AD-A114 385 ARMY COMMUNICATIONS COMMAND FORT HUACHUCA AZ F/G 17/2

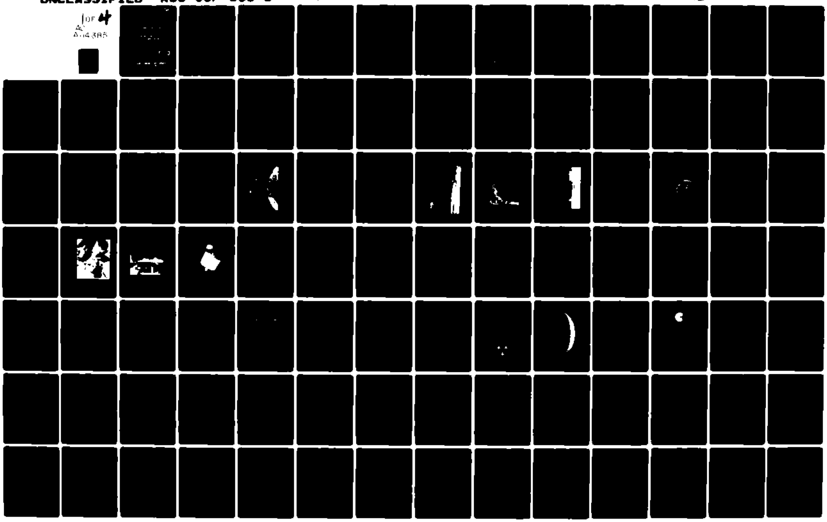
INTRODUCTION TO SATELLITE COMMUNICATIONS.(U)

APR 74

UNCLASSIFIED ACC-CCP-105-5 NL

1 OF 4

AD A114385



CCP 105-5

SACC

VOICE OF THE ARMY

COMMUNICATIONS ELECTRONICS

INTRODUCTION TO SATELLITE COMMUNICATIONS

THIS COPY IS A REPRINT WHICH INCLUDES CHANGE 1

15 APRIL 1974

HEADQUARTERS
U.S. ARMY COMMUNICATIONS COMMAND
FORT HUACHUCA, ARIZONA 85613

DEPARTMENT OF THE ARMY
US ARMY COMMUNICATIONS COMMAND
Fort Huachuca, Arizona 85613

Change 2
USACC Pamphlet
No. 105-5

22 July 1976

Communications-Electronics

INTRODUCTION TO SATELLITE COMMUNICATIONS

CCP 105-5, dated 15 April 1974, is changed as indicated below.

1. Remove old pages and insert new pages as follows:

Remove pages	Insert pages
7-17 and 7-18	7-17 and 7-18

2. New or changed material is indicated by a side bar.

3. File change sheet in front of publication for reference purposes.

(CC-OPS)

FOR THE COMMANDER:

OFFICIAL:

M. K. LABAR
Colonel, AGC
Adjutant General

C. E. McKNIGHT, Jr.
Colonel, GS
Chief of Staff

DISTRIBUTION:

HQ, USACC

25 - CC-OPS-S
5 - CC-LOG
1 - CC-PER
1 - CC-COMPT
1 - CC-ENGR
1 - CC-IG
5 - CC-PA-AS
3 - CC-PA-AMP

10 - 5th Signal Command, APO New York 09056
2 - USACC-EUR Signal Support Agency Mediterranean, APO New York 09019
2 - USACC-EUR Signal Facility Tehran, APO New York 09205
2 - HHD Signal Group 106, USACC-EUR, APO New York 09131
2 - USACC-EUR Area Maintenance and Supply Facility, APO New York 09019
5 - Berlin Bde, C-E Div, ATTN: SATCOM Terminal, APO New York 09742
10 - 6th Signal Command, APO San Francisco 96558
5 - 1st Signal Brigade (USACC), ATTN: CCP-K, APO San Francisco 96557
10 - USACC-JAPAN, APO San Francisco 96343
10 - USACC-THAILAND, APO San Francisco 96346
2 - USACC Signal Support Agency, Hawaii, APO San Francisco 96557
2 - USACC Long-Lines, Okinawa, APO San Francisco 96331
15 - 7th Signal Command, Fort Ritchie, MD 21719
5 - USACC-FORCES, Ft McPherson, GA 30330
10 - USACC-TRADOC, ATTN: ATCD-IC, Ft Monroe, VA 23651
5 - USACC-AMC, ATTN: CCNC-CS, Alexandria, VA 22333
10 - Satellite Communications Station, Camp Roberts, CA 93451
10 - USACC-East Coast Telecommunications Center, ATTN: CCNI, Ft Detrick, MD 21701
10 - USACC-Satellite Station, Lakehurst, PO Box 85, Lakehurst, NJ 08733
1 - 11th Signal Group (USACC), Ft Huachuca, AZ 85613
5 - 521st Signal Company, Ft Huachuca, AZ 85613
100 - HQ, USACEEIA, Ft Huachuca; AZ 85613
5 - USACEI Bn, Ft Huachuca, AZ 85613
25 - USACEEIA-CONUS, Ft Ritchie, MD 21719

2 - USACEEIA-CONUS Field Office EAST, Ft Meade, MD 20755
2 - USACEEIA-CONUS Field Office WEST, Presidio San Francisco 94129
10 - USACEEIA-EUR, APO New York 09056
10 - USACEEIA-PAC, APO San Francisco 96557
5 - US Army Communications Systems Agency, Ft Monmouth, NJ 07703
2 - USACSA(H), Ft Huachuca, AZ 85613
360 - HQ, Fort Huachuca, ATTN: CCH-PCA-ASP, Ft Huachuca, AZ 85613
1 - USACC Ballistic Missile Defense Communications Activity, Ft Huachuca, AZ 85613
1 - USACC Liaison Office, ATTN: Mr. Robert Geier, Pentagon, Washington, DC 20310
1 - USACC LNO, SATCOMA/TRI-TAC, ATTN: CC-LNSAT/Jerry Lucas, Bldg 209, Ft Monmouth, NJ 07703
1 - USACC Liaison Office, ATTN: CC-LNSIG/CPT Enfinger, Ft Gordon, GA 30905
2 - HQDA (DAMO-TCS) Washington, DC 20310
2 - HQDA (DAMA) Washington, DC 20310
5 - DCA, Washington, DC 20305
5 - US Army Electronics Command, Ft Monmouth, NJ 07703
3 - US Army Satellite Communications Agency, Ft Monmouth, NJ 07703
420 - US Army Southeastern Signal School, ATTN: ATSN-TE-MS(400); ATSN-TE-SC(20), Ft Gordon, GA 30905
2 - US Army Infantry School, Ft Benning, GA 31905
2 - US Army Artillery School, Ft Sill, OK 73503
2 - US Military Academy, West Point, NY 10996
2 - US Army Air Defense Artillery School, Ft Bliss, TX 79906
5 - Lexington-Bluegrass Army Depot, Lexington, KY 40507
2 - SATCOM Detachment, Landstuhl Germany, APO NY 09180
2 - SATCOM Detachment, Korea, APO San Francisco 96301
5 - Defense Communications Engineering Center, Reston, VA 22070
5 - US Naval Telecommunications Command, 4401 Mass Ave, Washington, DC 20390
5 - US Naval Electronics Systems Command, Ft Monmouth, NJ 07703
2 - US Naval Academy, Annapolis, MD 21402
5 - Air Force Communications Service, Richards-Gebaur AFB, MO 64030
2 - Rome Air Development Center, Rome, NY 13440
2 - US Air Force Academy, Colorado Springs, CO 80900

b. With P_{TC} defined as the rf carrier power output delivered to the transmission line leading to the duplexer (either in the earth terminal or aboard the satellite) and P_{RC} defined as the carrier power delivered to the receiver amplifier input at either the earth terminal or satellite receiver, the ratio TC/P_{RC} can be called the transmission loss (L_T) of a single communications space link.

$$L_T = \frac{P_{TC}}{P_{RC}} \qquad L_T(\text{dB}) = 10 \log_{10} \left(\frac{P_{TC}}{P_{RC}}\right) \tag{7.11}$$

It is convenient to divide L_T into the sum of five losses for further study, thus:

$$L_T = L_S + L_a + L_r + L_t + L_{axr} \tag{7.12}$$

where

L_S is an isotropic space loss (dB)

L_t is the loss between the output of the transmitter power amplifier and the antenna (dB)

L_a is an absorption loss due to normal (clear) weather (dB)

L_{axr} is excess atmospheric absorption loss during periods of rain (dB)

L_r is a loss similar to L_t between the receiver antenna input and the receiver amplifier (dB)

(1) *Isotropic free space loss.*

(a) For any free space link, the ratio of the power radiated by the transmitter antenna P_T to the power absorbed by the receiving antenna P_R (if both antennas are assumed to be isotropic and in free space) is called the isotropic free space loss (L_S).

$$L_S = \frac{P_T}{P_R}$$

$$L_S(\text{dB}) = 10 \log_{10} \left(\frac{P_T}{P_R}\right) \tag{7.13}$$

(b) This isotropic free space loss is related quantitatively to the distance between the transmitter and receiver and to the frequency, or wavelength, of the signal. The free space loss formula (para 7-2) has been used to prepare the graph shown in figure 7-11. It can be seen from this graph that for a distance R of $4 \cdot 10^4$ km, which is typical of a ground-to-satellite link, and a frequency of 8 GHz (shf band), the isotropic free space loss is 203 dB. The isotropic loss L_S increases as the square of both distance and frequency. Expressed as a formula:

$$L_S = \left(\frac{4\pi Dr}{\lambda}\right)^2$$

$$L_S(dB) = +92.45 + 20 \log f + 20 \log Dr \qquad (7.14)$$

where

f is frequency in GHz

Dr is path length in km

(2) *Atmospheric and rain loss.*

(a) The atmosphere has a selective absorption of radiation in the microwave millimeter portions of the spectrum between 10^9 and 10^{11} Hz, as diagramed in figure 7-12. The right-hand curves represent absorption by oxygen and water vapor which occurs under even the best of conditions; i.e., clear weather. The left-hand curves represent excess rain loss in the atmosphere when rain or fog formations are dominant factors. The curves of figure 7-12 have been theoretically derived and experimentally verified.

(b) The amount of absorption due to water vapor and excess rain loss (L_{exr}) varies considerably depending on the weather. Both excess rain loss and clear weather atmospheric loss (L_a) are dependent to some extent on antenna elevation pointing angle. This is because, at low elevation angles, the radio signal traverses a longer path through the earth's atmosphere than at high elevation angles. At an antenna elevation angle of 5 degrees, the path length through the atmosphere is approximately 50 km in length. The constituents of atmosphereic absorption and how they vary with frequency in the case of a 50-km path length are shown in figure 7-12. Figure 7-13 shows the total one-way clear weather absorption as a function of frequency for different antenna elevation angles.

(c) Since excess rain loss becomes a factor only during periods of rain or fog, it is expressed as a percentage of time that it exceeds a certain value. A curve plotted to show this type of relationship is known as a distribution curve. Figure 7-14 is a typical distribution curve of excess rainfall attenuation for a relatively dry climate such as San Diego, California.

DEPARTMENT OF THE ARMY
US ARMY COMMUNICATIONS COMMAND
Fort Huachuca, Arizona 85613

Change 1
USACC Pamphlet
No. 105-5

10 November 1975

Communications-Electronics

INTRODUCTION TO SATELLITE COMMUNICATIONS

CCP 105-5, dated 15 April 1974, is changed as indicated below.

1. Remove old pages and insert new pages as follows:

Remove pages	Insert pages
2-1 through 2-4	2-1 through 2-4
2-19 and 2-20	2-19 and 2-20
7-3 and 7-4	7-3 and 7-4
7-17 and 7-18	7-17 and 7-18
7-25 through 7-28	7-25 through 7-28
7-31 and 7-32	7-31 and 7-32

2. New or changed material is indicated by a side bar.

3. File change sheet in front of publication for reference purposes.

(CC-OPS)

FOR THE COMMANDER:

OFFICIAL:

M. K. LABAR
Colonel, AGC
Adjutant General

C. E. McKNIGHT, Jr.
Colonel, GS
Chief of Staff

DISTRIBUTION:

HQ, USACC

25 - CC-OPS-S
5 - CC-LOG
1 - CC-PER
1 - CC-COMPT
1 - CC-ENGR
1 - CC-IG
5 - CC-PA-AS
3 - CC-PA-AMP

10 - 5th Signal Command, APO New York 09056
2 - USACC-EUR Signal Support Agency Mediterranean, APO New York 09019
2 - USACC-EUR Signal Facility Tehran, APO New York 09205
2 - HHD Signal Group 106, USACC-EUR, APO New York 09131
2 - USACC-EUR Area Maintenance and Supply Facility, APO New York 09019
5 - Berlin Bde, C-E. Div, ATTN: SATCOM Terminal, APO New York 09742
10 - 6th Signal Command, APO San Francisco 96558
5 - 1st Signal Brigade (USACC), ATTN: CCP-K, APO San Francisco 96557
10 - USACC-JAPAN, APO San Francisco 96343
10 - USACC-THAILAND, APO San Francisco 96346
2 - USACC Signal Support Agency, Hawaii, APO San Francisco 96557
2 - USACC Long-Lines, Okinawa, APO San Francisco 96331
15 - 7th Signal Command, Fort Ritchie, MD 21719
5 - USACC-FORCES, Ft McPherson, GA 30330
10 - USACC-TRADOC, ATTN: ATCD-IC, Ft Monroe, VA 23651
5 - USACC-AMC, ATTN: CCNC-CS, Alexandria, VA 22333
10 - Satellite Communications Station, Camp Roberts, CA 93451
10 - USACC-East Coast Telecommunications Center, ATTN: CCNI, Ft Detrick, MD 21701
10 - USACC-Satellite Station, Lakehurst, PO Box 85, Lakehurst, NJ 08733
1 - 11th Signal Group (USACC), Ft Huachuca, AZ 85613
5 - 521st Signal Company, Ft Huachuca, AZ 85613
100 - HQ, USACEEIA, Ft Huachuca, AZ 85613
5 - USACEI Bn, Ft Huachuca, AZ 85613
25 - USACEEIA-CONUS, Ft Ritchie, MD 21719

2 - USACEEIA-CONUS Field Office EAST, Ft Meade, MD 20755
2 - USACEEIA-CONUS Field Office WEST, Presidio San Francisco 94129
10 - USACEEIA-EUR, APO New York 09056
10 - USACEEIA-PAC, APO San Francisco 96557
5 - US Army Communications Systems Agency, Ft Monmouth, NJ 07703
2 - USACSA(H), Ft Huachuca, AZ 85613
360 - HQ, Fort Huachuca, ATTN: CCH-PCA-ASP, Ft Huachuca, AZ 85613
1 - USACC Ballistic Missile Defense Communications Activity, Ft Huachuca, AZ 85613
1 - USACC Liaison Office, ATTN: Mr. Robert Geier, Pentagon, Washington, DC 20310
1 - USACC LNO, SATCOMA/TRI-TAC, ATTN: CC-LNSAT/Jerry Lucas, Bldg 209, Ft Monmouth, NJ 07703
1 - USACC Liaison Office, ATTN: CC-LNSIG/CPT Enfinger, Ft Gordon, GA 30905
2 - HQDA (DAMO-TCS) Washington, DC 20310
2 - HQDA (DAMA) Washington, DC 20310
5 - DCA, Washington, DC 20305
5 - US Army Electronics Command, Ft Monmouth, NJ 07703
3 - US Army Satellite Communications Agency, Ft Monmouth, NJ 07703
420 - US Army Southeastern Signal School, ATTN: ATSN-TE-MS(400); ATSN-TE-SC(20), Ft Gordon, GA 30905
2 - US Army Infantry School, Ft Benning, GA 31905
2 - US Army Artillery School, Ft Sill, OK 73503
2 - US Military Academy, West Point, NY 10996
2 - US Army Air Defense Artillery School, Ft Bliss, TX 79906
5 - Lexington-Bluegrass Army Depot, Lexington, KY 40507
2 - SATCOM Detachment, Landstuhl Germany, APO NY 09180
2 - SATCOM Detachment, Korea, APO San Francisco 96301
5 - Defense Communications Engineering Center, Reston, VA 22070
5 - US Naval Telecommunications Command, 4401 Mass Ave, Washington, DC 20390
5 - US Naval Electronics Systems Command, Ft Monmouth, NJ 07703
2 - US Naval Academy, Annapolis, MD 21402
5 - Air Force Communications Service, Richards-Gebaur AFB, MO 64030
2 - Rome Air Development Center, Rome, NY 13440
2 - US Air Force Academy, Colorado Springs, CO 80900

DEPARTMENT OF THE ARMY
HEADQUARTERS, US ARMY COMMUNICATIONS COMMAND
Fort Huachuca, Arizona 85613

USACC Pamphlet
CCP 105-5 15 April 1974

Communications-Electronics

INTRODUCTION TO SATELLITE COMMUNICATIONS

*This pamphlet supersedes CCP 105-5, 15 February 1968.

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1-1. **THE GROWING NEED FOR COMMUNICATIONS.** Perhaps the most spectacular evolution in all branches of technology during the last half century has occurred in the field of communications. Radio has evolved from an experimental curiosity to a definite necessity. Since World War II, advances in manufacturing techniques and the circuitry design of communications equipment have led to greater reliability and capability, together with considerable decreases in size and weight.

a. At the same time the demands made on the techniques and capabilities of communications have increased enormously. A world population once separated by weeks, or months, is now only minutes, or at most, hours apart. Peoples and whole areas that were little more than question marks on maps are now in close contact with the major population centers of the world. Modern military equipment and weapons, that permit an army to be moved halfway around the globe in a matter of hours and strikes to be made against any point on the earth's surface in minutes, demand an increase in the ability to communicate rapidly and accurately with such forces, wherever they may be located.

b. The constant need to improve military communications has long been recognized. Unfortunately, such improvements often have been hampered by the crowded frequency spectrum, budget limitations, and the regimented pace of the military to adopt advances in the state-of-the-art. Although the communications facilities of the various military departments have been able to support their communications requirements in the past, the predictable demands of the future will require large-scale improvements to be made more rapidly than in the past.

c. Experience over the past 20 years has shown that the usage rate of both commercial and military systems increases by approximately 10 percent per year. Also, when an improved service is offered, the traffic tends to increase. An example of the latter is the increase in the number of long-distance telephone calls following the introduction of direct distance dialing service.

d. New facilities, particularly those for long haul communications, will be made available in the near future to areas where they are now either inadequate or nonexistent. This new long haul traffic will constitute an increasing percentage of the total traffic. Also, the increasing use of data processing equipment and computers will result in an increase of digital machine-to-machine traffic volume. This will open new areas and requirements and bring additional users into the Defense Communications System (DCS).

e. A significant example of digital traffic growth will be the widespread use of digitized voice to provide secure voice communications and bulk encryption. At present, digitized voice with speaker recognition cannot be transmitted within the standard 4-kHz telephone channel. While digital transmission may require wider bandwidths, it can use transmission media more efficiently and be multiplexed more cheaply than comparable analog systems. Digital system performance is more stable and allows better system design than analog systems.

f. The requirements for pictorial and analog voice communications are expected to increase gradually. The requirements for digital traffic are, however, forecast to increase at a rapid rate.

g. The growth in communications is not confined to the military and commercial systems of the United States. The systems of other countries are experiencing the same growth of traffic and new fields of technology are making possible additional communication services. These are adding their demands for frequencies to those of the existing military and the commercial communications users.

h. These worldwide demands placed against the limited number of usable frequencies available are rapidly exhausting the potential for new conventional radio communication systems. When coupled with the need for increased capability in military communications systems, the need to develop and implement new systems becomes even more pressing.

1-2. REQUIREMENTS FOR A MILITARY COMMUNICATIONS SYSTEM. While many of the requirements to be met by future long haul communications systems are the same for commercial and military use, the military systems must generally be more flexible. Military systems must provide communications that will permit early warning of high-speed attacks; that are reliable, flexible, and secure for transmitting decisions; and that can maintain absolute control of forces and weapons in nuclear, limited, or conventional war. In order to meet these requirements, military long haul communications systems must be reliable, invulnerable, secure, flexible, and of adequate capacity and quality. These requirements are discussed below.

a. *Reliability.* In order for a military communication system to fulfill its purpose it must be always available. The availability of a system depends on several interlocking factors; the reliability of the equipment and components employed, the reliability of the particular communication media employed, and the skill and knowledge of the personnel operating and maintaining the system. Each of these availability requirements is important, for reliable equipment is useless without skilled operating personnel, and skilled operating personnel cannot obtain dependable performance from poorly designed equipment.

(1) The requirements for reliability are much more stringent for a military system than for a comparable commercial system, for even a brief failure of the military system might have disastrous political and international consequences.

(2) The military system is often forced to provide this reliability while operating in a much more difficult environment than would ever be selected for a commercial system.

b. *Invulnerability.* A second distinguishing requirement of the military system as opposed to the purely commercial system is that the military system must be as invulnerable as possible to overt enemy action. This enemy action may take various forms, including physical destruction, capture of parts of the system, or the transmission of high powered jamming signals.

(1) To meet these invulnerability requirements, the military system must be designed and installed so that it can be protected from military action and be as resistant as possible to physical destruction. This is often accomplished through the hardening of installation sites. This is not economically feasible for commercial systems, although some commercial systems have been designed with the survival of enemy attack in mind.

(2) Provisions for combating jamming must also be included in the military system. These provisions normally take the form of incorporating, within the system, an ability to tradeoff system capacity during jamming in order to use techniques that make the system more resistant to jamming signals. Such equipment generally requires wider bandwidths than those normally used in commercial practice and is; therefore, usually incompatible with commercial systems.

c. *Security.* A third requirement placed on the military system is to deny enemy access to the information being transmitted. To ensure that the enemy is unable to decipher the information, encryption techniques are necessary. This imposes additional requirements on the communications system which are not normally satisfied by commercial communications systems.

d. *Flexibility.* Commercial communications systems grow in a fairly predictable pattern with the growth of large population centers. Military communications systems do not have a directly predictable pattern.

(1) Considering the present world political situation, requirements for reliable, invulnerable, secure communications for military purposes possibly may arise tomorrow in areas where there are no existing communication systems. No commercial system would, nor could, reasonably be expected to supply such flexibility.

(2) To meet this requirement, the military system must include not only fixed sites which will, as with commercial systems, be located in selected politically stable areas, but also must include transportable equipment which can be moved rapidly to areas where emergency demand might arise.

e. *Capacity.* As discussed in paragraph 1-1, the military services require more and more communications channels. This increase in requirements is especially true of the channels capable of handling digital traffic.

f. *Quality.* The quality of the transmission required in the military system differs from commercial practice. While the military may relax quality requirements for tactical circuits, the strategic long haul circuits are comparable to those in commercial systems.

1-3. **LIMITATIONS OF PRESENT TECHNIQUES.** The facilities for long haul trunking in use in present-day systems consist of high frequency (hf), ionospheric scatter, tropospheric scatter, multiple-hop line of sight (LOS), cable, and satellite. These methods have inherent limitations that prevent them from entirely meeting the requirements for which they were designed. These limitations are discussed below.

a. *High Frequency Limitations.* The propagation of signals in the hf band (3 to 30 MHz) over long distances is dependent upon the reflection of hf signals in the ionosphere. Ionization in this layer is due principally to ultraviolet radiation from the sun; as a result, the height of the ionized layer and the degree of ionization are subject to pronounced daily and seasonal variations. The degree of ionization is also influenced by sunspot activity; the effect of increased sunspot activity is to increase the maximum usable frequency. Since there is an observed cycle of approximately 11 years in sunspot activity, long distance communications in the hf band will also vary in an 11-year cycle. The variation in the usable frequency spectrum is on the order of 2:1.

(1) Figure 1-1 illustrates the important variations in maximum usable frequency of the hf band for single-hop transmission. Multiple-hop transmission is subject to the same variations but the effect is exaggerated with each hop. After the low point of the sunspot cycle, the maximum usable frequency gradually increases until the next maximum sunspot activity is reached. At this time the usable portion of the hf band will be more than twice as great as that at the sunspot activity minimum. Sunspot cycle peaks, from now until the end of this century, are predicted to be less active than recent peaks.

