. Rev., Vol.10, No.1, (Spring 1980), 1-26.

Duncan, J., et al.

(1) "Dual Polarization Multiple Beam Antenna for Frequency Reuse Satellites," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-248.

Easton, M.C.

(1) "Batch Throughput Efficiency of ADCCP/HDLC/SDLC Selective Reject Protocols," IEEE Trans. Comm., Vol.COM-28, No.2, (Feb 1980), 187-195.

Eckhian, L.K.

(1) "State Vector Estimation in the Presence of Measurement Uncertainty," S.M. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1980), 218 p.

Ellington, T.D.

(1) "DSCS III - Becoming an Operational System," IEEE Trans. Comm., Vol.COM-28, No.9, (Sept 1980), 1499-1504.

Esserman, C.H.

(1)"The Decomposition and Modularization of the Investment Problem," S.B. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1979), 93 p.

Fabrycky, W.J. and Thuesen, G.J.

(1) Economic Decision Analysis, Prentice-Hall, New Jersey, 1974.

Faillon, G. and Devillenger, J.P.

(1) "Improvements in the Performance Characteristics of of 6 GHz Earth-Station Klystrons," Microwave Jnl., Vol.23, No.7, (July 1980), 57-60.

Fang, R.J.F.

(1) "Modulation Transfer from TDMA On-Off Bursting to FM Carriers in Memoryless Nonlinear Devices, Part II: Baseband Performance," IEEE Trans. Comm., Vol.COM-26, No.4, (April 1978), 439-448.

Forney, G.D.

(1) "Coding and Its Application in Space Communications," IEEE Apectrum, Vol.7, No.6, (June 1970), 47-58.

(2) "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," IEEE Trans. Info. Theory, Vol.IT-18, No.3, (May 1972), 363-378.

(3) "The Viterbi Algorithm," Proc. IEEE, Vol.61, No.3, (March 1973), 268-278.

Fuenzalida, J.C.

(1) "Summary of the INTELSAT V Communications Performance Specifications," COMSAT Tech. Rev., Vol.7, No.1, (Spring 1970), 311-326.

Fugono, N., Hayashi, R. and Ishizawa, Y.

(1) "ETS-II Experiments Part I: Japan's First Geostationary Satellite," IEEE Trans. Aero. and Elect. Syst., Vol.AES-16, No.5, (Sept 1980), 549-557.

Gabriel, W.F.

(1) "Adaptive Arrays - An Introduction," Proc, IEEE, Vol.64, No.2, (Feb 1976), 239-272.

Gagliardi, R.M. and Karp, S.

(1) Optical Communications, Wiley, New York, 1976.

Gagliardi, R.M.

(1) Introduction to Communications Engineering, Wiley, New York, 1978.

Gallager, R.G.

(1) "A Simple Derivation of the Coding Theorem and Some Applications," IEEE Trans. Info.Theory, Vol.IT-11, No.1, (Jan 1965), 3-18.

(2) Information Theory and Reliable Communication, Wiley, New York, 1968.

Gatfield, A.G.

(1) "ARQ Error Control on the Satellite Channel," IEEE International Conference on Communications, Minneapolis, MN, (June 1974),

(2) "Error Control on Satellite Channels Using ARQ Techniques," COMSAT Tech. Rev., Vol.6, No.1, (Spring 1976), 179-188.

Gerber, E.A. and Sykes, R.A.

(1) "State of the Art - Quartz Crystal Units and Oscillators," Proc. IEEE, Vol.54, No.2, (Feb 1966), 103-115.

Gould, R.G. and Lum, Y.F., editors

(1) Communications Satellite Systems: An Overview of the Technology, IEEE Press, New Jersey, 1976.

Gray, J.P.

(1) "Line Control Procedures," Proc. IEEE, Vol.60, No.11, (Nov 1972), 1301-1312.

Gray, R.M. and Tausworthe, R.C.

(1) "Frequency-counted Measurements and Phase Locking to Noisy Oscillators," IEEE Trans. Communications Technology, Vol.COM-19, No.1, (Feb 1971), 21-30.

Greene, K., Et al.

(1) "Automatic Seismic Observatory Communications System," COMSAT Tech. Rev., Vol.10, No.1, (Spring 1980), 179-222.

Greenhaigh, S.J., Pengelly, R.S. and Turner, J.

(1) "Receivers Evolving for TV-by-Satellite," Microwave Systems News, Vol.10, No.9, (Sept 1980), 82-92.

Han, C., and et al.

(1) "A General Beam Shaping Technique - Multiple-Feed Offset Reflector Antenna System," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-249.

Heller, J.A. and Jacobs, I.M.