(2) Channel capacity in the hf band is limited most directly by the crowded conditions of the band. This band must be shared by many foreign and domestic

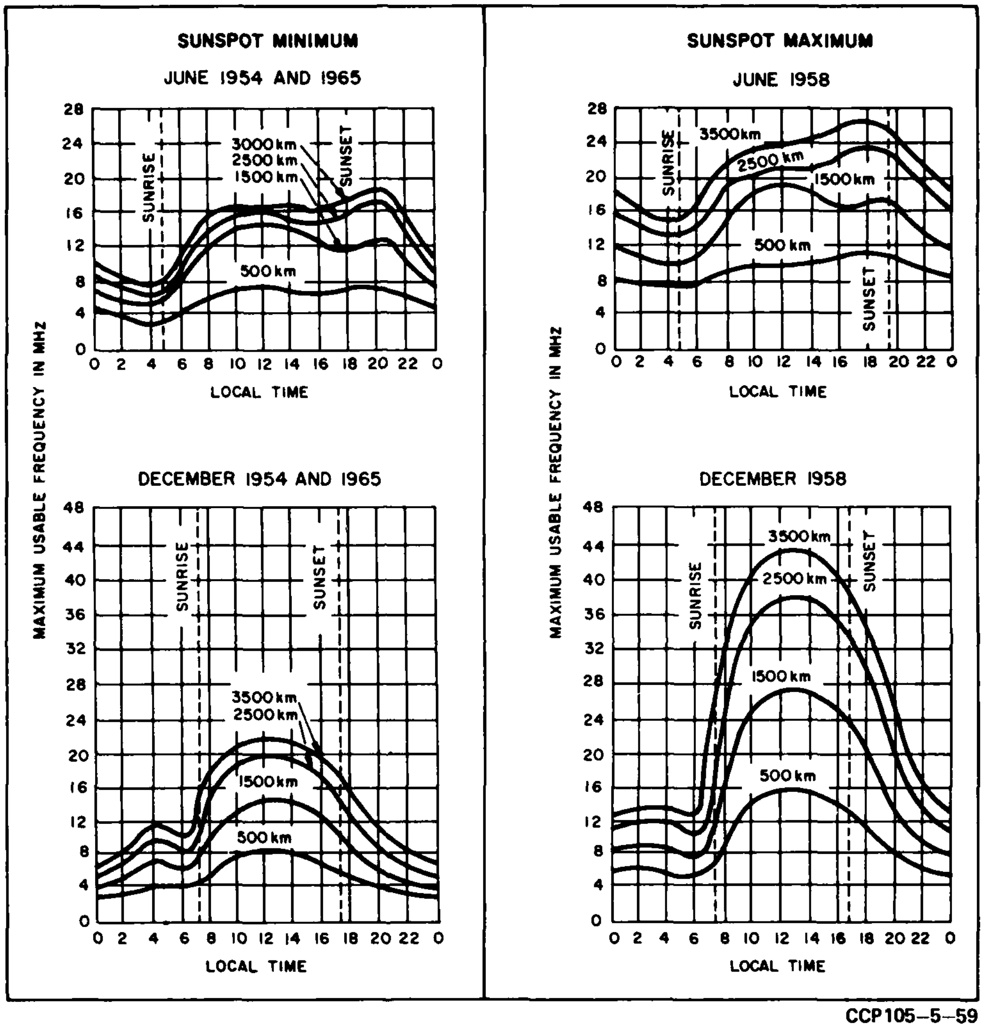


Figure 1-1. Graphic presentation of maximum usable frequency variation during sunspot cycle.

users and only a small portion is allocated to the DCS. New developments have permitted more efficient use of the available bandwidth slots; however, circuit requirements have grown coincidentally and the band is more crowded than ever. Also, only a small portion of this band is useful for point-to-point long-distance communication at any time.

(3) Reliability of long-distance hf propagation, in terms of military requirements, is considered poor. In addition to the variations in the usable spectrum that occur daily, seasonally, and over the 11-year sunspot cycle, there are occasions where the entire hf band becomes relatively useless due to magnetic storms caused by solar flares. These latter effects are particularly disrupting to communications to and from the polar areas. In addition, hf transmissions are subject to fading due to multipath effects and variations in the ionosphere. Service at hf is also subject to blackout as the result of nuclear detonations in the upper atmosphere.

b. *Frequency Limitations.* Point-to-point radio communications in the very high frequency (vhf) band (30-300 MHz), the ultra high frequency (uhf) band (300-3000 MHz), and the super high frequency (shf) band (3-30 GHz) may be accomplished by two basically different techniques.

(1) One technique uses numerous repeaters or radio relays spaced at, or near, the optical horizon. These systems are referred to as LOS systems. Where vhf is employed, the relays may be beyond the optical horizon (fringe area) due to refraction of vhf signals in the atmosphere.

(2) The second technique employs the phenomenon of radio wave scattering. Scatter propagation is not to be confused with reflection or refraction. The path loss is great in the scatter modes, and scatter propagation generally requires high-power transmitters and sophisticated receiving techniques. The principal characteristics of LOS and scatter systems are compared in table 1-1.

(a) While having numerous advantages over hf, the LOS and scatter systems do have limitations. Neither LOS, with the exception of satellite communications, nor scatter propagation will span a large body of water unless sites for repeater stations are available. In addition, ionospheric scatter is too limited in bandwidth to provide more than a fraction of the service needed for the DCS.

(b) Terminal and repeater stations are vulnerable to attack and sabotage. The vulnerability of LOS systems is particularly great because of the large number of repeaters. Jamming is possible with only moderate effort. Ground-based or airborne jammers can jam the side lobes of LOS receiving antennas or tropospheric scatter receiving antennas. Transportable terminals of LOS equipment are easily and quickly

Table 1-1. LOS versus scatter propagation

Parameter	LOS microwave relay	Ionospheric scatter	Tropospheric scatter
Frequencies	Uhf and shf (300 MHz-15 GHz)	Low vhf (30-50 MHz)	Uhf and shf (250-8000 MHz)
Transmitter power	Low 1/2-10 W	High 1-50 kW	High 1-100 kW
Distance to next repeater	Average of 30 mi for uhf, shorter at high shf frequencies	800-1300 mi	60-700 mi (usually about 100-200 mi)
Baseband band-width in terms of voice channels	Typical: 60-1800 voice channels	Several telegraph channels plus 1 voice channel	12-200 voice channels
Antennas	Horns or dishes on tower	Very large rhombic corner reflectors	Large parabolic dishes or reflectors 30-120 feet
Manned/unmanned repeaters	Unmanned	Manned	Manned
Diversity receivers	Desirable	Required	Required
Propagation reliability	99.99 percent +	99 percent	99 percent +

set up. It is possible, but not as easily accomplished, to use them as tropospheric scatter terminals as well. Such transportable terminals may be very large and require considerable setup time.

c. *Cable Limitations.* The use of cables for long-distance communications is expensive, especially deep sea cables. Cables are ordinarily used when radio techniques are not feasible or adequate. Long haul cables for wideband (on the order of several MHz) communications systems are generally coaxial transmission lines. Coaxial lines exhibit a nonlinear attenuation characteristic with a cutoff frequency. Other transmission line characteristics that affect the use of cable are the varying speeds of propagation at different frequencies and the relatively high attenuation. These characteristics require the use of equalizing repeater-amplifiers at frequent intervals along the line.

(1) Laying wideband cables over long distances is a long-term project, and once in place, the cable systems are relatively inflexible. Land cables are generally buried. Even when property acquisition is not considered, this is a long, involved operation. The design, manufacture, and laying of an intercontinental undersea cable may require 5 years or more, as such cables are definitely not off-the-shelf items.

(2) From a military standpoint, the greatest weakness of long haul cables is their vulnerability. A cable system offers an infinite number of points at which it may be attacked or sabotaged. An overland cable may be made fairly survivable if deeply buried. An undersea cable can easily be cut by either surface ship or submarine. There have been numerous instances where fishing trawlers have unintentionally snagged an undersea cable with their nets; and, in an effort to free the nets, have broken or cut the cable. Such breaks in the cable require considerable time and skill to repair. Cables, accessible to the enemy, are subject to interception and jamming, and to the insertion of spurious messages. This renders the security of the cable questionable.

1-4. **SATELLITE COMMUNICATIONS ADVANTAGES.** From the preceding discussion the need for new techniques and equipment is apparent. The military departments have long recognized the potential advantages inherent in the use of satellites and have had research and development programs in effect for a number of years. These programs are designed to result in a practical military satellite communications system that will incorporate all of the features required by the military. Additionally, communications systems employing satellite links offer a unique combination of advantages for long haul trunking. These advantages are discussed below.

a. *Capacity and Reliability.*

(1) Satellite systems are capable of handling thousands of voice frequency (vf) channels, although first-generation satellite systems were limited to less than a dozen.

(2) The reliability of active satellite communications systems is limited, essentially, only by the reliability of the equipment employed and the skill of the operating and maintenance personnel.

b. *Invulnerability.* While it is possible to destroy an orbiting vehicle, present developments in rocketry indicate that this would be quite difficult and expensive, when balanced against the tactical advantage to be gained. It would be particularly difficult to destroy an entire multiple-satellite system, primarily because of the number of vehicles involved. Owing to the relatively small number of ground stations in a satellite communications system, these offer a more advantageous target, but hardened ground sites can decrease system vulnerability considerably. Theoretically, satellite communications ground stations in certain instances are vulnerable to jamming attacks. However, by judicious planning and systems design, an enemy can be forced to make such great expenditures in time, effort, and material in order to jam a communications facility effectively that it would be unwarranted from his point of view.

(1) A rocket or satellite-based jammer, being power-limited, must be positioned within the main beam of the ground station to be effective. This implies that in a multisatellite system, a rocket or satellite jammer must be in approximately the same position and orbit as each communications satellite to be jammed. Considerable effort and precision on the part of the enemy would be required to achieve jamming.

(2) In any event, the bandwidth of satellite receivers is great enough to accommodate the use of spread spectrum techniques. Such techniques are very effective, even in the face of heavy jamming.

c. *Flexibility.* Van-housed satellite ground station equipment can be flown to remote areas and placed in operation in a matter of hours. Communications can be established with other satellite ground stations even though they are thousands of miles apart.

1-5. **APPLICATION OF THE DEFENSE COMMUNICATIONS SYSTEM.** The DCS is a communications network established by order of the Department of Defense (DOD) and placed under the direction of the Defense Communications Agency (DCA). The mission of the DCA is to ensure that the DCS will be so established, improved, and operated as to meet the long haul, point-to-point telecommunications requirements of the DOD and other governmental agencies, as directed.

a. The DCS meets the long haul, point-to-point telecommunications requirements of the DOD and provides facilities for command, intelligence, weather, logistics, and administrative purposes. The DCS provides the quality and quantity of communications capabilities required for these purposes.

b. The DCA network consists of switching centers located in various areas around the world and the links interconnecting them. Individual service units need only provide communications to the nearest switching center to become network subscribers and have access through the network to the entire system. Satellite communications form a part of the long haul links between the switching centers. These links employ the satellites in addition to other forms of existing communications media (hf, tropospheric scatter, ionospheric scatter, LOS microwave, and cable). The satellite trunks are placed in parallel with the trunks that use conventional means of communication. This provides added capacity between various points in the network and allows the various trunks to back up each other. This also provides the important reliability factor necessary to military communications.

c. The DCA is presently implementing the Phase II Defense Communications Satellite Program to link military headquarters, field commanders, and logistic centers throughout the world. Previously, the Defense Satellite Communications System (DSCS) consisted of less than 18 random orbiting satellites in subsynchronous orbits. Starting in 1972 the Phase II synchronous satellites were launched to provide additional channel capacity. Major satellite terminals are located in the United States, Okinawa, Hawaii, Guam, Australia, Korea, Thailand, West Germany, Turkey, and Ethiopia.

1-6. PAST EFFORTS IN SPACE COMMUNICATIONS.

a. *Background.* On 4 October 1957, the USSR successfully launched the first manmade earth satellite. This demonstrated man's ability to place objects into an orbit around the earth and brought to fruition a long-nurtured ambition. The tremendous potential of the specialized field of satellite communications had long been realized, and many plans had already been offered for such a communications system. In the October 1945 issue of *Wireless World,* Arthur C. Clark discussed the potential of satellite communications. In April 1955, John R. Pierce published a paper entitled *Orbital Radio Relays.* The United States has long been interested in satellites for a variety of reasons, and research in this field has been accelerated since the conclusion of World War II.

b. *Early Satellite Projects.* The first demonstration of the possibility of reflecting detectable electromagnetic signals from the moon was made by the US Army Signal Corps on 10 January 1946. A high-power radar set at the Evans Signal Laboratory, Belmar, New Jersey, was used for the test (fig. 1-2).

(1) The US Navy demonstrated, in 1951, that it was feasible to use the moon as a reflector of electromagnetic radiation (fig. 1-2). In 1957, the Chief of Naval Operations directed the establishment of a two-way telegraph and facsimile communications moon relay (CMR) between Washington and Hawaii. The CMR system was successfully used until November 1959, when solar disturbances in the ionosphere disrupted conventional hf circuits.

(2) On 18 December 1958, the Signal Corps Orbiting Relay (SCORE) communications satellite was successfully launched. As a practical demonstration of its capability, President Eisenhower recorded a Christmas message which was rebroadcast to the world via the satellite. SCORE operated for 12 days, during which time 97 contacts were made. SCORE was designed to receive message traffic as it passed over one station, record it, and retransmit the traffic as it passed over another station. This method of recording messages for later transmission is known as the store-and-forward technique.

(3) The Courier communications satellite was the joint responsibility of the US Army Signal Corps and the US Air Force. Courier was designed with a dual capability, permitting it to record and retransmit traffic and to function as a direct relay. The first launch attempt failed because of booster difficulties, but Courier IB was successfully injected into orbit on 4 October 1960. The experiment demonstrated many new techniques, provided valuable experience, and proved the feasibility of high capacity store-and-forward satellite communications.

(4) The Echo satellites (fig. 1-2) were essentially large inflated spheres with highly reflective surfaces. These were termed "passive satellites" as they contained no active electronic circuits and merely reflected all impinging energy. These were the largest-diameter satellites ever launched and were easily visible to the naked eye under favorable viewing conditions.

(5) The primary objective of Project Westford was to place an orbital belt around the earth consisting of millions of tiny dipoles. The first launch took place on 21 October 1961, but the experiment failed when the dipoles did not disperse in orbit. On 21 May 1963, a second Westford experiment package was launched and dipole dispersal took place as planned (fig. 1-2).

(6) Project Advent was a major effort directed at placing a multichannel active repeater satellite at synchronous altitude. It was planned to launch Advent,

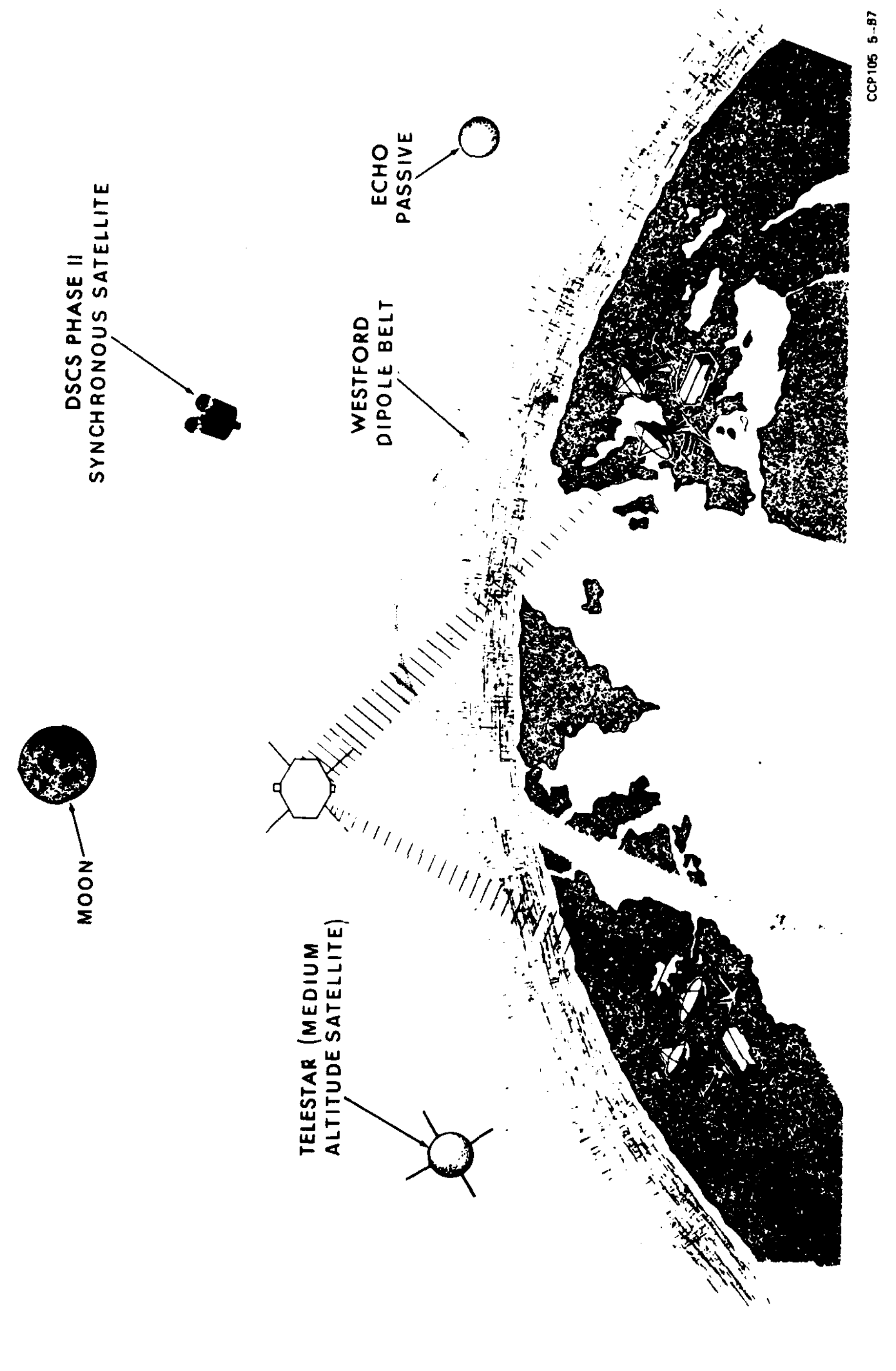


Figure 1-2. Satellite communication facilities.

which was to have weighed 1,300 pounds, using the Atlas-Centaur booster. Slippage in the Centaur program combined with other complications caused cancellation of the entire program, and the effort was reoriented into a medium altitude military communication satellite program, details of which are classified.

(7) Telstar was a medium altitude communications satellite project funded by the American Telephone and Telegraph Company. Telstar I was orbited on 10 July 1962, and used to transmit a live television program between North America and Europe. Telstar II, orbited on 7 May 1963, was placed in a higher orbit (6,702 miles) and equipped with some evacuated devices in an attempt to reduce ionization effects experienced with Telstar I.

(8) Relay, a medium altitude (820 to 4,612 miles) National Aeronautics and Space Administration (NASA) communications satellite weighing 172 pounds, was launched by a Thor-Delta booster on 13 December 1962. Relay's radio frequency power (10 watts compared to Telstar's 1.5 watts), its differences in circuitry, and various differences in construction allowed an evaluation to be made of various techniques by comparing the Telstar and Relay performance. Relay was used to transmit television from the United States to Europe and established the first link via satellite between the United States and South America.

(9) The synchronous communications (Syncom) satellites were designed to have orbits synchronized with the rotation of the earth. This simplifies ground station tracking equipment and provides a permanent link between ground stations able to see the satellites.

(10) Because the booster rocket could only place a limited amount of weight in a synchronous orbit, the Syncom satellites were relatively small and lightweight. This resulted in a lower communications capacity, compared to that of the low- and medium-orbit nonsynchronous satellites. As the state-of-the-art advanced, synchronous satellites with higher communications capacities became available.

(11) The Early Bird satellite, a commercial venture of the Communications Satellite Corporation, was a synchronous satellite stationed over the Atlantic Ocean. Launched on 4 April 1963, it was used to relay live TV coverage of the Gemini space vehicle recovery. It was also used in an extensive test program, which has served to pave the way for future designs.

(12) Since then, satellite performance has constantly improved. The newest International Telecommunications Satellite Consortium (INTELSAT) satellites can carry thousands of voice channels, as well as television and data traffic.

1-7. PRESENT AND FUTURE EFFORT IN SPACE COMMUNICATIONS.

a. *Initial Defense Communication Satellite Program.* The first seven Initial Defense Communication Satellite Program (IDCSP) satellites were orbited on 16 June 1966, the next eight on 18 January 1967, and the third launch of three satellites on 1 July 1967. On 13 June 1968, the DCA orbited the last 8 of 26 IDCSP satellites. Along with a worldwide system of ground stations and a waterborne terminal (USNS Kingsport), the satellites provide the DCS with a reliable, fairly high capacity worldwide communications network.

(1) The IDCSP was a triservice program with each of the military services having specific responsibilities under the direction of the DCA. The Air Force, as project manager, developed and launched the communications satellites with the Navy responsible for shipboard terminals. The Army, with the Satellite Communications Agency (SATCOMA) as project manager, developed the ground terminals and conducted the communications technical test program.

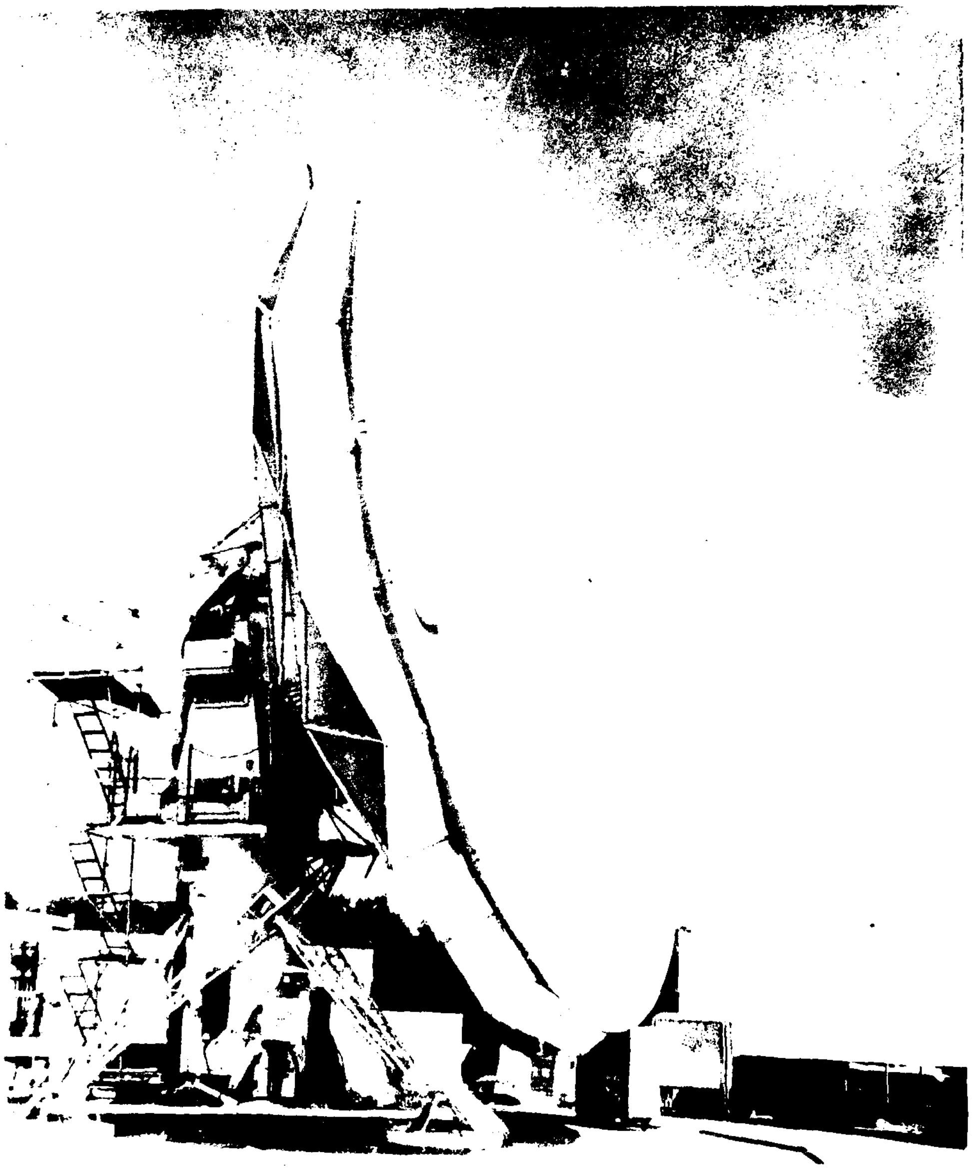
(2) Paragraphs b through d describe the IDCSP ground terminals and satellites and explain how the satellites were launched and deployed. A discussion of satellite orbiting considerations and launch and deployment techniques is contained in chapter 2.

b. *Earth Terminals.* The Fort Dix, New Jersey, and Camp Roberts, California, fixed stations served as the principal entry points for the satellite communications links from the Pacific and Europe. Originally built for the Advent program, these earth terminals were modified first for Syncom, then for the IDCSP, and finally the DSCS Phase II program.

(1) The surface complex is a mixture of fixed and transportable terminals with planned future augmentation by a new, smaller, transportable design. Included are the Satellite Earth Terminal AN/FSC-9 (fig. 1-3) used as fixed stations at Fort Dix, New Jersey, and Camp Roberts, California, the Air Transportable Satellite Communication Terminal AN/MSC-46 (fig. 1-4) deployed throughout the globe, and the lightweight terminal AN/TSC-54 (fig. 1-5).

(2) The AN/MSC-46 is the first terminal to be specifically designed for operational military satellite communications. The terminal is transportable by military cargo aircraft and consists of a 40-foot diameter parabolic antenna, three 30-foot vans, and three 100-kVA diesel generators. Semipermanent sites enclose the antenna with a rigid radome. AN/MSC-46 installations are located in the Continental United States, Hawaii, Korea, Okinawa, Ethiopia, and West Germany. A terminal has been assigned to the signal school at Fort Monmouth, New Jersey, for training personnel. The AN/TSC-54 can be transported along with its six-man

Figure 1-3. Satellite Earth Terminal, AN/FSC-9, Camp Roberts, California.



CCP105 5 89

Figure 1-4. Air Transportable Satellite Communications Terminal, AN/MSC-46.

Figure 1-5. Lightweight Terminal, AN/TSC-54.

crew by a C-130 aircraft and can be set up for communications via satellite within 2 hours after unloading.

c. *Satellite Description.* Each IDCSP satellite (fig. 1-6) is approximately 36 inches in diameter and weighs about 100 pounds. Each satellite consists of a mechanical structure, power system, communications system, and a telemetry system.

(1) The orbital periods of the satellites are between 22 and 23 hours, with an average rate of drift of about 1.3 degrees per hour. The mutual visibility time for a link terminal pair through a given satellite varies from a minimum of 0.25 days to a maximum of 5 days. Eclipse periods will cause downtime of up to 1.5 hours per day in the spring and fall, for a 45-day period. The orbit injection technique causes bunching and spreading in 18-month cycles. The satellites now are functioning operationally for the DCS with operational control exercised by the Satellite Communications Control Facility (SCCF).

(2) The satellite structure provides for the housing and passive temperature control of all vehicle elements, for satellite separation from the dispenser, and for spin stabilization of the satellite. Solar cell panels cover the outside of the structural frame. The internal surfaces are thermal coated to maintain the in-orbit operating temperatures for the communications, telemetry, and power supply subsystems.

(3) The power system, which contains no batteries, consists of a power control unit, an array of solar cells, and the radio frequency (rf) radiation termination switch. The solar array provides electrical power while the power control unit regulates and distributes the solar array power.

(4) The satellites operate on the same frequency (8-GHz range) with a bandwidth of 50 MHz. Beacon signals (modulated with 110- to 130-kHz tones) generated by each satellite are used by the communications ground stations for satellite acquisition and tracking, and for individual satellite identification. There are two traveling-wave tube (TWT) power amplifiers, each providing a maximum of 3 watts to the communications antenna. Automatic changeover to the second TWT occurs when the performance of the first TWT deteriorates.

(5) The telemetry system includes a telemetry generator, transmitter, sun angle sensor, antenna, and instrumentation sensors that measure critical satellite parameters. Outputs from the sensors, along with individual satellite identifications, are processed as 56 analog telemetry signals and 18 binary signals. Each of the launch groups has its own 400-MHz telemetry frequency. When the power level deteriorates, the telemetry transmitter is automatically turned off to conserve power for the communications transponder. The telemetry transmitter is automatically turned back on when the power level increases above a preset level.

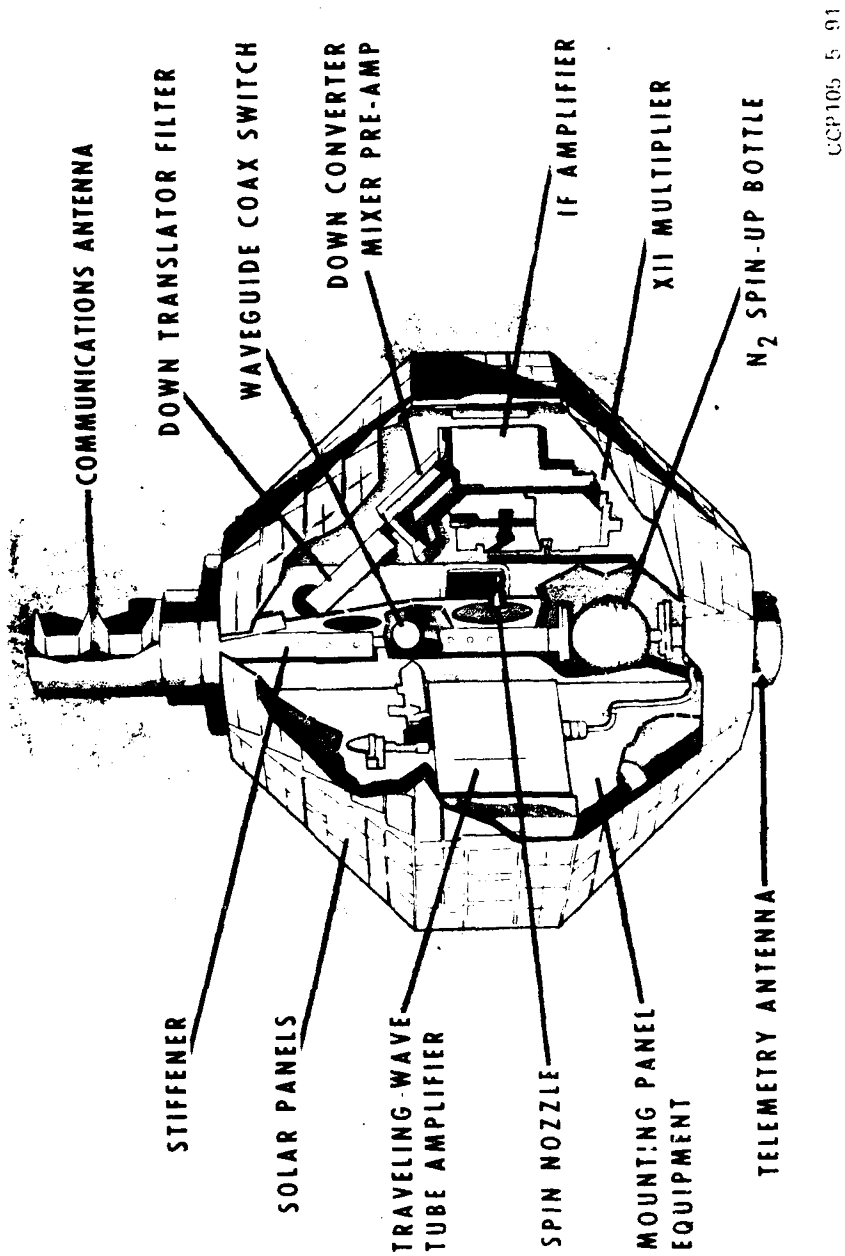


Figure 1-6. IDCSP spin-stabilized communications satellite, cross sectional view.

d. *Launch and Deployment.* The IDCSP satellites were contained in a dispenser that was launched by a Titan IIIC rocket.

(1) Following lift-off, the Titan IIIC rolled counterclockwise from a 100.2-degree launch azimuth to a flight azimuth of 93 degrees. The solid motors were jettisoned 2 minutes later and fell into the Atlantic Ocean. After another 2 minutes and 20 seconds, the first-stage engines burned out and were jettisoned. In another 2 minutes the fairing was jettisoned and after another 3 minutes, the second stage burned out and was jettisoned.

(2) After 15 seconds, the transtage motors were operated for 17 seconds to place the transtage and spacecraft dispenser in a parking orbit of 90 nautical miles (nmi). After 58 minutes in the parking orbit the transtage motors were ignited for the second time to place the spacecraft in the transfer orbit.

(3) The spacecraft then coasted for a little less than 5 hours. During this time it climbed from 90 to 18,200 nmi. In the transfer orbit, the attitude control system reversed the position of the spacecraft relative to the sun approximately once every hour in order to distribute the solar radiation heat. These sun orientation maneuvers were designed to protect the spacecraft and its payload of communications satellites from the temperature extremes that occur between the sun and shade sides of a spacecraft.

(4) At 18,200 nmi, the transtage was ignited for the third time for approximately 2 minutes, to inject the spacecraft into orbit.

(5) The eight satellites were sequentially ejected 2 minutes later. There was an interval of 17 seconds between the first and second satellite ejections. This interval gradually increased to 26 seconds between the seventh and eighth ejections. During these times, the vehicle attitude was stabilized and the attitude control system motors fired to obtain required incremental velocities. The first satellite had an orbital velocity of approximately 10,311 feet per second (ft/s). The other satellites had increasingly larger velocities, which served to gradually disperse the satellites and placed each one in an orbit with a higher apogee.

(6) The satellites were carried into orbit aboard a satellite dispenser structure attached to the transtage. Behind each satellite was a dispensing mechanism that consisted primarily of a base plate and four springs. For separation, a signal from the guidance system started a sequence that released a clamp restraining the satellite. The springs forced the satellite laterally away from the dispenser at about 3 ft/s. The satellites were ejected, either north or south, perpendicular to the eastward line of flight (orbit plane).

(7) On ejection, the nitrogen spin-up system in each satellite imparted a rotation of 150 revolutions per minute to the satellite. The satellites spin about the communications and telemetry antennas axes keeping these antennas always in sight of the earth. Spinning also continually exposes different sides to the sun, providing nearly uniform temperatures within the satellite.

(8) A final 6-second burn of the attitude control system motors began 1.5 seconds after the last satellite had been released. This burn removed the transtage and payload dispenser from the vicinity of the satellites. The transtage and payload dispenser will orbit indefinitely at an altitude of 18,200 to 18,500 nmi.

e. *Defense Satellite Communications System.* Starting in 1972, the IDCSP satellites were phased out and replaced with the newer synchronous satellites. The new DSCS Phase II satellites have the potential of carrying up to 1 300 full-duplex voice channels, which is much greater than the older IDCSP satellites.

f. *Earth Terminals.* The three types of terminals used in the IDCSP program were modified to take advantage of the increased capabilities of the newer synchronous satellites. The AN/FSC-9, AN/MSC-46, and the AN/TSC-54 terminals could operate on both the old IDCSP frequencies and the new Phase II frequencies. The program will be carried out in four steps as follows:

(1) *Stage 1-A.* The terminals will be modified to operate on the new satellite frequencies. Operation through the new satellites will be by frequency division multiple access (FDMA), with a central controller to coordinate channel assignments.

(2) *Stage 1-B.* In this phase the new synchronous satellites will be used in a network configuration employing a multiple carrier capability in the AN/MSC-46 and AN/FSC-9 terminals. Within a network any terminal can communicate with several others. Spread spectrum techniques will provide an additional multiple access means, as well as provide a jam-resistant communications channel. The new heavy transportable (HT) and medium transportable (MT) terminals will be introduced.

(3) *Stage 1-C.* This stage commences with the introduction of the developmental pulse code modulation (pcm), time division multiplex (tdm), and phase shift key (psk) modulation equipments on the first Phase II satellite communications link. Since this new digital equipment will be phased into the Phase II DSCS over an extended period of time, the initial period of Stage 1-C will be a "hybrid" FDMA operation; that is, a mixture of analog and digital operation on separate rf carriers.

(4) *Stage II.* In this stage, time division multiple access (TDMA) techniques will be used. Many stations will use the same frequencies, but will transmit in

specific time slots. For a network involving only a few large stations, TDMA is much more efficient than any other multiple access method.

(5) *Future terminals.* Three new types of terminals are being developed for future use. One is an HT AN/MSC-60 terminal station which will be used as a major nodal terminal on long haul circuits (fig. 1-7). The terminal will have a 60-foot parabolic antenna and weigh less than 400,000 pounds. The second station is the MT AN/MSC-61 terminal similar to the HT but uses a 35-foot diameter antenna. The electronic equipment is basically identical to that for the HT with the antenna's weight, complexity, and gain constituting the main differences. The third is the lightweight (LT) AN/TSC-54 terminal (fig. 1-8) which will be developed with two antenna subsystems.

g. *Satellite Description.* The new satellite will be much larger and more complex than the IDCSP satellites. The satellites weigh 1,000 pounds and have the potential for up to 1,300 full-duplex voice channels (fig. 1-9).

(1) The antennas are mechanically despun. This allows the antennas to remain pointed towards earth while the other components of the structure rotate to spin-stabilize the satellite. Two of the antennas are earth coverage horn antennas. They have a gain of 16.8 decibel/isotropic (dBi) antenna and a beamwidth of 18 degrees. The other two antennas are parabolic dishes for narrow beam coverage of a particular area on the earth. The gain of the dishes is 33 dBi, and the beamwidth 2.5 degrees.

(2) The communications portion of the electronic equipment contains redundant components for increased service life and reliability. Four TWT amplifiers are carried. One TWT feeds the horn antenna, another the parabolic antenna, with the remaining two being standby units. The amplifiers are designed to operate as wideband low-distortion amplifiers with power outputs of 20 watts.

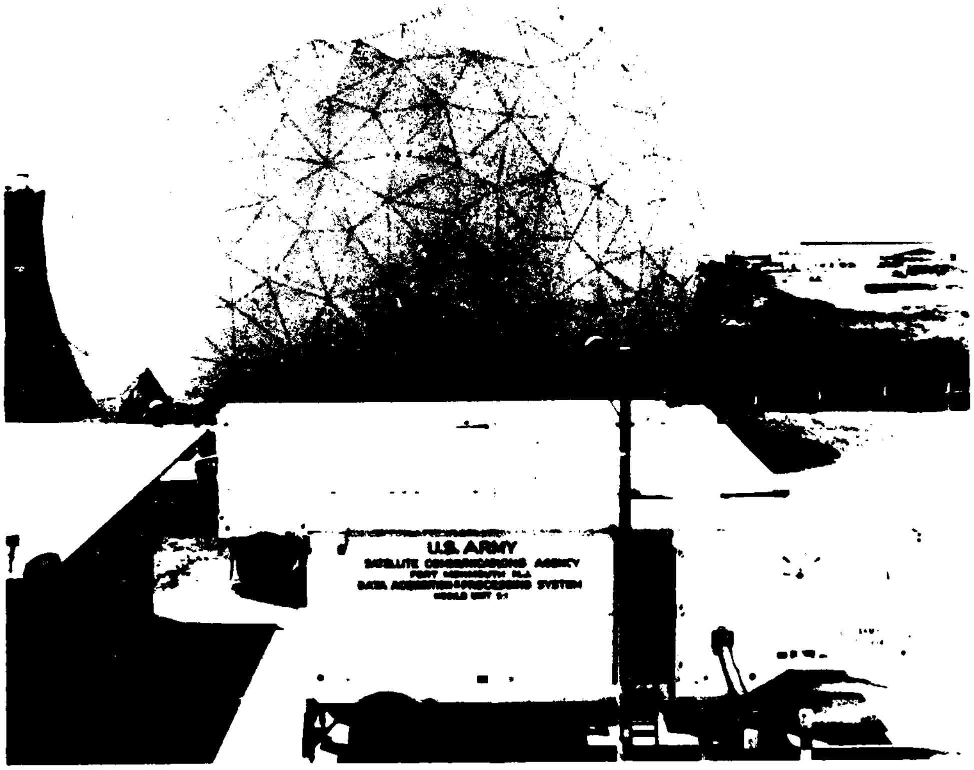
(3) The satellite operates with four different frequency channels (table 1-2, part 4). These channels can be used in various combinations to provide for flexible operation.

h. *Launch and Deployment.*

(1) The new Phase II satellites will be launched by a Titan IIIC rocket into a synchronous equatorial orbit. Because of their larger size and weight only two of the newer satellites can be launched with one booster. The launch sequence will be similar to that of the IDCSP launches, except that the booster will place the satellites at a synchronous altitude. The initial positioning is expected to be within 3 degrees of the desired equatorial subsatellite point.

CCP105-5-9

Figure 1-7. Heavy transportable terminal, exterior side view.



CCP105–5–93

Figure 1-8. Medium transportable terminal, with radome erected.

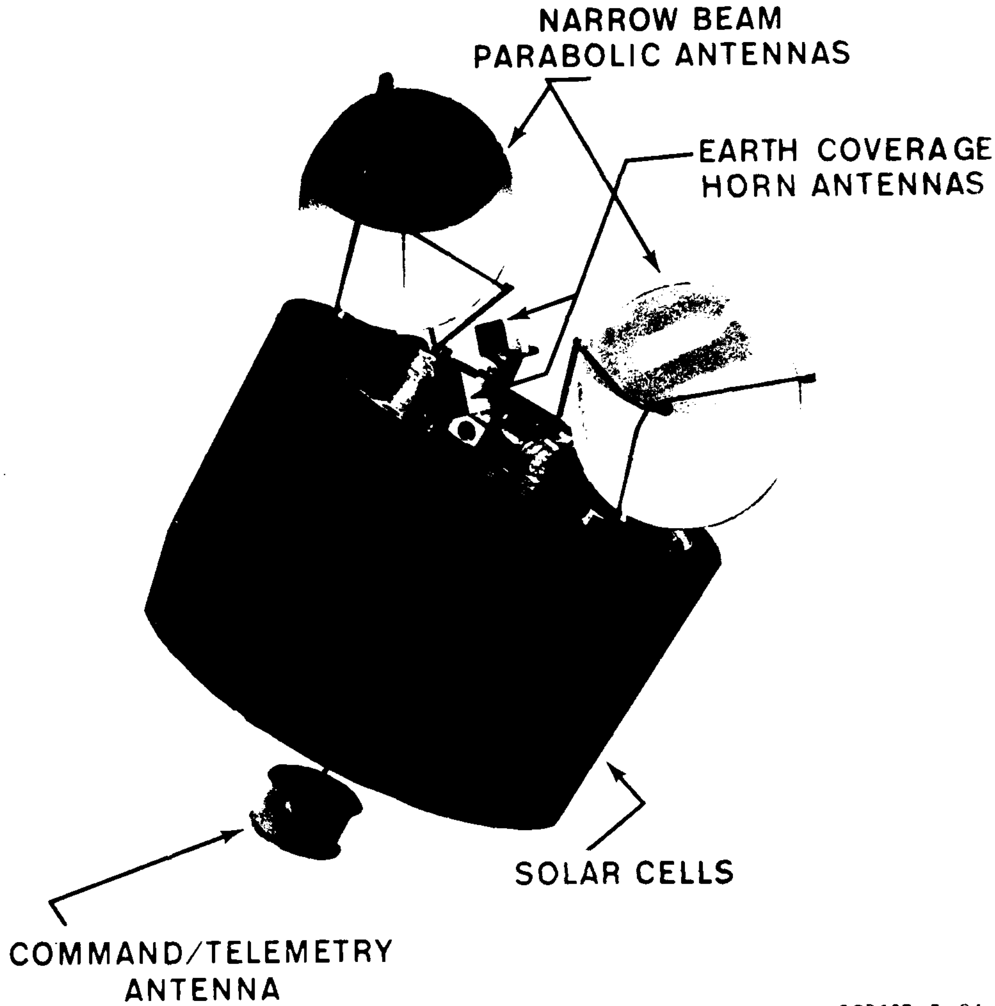


Figure 1-9. Phase II synchronous satellite, exterior view.

(2) The satellites can reposition themselves once during their operational life to another equatorial subsatellite point. This would allow a satellite to replace one that had become inoperative or to provide communications with a remote area not normally covered by DSCS satellites.

i. *Tactical Satellite Communications.* The DOD is currently investigating the use of satellites to provide tactical communications. The Tactical Satellite Communications (TACSATCOM) program involves all military departments (MILDEPS) and is still in the development stage.

(1) The MILDEPS have developed terminals for their needs; the Air Force is developing airborne terminals; the Navy is developing shipboard terminals, and the Army and Marines are developing ground based terminals.

(2) The Navy will be the first MILDEP to use the satellites for operational tactical use. The AN/WSC-2 shipboard terminal will allow the Navy to communicate directly with its worldwide fleet and eliminate its dependence on low, medium, and hf radio.

j. *Types of Terminals.* The MILDEPS have developed a series of terminals for tactical use to meet three basic needs. One terminal is a broadcast warning receiver that will have no transmitting equipment. The satellite will be used to transmit warnings or general information type messages to a large number of units scattered throughout a large area. A second terminal will be for netting. The netting capability will allow a number of units within an area to communicate with each other. Only single channel communications would be necessary. The third terminal would be for multichannel links. This use would be limited to higher headquarters that require a number of communications channels to other headquarters.

k. *Frequencies.* TACSATCOM plans to use two separate frequency bands. One is at uhf (235-400 MHz) and the other at shf (8 GHz). The frequency selected will depend on the ultimate use of the equipment. Multichannel equipment will use the wider bandwidths available at shf while small field terminals, which require simple, rugged, and lightweight equipment, will use uhf. The TACSATCOM satellites themselves will handle both frequency bands and can be commanded to operate cross band (uhf to shf or shf to uhf), if necessary. At present, plans for TACSATCOM are not yet firm. Future development and research work, as well as the needs of the MILDEPS, will determine the ultimate course of the program.

l. *Civilian Satellite Communications.* The civilian applications of satellite technology have kept up with or have been ahead of the military. The latest series of civilian communications satellites, the INTELSAT IV's, are as advanced as the DSCS Phase II satellites. Besides their use in fixed point-to-point communications, some expected civilian use of satellites includes television broadcast and distribution, aeronautical communications and navigation, maritime communications and navigation, deep space research and data relaying, satellite-to-satellite relaying, and air surveillance.

m. *Operating Systems.* Numerous proposals for systems exist, and many will be implemented in the near future. At present the uses are restricted to point-to-point communications. The INTELSAT that controls satellite communications has set up an international system using the 4- and 6-GHz frequency bands. Canada is starting its own domestic satellite system with the launching of the Anik satellite. Besides carrying point-to-point communications, the Anik satellite will be used for distributing television programs to outlying areas presently not serviced by TV networks. The Soviet Union is also developing a system similar to the Canadian program. Both face similar problems of covering large land areas that are sparsely populated.

n. *Proposed Systems.* In the United States, numerous proposals exist for using satellites. In the area of point-to-point communications the newer systems will use the higher frequency bands of 11 to 14 GHz and 20 to 30 GHz. The American Telephone and Telegraph Company has proposed a system that would provide 100,000 full-duplex voice channels, or their equivalent, in the 20- to 30-GHz bands. The television networks have proposed a system that would allow much greater flexibility in distributing television programs, as well as picking up remote video feeds. The frequencies planned for use would either be 7 or 12 GHz and would permit the networks to broadcast directly to their affiliate stations

(1) On an international scale, work is being done on the proposed Aerosat system. This would use a satellite to communicate with aircraft when they are over ocean areas. The links would also allow route supervision by air controllers and more precise navigation by aircraft. A final Aerosat system has not been selected yet, although propagation tests at various frquencies have been performed to evaluate equipment.

(2) As equipment and technology improve, higher frequencies will be used for satellite communications. The higher frequencies in the extreme high frequency (ehf) range will provide the wide bandwidths necessary for the communications needs of the future. Table 1-2 shows the frequencies allocated for the various applications of satellite communications.

Table 1-2. Allocated satellite communications frequencies

1. Point-to-point fixed communications:

   4/6 GHz bands
   3.7 - 4.2 GHz downlink
   5.925 - 6.425 GHz uplink

   11/14 GHz bands
   10.95 - 11.70 GHz downlink
   14.0 - 14.5 GHz uplink

   20/30 GHz bands
   17.7 - 21.2 GHz downlink
   27.5 - 31.0 GHz uplink

   11.7 - 12.2 GHz downlink or broadcast Region 2

   40 - 41 GHz downlink

   50 - 51 GHz uplink

   92 - 95 GHz uplink

   102 - 105 GHz downlink

   140 - 142 GHz uplink

   150 - 152 GHz downlink

   220 - 230 GHz no direction specified

   265 - 275 GHz no direction specified

2. Television broadcast and distribution:

   620 - 790 MHz broadcast service (limited)

   2500 - 2690 MHz distribution

   6625 - 7125 MHz TV network distribution (only in the US, Canada, and Brazil)

Table 1-2. Allocated satellite communications frequencies--continued

11.7 - 12.5 GHz broadcast Region 1

11.7 - 12.2 GHz broadcast Region 3

3. Aeronautical services:

1543.5 - 1558.5 MHz communications

1558.5 - 1636.5 MHz navigation

1645 - 1660 MHz communications

43 - 48 GHz aircraft/maritime, communications/navigation

66 - 71 GHz aircraft/maritime, communications/navigation

95 - 101 GHz aircraft/maritime, communications/navigation

142 - 150 GHz aircraft/maritime, communications/navigation

190 - 200 GHz aircraft/maritime, communications/navigation

250 - 265 GHz aircraft/maritime, communications/navigation

4. Defense Communications System:

235 - 328.6 MHz communications tactical/nautical

335.4 - 399.9 MHz communications tactical/nautical

DSCS Phase II:

Channel 1:	7975 - 8100 MHz uplink earth coverage 7250 - 7375 MHz downlink earth coverage 7250.1 - beacon, earth coverage
Channel 2:	8125 - 8175 MHz uplink narrow beam 7400 - 7450 MHz downlink earth coverage

Table 1-2. Allocated satellite communications frequencies--continued

Channel 3:	8215 - 8400 MHz uplink narrow beam 7490 - 7675 MHz downlink narrow beam 7675.1 - beacon, narrow beam
Channel 4:	7900 - 7950 MHz uplink earth coverage 7700 - 7750 MHz downlink narrow beam

5. NASA deep space research and data relay:

   13.25 - 14.2 GHz

   14.4 - 15.35 GHz

6. Satellite-to-satellite relay:

   54.25 - 58.2 GHz

   59 - 64 GHz

   105 - 130 GHz

   170 - 182 GHz

   185 - 190 GHz

CHAPTER 2

ORBITAL MECHANICS

2-1. **INTRODUCTION.** A manmade satellite, regardless of its tasks and functions, has orbital characteristics the same as those of other astronomical bodies. Orbital characteristics are dictated by physical principles which govern the choice of an orbit and the weight of equipment that can be incorporated in the manmade satellite. In order to understand how orbits are chosen and what limitations exist, a background in the basic mechanics of orbiting bodies is provided in this chapter. Orbital mechanics are also involved in the launch of satellites and the placing of a satellite in orbit.

2-2. **PHYSICAL LAWS AND FORCES GOVERNING SATELLITES.** The basic physical laws governing satellites (natural or manmade) were formulated from observations of celestial bodies made over many centuries. Launching and orbiting a manmade satellite requires a knowledge of these laws since they govern the shape, altitude, and location of orbits. The booster rocket size and the type of orbit desired limits the payload weight and the amount of equipment a satellite can carry.

a. Astronomers and other scientists, working from observations of natural phenomena, postulated laws describing these observations and provided the foundation on which the science of celestial mechanics is built.

b. The most prominent of these men were Sir Isaac Newton, Johannes Kepler, and Galileo Galilei. These principles are basic to all physical phenomenon and were not formulated specifically for the study of satellites. These principles are stated briefly in the following paragraphs and are followed by specific applications to the problems of launching and orbiting a satellite.

(1) *Newton's laws.* Newton's first law was stated as, "Every body persists in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by forces impressed on it."

(a) The second law was stated as, "The change of motion is proportional to the motive force impressed, and is made in the same direction as that of the impressed force." Stated simply in equation form the law becomes:

$$\vec{F} = \frac{d}{dt}(m\vec{v}) = m\frac{d}{dt}\vec{v} = m\vec{a} \qquad (2.1)$$

($\rightarrow$ indicates vector quantities)

where

$\vec{F}$ is force

m is mass

$\vec{v}$ is velocity

$\vec{a}$ is acceleration

(b) Newton stated his third law of motion as, "To every action there is always an equal and opposite reaction; or, the mutual actions of two bodies upon each other are always equal, and oppositely directed."

(c) These laws of motion, action, and reaction apply to any object whether or not it is an astronomical body.

(2) *Kepler's laws.* Kepler's laws are the result of astronomical observations reduced to mathematical formulas. They apply directly to the mechanics of the solar system and the calculation of satellite orbits.