(1) "Viterbi Decoding for Satellite and Space Communication," IEEE Trans. on Communications Technology, Vol.COM-19, No.5, (Oct 1971), 835-848.

Hellman, M.E.

(1) "A Cryptanalytic Time - Memory Trade-Off," IEEE Trans. Info. Theory, Vol.IT-26, No.4, (July 1980), 401-406.

Herman, J.H.

(1) "Travelling Wave Tube Amplifiers Above 10GHz for Space Applications," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-294.

Hills, M.T.

(1) Telecommunications Switching Principles, MIT Press, Massachusetts, 1979.

Hogg, D.C. and Chu, T-S.

(1) "The Role of Rain in Satellite Communications," Proc. IEEE, Vol.63, No.9, (Sept 1975), 1308-1331.

Huang, T.C., Omura, J.K. and Lindsey, W.C.

(1) "Analysis of Coherent Satellite Communication Systems in the Presence of Interference and Noise," IEEE Trans. Comm., Vol.COM-29, No.5, (May 1981), 593-604.

Huber, W.G. and Huffaker, C.F.

(1) "IUS Capability for Earth Orbital Spacecraft," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-276.

INTELSAT

(1) "Standard A: Performance Characteristics of Earth Stations in the INTELSAT System IV, IV-A and V Having a Received G/T of 40.7 dB/K," INTELSAT Doc. No. BG-28-72E, (August 23, 1977), 130 p.

(2) "Standard B: Performance Characteristics of Earth Stations in the INTELSAT System IV, IV-A and V Having a Received G/T of 31.7 dB/K," INTELSAT Doc. No. BG-28-74E, (July 27, 1977), 84 p.

(3) "Standard C: Performance Characteristics of Earth Stations in the INTELSAT V System (14 and 11 GHz Frequency Bands)," INTELSAT Doc. No. BG-28-73E, (August 18, 1977), 73 p.

Inukai, T.

(1) "An Efficient SS/TDMA Time Slot Assignment Algorithm," IEEE Trans. Comm., Vol.COM-27, No.10, (Oct 1979), 1449-1455.

Ivanek, F.

(1) "Oscillator Alternatives Examined," Microwave Systems News, Vol.9, No.11, (Nov 1979), 69-80.

(2) "Solid-State Amplifiers as TWT Substitutes Part I," Microwave Jnl., Vol.23, No.9, (Sept 1980), 65-72.

(3) "Solid-State Amplifiers as TWT Substitutes Pert II," Microwave Jnl., Vol.23, No.10, (Oct 1980), 23-25.

Jordan, K.L.

(1) "The Performance of Sequential Decoding in Conjunction with Efficient Modulation," IEEE Trans. Communications Technology, Vol.COM-14, No.3, (June 1966), 283-297.

Kaplan, M.H.

(1) Modern Spacecraft Dynamics and Control, Wiley, New York, 1976.

Kaul, A.K.

(1) "Performance of High-level Data Link Control in Satellite Communications," COMSAT Tech. Rev., Vol.8, No.1, (Spring 1978), 41-87.

(2) "A Model for a Packet Switching Node Including Packet Retransmission Effects," COMSAT Tech. Rev., Vol.9, No.2B, (Fall 1979), 669-703.

Kittiver, C. and Zitzmann, F.R.

(1) "The SBS System - An Innovative Domestic Satellite System for Private-Line Networks," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-307.

Klein, P.I. and Soifer, R.

(1) "Intersatellite Communication Using the AMSAT-OSCAR 6 and AMSAT-OSCAR 7 Radio Amateur Satellites," Proc. IEEE, Vol.63, No.10, (Oct 1975), 1526-1527.

Kleinrock, L.

(1) Queueing Systems, Vol. I: Theory, Wiley, New York, 1975.

(2) Queueing Systems, Vol. II: Computer Applications, Wiley, New York, 1976.

Kloves, T.R.

(1) "Space Shuttle Program - An Overview," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-275.

Konheim, A.G.

(1) "A Queueing Analysis of Two ARQ Protocols," IEEE Trans, Comm., Vol.COM-28, No.7, (July 1980), 1004-1014.

Kreutel, R.

(1) "A Multiple Beam Torus Reflector Antenna for 20/30 GHz Systems," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-302.

(2) "Optical Transmission Technology in Satellite Communications," COMSAT Tech. Rev., Vol.10, No.2, (Fall 1980), 321-368.

Kwan, R.K.

(1) "The Telesat TDMA System," IEEE International Conference on Communications, San Francisco, CA, (June 1975), 44-1 to 44-5.

Lam, S.S.