(a) Kepler's first law is that the orbit of every planet is an ellipse with the sun at one focus. This may also be read that the orbit of any satellite is an ellipse with the parent at one focus.

(b) Kepler's second law states that a line joining any planet to the sun sweeps out equal areas in equal time (the law of areas).

(c) Kepler's third law states that the cube of the mean distance of a planet from the sun is proportional to the square of its period (the law of periods).

(d) The term, *mean distance,* refers to the average distance from the center of the sun. In the case of a near-earth satellite there is a big difference between its altitude and its distance from the center of the earth. The average distance from center to center is the same as one half the length of the major axis. The long axis of an ellipse is called its major axis, while the short axis is called the *minor axis.*

(3) *Law of universal gravitation.* This law was derived by Newton and is stated, "Each particle of matter attracts every other particle of matter with a force which is directly proportional to the product of their masses and inversely proportional to the square of the distance between them." Expressed mathematically the law states that the attractive force between two bodies is

$$F = \frac{Gm_1 m_2}{r^2} \quad (2.2)$$

where

F is force (Newtons)

m_1 and m_2 are the masses of the bodies (kilograms)

r is the distance between the bodies (meters)

G is the gravitational constant (6.67×10^{-11} Nm^2/kg^2)

(4) *Centrifugal force.* For an understanding of orbital motion, it is helpful to introduce the following principle: The rate of change of radial speed of a satellite with respect to its parent is proportional to the difference between the gravitational force and centrifugal force. This principle is concerned strictly with changes in the distance from parent to satellite. It says that when centrifugal force is greater than gravitational pull, the radial velocity of the satellite tends to increase in magnitude. This principle can be derived from Newton's second law.

(a) Centrifugal force can best be illustrated by the following example: If an object is tied to a string and whirled in a circular path, then (excluding the presence of gravity) the pull, or force, felt against the string is the centrifugal force. In short, it is the outward force from the center, exerted as a result of the velocity of the object. The velocity is transverse, or perpendicular to the string. To maintain the circular orbit, the inward pull, or force, by the string must exactly equal the centrifugal force. This outward force (centrifugal force) can be shown to be:

$$F = \frac{mv^2}{r} \quad (2.3)$$

where

F is force (Newtons)

r is radius of circle (meters)

v is velocity of the object in the direction perpendicular to its radius vector (meters/second)

m is mass of the object (kilograms)

(b) The mass of a body is always constant, even in "weightless" space. The weight of a body changes, depending on the gravitational field. The result is that a man weighs less while on the moon even though his actual mass has not changed.

(c) As long as circular motion is maintained, centrifugal force will equal the inward force but will be exerted in the opposite direction. Equation (2.3) for centrifugal force holds for noncircular orbits as well, as long as one is careful to regard v as the component of velocity perpendicular to the radius vector. This velocity component is called the transverse velocity. For circular motion it is the same as the total velocity.

(5) *Application of physical laws.* From the previous paragraphs a relationship between the velocity, mass, and orbital radius necessary to maintain a satellite in orbit can be derived. From equation (2.2) it can be seen that gravity, or inward force, is proportional to the product of the masses divided by the square of the distance between the parent and the satellite. Also, from equation (2.3) the outward or centrifugal force is equal to the product of the mass of the satellite and the square of its transverse velocity divided by the distance between the bodies. It is concluded that in order to maintain an orbit, these two forces must be equal. Upon equating these two forces,

$$\frac{m_2 v^2}{r} = \frac{G m_1 m_2}{r^2} \qquad (2.4)$$

where

m_2 is mass of satellite (kilograms)

v is transverse velocity (meters/second)

r is distance between bodies (meters)

G is gravitational constant (6.67 × 10^{-11} Nm^2/kg^2)

m_1 is mass of parent (mass of earth = 5.98 × 10^{24} kg)

(a) Cancelling terms, we find the necessary transverse velocity to maintain an orbit is:

$$v = \sqrt{\frac{G m_1}{r}} \qquad (2.5)$$

(b) It can be seen that if v is [illegible] the satellite toward the parent until a stable [illegible] if v is too high the satellite will move [illegible] at a higher altitude. In all cases, the satellite [illegible] and centrifugal forces balance. [illegible] the velocity changes during the orbit. A [illegible] therefore, a constant value of v. With [illegible] bodies is always changing, so by equation [illegible] always be changing.

(c) Kepler's second [illegible] states [illegible] center of the parent planet to the satellite [illegible] in equal times. In figure 2 1, area COD includes the apogee (the point in the orbit at which the satellite is most distant from the parent). Assume [illegible] takes the satellite 20 minutes to transverse the distance from point C to point D. Now moving to the other end of the orbit which includes the perigee (or the point at which the satellite most closely approaches the parent), we can construct an area QOR which will exactly equal that of COD. Kepler' law states that the satellite must transverse the portion of the orbit between Q and R in the same length of time that it took to cross from C to D. It is obvious that the satellite must move at considerably greater speed, since the peripheral distance along the path from Q to R is considerably greater.

(d) By applying this same relationship to other segments of the orbit, a generalization may be made about satellites. That is, the greater the altitude the less the velocity of the satellite. This can be related to the opposition of gravity and centrifugal force. The farther from the sun the planet is, as it approaches

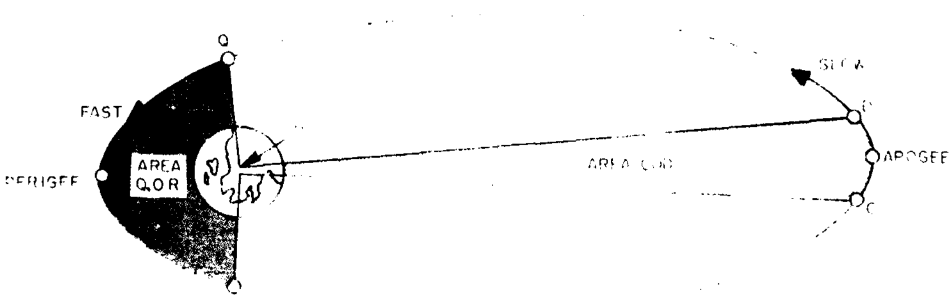


Figure 2 1. Representation of Kepler's second law, orbital diagram apogee, for example, the less inward force is exerted on it in terms of gravity from the sun. At the same time, when the velocity and particularly the transverse velocity of the planet is less, it requires less inward force to hold it on its path.

(e) As the earth approaches the sun at perigee the gravitational attraction between the earth and the sun increases, but so does the transverse velocity of the planet; as a consequence, a balance is established that prevents the earth from either being drawn into the sun by the increased inward force, or from shooting off into space because of its increased velocity.

(f) The question may arise that if gravitational and centrifugal forces are always balanced, why do satellites assume elliptical orbits where r and v are always changing? The answer lies in looking at all the forces acting on the satellite. For circular orbits the velocity of the satellite is always perpendicular to the line connecting the parent and the satellite. With an elliptical orbit the satellite has an additional velocity component either toward or away from the parent. The additional motion was given to the satellite when it was placed in orbit. This is shown in figure 2-2.

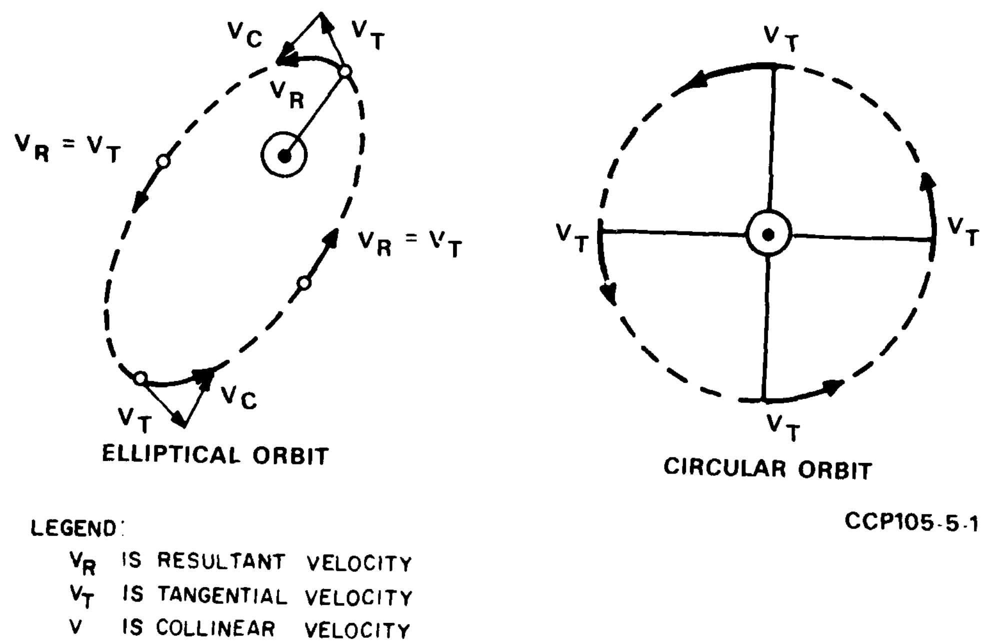


Figure 2-2. Forces acting upon a satellite, orbital diagram.

(g) *The elliptical orbit* has a radial or colinear component of velocity v_c, in addition to the tangential component v_t. The resultant velocity, v, is the vector sum of these two velocities.

(6) *Artificial satellites.* Artificial satellites obey the same set of laws and follow the same relationships as the planets in the solar system. Orbit shapes, orbit selection, and the mechanics of launching an artificial satellite will be discussed.

(a) Before these specifics are discussed it is well to present in more detail the calculations bearing on the ellipse and the shape of all elliptical orbits. These shapes are generally referred to as Keplerian orbits and describe all satellite orbits. An ellipse is a geometrical shape described as a plane curve generated by the path of a point, the sum of whose distances from two fixed points is constant. Figure 2-3 illustrates the geometrical ellipse.

(b) The terms applied to the ellipse in this discussion are the major axis and the minor axis as illustrated in figure 2-3. The major axis is a line drawn through the two foci (F and F') and intersecting the curve at A and A'. The minor axis is a line at right angles to the major axis at a point midway between the foci; it intersects the curve at B and B'. The point where the axes intersect is indicated as O and the lines joining this intersection with A and B are designated a and b, respectively.

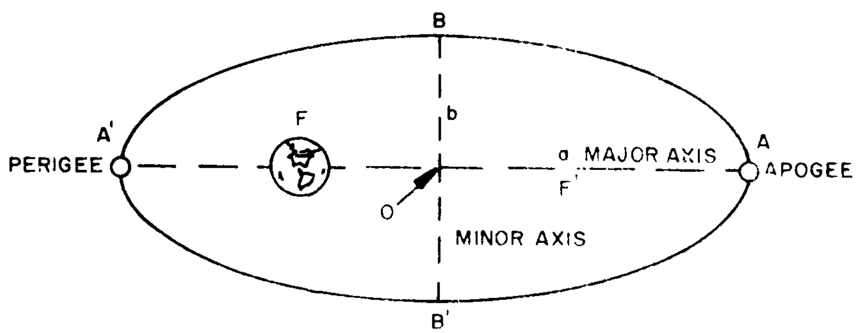


Figure 2-3. Diagram of elliptical satellite orbit

(c) An ellipse may be plotted as a graph by using the formula:

$$y = b\sqrt{1 - (x/a)^2} \tag{2.6}$$

Another way of writing the equation is:

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{2.7}$$

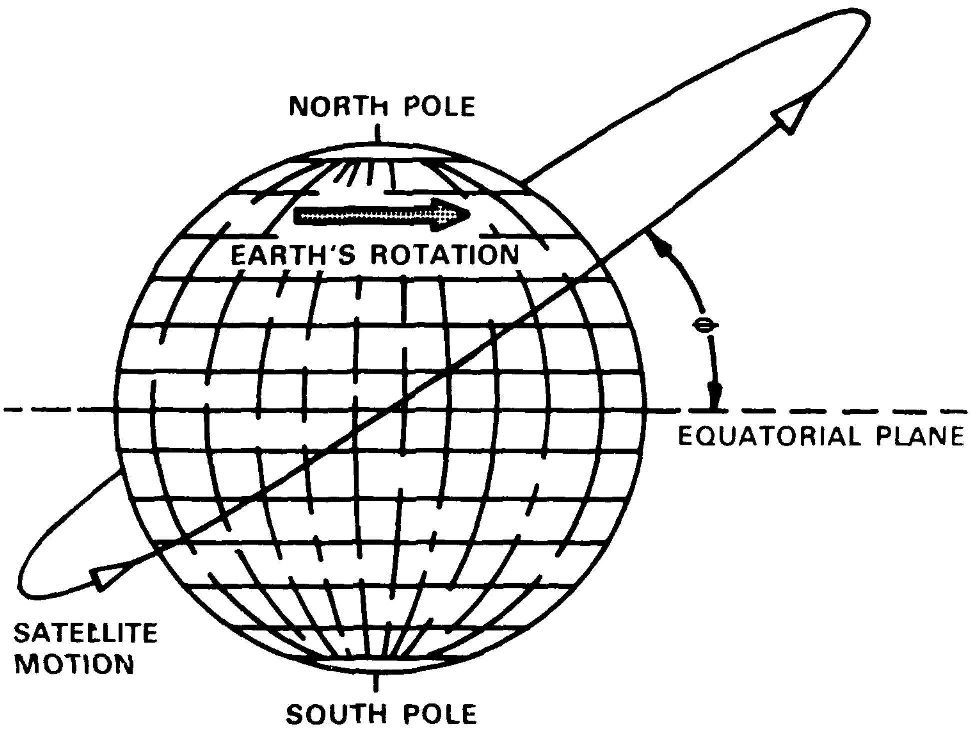
(d) Of special interest is the location of the foci. If they are coincident so that a and b are equal, the result will be a circle. As the foci move further apart, so that the difference between the lengths of a and b becomes greater, the ellipse becomes more elongated (eccentric). In specifying the shape of an orbit the values for a and b, together with the value of e, are usually given. The term e is called the eccentricity of an orbit and is defined mathematically as:

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \tag{2.8}$$

The larger e is, the more eccentric the orbit, until, in extreme cases, e is just slightly less than 1. For circular orbits e is zero.

(e) The shape of the orbit is important to the application in the communication system of the satellite. The types of orbits are designated as either synchronous or low-altitude random orbits. The synchronous orbiting satellite is one that has been located at such an altitude that it appears to hang more or less motionless over some point on the earth's surface. It can be seen that, in order to fulfill this requirement and to be synchronized with the rotation of the earth, the satellite will have to be in a very nearly circular orbit and will require its orbital plane to be coincidental with the plane of the earth's equator. The velocity of the satellite must remain constant and must equal the rotational velocity of the earth at the equator.

(f) Low-altitude satellites may have a very eccentric orbit and be inclined to the plane of the equator. This inclination angle θ is another of the elements that must be specified in describing an orbit. It is important that inclination not be confused with the plane of the earth's orbit and remember that it is referenced to the plane of the equator. Therefore, a statement about a satellite orbit is with reference to the earth and not to fixed stars or the sun. Figure 2 4 illustrates an inclined orbit, in which the angle θ represents the angle of inclination. Figure 2-4 illustrates a satellite in orbit about the earth (in an elliptical plane with the center of the earth at one focus) with one additional factor introduced, the rotation of earth itself. The earth's rotation has a distinct bearing on the use of satellites for communications.



CCP105-5-2

Figure 2-4. Representation of inclined satellite orbit.

(g) If the plane of the satellite is visualized as being fixed, with the earth rotating through it, it becomes apparent that all points on the equator will pass through the plane of the orbit twice each day (unless the orbit is equatorial, or at an inclination of 0 degrees, in which case the entire equator is in the plane of the orbit). As the inclination increases to 90 degrees progressively more and more of the earth's surface is within the area crossed by the satellite path. A representation of this is shown in figure 2-5. At 90 degrees the orbit is called a polar orbit and all points on the earth pass through the plane twice a day.

(h) It is now necessary to introduce two additional terms into the discussion; ascending node and descending node. The ascending node is the point at which the satellite passes through the plane of reference (the equatorial plane for earth

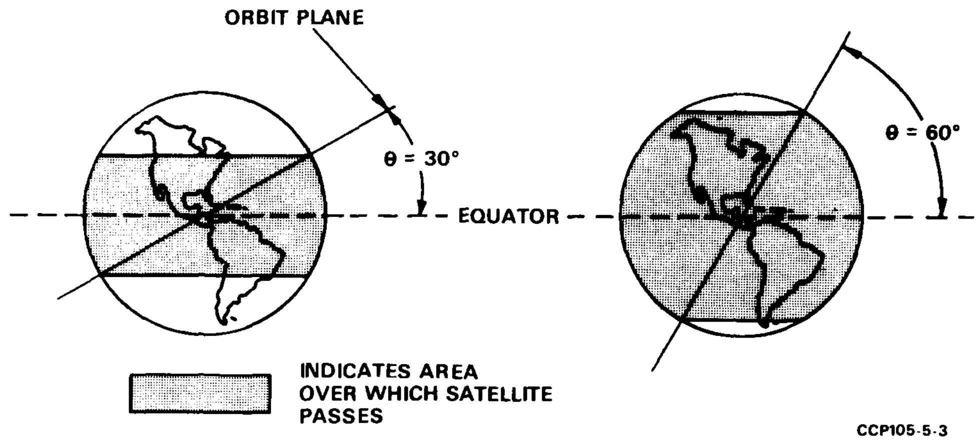


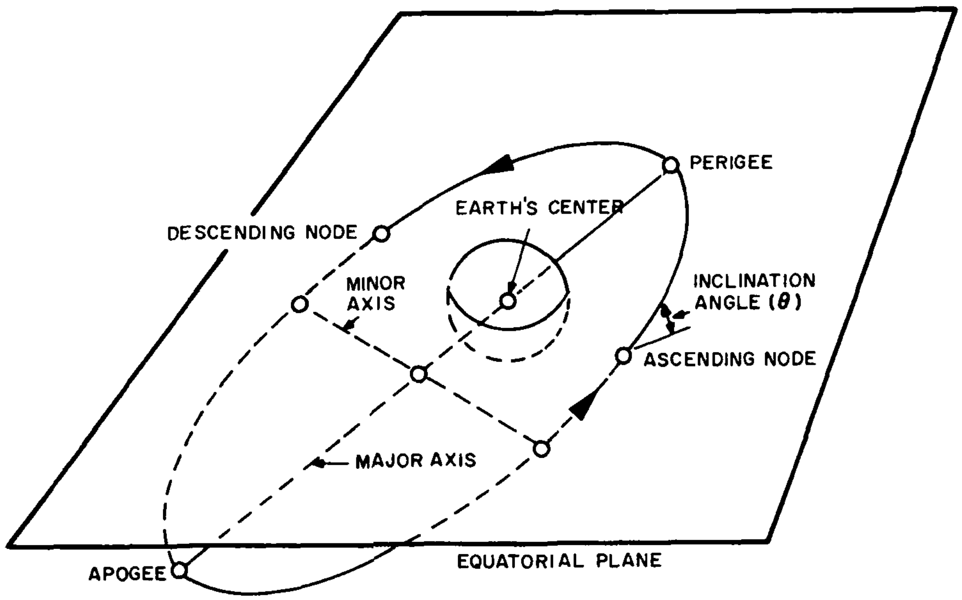
Figure 2-5. Representation of effect of orbit plane inclination on satellite coverage.

satellites) from the south. The descending node is the point at which the satellite passes through the equatorial plane from the north. This is illustrated in figure 2-6. Note that the nodes are shown as lying on the end of the orbit toward the perigee. This is not always the case, nor is it necessary that the line connecting the nodes be at right angles to the major axis of the orbit ellipse.

(i) The illustrations so far have depicted satellites whose apparent motion is toward the east. The inclination of such satellites can range from near equatorial (near 0 degrees) to 90 degrees. If the angle that the orbit path makes with the equator is greater than 90 degrees when measured at the ascending node, the satellite will have a westward motion (the orbit in this case being termed retrograde). This angle may be increased until the orbit is again equatorial but with the satellite moving in a westward direction around the equator.

(j) The definition of inclination must be modified so that the angle is measured at the ascension.

(k) Kepler's third law when applied to satellites yields the following relation: The squares of the orbital periods of satellites are proportional to the cubes of their major axes. Thus, if the period and semimajor axis (mean distance from focus) of one satellite are known, the period of another satellite may be calculated from a comparison of the major axes. Let 2a equal the major axis of a satellite and t denote the satellite period. Then, if the period of a reference satellite is represented by t_0, and its major axis by $2a_0$, an equation can be developed from Kepler's



CCP105–5–97

Figure 2-6. Elements that specify the shape and location of orbits.

third law. This equation will permit determination of the period if the major axis is known, or the major axis if the period is known. The equation is:

$$\frac{(t)^2}{(t_o)^2} = \frac{(2a)^3}{(2a_o)^3} \qquad (2.9)$$

In using this equation it is necessary to remember that some characteristics of an orbit must be known.

2-3. ORBIT SELECTION AND EFFECTS.

a. *Available Choices.* A very broad range of choices is available when selecting a type of orbit for application to a communications system. These include the equatorial orbit, the polar orbit, and orbits of various inclinations which may range from the highly elongated through the circular. These orbits impose limitations on the initial design of the satellite and the vehicle used to place it in orbit. Some may require extreme altitude, thereby severely limiting the weight of the satellite. Others require the use of a considerable amount of power in maneuvering to attain the desired inclination.

(1) *Synchronous satellites.* The synchronous equatorial orbit is one in which the satellite is synchronized with the speed of rotation of the earth and appears to hang motionless over some point on the equator. However, there are problems associated with the launch and injection of such a satellite. For a synchronous satellite, calculations of the velocity and altitude necessary to balance exactly the earth's gravity at the equator indicated that the satellite had to be 22,248 miles above the earth's surface and required a velocity of 10,066 ft/s. These figures are large in themselves, but further calculations showed that the satellite required a velocity of approximately 57,000 ft/s by the time it reached booster burnout altitude of approximately 115 miles, to enable it to coast to synchronous orbit altitude.

(a) The extreme velocity required for the launch severely limits the weight available for the satellite and the amount of equipment carried in the satellite.

(b) A further difficulty associated with the launching of synchronous satellites is the need to dogleg the orbit of the satellite into an equatorial plane. This is due to the lack of launching sites on the equator. Also, a dogleg launch requires additional control equipment in the vehicle which tends to reduce the weight available for the satellite itself.

(2) *Low- and medium-altitude.* Low- and medium-altitude, inclined-orbit satellites are much less difficult to launch. Their ultimate payload capacity is considerably greater than the synchronous satellites. The low altitudes and high velocities present tracking and acquisition problems which are not present with the synchronous satellites.

b. *Communications Satellite Considerations.* Because of launching limitations, certain restrictions are imposed on communications satellites. From the standpoint of the military system, the selection of the orbit and the equipment to be carried aboard the satellite form a considerable part of the planning and system design.

(1) The equatorial synchronous satellite is a high altitude communications relay which is stationed practically motionless relative to the earth's surface. Figure 2-7 illustrates the coverage (illumination) available from a synchronous satellite. Note that a third of the earth's surface is in view of the satellite and that all areas within that illunination could use the satellite as a link with all other areas so illuminated. The usefulness is somewhat limited at the edge of the illuminated area, where the radiation from the satellite becomes tangential to the surface of the earth, but for the most part, a satellite will provide a link between most points in the hemisphere it illuminates. Figure 2-8 depicts a projected worldwide system using synchronous satellites. This figure shows the illumination of the earth's surface as viewed from the North Pole. Note that the areas overlap slightly so that terminals in these overlapping areas could relay communications from two satellites. This would permit around the world message transmission through two or more satellites

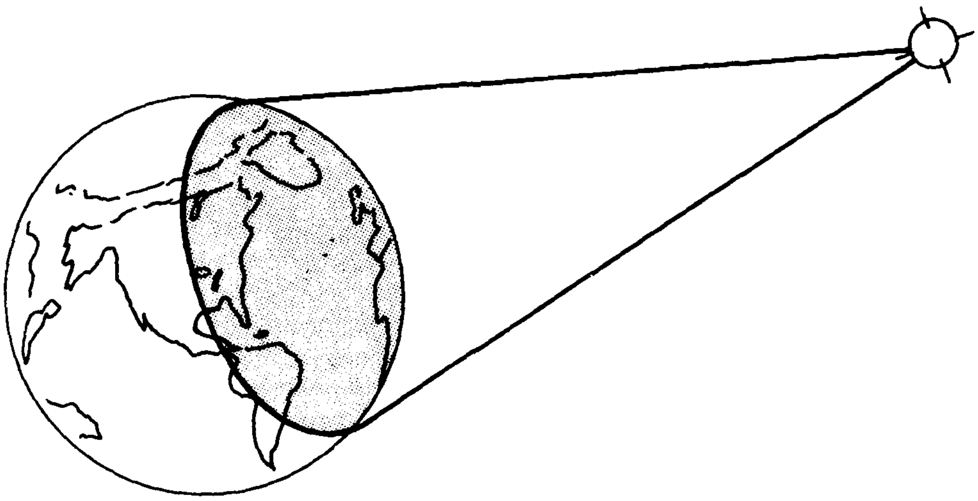
(2) The drawbacks to this type of system lie in the technical difficulties experienced in orbiting the satellites, the satellite's vulnerability to attack, and the weight limitations imposed by the high altitudes required. The limited weight reduces the amount of communications equipment and, thereby, limits the amount of information that could be handled by such a system. The advantage of this system is that the satellite would always be available for use by all ground stations.

(3) The low- and medium-altitude satellites have a much more restricted area of illumination. For direct communications, the ground stations must simultaneously have an LOS radio path to the satellite. A worldwide communications system using random orbit satellites (fig. 2-9) requires a considerable number of satellites in operation, so that the ground stations have a mutually visible satellite available almost all of the time. Random orbit systems also require ground stations to relinquish contact with a satellite as it passes from view and acquires another.

2-4. BOOST TO ORBITAL ALTITUDE.

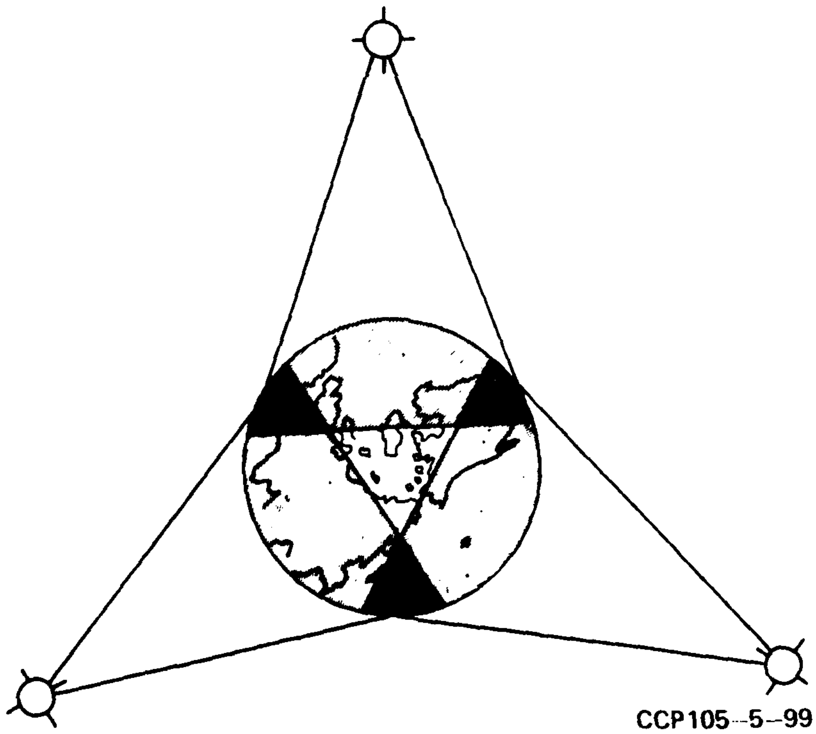
a. *Rocket Motors.* Communications satellites are lifted into orbit by space vehicles, which are launched vertically for structural and aerodynamic reasons. Rocket motors provide the motive power for these vehicles. A rocket motor is a reaction motor that works on the action-reaction principle with thrust being produced by the expansion of gases in an enclosed space. The expansion of gases is caused by the reaction of a fuel with an oxidizer. A jet aircraft engine is also a reaction motor· however, it uses oxygen from the atmosphere as the oxidizer. This limits the jet engine to altitudes at which oxygen is available to support the combustion process. Rocket motors are classified as solid propellant or liquid propellant types according to the fuel that they use.

b. *Solid Propellant Rockets.* The solid propellant rocket works on a principle similar to the small gunpowder rockets used for fireworks displays. It contains



CCP105–5–98

Figure 2-7. Illumination of earth from synchronous satellite.



CCP105–5–99

Figure 2-8. Worldwide synchronous satellite systems illumination viewed from North Pole.

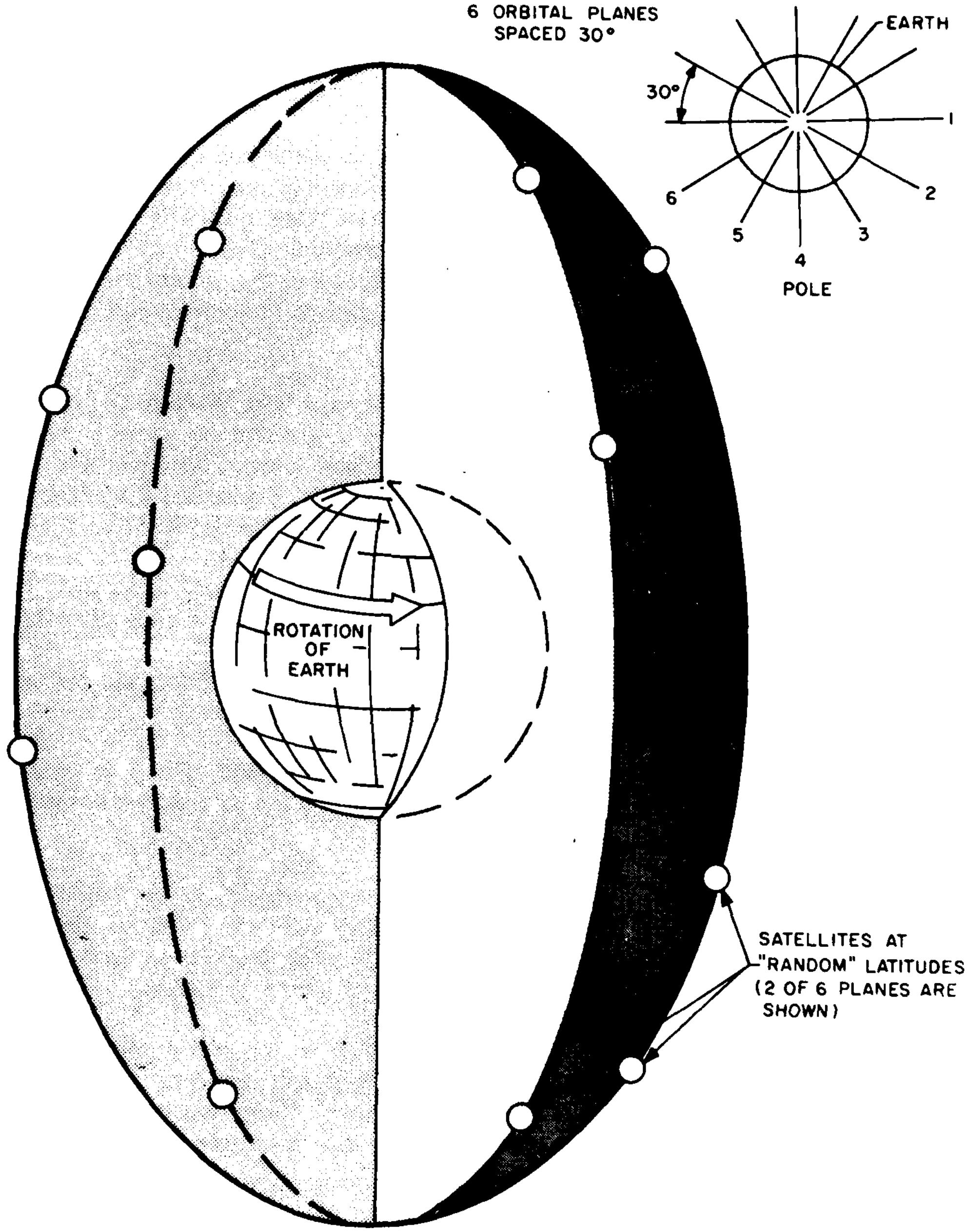


Figure 2-9. Representation of polar orbit satellite system.

a fuel that is a combination of burnable material and an oxidizing element. The fuel is burned in a restricted combustion space with nozzles to direct and accelerate the outgoing gas flow (fig. 2-10).

(1) Unlike the simple black powder rocket, in which the thrust is obtained by simply packing a tube or cylinder with the propellant and allowing the gases to exit from one end, today's solid propellant motors are extremely complex in design. The fuel, often referred to as grain, is generally prepared in the semiliquid form and cast into the motor tube. Internal shapes and the lamination of fuels with different burning characteristics are used to change the burning rate and the flight characteristics of the rocket.

(2) One method of achieving directional control is to place restrictors, or deflectors, directly in the flow of exhaust from the motor and use them to change the direction of thrust. Until recently, solid propellant motors were restricted by their own weight/energy ratio and were not capable of developing the extremes of thrust in proportion to their weight that could be achieved by the liquid fueled types. However, breakthroughs in the chemistry of the solid propellants now indicate that the solid propellant rocket will achieve considerably wider use. The POLARIS missile is an outstanding example of the use of this type of motor.

(3) In application to satellite launching, solid propellant motors are frequently used in the upper stages of multiple-stage vehicles.

c. *Liquid Fueled Rockets.* The liquid propellant rocket (shown in greatly simplified form in figure 2-11) carries two tanks outside the combustion chamber, one for fuel and one for oxidizer.

(1) The oxidizer and fuel, in liquid form, are metered to the combustion chamber and ignited; from that point on, the function of the rocket is identical to that of the solid propellant motors. In practice, the liquid propellant motor is much more complicated and presents enormous design problems.

(2) An advantage of the liquid fueled type of motor is that the thrust of the motor may be controlled by varying the amount and proportion of fuel fed into the combustion chamber. Current boosters using liquid fuel have been able to achieve more thrust and higher specific impulse than those using solid propellant fuel.

d. *Escape Velocity.* The techniques of rocketry are based on an application of Newton's third law: For every action there is an equal and opposite reaction. The action is provided by the hot expanding gases and the reaction is that of movement of the vehicle and its payload. From this law has evolved a calculation fundamental to all rocket theory and represented by the following formula:

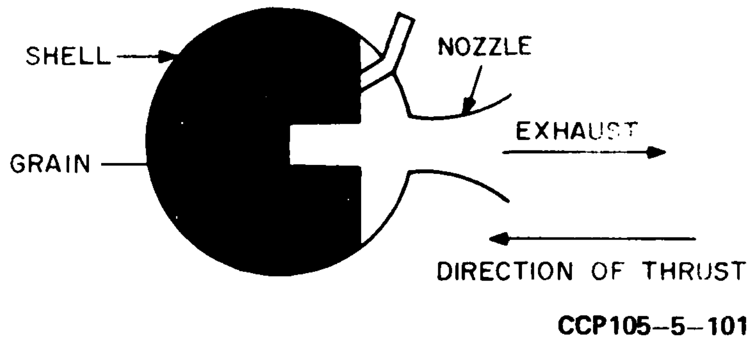


Figure 2-10. Solid propellant rocket motor functional diagram.

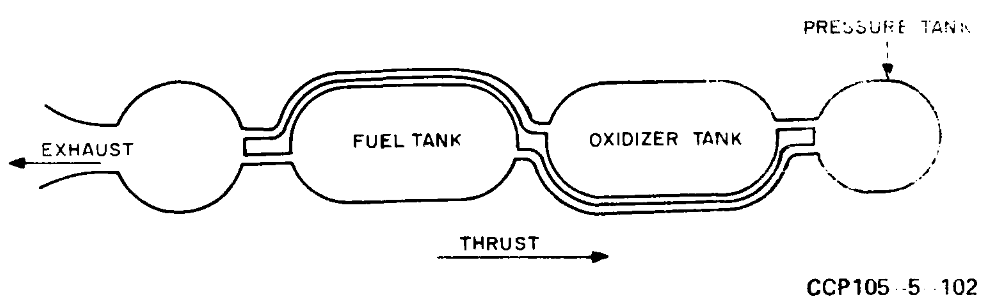


Figure 2-11. Liquid propellant rocket motor functional diagram.

$$V_m = V_g \ln \left(\frac{M_t}{M_t - M_P} \right) \tag{2.10}$$

where

V_m is maximum velocity of the missile

V_g is velocity of gases escaping from rocket nozzle

$\ln$ is natural logarithm

M_t is takeoff weight of rocket

M_P is weight of propellant

(1) In order to overcome the earth's attractive forces, an escape velocity of 37,000 ft/s must be achieved. Equation (2.10) indicates that in order to attain this velocity the vehicle must have either a high exhaust velocity (V_g) or a high ratio of takeoff weight to burned-out weight. This ratio, known as the mass ratio, is represented in equation (2.10) by the term:

$$\frac{M_t}{M_t - M_P} \tag{2.11}$$

Since an exhaust velocity of 9,000 ft/s is near the upper limit for presently available rockets, a mass ratio of 60:1 is required to enable the vehicle to attain escape velocity.

(2) With the weight of the rocket and structure, this mass ratio seems impossible to attain. However, if a multiple-stage booster rocket is used and each booster and its accompanying structure is discarded as its fuel is expended, the required mass ratio can be achieved and escape velocity attained. Satellite vehicles are usually designed so that the thrust exceeds the weight of the vehicle by 30 to 50 percent. This is the same as a thrust-to-weight ratio of 1.3 or 1.5 to 1 on the ground. As the vehicle ascends, vehicle weight is reduced by fuel consumption. Therefore, the ratio of thrust-to-vehicle weight is improved. Further acceleration with the same thrust is gained at high altitudes as the resistance of the atmosphere is reduced.

(3) After the first rocket booster burns, it is jettisoned and the second booster is ignited. Each stage provides a separate propulsive interval. The thrust of the second stage need not be as great as that of the first stage, since the total weight of the vehicle has now been diminished by the weight of the first stage.

(4) When the vehicle attains altitudes above 30 miles the atmosphere no longer presents any appreciable drag. The vehicle is now turned away from the vertical and begins to acquire the velocity tangential to the earth that is necessary to establish the satellite orbit. This is usually a relatively early maneuver, since the tangential velocity acquired during the boost stages lessens the additional velocity required at injection.

(5) When all but the upper stage has been jettisoned, the satellite and the upper stage coasts along in an orbit; however, it is not usually the desired orbit for the satellite. Generally, the perigee will lie too close to the earth's surface, or even under the earth's surface if the satellite continued in this path. The apogee, however, will have been chosen to coincide with the desired altitude for injection. The vehicle, consisting of the satellite and injection motor, is allowed to coast in this orbit until it attains the injection altitude, at which point injection occurs.

2-5. **INJECTION INTO ORBIT.** When the satellite reaches the desired altitude it is injected into the desired orbit. This is the most critical of the maneuvers in orbiting a satellite in that the final increment of velocity must occur at exactly the right altitude, be of exactly the right magnitude, and be in the proper direction. To achieve these conditions, some form of guidance system must be provided. This system may be entirely self-contained within the injection motor and programmed before the vehicle is launched, or it may involve response to commands from the ground. Most of the systems used today involve radar tracking and command from the ground. A radar beacon is installed in the injection vehicle and is tracked by several ground radars. This provides the triangulation necessary to determine the satellite path and the Doppler shift. The radar-derived data is fed into a computer which designates where the thrust is to be applied and the amount of thrust to be used. The direction of the thrust is controlled by gyroscopes within the vehicle. Directional changes are made by using small gas jets. The jets are formed by a small tank of compressed gas or generated by the decomposition of hydrogen peroxide. The nozzles are situated in opposing pairs so that the net change of thrust to the entire vehicle is zero, but the torque will rotate it to the desired position.

a. *Required Velocity.* To determine how these factors apply to the preceding discussion, consider a satellite with a mass of 1 kg which is to be injected into a circular orbit at 8,000 miles altitude. At this height the gravitational attraction in Newtons (N) will be:

$$F = \frac{G_1 M_1 M_2}{r^2} \tag{2.12}$$

$$= \frac{(6.67)10^{-11}\ \frac{Nm^2}{kg^2}\ (1kg)\ (5.98)^{24} kg}{[12.9 \times 10^6\ m + (6.37)10^6\ m]^2} = 1.08\ N$$

where

r *is altitude of satellite + radius of earth in meters*

The centrifugal force necessary to balance this will be created by a v of:

$$1.08\ N = \frac{M_1 v^2}{r} \qquad v = \sqrt{\frac{(1.08\ N)r}{M_1}}$$

$$v = \sqrt{\frac{1.08\ N\ (19.3)10^6\ m}{1\ kg}} = (4.56)10^3\ m/sec \tag{2.13}$$

In general, a circular orbit is achieved by giving the satellite the proper velocity at the desired altitude, so that the force of gravity will equal the centrifugal force. The equations can be simplified as:

$$v = \sqrt{\frac{GMe}{r}}$$

$$= \frac{\sqrt{(6.67)10^{-11} \left(\frac{Nm^2}{kg^2}\right)\ (5.98)10^{24} kg}}{\sqrt{r}} \tag{2.14}$$

$$= \frac{(2)10^7}{\sqrt{r}}\ m/sec\ \text{(where r is in meters)}$$

(1) When the formula for the force of gravity was equated to the formula for the centrifugal force, the mass of the satellite canceled out. That is why equation (2.14) does not involve the satellite's mass. Velocity for a circular orbit depends only on the altitude of the satellite, whether the satellite mass is 1 or 1,000 kg. When the satellite attains the proper velocity at a given height, it will remain in a circular orbit. However, this does not discard the mass factor when considering the booster rocket. The greater the mass, the greater the thrust must be in order to achieve the desired altitude. With regard to the energy required for launch, the mass of the satellite makes a vital difference.

(2) Equation (2.14) also shows that the greater the altitude, the lower the velocity needed to maintain the orbit. However, greater altitudes always cost more in terms of boost energy. The total thrust energy needed to orbit a satellite at an extreme altitude is greater than for a lower altitude, even though the necessary orbital velocity is less. Present booster rocket capabilities limit the weight of the payload that can be lifted to synchronous altitude; therefore, the weight of the equipment that can be carried aboard is limited.

(3) Although these considerations have been based on the illustrations of a circular orbit, it is possible to vary the shape of the orbit to any desired degree of eccentricity. This is accomplished by adjusting the magnitude and direction of the thrust velocity at injection. There are practical limits and in extreme cases, where the elongation is such that the velocity at perigee is extreme, the orbit will gradually decay. The decay is primarily due to the satellite's high-speed encounter with atmospheric particles at perigee. The satellite will eventually enter denser atmosphere where air friction will cause it to burn.

b. *Latitude of Injection.* If injection is well controlled, the resulting orbit can have any desired major axis length and eccentricity. However, the ease with which a particular orbit inclination can be achieved is dependent on the latitude of the launch site.

(1) To obtain a 28-degree inclination launching from Cape Kennedy the boost phases and injection should all be aimed straight east (or straight west for a retrograde orbit). When this is done the satellite will never pass directly over any point on earth farther north than 28 degrees N latitude. To obtain less than a 28-degree inclination from Cape Kennedy an intermediate orbit (parking orbit) is first required. The intermediate orbit is achieved when a due east or west shot is made from the launching station. As a result, the inclination angle of the parking orbit will be the same as the latitude of Cape Kennedy. When the satellite is above the equator, the thrust of the remaining rocket engine is applied at right angles to the parking orbit. In reaction to this thrust, the satellite veers into an orbit of the desired inclination with no change in either altitude or velocity. This is the technique known as doglegging. An equatorial orbit could be achieved without doglegging if the satellite were launched along the equator.

(2) Unless the technique of doglegging is used, the angle of inclination of the orbit cannot be any less than the latitude of the launch site. For example since the latitude of Cape Kennedy is 28 degrees N the orbit inclination angle of a satellite launched from Cape Kennedy cannot be less than 28 degrees because the point of injection is a point on the orbit. Hence, a satellite must always return to the latitude at which it was injected.

2-6. **SATELLITE CONTROL.** The two main aspects of satellite control are orientation control and position control. Orientation control is required in a communications satellite for initial orientation and for all subsequent adjustments so the antenna will constantly view the earth and the solar cells will view the sun. Position control is required for initial satellite emplacement and to compensate for forces which tend to divert the satellite from its orbit. Such forces include lunar and solar gravitational action, longitude-dependent variations in the earth's gravitational and magnetic fields, and solar radiation pressure. Some orientation and position control systems are completely contained within the satellite and require no outside control. Other control systems allow ground controllers access to the control loop which affords more flexibility.

a. *Orientation Control.* Orientation control (attitude control) is defined as the control of a vehicle about any, or all, of its axes (roll, pitch, and yaw).

(1) *Spin stabilization.* Spin stabilization operates on the principle that the direction of the spin axis of a rotating body tends to remain fixed in inertial space. A natural example of spin stabilization is the effect of the earth's rotation in keeping the earth's axis fixed in inertial space. A satellite having a spin axis parallel to the earth's axis will maintain this position since both axes are fixed in inertial space. Figure 2-12 illustrates the use of this principle with an equatorial orbit satellite. Spin stabilization requires virtually no additional energy or expenditure of mass once the system is in motion. A spin-stabilized satellite is usually shaped as shown in figure 2-13. The satellite is constructed like a flywheel with the heavier equipment mounted in the same plane and as close to the periphery as possible.

(a) After orbital injection the satellite spin axis is oriented with the earth axis by means of the axial jets. Radial jets are pulsed to spin up the satellite.

(b) Jets are pulsed when necessary to provide orbit position and attitude correction (fig. 2-13). Reference and error signals for the control loop are provided by sun, star, and earth sensors in conjunction with preprogrammed ephemeris data and/or ground commands.

(2) *Gravity gradient stabilization.* The gravity gradient stabilization method requires no power, or expendable mass, once it has been deployed. This method allows more of the satellite electrical power and weight to be available for communications purposes. In addition, a satellite stabilized by the gravity gradient method can use a highly directional antenna with its inherent gain advantage, resistance to jamming, and frequency reuse capabilities.

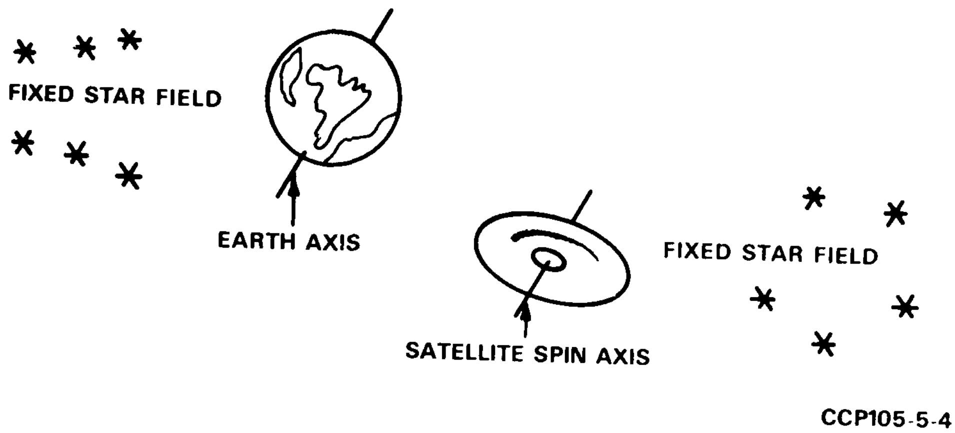


Figure 2-12. Representation of spin-stabilized satellite.

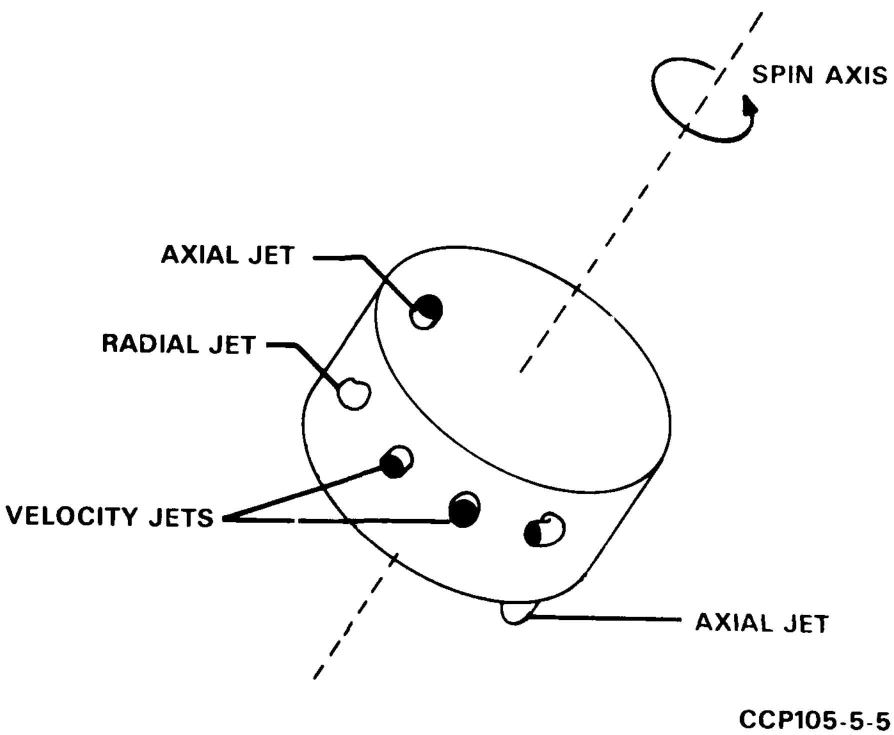


Figure 2-13. Spin-stabilized satellite controls functional diagram.

(a) The gravity gradient stabilization system uses as a control element one or more rods that may extend several hundred feet from the satellite, as illustrated in figure 2-14.

(b) In this illustration, point C (in the satellite) is at the center of gravity of the system and in an orbit. Therefore, the force of gravity and the centrifugal force are balanced at that point. Since the protuberances from A to C and C to B are tied together, A, B, and C all must make a revolution around the earth in the same elapsed time and have identical periods. For a given period, centrifugal force and gravity will be balanced at only one altitude, in this case the orbit of C. At B, the period (being that of C) is longer than required to balance gravity and a slight force toward the center of the earth is felt at B. At A, the period (also being that of C) is shorter than that required for balance, and A experiences a net force [illegible] center of the earth. If the line connecting AB is somehow rotated around C (displaced from the line through C and the center of the earth) the attraction of B^1 toward the center of the earth and the repulsion of A^1 from the center of the earth results in a torque being applied to the satellite. This torque tends to restore the line AB to a position which points toward the center of the earth.