(1) "Satellite Packet Communication - Multiple Access Protocols and Performance," IEEE Trans. Comm., Vol.COM-27, No.10, (Oct 1979), 1456-1466.

Lebow, I.L., et al.

(1) "Satellite Communications to Mobile Platforms," Proc. IEEE, Vol.59, No.2, (Feb 1971), 139-159.

Lipke, D.W. and Swearingen, D.W.

(1) " Communications Systems Planning for MARISAT," IEEE International Conference on Communications, Minneapolis, MN, (June 1974), 29B-1 to 29B-5.

Lin, S. and Yu, P.S.

(1) "An Effective Error Control Scheme for Satellite Communications," IEEE Trans. Comm., Vol. COM-28, No.3, (March 1980), 395-401.

Martin, J.

(1) Communications Satellite Systems, Prentice-Hall, New Jersey, 1978.

Maseng, T. and Bakken, P.M.

(1) "A Stochastic Dynamic Model of Rain Attenuation," IEEE Trans. Comm., Vol.COM-29, No.5, (May 1981), 660-669.

Maxemchuk, N.F. and Schiff, L.

(1) "Third Order Intermodulation Interference - Bounds and Interference-free Channel Assignment," IEEE Trans. Comm., Vol.COM-25, No.9, (Sept 1977), 1041-1045.

McCormick, K.S. and Maynard, L.A.

(1) "Measurements of S.H.F. Tropospheric Fading Along Earth-Space Paths at Low Elevation Angles," Electronic Letters, Vol.8, (     1972), 274-276.

McEliece, R.J.

(1) The Theory of Information and Coding Vol. 3: Section, Probability (Encyclopedia of Mathematics and Its Applications, Editor: G-C. Rota), Addison-Wesley, Massachusetts, 1977.

McGarty, T.P.

(1) "The Estimation of the Center of Gravity of a Photon Density Profile in Noise," IEEE Trans. Aero. and Elect. Syst., Vol.AES-5, No. , (Nov 1969), 974-979.

(2) "On the Structure of Random Fields Generated by Multiple Scatter Media," Ph.D. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1971), 198 p.

(3) "Models of Multipath Propagation Effects in a Ground to Air Surveillance System," MIT Lincoln Laboratory, Lexington, MA, Available NTIS, AD-777 241, (Feb 1974), 157 p.

(4) Stochastic Systems and State Estimation, Wiley, New York, 1974.

(5) "Adaptive Decoding for a Specular Multipath Channel," IEEE National Telecommunications Conference, Los Angeles, CA, (Dec 1977), 20.1-1 to 20.1-5.

(6) "Multiple Access Techniques for Low Data Rate Satellite Communications Systems," IEEE National Telecommunications Conference, Los Angeles, CA, (Dec 1977), 5.2-1 to 5.2-6.

(7) "Multiple Beam Satellite Optimization," IEEE Trans. Aero. and Elect. Syst., Vol.AES-13, No.5, (Sept 1977), 504-511.

(8) "Processing Voice and Data in Mobile Satellite Communications Systems," IEEE International Conference on Communications, Chicago, IL, (June 1977), 63-77.

(9) "Unattended Earth Stations for Global Data Collection," IEEE International Conference on Communications, Boston, MA, (June 1979), 6.1-1 to 6.1-5.

(10) "Financial Data Networking and Technological Impacts," International Telecommunication and Computer Exposition (INTELCOM '80), Los Angeles, CA, (Nov 1980), 411-416.

Merkle, R.C. and Hellman, M.E.

(1) "Hiding Information and Signatures in Trapdoor Knapsacks," IEEE Trans. Info. Theory, Vol.IT-24, No.5, (Sept 1978), 525-530.

Moesl, G. and Muller, J.

(1) "In-Orbit Performance and Experimental Utilization of the Symphonie Satellites," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-306.

Mohanty, N.C.

(1) "Spread Spectrum and Time Division Multiple Access Satellite Communications," IEEE Trans. Comm. Vol.COM-25, No.8, (Aug 1977), 810-814.

Monzingo, R.A. and Miller, T.W.

(1) Introduction to Adaptive Arrays, Wiley, New York, 1980.

Morgan, W.L.

(1) "Satellite Utilization of the Geosynchronous Orbit," COMSAT Tech, Rev., Vol.6, No.1, (Spring 1976), 195-205.

Ohta, K. et al.

(1) "5-Watt 4-8 GHz GaAs-FET Amplifier," Microwave Jnl., Vol.22, No.11, (Nov 1979), 66-67.

Posner, R.D.

(1) "Design and Specification Considerations of GaAs-FET LNAS," Microwave Jnl. Vol.22, No.5, (May 1979), 69-75.