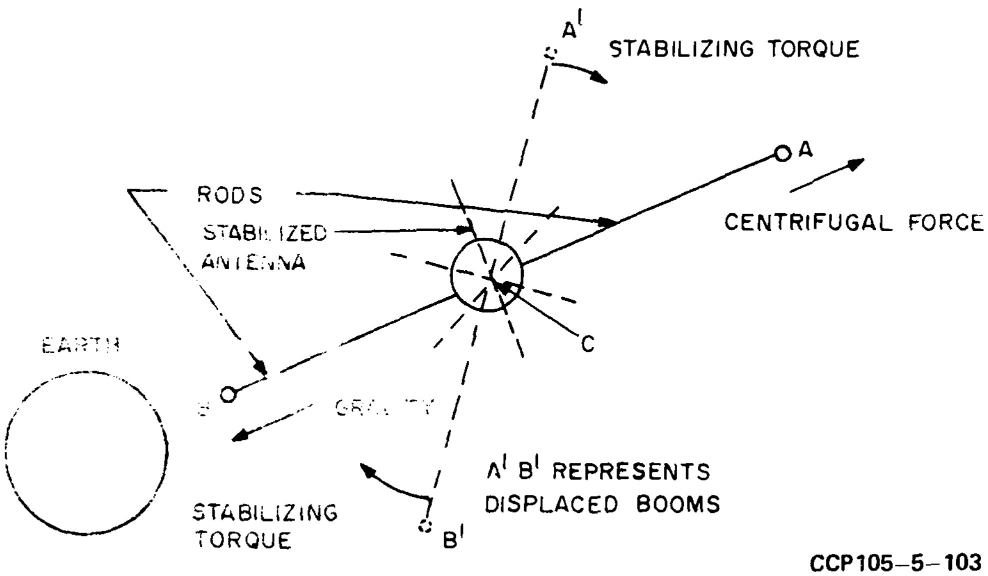


Figure 2-14. Graphic representation of gravity gradient stabilization.

(c) An interesting example of natural gravity gradient stabilization is given by the earth's natural satellite, the moon. As a result of the net inward force on the near side and the net outward force on the back side of the moon, the moon always keeps one face toward the earth.

(d) Since a satellite could not be launched with two large rods protruding, the rods are extended by a mechanism when in space. The rods consist of preformed tapes of silver plated beryllium-copper alloy wound on motor-driven storage drums. To extend the rods, the drums are unwound by a ground command or by a programmed timer within the vehicle or satellite. As many as six rods may be used to provide stabilization on all axes.

(e) There are still several problems to be overcome in the gravity gradient stabilization system.

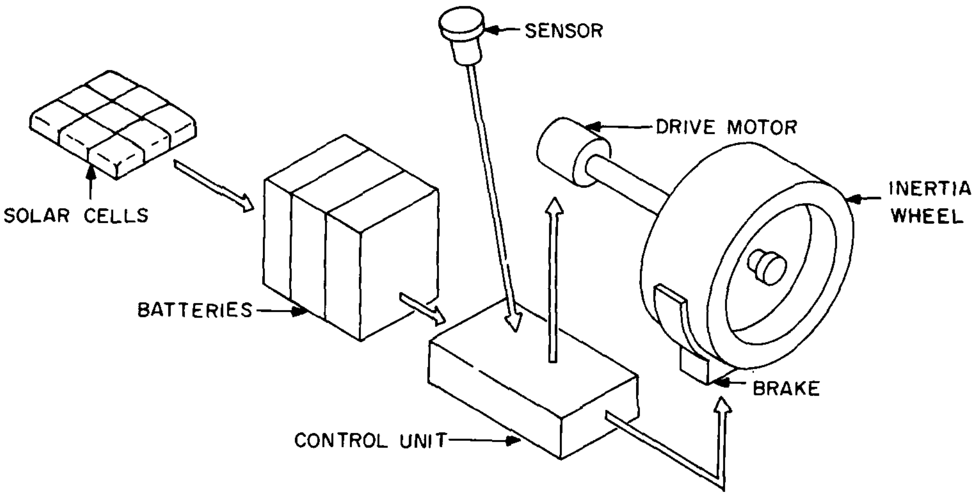
1. Thermal bending of the gravity gradient rods results from the fact that one side of the rod is exposed to the sun. The differential in thermal expansion causes the rod to bend. Rod bending affects stabilization by changing the location of the axes of inertia and the center of mass of the system.

2. Residual magnetic fields in the satellite tend to align to the local magnetic field produced by the earth's magnetic forces and solar-induced magnetic fields. This produces torques that tend to cause orientation errors.

3. After injection into orbit and until stabilized, the satellite will physically oscillate. Since the space environment offers little resistance to this oscillation, decay time will be excessively long. Damping techniques to shorten this decay time are being developed and tested. These include rods of various configurations, contraction and expansion of rods, use of reaction jets, and the earth's magnetic field.

(3) *Inertia wheel stabilization.* In the inertia wheel stabilization system, masses attached to the satellite are rotated to create a reaction force that causes the satellite to rotate in the opposite direction. An inertia system is illustrated in figure 2-15. In this system the satellite's attitude is conveyed by a sensor to a control unit, which compares the actual attitude with the required attitude. If an error exists, the control unit releases the brake and supplies power to the drive motor, which rotates the inertia wheel at high speed in the direction opposite to that in which the satellite must move. This causes the satellite to rotate to the desired attitude. When an error no longer exists, the control unit removes power from the drive motor and applies the brake. The sensors used to supply the attitude signals to the control unit may be infrared horizon sensors, star trackers, or gyroscopes.

b. *Position Control.* Synchronous satellites require a means of providing position control or station keeping. The axial and radial jets used to place the satellite in its final position are also used for station keeping. The new synchronous satellites carry enough fuel to reposition themselves to a completely new subequatorial point, if necessary. Station keeping can be accomplished by inertial controls but is usually performed upon command by the ground controllers.



CCP105–5–104

Figure 2-15. Inertia wheel stabilization functional diagram.

CHAPTER 3

SATELLITE TRACKING

3-1. INTRODUCTION.

a. Satellite acquisition and tracking is a major element in the operation of a satellite communications link. Ground stations are required to locate the satellite, achieve radio contact, and maintain contact while the satellite is in view. Once acquisition of a synchronous satellite has been achieved by a ground station antenna, only small changes in antenna position are required to maintain communications. With the random orbit satellite, the ground station antenna is constantly in motion. Because of the comparatively narrow beamwidth of the antenna, it must track with high accuracy. This requires comparatively sophisticated equipment.

b. There are several elements in the process of establishing and maintaining a link. In sequence, these are as follows: determining the satellite orbit, predicting the satellite location, pointing of the ground station antenna, acquisition of the satellite, and tracking. The steps in this sequence are discussed in paragraphs 3-2 through 3-5. The satellite ground station complex, while varying somewhat from location to location, is essentially a transmitting and receiving station equipped with a highly directional antenna. The antenna is power driven and controlled with extreme accuracy with respect to direction of pointing and rate of motion.

c. Beacon and telemetry equipment are important elements in the tracking system. The beacon is a transmitter that transmits a relatively low power signal from the satellite. This signal is used as an aid in satellite location by the ground station. Telemetry refers to the data transmitted by the satellite to provide the ground station with information on its condition and operational state. Each satellite transmits its individual beacon code imposed on the beacon signal to permit positive identification of a particular satellite. The equipment and frequencies used for beacon and telemetry functions are usually separate from the communications equipment and frequencies in the satellite.

3-2. ORBIT PREDICTIONS AND EPHEMERIS DATA.

a. *General.* In considering the operation of a system consisting of as many as two dozen satellites in a variety of orbits, the initial problem of knowing where to look for a satellite becomes extremely important. Once an orbit has been established and the orbital parameters are known, it becomes relatively easy to predict when and where a satellite will come into view. This prediction capability is the basis for the initial step in the acquisition and tracking of communications satellites by the ground stations.

b. *Ephemeris Data.*

(1) An ephemeris is a table showing the calculated positions of a satellite (or any heavenly body) at regular intervals of time. The ephemeris of a satellite is calculated from its orbital parameters and a knowledge of the laws of motion. Once the orbit has been determined the ephemeris data for a station may be computed from knowing the satellite equatorial crossing times and the longitudes at which the satellite track (the path on the earth's surface directly beneath the satellite's orbit) crosses the equator. The information on the orbit is constantly updated by tracking stations, so that precise ephemeris data will always be available.

(2) While the satellite moves at high speeds, its orbital characteristics will change slowly. The period, inclination, and dimensions of the major and minor axes will vary slowly or not at all for a stable orbit.

c. *Orbital Prediction.*

(1) The constants defining an orbit are initially obtained by the process of tracking. The rocket is tracked by radar from time of launch until it passes out of sight. The recorded tracking data is sufficient for making rough predictions of the orbit. These predictions are made rapidly with a computer and sent to other tracking stations in other parts of the world. The other tracking stations around the world watch for the satellite during its first trip and record additional data, which enables more precise predictions to be made. Thus, during the first few orbits, tracking stations all around the world are progressively obtaining more accurate data concerning the satellite. These data are put into a computer and corrections are repeatedly made to earlier estimates of the orbit.

(2) Once the initial predictions are complete and the satellite link becomes operational, there is very little change in these calculations. The orbits will change slightly over a period of time, but these changes are so gradual that predictions will be accurate enough to be used for weeks or months without further corrections.

3-3. ANTENNA POINTING.

a. When the orbits are known precisely, an ephemeris can be calculated for each satellite of the system. These ephemerides can then be distributed to the ground stations.

b. Antenna pointing instructions are derived from the ephemeris of a satellite. These instructions must be computed separately for each ground station location. A satellite which bears due north of station A at an antenna elevation of 25 degrees may simultaneously bear due east of station B at an antenna elevation of 30 degrees. While the orbit of the satellite is determined by applying the considerations discussed in chapter 2, pointing instructions are determined by taking into consideration both the orbital predictions, and the latitude and longitude of each ground station.

(1) *Coordinate transformations.*

(a) It is convenient to express the ephemeris in terms of a geocentric coordinate system; that is, a coordinate system whose origin is the center of the earth rather than some point on the surface of the earth. Pointing instructions are obtained by converting the geocentric coordinates to local coordinates by a further calculation.

(b) A natural geocentric coordinate system is the one which gives the position of the satellite in terms of spherical coordinates. The position is described by the distance from the center of the earth, and the latitude and longitude of the point on the earth, which is directly beneath the satellite at a given instant (satellite point).

(c) While the use of modern computers for orbital calculations permits rapid calculations in any coordinate system desired, an ephemeris should be considered to be a table giving satellite position relative to the earth as a whole. The calculations that convert geocentric coordinates to local coordinates are called coordinate transformations.

(d) From the standpoint of acquiring radio contact with a satellite, only the local azimuth and elevation angles are important. Knowledge of the bearing and elevation of a satellite, at the time planned for acquisition, permits the antenna to be properly pointed. In addition to position, the ground station requires knowledge of the velocity at which the satellite is approaching in order to properly adjust its receiver to compensate for the Doppler shift. Thus, predictions of both position and velocity must be made from the ephemeris and transformed into local coordinates.

(2) *Survey and alignment.* To perform a coordinate transformation, the location of the ground station antenna must be known accurately. The exact location is determined by surveys. The antenna is aligned mechanically and electrically so that it points exactly in accordance with pointing instructions.

(3) *Preparation of pointing instructions.*

(a) There are two steps in the preparation of pointing instructions. The first step is the planning that concerns the selection of a satellite to be used at a given time between each pair of ground stations. The second step is the calculation of pointing instructions for the acquisitions that have been planned.

(b) The use of satellites to set up particular communications links requires planning. The varying and contingency needs of users must be considered. With a limited number of random orbit satellites, it is possible that there may be no satellite in the common view of certain pairs of ground stations for minutes or hours at a time. Also, failures of electronic equipment will sometimes occur. Planning must take all of these things into consideration in order to make best use of the satellites.

(c) Antenna pointing instructions are calculated for planned satellite acquisitions and for additional acquisitions to provide reliability in event of satellite equipment malfunction. Calculations are performed at the SCCF by a central computer and forwarded to the ground stations. The SCCF provides a program, several weeks in advance, for planned and contingent satellite use. The SCCF receives data on satellite acquisition and tracking from each ground station. This information is used to update and check the computer's ephemeris data bank.

3-4. ACQUISITION.

a. The acquisition of satellite signals by a ground station equipped with large antennas and operating at super high frequencies places severe requirements on the acquisition system. These requirements can be divided into two problem areas; spatial-time uncertainties and frequency variations.

b. The spatial-time acquisition of a satellite involves knowing both its position in space at some instant of time and its signal frequency. Nothing would be accomplished if the antenna locates the right point in space and the receiver is not tuned to the right frequency. These problems are discussed below.

(1) *Spatial-time acquisition.*

(a) Very accurate pointing data is supplied to the ground station by the SCCF. However, due to equipment limitations it is necessary to conduct an area search in the predicted location of the satellite in order to make initial contact. This searching consists of a manual or automatic scanning of a small area around the point where the satellite appearance is predicted.

(b) Once the beacon signal from the satellite has been received, the tracking receiver generates error signals for the servomechanism of the antenna. The signals position the antenna in the direction of maximum possible signal and allow the system to automatically track the satellite.

(2) *Timing control.*

(a) Timing signals for the entire system are provided by the SCCF. Each station has a local time standard or clock that is synchronized with all other timing standards throughout the global system.

(b) The local time standard is the focal point for fulfilling all timing requirements within the station.

(3) *Frequency control.*

(a) Frequency control compensates for ground and satellite equipment instabilities and for frequency shifting resulting from the Doppler effect. (Doppler effect (in this case) is the apparent change in frequency of any electromagnetic radiation when the distance between the source of the radiation and the observer is increasing or decreasing owing to the motion of either or both.) Frequency instabilities due to equipment are negligible in comparison with frequency shifts caused by Doppler effect since the oscillators in the satellite and ground station are extremely stable. Doppler effect shifts the frequency of the beacon signals from the satellite to the ground in the acquisition mode and shifts the frequency of the communication signals going from one ground station to another through the satellite.

(b) In the acquisition mode, when the ground station is listening for the satellite beacon, the relative motion between the satellite and the ground station creates a condition that causes the frequency of the satellite signal to change during the satellite's orbital pass. This makes it necessary to control the tracking receiver so that it will change frequency to follow the change in the incoming signal. This change is calculated from the ephemeris data and is used to control the receiver during acquisition.

(c) During tracking, frequency control is accomplished by comparing the incoming signal with a signal of known frequency generated by a highly accurate standard. As the incoming signal varies, due to Doppler shift, the variation from the standard signal is followed as a differential frequency. This differential frequency is used to generate electronic commands that shift the receiver frequency to compensate for the Doppler shift. Details of time and frequency control are discussed in chapter 8.

3-5. TRACKING METHODS.

a. When a satellite has been acquired, the ground station antenna must continue to track that satellite as long as it is to be used as the communications relay.

b. Two methods of tracking are being used; these are programmed tracking and automatic tracking.

(1) *Programmed tracking.*

(a) In programmed tracking, the known orbit parameters of the satellite are fed into computation equipment to generate antenna pointing angles. The antenna pointing angles are fed as commands to the antenna positioning servomechanisms (chapter 9) which point the antenna in the required direction.

(b) In using programmed tracking with narrow beamwidth antennas, as used in satellite communications, the amount of data and computation involved in pointing the antenna is extensive. In addition, some deviations from calculated pointing angles arise as a result of antenna mount flexure and atmospheric and ionospheric bending of radio waves. Since these uncertainties exist, programmed tracking is not totally satisfactory.

(2) *Automatic tracking.*

(a) Automatic tracking systems provide advantages over programmed tracking in that they track the actual signal received from the satellite. Since automatic tracking systems are following the apparent position of the satellite; that is, the direction of arrival of the radio signal, the real position of the satellite is not required. The automatic tracking system is a servomechanism, and once acquisition has been accomplished, it continually generates its own pointing data, thus eliminating the requirement for data input and computation.

(b) Two varieties of tracking antenna systems are used. Antenna systems can be divided into those which physically move to point in the right direction and those which process the signal from various antenna elements to keep the antenna beam properly oriented. Examples of the first type are parabolic dishes and examples of the second type are phased arrays. Phased array techniques have recently been developed for the antiballistic missile program and multipurpose aircraft radar. Because the antenna does not physically move, the phased array scans much more rapidly than a mechanically scanned radar. The computers that control the beam steering can also develop more than one beam to track multiple targets while scanning. The Air Force is using a phased array at Eglin AFB, Florida, for satellite tracking.

(c) Sequential lobing techniques such as conical scanning and simultaneous lobing techniques such as monopulse are used for automatic tracking. Both of these systems are employed in satellite communications applications. Both depend on generating an error signal when the satellite is not in the desired part of the antenna pattern and using this error signal to drive the antenna pointing servomechanism. Examples of both systems are described in chapter 8.

CHAPTER 4

SYSTEM DESCRIPTION

4-1. OVERALL SYSTEM DESCRIPTION.

a. A satellite communications link between two DCS subscribers is shown in figure 4-1. The DCS switches and local loops-to-DCS subscribers shown are not actually a part of the satellite communications system but are included to show the relationship of the satellite communications system to the total system.

b. The satellite communication ground complex can logically be separated into the following: (1) The satellite link terminal, (2) the interconnect link, and (3) the terminal operation and control area. The satellite link terminal provides a means of receiving the information relayed by the satellite. The interconnect link provides the means for relaying this information to the DCS switch. The terminal operation and control area ensures that the correct connections are made within the network at the proper time; i.e., the link terminal antenna is pointed at the appropriate satellite and is, in turn, connected by the proper interconnect link to the correct DCS switch.

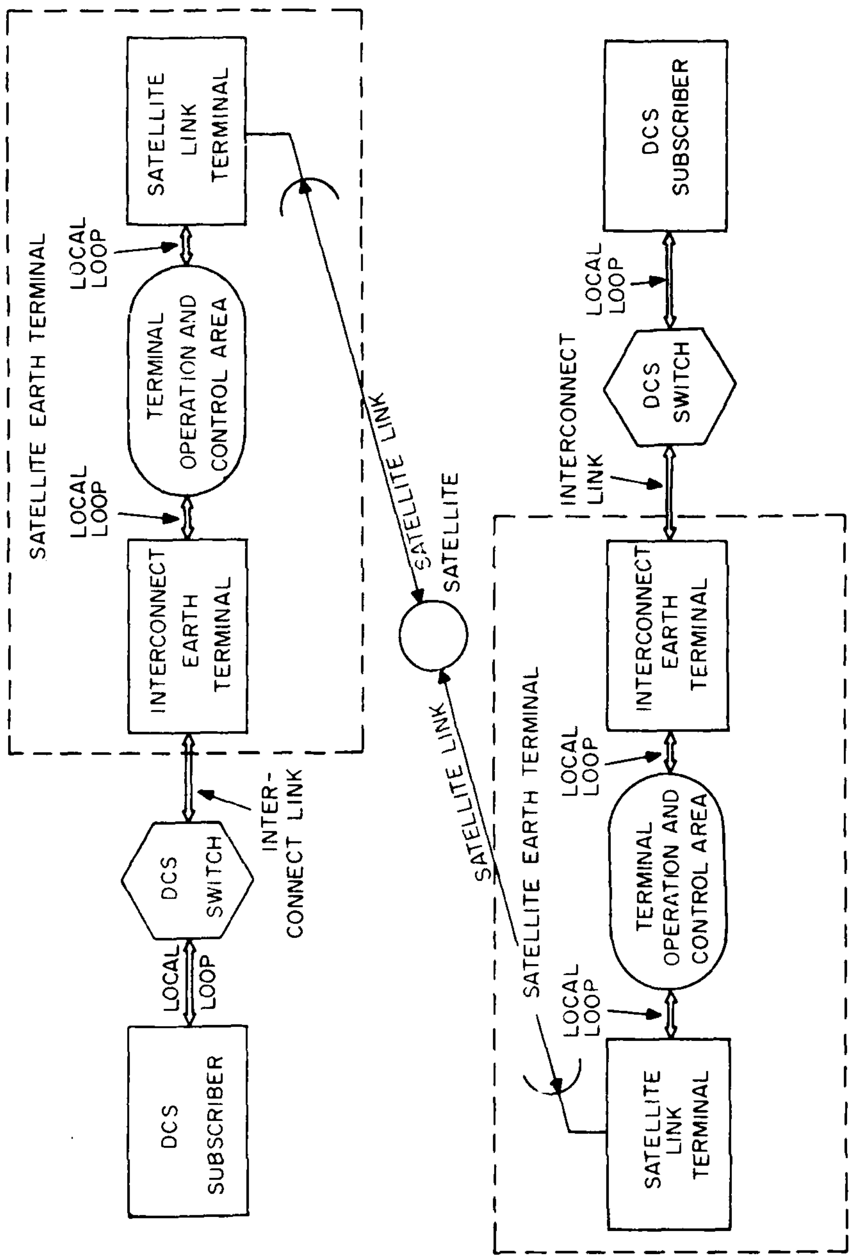
c. The interconnect links consist of conventional communications systems (LOS microwave, tropospheric scatter, cable systems). The satellite link terminals consist of equipment specifically designed to maintain radio contact with the satellite (tracking antennas, computer, etc.).

d. The terminal operation and control area may actually be a separate control facility (in the case of earth terminals employing a number of link terminals) or merely some additional equipment in the satellite link terminal.

4-2. FUNCTIONAL DESCRIPTION.

a. The basic function of a satellite earth terminal is to provide the means by which one DCS switch can transfer information to another DCS switch. Existing satellite earth terminals designed to perform this function and those planned for the future will vary somewhat in detail. Some existing terminals were originally designed and built for specific assignments and include only limited operational characteristics.

b. These terminals are relatively simple, while other stations have additional equipment incorporated into the design. However, the principles of operation are essentially the same for all the stations. Thus, a single basic station can be visualized



CCP105–5–117

Figure 4-1. Satellite communications system, block diagram.

in which the functional requirements will remain equally true for all specific equipments, though the details of equipment and layout may vary between some stations.

(1) *Link terminal.*

(a) The link terminal consists of the equipment designed to communicate with, and through, a satellite. This normally consists of an antenna and associated transmitting, receiving, telemetry, modulating, demodulating, and multiplex equipment.

(b) The link terminal is equipped to accomplish some, or all, of the following functions:

1. Receive signals and process these signals to the proper form for transmission to the satellite.

2. Transmit the processed signals to the satellite.

3. Receive and demodulate signals from the satellite.

4. Generate and record tracking data.

5. Acquire and track the satellite.

6. Process the demodulated information received from the satellite to put it in the desired form for transmission to the DCS switch.

7. Provide timing signals to all station equipment.

8. Provide performance data to operating and control personnel.

(2) *Interconnect link.*

(a) An interconnecting link usually exists between the satellite terminal and the source of communications to be transmitted and received via the satellite. This source is generally a DCS facility (communications center or switching center); however, in some cases it may be the actual user. The interconnection equipment provides a link to the source as well as providing proper input/output signal levels and signal processing.

(b) Specialized processing equipment (inband signaling equipment, delay equalizers, echo suppressors, and order wire facilities) between the two sources may also be included, as would any other device necessary to fulfill the communications relationship between the user and the satellite link. The interconnect link is terminated at one end by the DCS (or user) and at the other by the satellite terminal.

(3) *Terminal operation and control area.*

(a) The terminal operation and control area includes the equipment to establish and exercise operational control over all functions of the earth terminal. This consists of signal processing and buffering equipment; display, timing and performance monitoring systems; equipment command and control systems; the local computer system; and order wire facilities, as required.

(b) The terminal operation and control area is equipped to provide the following functions:

1. Receive, process, and disseminate all data from control sources (orbital predictions, telemetry requirements, and schedules).

2. Process and provide such data received from the satellite or from the link terminal as may be required by control sources.

3. Process performance data received from the link terminal and interconnect link.

4. Record and process telemetry data received from the satellite via the link terminal.

4-3. EARTH TERMINAL SUBSYSTEM.

a. Figure 4-2 represents a typical earth terminal. The major objective of the earth terminal is to transmit signals from the interconnect link to a satellite for relay to another earth terminal and to receive signals relayed from the other earth terminal via the satellite. The paths for this transmission and reception process are indicated by wide lines (black for transmission and white for reception). Secondary functions that support the performance of the ground station's primary function are indicated by narrow lines.

b. Signals to be transmitted are received from the interconnect link. The signals are processed, or stored, in the signal processing and buffering subsystem, which is part of the operation and control function. The buffering subsystem compensates for the changing path length during a pass of the satellite. The signals held in the storage subsystem are read out at the proper time, as determined by the timing subsystem.

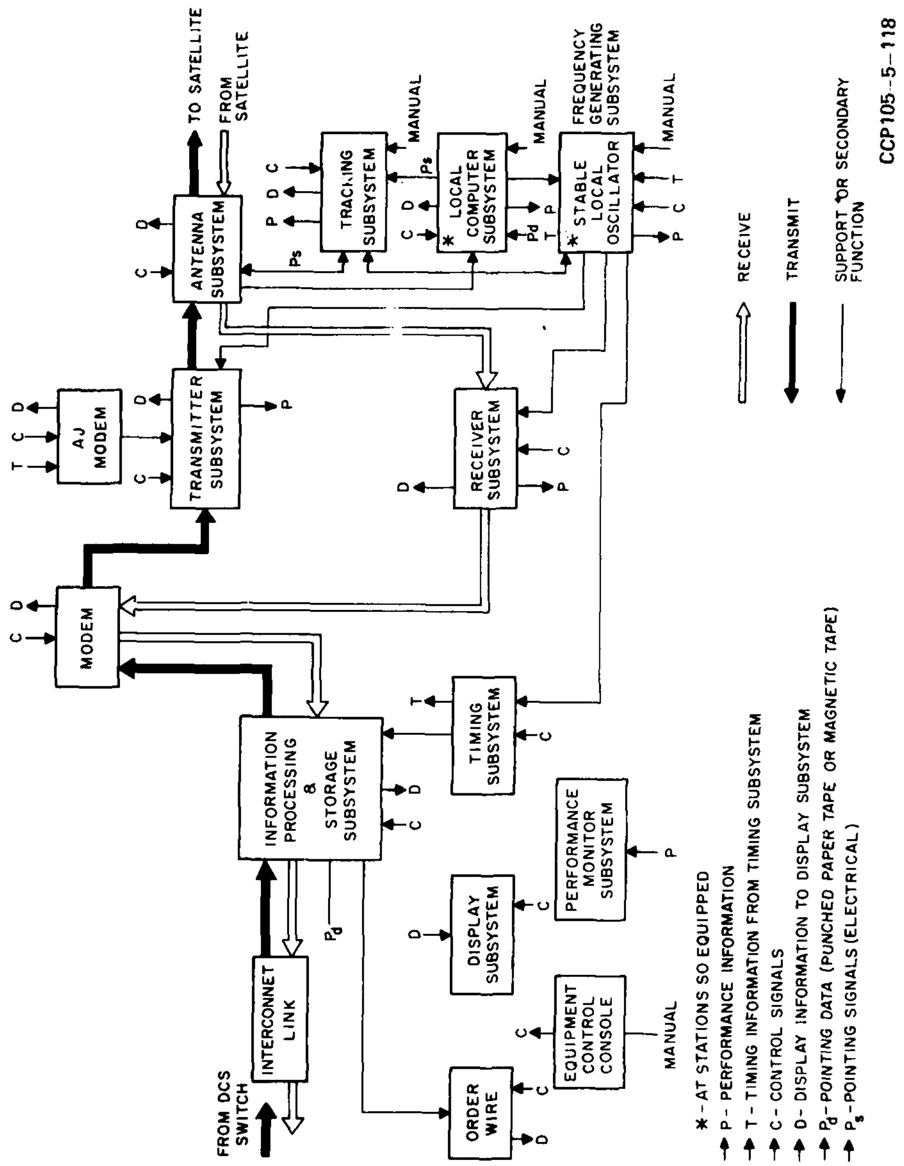


Figure 4-2. Typical earth terminal block diagram.

c. The signal is sent to the modulator-demodulator (modem) where it is multiplexed with other information and put into proper format for transmission. The multiplexed signals are sent to the transmitter subsystem where they generate a modulated rf signal. This modulated rf signal is directed through the antenna subsystem to the satellite. The antenna beam is directed toward the satellite by the tracking subsystem, the local computer subsystem, or both.

d. The signal being transmitted from the other end of the link is received by the antenna subsystem. Information to position the antenna is derived from this received signal by the tracking subsystem. The received signal is also processed by the receiver subsystem, which amplifies the received signal and translates it down to the frequency required for the modem. Since the received signal will be shifted in frequency by the Doppler effect, the local oscillator frequency supplied to the receiver subsystem must also vary with time. The frequency supplied will be predicted by the local computer subsystem or will be controlled by a Doppler tracking loop in the tracking subsystem. The received signal is sent to the modem for demodulation and demultiplexing. The signal is processed and transmitted via the interconnect link to the switching center.

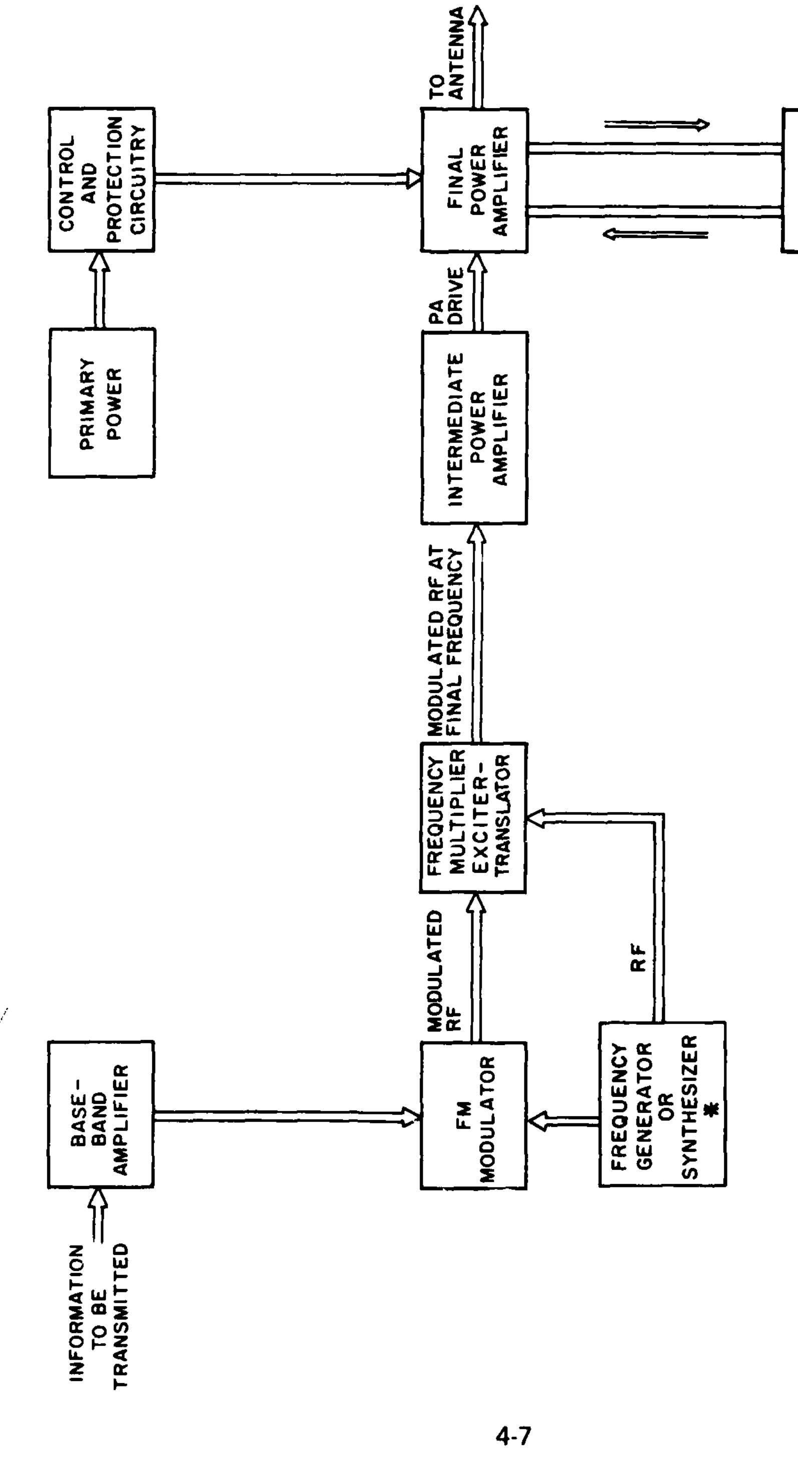
e. Other subsystems essential to the operation of the link terminal, but not involved in its primary function are the display subsystem, equipment control console and performance monitor subsystem, and the order wire which provide the indications and controls necessary to operate the system from one location. Arrows labeled with C or P indicate that the equipment control console and performance monitor are connected to practically every other major subsystem. The order wire allows the operators of the link to use the link for coordination or checkout procedures. The antijamming (aj) modem will not be used normally but will replace the normal modem when the system requires protection against jamming.

(1) *Transmitter subsystem.*

(a) The transmitter in the satellite earth terminal converts the baseband, or signals to be transferred, to the proper rf for transmission and amplifies the signal to a level great enough to overcome losses in the transmission path.

(b) A functional block diagram of a typical ground station transmitter is shown in figure 4-3. Frequency generation and control may be accomplished within the transmitter subsystem or may be achieved through the use of frequencies derived in other subsystems devoted exclusively to frequency control.

(c) The baseband amplifier takes the information to be transmitted (voice signals and teletype signals combined, or a video signal) and amplifies it to the



* AT STATIONS NOT CONTAINING SIMILAR DEVICES AS PART OF THE MODULATOR AND EXCITER.

CCP105–5–119

Figure 4-3. Earth terminal transmitter block diagram.

level needed for modulation. The baseband amplifier must have sufficient bandwidth to pass the total baseband signal and must be linear to avoid crosstalk and distortion of the separate signals contained in the composite baseband signal.

(d) The outputs of the baseband amplifier and the frequency generating unit are fed to the modulator, which imposes the information to be transmitted upon the carrier generated by the frequency generator. The output of the modulator goes to a frequency multiplier or a translator. The purpose of this unit is to raise the frequency of the signal to the proper frequency for transmission. For the transmitter illustrated, a combination of multiplication and translation is used; the unit serves the additional purpose of supplying the desired deviation. For instance, if a carrier that is being deviated 1 kHz is tripled, the deviation of the tripled carrier will be 3 kHz. A frequency multiplier multiplies the carrier frequency and the deviation by the same amount. A frequency translator (or mixer) merely changes the frequency with no change in deviation. By the proper selection of multiplying the translating stage the correct frequency and deviation can be provided: this is an important factor in the performance of the frequency modulation (fm) system. When the signal is at the proper transmission frequency and has the correct deviation, the signal is amplified to the proper power level to drive the final power amplifier.

(e) Satellites normally are at a considerable distance from the earth terminal transmitter and; therefore, require a large power output from the transmitter. At present, klystrons and TWT's give large power outputs at the frequencies required and are discussed in chapter 8. The final power amplifier, with its primary power supply, heat exchanger, and control and protection circuitry make up the bulk of the transmitter.

(2) *Receiver subsystem.*

(a) The receiver in a satellite communications earth station selects the desired signal from other signals and noise, amplifies it to a suitable level, and translates it from the rf employed for transmission by the satellite to the baseband frequency. Figure 4-4 is a block diagram of a typical ground station receiver.

(b) In satellite communications, the performance of the receiving system is determined mainly by the noise introduced in the first stage of amplification. To keep this noise to a minimum, a low-noise device, such as a parametric amplifier, is employed as the first rf amplifier. The parametric amplifier (chapter 8) is cooled to keep thermal noise to a minimum. In addition to a low noise characteristic, the parametric amplifier can handle wide bandwidths and tune over wide ranges.

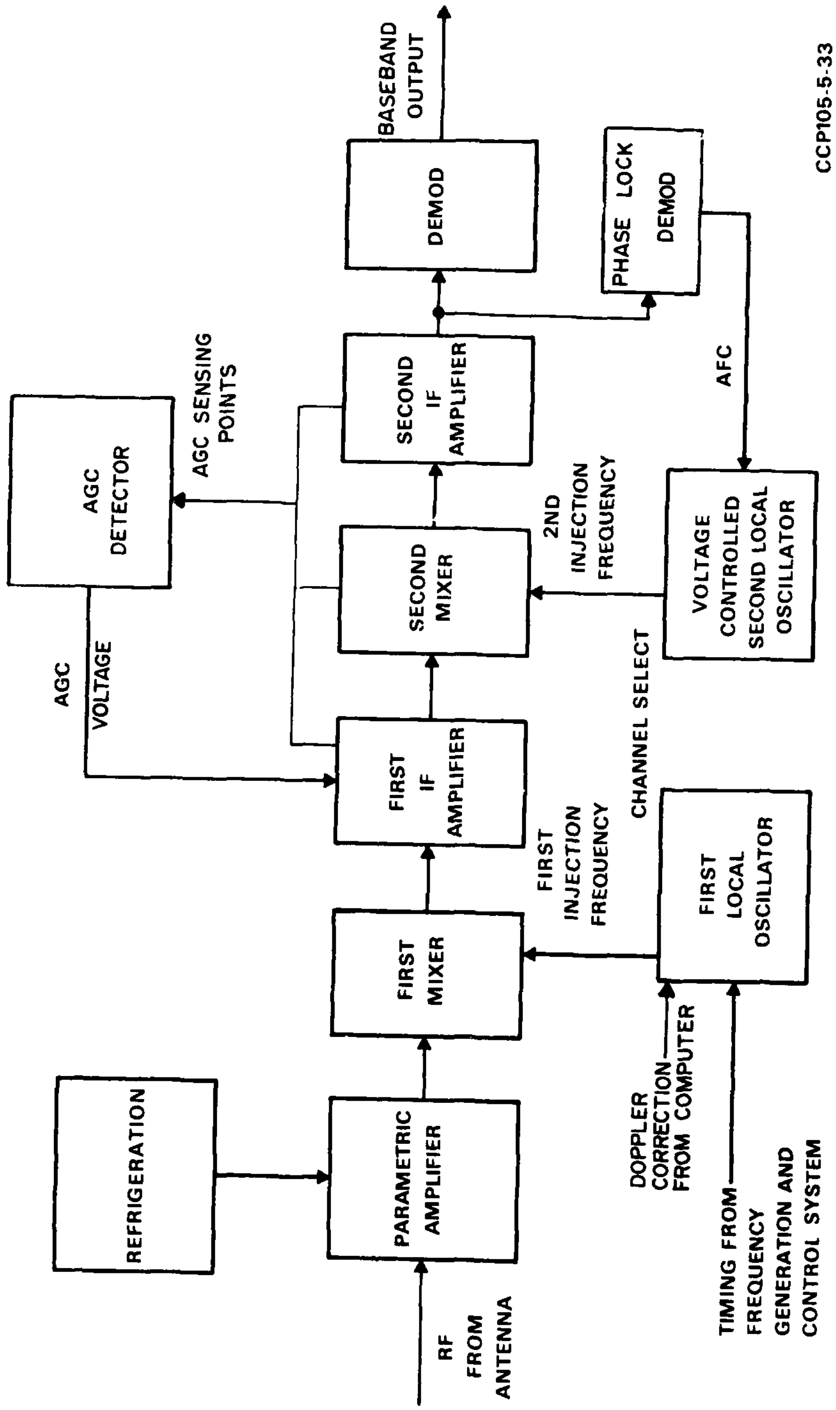


Figure 4-4. Earth terminal communications receiver block diagram.

(c) The output of the parametric amplifier is fed to the first mixer. The signal is mixed with injection frequency from the first local oscillator to yield an intermediate frequency (IF). Typically, the IF is in the 30 to 90 MHz range. The injection frequency must have the same stability requirements as the transmitter. In use, many systems employ the same frequency generating unit for both the transmitter and receiver, or may use a separate frequency generating subsystem. The first mixer must also have a fairly low noise figure since the noise generated in the first mixer stage contributes to the overall noise figure (chapter 8)

(d) The first local oscillator receives timing signals from the earth terminal frequency generation and control subsystem. In addition, a Doppler correction voltage which keeps the receiver on the transmitter frequency (in spite of Doppler shifts caused by satellite velocity) is received from the computer. This Doppler correction voltage is developed from a computer predicted Doppler value based on past satellite orbits. The output of the first mixer is connected to the first IF amplifier.

(e) The first IF amplifier stage amplifies the signal and feeds it to the second mixer. By this time, the signal level is high enough and noise considerations are not of primary concern. The signal is mixed at the second mixer with the frequency from the second local oscillator (a voltage controlled oscillator). To maintain a constant frequency output from the second mixer, a phase-lock loop is used to track the incoming frequency. This consists of the second mixer, second IF amplifier, phase-lock demodulator, and a voltage-controlled second oscillator. The phase-lock loop is explained in chapter 8. In addition to the phase-lock loop correction voltage (automatic frequency control (afc)) applied to the second local oscillator, channel select and slew voltages are also applied. The channel voltage varies according to the portion of the first IF band used. The slew voltage is varied to provide vernier control for initial acquisition and lockup.

(f) Since received signals may vary greatly in amplitude, some provision must be made to prevent extremely strong signals from saturating the amplifiers. This is accomplished by means of automatic gain control (agc). The agc circuit senses the signal level and provides a voltage to reduce amplifier gain, if required. As shown in figure 4-4, the first and second IF amplifiers, the second mixer, and the demodulator are sensing points. The agc generates a control voltage, which is applied to the first IF amplifier to adjust the gain to the optimum, over rather large excursions of signal strength. Since the parametric amplifier gain should be as large as possible to give good noise performance, agc is not used to reduce its gain. The receiver employes an fm signal.

(g) In order to use amplitude modulation (am), a similar system could be employed. In this case the detector would be a simple am detector. The amplification in the parametric amplifier, first IF amplifier, and second IF amplifier must be linear (output proportional to input). In the fm system, limiting (in the form of nonlinear amplification) would be intentionally employed to eliminate any am of the signal.

(3) *Antenna subsystem.*

(a) The antenna subsystem, used for transmitting and receiving at the satellite communications earth terminal, consists of a reflector (to shape a beam of rf energy), a feed (to match the impedance of the atmosphere to that of the transmission line), and the transmission line system (to couple the rf energy from the transmitter-receiver to the feed). Either Cassegrainian or prime-focus feed antennas are used at the earth terminal. For a description of these antennas and their advantages, refer to paragraph 7-4.

(b) In reception, the received signals are focused by the antenna reflector on the feed, propagated along the transmission line, and passed through a receiver filter to the receiver. The main purpose of the receiver filter is to reject the transmitter frequency of the local transmitter. The receiver provides discrimination against unwanted frequencies, but the high level of the transmitted power, if not attenuated, would literally burn up the first stage of the receiver. The receiver filter normally provides about 100 dB isolation at the transmitter frequency.

(c) For transmission, the signal is generated in the transmitter and passes through the harmonic filter, transmitter filter, and out the feed. The filters in this line prevent radiation of unwanted components of the transmitter output. The harmonic filter dissipates any harmonics of the transmitter frequency generated by nonlinearities in the power amplifier. This prevents radiation and interference with the station's own, or other, receivers. In addition to the desired transmitter output, noise is amplified in the transmitter power amplifier and transmitted. This noise is generally many dB below the desired transmitter output and decreases with separation from the carrier frequency. As a result, the noise at the receiver frequency may be 120 dB below the transmitted frequency; however, with the high-power transmitters used, this noise might still be 80 dB above the receiver noise. This, of course, is intolerable. The transmitter filter rejects this noise and will have a rejection of more than 100 dB at the received frequency. To achieve the high rejection, a high Q waveguide filter is used. The high Q waveguide filter results in larger than normal peak currents in the walls of the waveguide, which causes more attenuation in the filter than in a normal piece of waveguide. As a result of the high currents in the filter, heating takes place and some method of cooling the filter is required. Cooling methods range from simple cooling fins to forced air systems. In extreme cases liquid cooling is used.

(4) *Tracking subsystem.*

(a) After the satellite has been acquired, it is tracked using a tracking system such as that illustrated in figure 4-5. The objective of a tracking system is to determine the elevation and azimuth angle corrections required to move the antenna so that it continues to point directly at the moving satellite. The outputs of the comparator are identical to those described in the discussion on the monopulse method (paragraph 8-7). When the azimuth and elevation angles, respectively, drift off the antenna's center beam, ϕ and θ have a nonzero value. The symbol Σ is a reference signal.

(b) All receiver microwave components are mounted in an equipment package at the apex of the antenna's parabolic reflector. The feed horns are connected to the microwave receivers by short lengths of waveguide. Since the equipment package moves with the antenna, no flexible joints with their potential loss coefficients are required between the feed horns and the receiver front ends.

(c) Normally, each of the three receivers has two IF stages. The signals enter the first mixer without preamplification. After mixing, they are amplified and passed to the second IF stage. In order to obtain high acquisition sensitivity, the reference signal Σ is detected in a narrowband phase-lock loop. The phase-locked oscillator of the reference channel acts as the local oscillator for all three tracking receivers, decreasing fm noise in the difference channels. Coherent detection with post detection filtering (error signal filters) is employed in the different channels to achieve a low noise tracking error signal.

(d) The amplitude of the error signal, generated by the antenna misalignment, is compared to the reference signal Σ in the error detector. For small displacements, this unit produces an output proportional to the angular displacement. This output signal is amplified in order to drive a servomotor that realigns the antenna pointing angles.

(e) To initiate the process of acquisition and tracking, ephemeris data are fed into the local computer. The computer calculates the initial search angles and transmits the necessary signals to the servomotors, which positions the antenna for acquisition. Essentially, this computer drives the servosystem in a manner that reduces the difference between a stored position reference in the computer and the output of the position encoder.

(f) In order that current knowledge of the satellite's orbit be available to the local ground station for possible reacquisition and for ephemeris updating, a computer-compiled record of the satellite's track may be kept. This is performed by recording the output of the position encoder in digitalized form and printing, or taping, the resultant timed record.

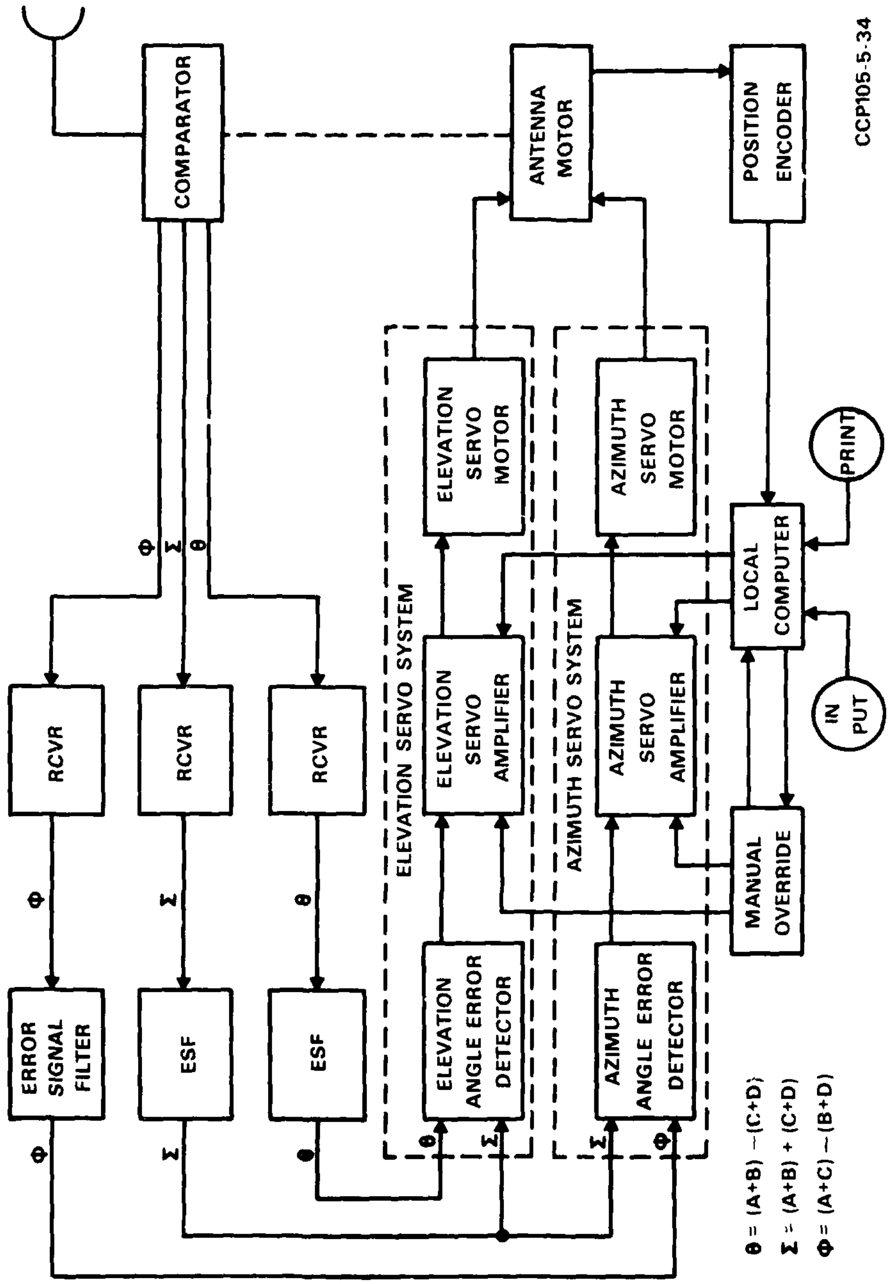


Figure 4-5. Antenna tracking system block diagram.

(g) A manual override is available to station personnel which will allow retention of a minimal workable system should any malfunction occur.

(5) *Modem and antijamming modem.*

(a) The signal to be transmitted is fed to the modem from the signal processing subsystem or alternately to the aj modem. The modem is employed in normal operation. The aj modem is employed when the system must operate while being jammed or in conditions of unusual interference.

(b) The modem transforms the information to the form required for transmission and also combines individual information channels to form one composite baseband signal. This operation, which may be accomplished on a time-division or frequency-division basis, is called multiplexing. (A discussion of multiplexing is given in paragraph 8-4.) When time-division multiplexing is employed, the modem will receive timing pulses from the timing subsystem. In cases where frequency-division multiplex is employed, frequencies derived from the frequency standard section of the stable local oscillator will control the frequencies used in multiplexing.

(c) The aj modem serves the same function as the normal modem but provides protection against enemy jamming by increasing the bandwidth of the signal.

(6) *Timing and frequency generating subsystem.*

(a) The timing and frequency generating subsystems are closely associated and will be discussed together.

(b) The acquisition of a signal at some point in space at some instant in time (spatial-acquisition) must also involve acquisition of the signal in the frequency sense. In other words, nothing would be accomplished if an antenna locates the right point in time and space but is not tuned to the right frequency. Further, if it is tuned to the right frequency initially, the communciations will be broken unless there is some means of staying on the right frequency as well as the right space-time location. These considerations point up the need for strict time and frequency control, not only within the local station, but throughout the entire system. Details of time and frequency standards are discussed in paragraph 8-5, and no attempt is made to identify specific items of equipment used in this type of control. This discussion is in general terms and only the broad, general methods are outlined.

(c) To meet the requirement for strict timing control throughout the entire system, it is obvious that synchronization of all time elements must be maintained

AD-A114 385 ARMY COMMUNICATIONS COMMAND FORT HUACHUCA AZ F/G 17/2

INTRODUCTION TO SATELLITE COMMUNICATIONS.(U)

APR 74

UNCLASSIFIED ACC-CCP-105-5 NL

2 OF 4

AD A114385

by a central source. This standard time is supplied to all of the local stations through one of the several techniques. Each station will have a station time standard or local clock that, though of extreme accuracy in its own right, is even more accurate when synchronized with all other timing standards throughout the global system. This local time standard is the focal point for fulfilling all timing requirements within the station, whether those requirements be the synchronization of timing commands or space-time instructions to the local computer.

(d) Frequency control primarily concerns two areas. These are listed as frequency instabilities (in ground transmitter, satellite translator, satellite transmitter, and ground receiver frequency references) and Doppler shifts.

(e) Equipment frequency instabilities are negligible. Oscillators designed and used in the satellite and ground station are extremely stable, and any slight instabilities are corrected by various compensating circuits as discussed in paragraphs 4-3(1) and 4-3(2) on the transmitter and receiver, respectively.