Poza, H.B.

(1) "TDRSS Telecommunications Payload: An Overview," IEEE Trans Aero. and Elect. Syst., Vol.AES-15, No.3, (May 1979), 414-429.

Pritchard, W.L.

(1) "Satellite Communications - An Overview of the Problems and Programs," Proc. IEEE, Vol.65, No.3, (March 1977), 294-307.

Puente, J.G. and Werth, A.M.

(1) "Demand-Assigned Service for the INTELSAT Global Service," IEEE Spectrum, Vol.8, No.1, (Jan 1971), 59-69.

Redisch, W.N.

(1) "ATS-6 Description and Performance," IEEE Trans. Aero. and Elect. Syst., Vol.AES-11, No.6, (Nov 1975, 994-1003.

Reed, M.A. and Smetanka, T.D.

(1) "How to Determine Message Response Time for Satellites," Data Comm., Vol.6, No.6, (June 1977), 42-47.

Rice, P.L. and Holmberg, N.R.

(1) "Cumulative Time Statistics of Surface-Point Rainfall Rates," IEEE Trans Comm., Vol.COM-21, No.10, (Oct 1973), 1131-1136.

Roberts, L.G.

(1) "Dynamic Allocation of Satellite Capacity Through Packet Reservations," AFIPS National Computer Conference, New York, NY, (June 1973), 711-716.

Rozec, X. and Assal, F.

(1) "Microwave Switch Matrix for Communications Satellites," IEEE International Conference on Communications, Philadelphia, PA, (June 1976), 35-13 to 35-17.

Rudge, A.W.

(1) "Multiple-Beam Antennas: Offset Reflectors with Offset Feeds," IEEE Trans. Antennas and Propagation, Vol.AP-23, No.3, (May 1975), 317-322.

Rudge, A.W. and Adatia, N.A.

(1) "Offset-Parabolic-ReflectorAntennas: A Review," Proc. IEEE, Vol.66, No.12, (Dec 1978), 1592-1618.

Scherer, D.

(1) "Today's Lesson - Learn About Low-Noise Design Part I," Microwaves, Vol.18, No.4, (April 1979), 116-122.

(2) "Today's Lesson - Learn About Low-Noise Design Pert II," Microwaves, Vol.18, No.5, (May 1979), 72-77.

Schmidt, W.G.

(1) "The Application of TDMA to the INTELSAT IV Satellite Series," COMSAT Tech. Rev., Vol.3, No.2, (FAll 1973), 257-276.

Schneider, K.S.

(1) "A Comparison of Satellite Processing Techniques: The Coherent AWGN Channel," MIT Lincoln Laboratory, Lexington, MA,

Schneider, K.S. and McGarty, T.P.

(1) "Channel Estimation and Decoding in a Multipath Environment," IEEE International Symposium on Information Theory, Ithaca, NY, (Oct 1977), 80-1 (Summary only).

(2) "Reliable Satellite Communications in the Presence of Specular Multipath," IEEE Trans.Comm., Vol.COM-26, No.2, (Feb 1978), 235-246.

Schwartz, J.G. et al.

(1) "Modulation Techniques for Multiple Access to a Hard-Limiting Satellite Repeater," Proc. IEEE, Vol.54, No.5, (May 1966), 763-777.

Schwartz, M.

(1) Computer-Communication Network Design and Analysis, Prentice-Hall, New Jersey, 1977.

Scott, W.G. et al.

(1) "Design Tradeoffs for Multiple Beam Antennas in Communication Satellites," IEEE International Conference on Communications, Philadelphia, PA, (June 1976), 4-1 to 4-6.

Shaft, P.D. and Roberts, J.A.

(1) "Optimum Allocation of Multibeam Communications Satellite Resources," IEEE Trans. Comm., Vol.COM-24, No.10, (Oct 1976), 1195-1200.

Silver, S., ed.

(1) Microwave Antenna Theory and Design, Dover Publications, New York, 1965. (Originally Published as Vol. 12 of MIT Radiadion Lab Series, McGraw-Hill, New York, 1949).

Sinha, A.

(1) "Optical Intersatellite Links," COMSAT Tech.Rev., Vol.10, No.2, (Fall 1980), 369-396.

Snyder, D.L.

(1) Random Point Processes, Wiley, New York, 1975.

Srinath, M.D. and Rajasekaran,

(1) An Introduction to Statistical Signal Processing with Applications Wiley, New York, 1979.

Stark, L.

(1) "Microwave Theory of Phased Array Antennas," Proc. IEEE, Vol.62, No.12, (Dec 1974), 1661-1701.

Steinberg, B.D.

(1) Principles of Apertaure and Array System Design, Wiley, New York, 1976.

Strickland, J.L., Olsen, R.L. and Werstiuk, H.L.

(1) "Measurements of Tropospheric Fading in the Canadian High Artic," URSI Commission F, LaBaule, France, (28 April-6 May 1977).

Swinton, W.G.

(1) "Optimization of Priority Queueing Systems in Delay-Tolerant Environments," S.M. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1980), 149 p.

Tengker, J.S.

(1) "Indonesian Domestic Satellite System," IEEE EASCON '76, Washington, DC, (Sept 1976), 11-A to 11-U.

Thorpe, D.G.

(1) "Evolution of the Telesat Canada System into the Mid-1980's," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-220.

Towsley, D.

(1) "The Stutter Go Back-N ARQ Protocol," IEEE Trans. Comm., Vol.COM-27, No.6, (June 1979), 869-875.

Towsley, D. and Wolf, J.K.

(1) "On the Statistical Analysis of Queue Lengths and Waiting Times for Statistical Multiplexers with ARQ Retransmission Schemes," IEEE Trans. Comm., Vol.COM-27, No.4, (April 1979), 693-702.

Traynham, K.C. and Steen, R.F.

(1) "Interpreting SDLC Throughput Efficiency Part I: 3 Models," Data Comm., Vol.6, No.10, (Oct 1977), 43-45, 48-51.

(2) "Interpreting SDLC Throughput Efficiency Part II: Results," Data Comm., Vol.6, No.11, (Nov 1977), 59-60, 63-66.

Tserng, H.Q.

(1) "Design and Performance of Microwave-Power GaAs-FET Amplifiers," Microwave Jnl., Vol.22, No.6, (June 1979), 94-100.

Turrin, R.H.

(1) "A Multibeam, Spherical Reflector Satellite Antenna for the 20- and 20-GHz Bands," Bell System Tech.Jnl., Vol.54, No.6, (July/Aug 1975), 1011-1026.

Van Trees, H.L.

(1) Detection, Estimation, and Modulation Theory Part I: Detection, Estimation, and Linear Modulation Theory, Wiley, New York, 1968.

(2) Detection, Estimation, and Modulation Theory Part II: Nonlinear Modulation Theory, Wiley, New York, 1971.

(3) "Future INTELSAT System (1986-1993) Planning," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-233.

(4) Satellite Communications, IEEE Press, New York, 1979.

Van Trees, H.L., Hoversten, E.V. and McGarty. T.P.

(1) "Communications Satellites: Looking to the 1980's," IEEE Spectrum, Vol.14, No.12, (Dec 1977), 42-51.

Viterbi, A.J.

(1) "Convolutional Codes and Their Performance in Communication Systems," IEEE Trans. Comm., Vol.COM-19, No.5, (Oct 1971), 751-772.

Viterbi, A.J. and Omura, J.K.

(1) Principles of Digital Communication and Coding, McGraw-Hill, New York, 1979.

Warner, T.

(1) "The Effects of Phase Noise on Low Bit Rate Data Transmission," S.M. Thesis, Massachusetts Institute of Technology, Electrical Engineering and Computer Science, (1979), 100 p.

Welsch, G.A. and Anthony, R.N.

(1) Fundamentals of Financial Accounting, revised edition, R.D. Irwin, Inc. Homewood, Illinois, 1977.

Weston, J.F. and Brigham, E.F.

(1) Managerial Finance, 5th edition, Dryden Press, Illinois, 1975.

Williams, H.C.

(1) "A Modification of the RSA Public-Key Encryption Procedure," IEEE Trans. Info Theory, Vol.IT-26, No.6, (Nov 1980), 726-729.

Winchell, D.

(1) "Single-Loop Synthesizer Upgrades SATCOM Links," Microwaves, Vol.19, No.3, (March 1980), 54-60.

Witherspoon, J. and Sherwin, R.

(1) "Real-Time Adaptive Control for DSCS," AIAA Satellite Communications Conference, Montreal, Canada, (April 1976), Paper No. 76-272.

Wolff, E.A.

(1) "ATS-6 Introduction," IEEE Trans. Aero. and Elect. Syst., Vol.AES-11, No.6, (Nov 1975), 982-983.

Wozencraft, J.M. and Jacobs, I.M.

(1) Principles of Communication Engineering, Wiley, New York, 1965.

Wozencraft, J.W. and Kennedy, R.S.

(1) "Modulation and Demodulation for Probabilistic Coding," IEEE Trans. Info. Theory, Vol.IT-12, No.3, (July 1966), 291-297.