(f) The relative motion between the satellite and the ground station will create a condition which, in effect, causes the frequency of the signal from the satellite to change (Doppler shift) during its orbital pass. This makes it necessary to control the receiver so that it will change frequency automatically to follow the change in the incoming signal. This is accomplished by comparing the incoming signal with a signal of known frequency generated by a highly accurate standard. As the incoming signal varies, the variation from the standard signal is followed as a differential frequency and is used to generate electronic commands that shift the receiver frequency. This same differential frequency is processed and used to drive display and recording devices and to shift the transmitter frequency.

(7) *Display and equipment control subsystem.*

(a) A typical satellite communications ground station has display and control subsystems as a part of the operations control center. The heart of the terminal operation and control center is the control console. It provides the personnel in operational control of the earth terminal with a means of maintaining central control over all subsystems.

(b) Also, the control console provides a means of inserting manual override signals to units normally automatically controlled and serves as a focal point for the origination and reception of verbal information to and from operating and maintenance personnel. The station supervisor has facilities available for communicating with the various units under his command.

(8) *Earth terminal computer.*

(a) Some earth terminals are equipped with a general purpose digital computer. This computer receives instructions from the system's master computer at the SCCF. Together, these computers provide data for antenna pointing and frequency control. The digital computer is also used for secondary functions, by means of time sharing and provision of appropriate programs. The SCCF computer computes data for several weeks in advance and sends this data to the earth terminal computer. The data pertains to time, position, and frequency information concerning the satellites in view of the station.

(b) A digital storage device (magnetic drum or magnetic tape machine) is connected to the computer for storing permanent or semipermanent data or programs. This storage device will read out and store information or feed in (control-instruct) information to the computer. Examples of its application are as follows:

1. Store ephemeris data (predictions) as received from the SCCF via data circuits, or airmail, and read out the data to the computer, as required.

2. Store precise ephemeris data as supplied by the tracking subsystem, when it is operating in the lock-on mode, as a matter of record and for possible later transmission to the SCCF.

(c) Earth terminals, not having the computer, receive the ephemeris data from the computer of another earth terminal or from the master computer at the SCCF.

(9) *Order wire subsystem.* The order wire subsystem provides earth terminal personnel with access to some of the channels relayed via satellites. This provides a channel in which link operation, or checkout, can be coordinated with other ground stations in the network.

(10) *Performance monitor subsystem.*

(a) The performance monitor subsystem monitors system parameters to provide operating and maintenance personnel with information concerning the performance status of the various subsystems. Typical examples of parameters monitored might be transmitter power, frequency and deviation, forward and reflected power into the antenna feed system, and receiver noise figure.

(b) To provide the required information, the performance monitor will accept inputs from almost every major subsystem. In turn, much of the information developed in the performance monitor subsystem will be sent to the display subsystem, where it will be presented visually or audibly to the system operator.

4-4. INTERSITE LINK SYSTEMS.

a. *Central Switching Centers.*

(1) Preceding parts of this publication have discussed the various systems within a typical satellite communications ground station. These have primarily pertained to the equipment groupings and internal functions of the ground station. All of these are absolutely necessary if the ground station is to accomplish its mission. However, unless the communications processed by the ground stations can be extended to the persons who require the use of these communications, no practical value can be realized from the impressive array of equipment stacked in a station. This section discusses the means of accomplishing this extension.

(2) It is impossible to give every subscriber a station that will communicate through a satellite to every other subscriber. Instead, central switching centers are established. These switching centers serve many subscribers and have a variety of means at their disposal for forwarding traffic from one subscriber to another, including satellite communications. Before this means can be used, the traffic must be sent to, or received from, the satellite communications earth terminal. The common designation, in referring to this portion of the communications traffic picture, is the interconnect link. This is the communications link that ties the earth terminal to the designated switching center.

(3) In a typical earth terminal, one or more of three general types of communications systems are used to link the station to the switching center. These are tropospheric scatter systems, LOS microwave systems, and landline systems. The choice of system depends upon many things; the location of the switch, the distance and terrain to be spanned, whether the link crosses friendly or enemy territory, the available frequencies, the existence of mutual interference between earth terminal equipment and intersite link equipment, and facilities presently installed. To familiarize the reader with possible configurations that may be encountered and to give some insight into why particular systems were designated in specific locations and how they work, the following discussion is included.

b. *LOS Microwave Systems.*

(1) The term LOS as used here, refers to a path between two terminals over which no obstruction exists to the rf path. Obstructions can include natural and manmade obstacles or the curvature of the earth. For the power and the frequency normally used, this LOS consideration determines the maximum path length that

can be spanned between terminals. Under average propagation conditions, using normal antenna tower heights and transmitter power, the LOS path will be approximately 30 miles with no relay.

(2) The carrier frequency for a microwave system generally ranges from 0.350 to 12 GHz. Power requirements are on the order of 1 to 10 watts. Figure 4-6 illustrates the configuration of the equipment in an LOS terminal with a frequency diversity application. This type of system has a 600-channel capacity and offers exceptional reliability.

c. *Tropospheric Scatter.*

(1) Tropospheric scatter system frequencies range from 400 MHz to 8 GHz. Path length ranges from 75 to 400 miles. Power requirements vary from 1 to 75 kW, depending on path length, propagation angle, type and degree of diversity employed, and the number of channels used (up to 600 channels). In addition to the fact that this type of system provides a high degree of reliability over a long period of time, tropospheric scatter is an ideal interim system capable of providing for initial or expansion needs.

(2) Figure 4-7 is a block diagram of a typical tropospheric scatter system terminal configuration. Following the signal flow for a tropospheric scatter terminal, the information inputs to the multiplexing equipment are in the form of separate 4-kHz channels. These channels are arranged side-by-side by the multiplexing equipment in the form of a composite baseband signal called a frequency-division multiplex (fdm) signal. The baseband signal is subsequently sent to the transmitter terminal equipment, where it frequency-modulates a subcarrier that is raised to the carrier frequency by means of frequency multiplication and frequency translation. The new fdm signal at the carrier frequency is amplified by the tropo exciter, which sends the signal to a high powered klystron amplifier. At this point, two methods are available for propagating the signal to the receiving terminal. If the system design is such that separate antennas are used for transmitting and receiving, the signal is sent directly to the transmitting antenna. If, as is normally the case, one antenna is used for both transmitting and receiving, a high power duplexer is included in the system. The function of the duplexer is to permit simultaneous operation of the transmitter and receiver using the same antenna. In this case, the signal is sent to the duplexer and then to the antenna.

(3) The receiving antenna intercepts the scattered signal and passes it through the duplexer. A low noise front end (normally a parametric amplifier) receives the signal, processes it, and passes it on to the tropospheric receiver (fig. 4-7). This particular system uses dual space diversity with horizontal and vertical polarization. This means that each receiver block can receive two signals of a particular polarization.

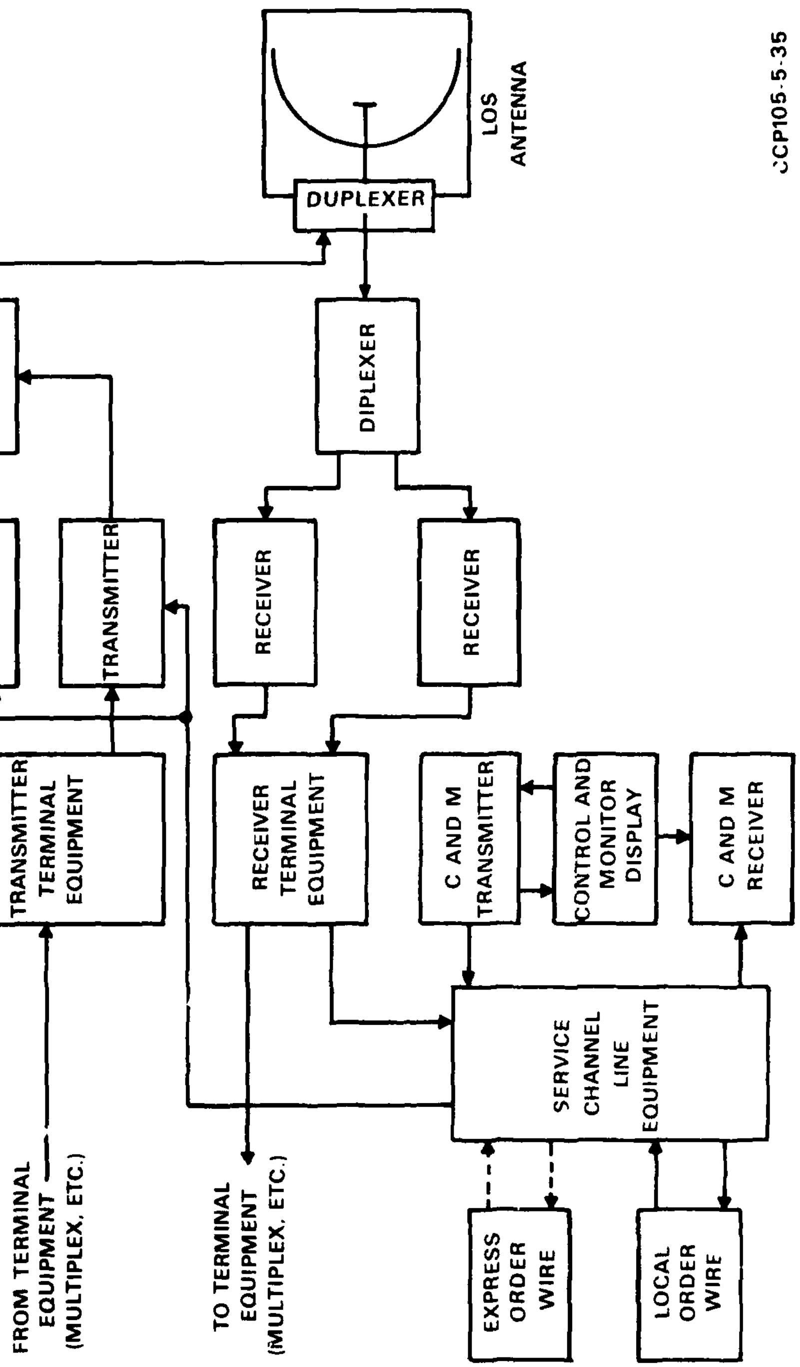


Figure 4-6. Dual frequency diversity system block diagram.

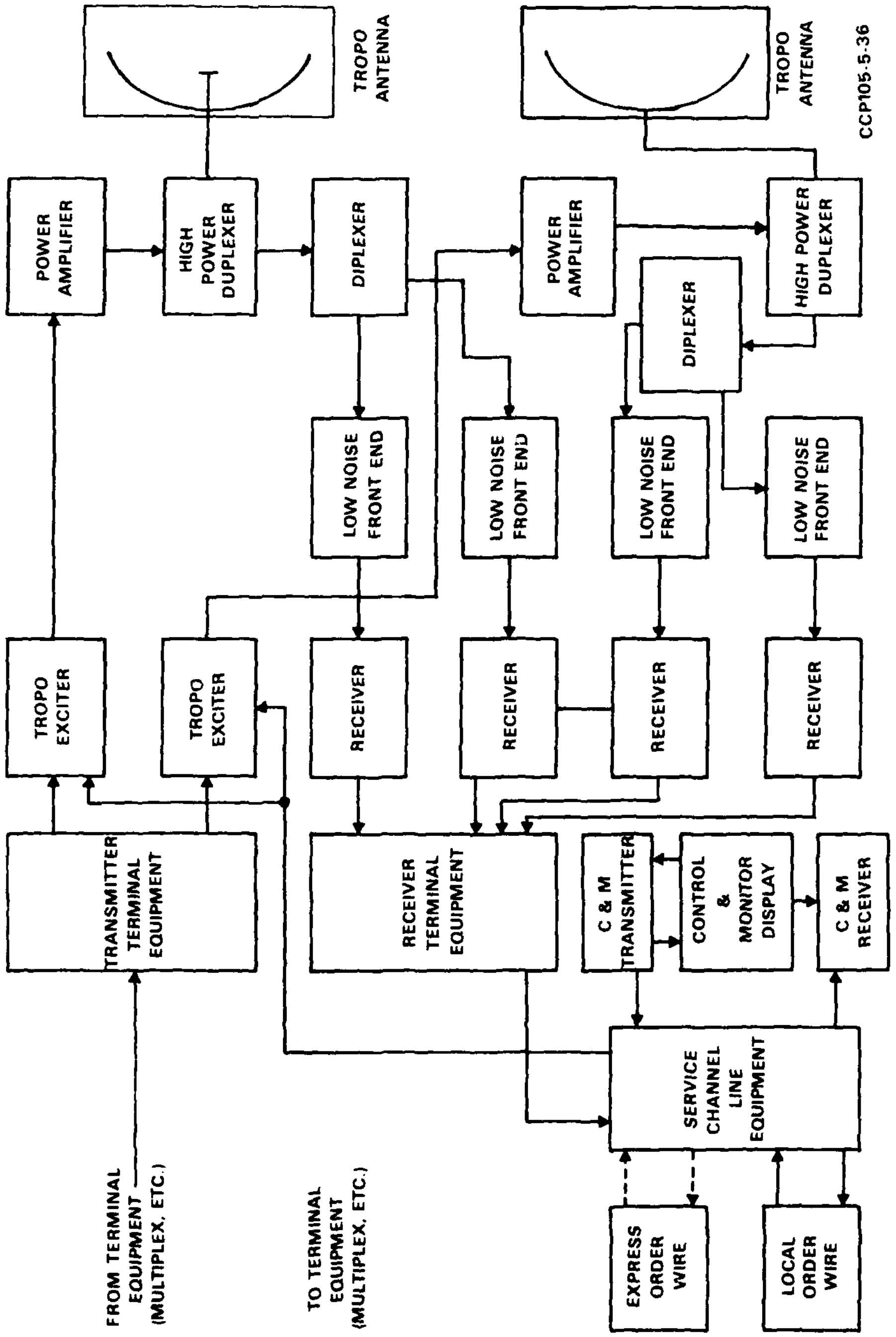


Figure 4-7. Quad diversity system block diagram.

(4) The receiver amplifies the rf signal from the front end (the parametric amplifier is frequency selective) and sends it to the receiver terminal equipment, where it is mixed with a signal developed by the local oscillator in the receiver terminal group. This mixing process produces an IF which carries the intelligence.

(5) The intelligence is obtained by passing the IF through a demodulator. Once this is accomplished, the intelligence is in the form of a baseband signal. This baseband is sent to the demultiplexing equipment, which separates the information contained in the baseband signal into separate communications channels.

(6) In the discussion of the signal flow, some of the blocks shown in the tropospheric scatter terminal diagram (fig. 4-7) are not mentioned. These blocks are ancillary equipment and are not necessarily a part of the signal flow. The basic functions of this type of equipment are described below.

(7) The control and monitor blocks shown for receiver and transmitter, together with their display console, perform the housekeeping functions for the terminal. Their functions consist of monitoring complete major units and critical circuits within the major units. The local order wire forms an independent communications channel, whereby control and monitor functions may be carried out without interfering with the operation of the individual communications channels. In addition, the order wire function is used to connect several radio links independent of the multichannel radio system, thereby establishing an independent communication path for control and monitor functions over several terminals.

d. *Signal Flow for a Microwave Terminal.*

(1) The signal flow through an LOS microwave terminal is essentially the same as that shown for the tropospheric scatter system with the exception that some equipment functions are deleted and several others are substituted. As an example, the relatively low power associated with an LOS terminal precludes the requirement for a power amplifier because the exciter can supply the rf carrier at the required level.

(2) In addition, a bipolar feed can be used to provide some discrimination between the transmitted and received signals on the same antennna. A branching filter network is used to separate the individual rf channels.

e. *Landlines.*

(1) Certain factors contribute heavily to any decision made for the use of landline connections for the intersite link. These factors are considered in the following discussion concerning the use of cable systems and open-wire systems.

(2) The frequency range of both cable and open-wire systems depends, for the most part, on the particular carrier system employed. The physical characteristics of the two differ to the extent that in open-wire systems both 2- and 4-wire applications are in use, and in addition, heavier gage wire, as opposed to the 19 AWG wire generally used in cable systems, is used. The large channel capacity of cable systems makes them more suitable for the satellite communications system than the smaller channel capability of open-wire systems. An important factor is that both methods provide good security from the interception of data by hostile forces. However, it must be pointed out that the landline path must be in friendly territory where the possibility of sabotage is virtually nonexistent. The use of amplifiers along the landline path can extend the range of a landline system to practically any path length desired. Figure 4-8 shows a multichannel cable system typical of those that could be employed for the intersite link.

(3) In a multichannel cable system, the multiplexing, modulation, demodulation, and demultiplexing are similar to the tropospheric scatter or microwave system. When using cable the signal is sent over wire. The radio system signal is propagated from an antenna. In the case of cable, low power amplification is possible at the source and amplifiers along the cable route can be used to increase the distance of the signal path.

(4) From these discussions of the various systems that could be used for the intersite link, it can be seen that all of them (microwave, tropospheric scatter, and landlines) provide a large number of channels with good reliability. Since a reasonably wide range of path lengths is available, any of the three systems could satisfy an individual intersite requirement, depending upon the distance between sites and the terrain involved. In addition, each system can be expanded to meet the needs of the future.

f. *Order Wire.*

(1) As with any other communications station, order wire units will be provided. An order wire is the name given to the link that is usually set up between communications stations for use by station personnel.

(2) This permits operations or maintenance personnel to talk with each other during the course of troubleshooting, while performing tests on a link, or while

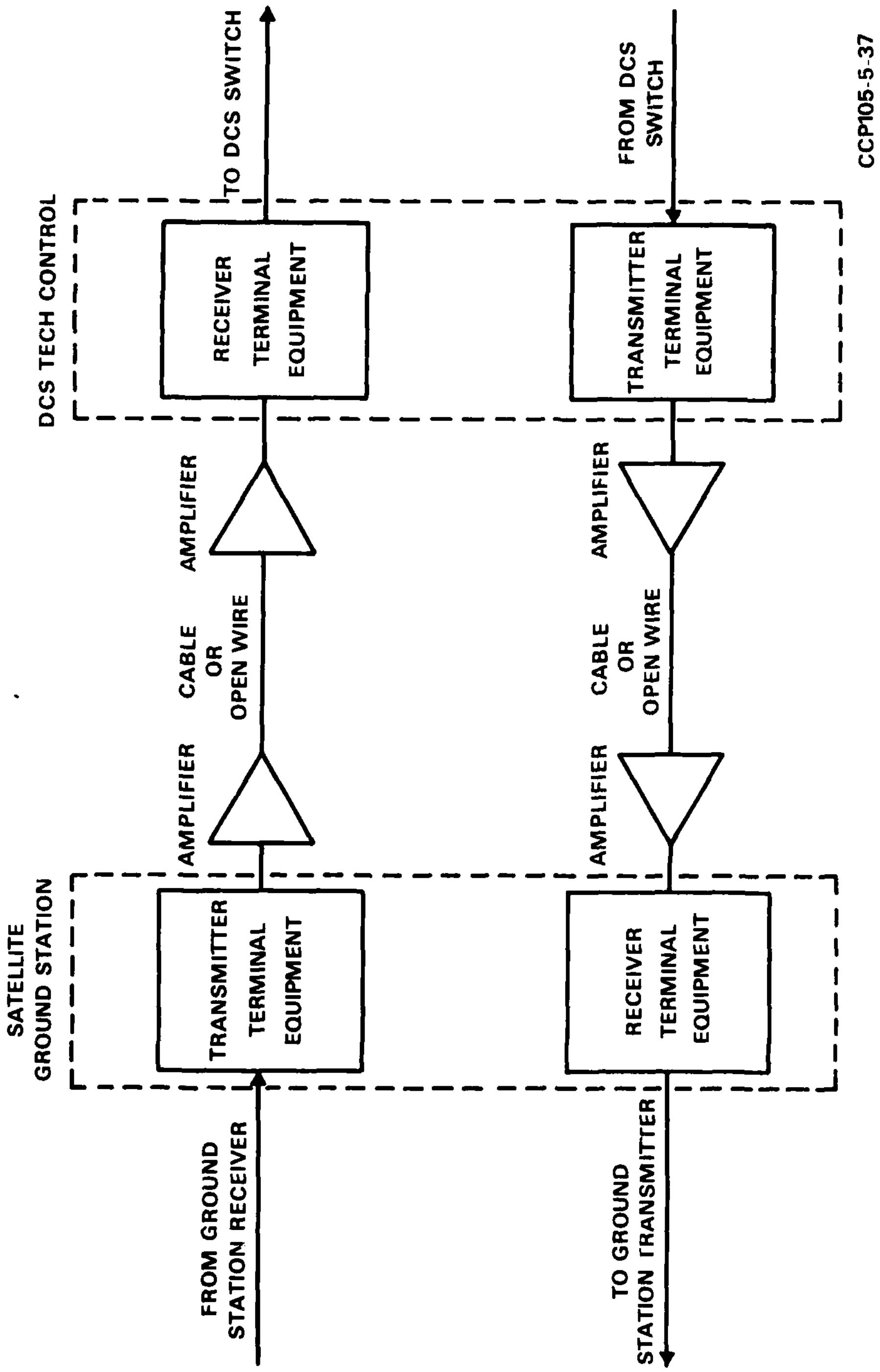


Figure 4-8. Multichannel wire system block diagram.

going through the formalities of establishing the link. Normally, the order wire link is a full-duplex link.

4-5. SYSTEM AVAILABILITY.

a. *Satellite Availability.*

(1) In order to establish a satellite communications link between a pair of ground stations, there must be a satellite present in the part of the sky that is mutually visible to the two stations. The probability of one satellite of a system being visible to a given pair of ground stations is determined by a number of parameters. These are as follows:

(a) Location of ground stations.

(b) Orbit altitude (or period, since one positively determines the other).

(c) Orbit inclination.

(d) Orbit shape (circular or elliptical).

(e) Number of orbital planes and uniformity of spacing.

(f) Number of satellites in each orbital plane.

(g) Minimum allowable elevation angle of ground station antennas.

(h) Minimum acceptable message time.

(i) Maximum acceptable waiting time.

(2) Referencing the preceding items ((a) through (i)), the location of the ground stations (a) is fixed by the communication requirements of the system and availability of suitable site. Orbit altitude (b) involves a number of considerations; e.g., high altitude orbits will make the satellite visible to more widely separated ground stations and for longer periods of time (h), but the longer period of a high altitude orbit will increase the waiting time (i) before its reappearance. Also, the longer path length will increase the path loss of the radio signal.

(3) Orbit inclination (c) involves the projected path of the satellite over the earth's surface. Polar orbits offer an advantage where most of the ground station pairs are 30 degrees or more in latitude, since a station at 30 degrees latitude can see the satellite (6,000 miles or more in altitude) until it passes over the nearer pole of the earth.

(4) The orbit shape (d) considers the eccentricity of the orbits. Variations in altitude resulting from ellipticity of the orbit do not offer any clear advantage; therefore, the orbits should be as nearly circular as is practicable. A greater number of orbital planes (e) will increase trunk availability, but this must be balanced against the cost of additional launches.

(5) The number of satellites in each orbital plane (f) is another consideration. It is desirable to have as many satellites in each plane as possible. This is accomplished by using a single booster to carry as many as eight satellites. Minimum elevation angle of ground station antennas (g) is a consideration, since satellite communications antennas will not normally be operated below 7-1/2 degrees of elevation. At lower angles, the rf signal must be propagated through the earth's atmosphere for greater distances, resulting in greater absorption of rf energy by the oxygen and water vapor in the atmosphere. Furthermore, the effective noise temperature of the system increases at low antenna elevation angles. This is due to the thermal radiation noise of the earth being picked up by either the side lobes or the main beam of the antenna. Additionally, antenna elevations below 7-1/2 degrees impose restrictions on the choice of ground station sites, since the ground station area would have to be very flat and free of obstructions.

(6) Other considerations that affect satellite availability include satellite failure, eclipses which prevent solar power generation, and environmental conditions (solar storms) that disrupt the link. These situations can be relieved by replenishment launches, storage battery capacity, and satellite redundancy, respectively.

b. *Satellite Bunching.*

(1) Where several orbital planes cross, satellite bunching can occur if more than one satellite is near the proximity point. This in itself is not serious, provided that the system has enough satellites to ensure that the bunching does not leave a void in another area, thus causing loss of capability.

(2) Bunching can provide additional channel capability on occasion, when the bunched satellites are far enough apart so that their signals can be separated by the earth terminal antennas.

c. *Satellite Hand Over.*

(1) With quasi-synchronous satellites, a given satellite will serve a particular pair of ground stations for a period of several days. The end of this time period is determined by the time that the satellite sinks below the antenna horizon (7-1/2 degrees) for either of the two ground stations operating as a link. When this occurs,

the ground station pair must switch their antennas to another available satellite. There will be some interruption or outage of service while the antennas acquire and lock on to the new satellite and the stations verify that a link has been established.

(2) A second antenna can be used to acquire and lock on to a new, visible satellite before the satellite in service has disappeared. In this way, outage time can be reduced to less than a second.

(3) The hand-over requirement is greatly eased with quasi-synchronous satellites, since hand over occurs only once every several days. With synchronous satellites there will be no reason to go through a hand-over process.

d. *Satellite Command and Control.*

(1) The term satellite command includes the coded signals which direct the satellite to turn on (or off) its transmitter or receiver, or switch between wideband and narrow band modes. These functions are the responsibility of the ground stations. The tracking stations are responsible for issuing commands to the vehicle to change altitude in relationship to the earth, to read out stored telemetry data, or as in the case of fixed or stable satellites, to change orbit.

(2) For those satellites incapable of automatic receiver on/off operation, commands are issued from ground stations for their control. Signals are transmitted in a coded form which, when matched with stored commands, cause the vehicle to respond to the command issued. A mismatch receives no response. Spurious codes and noise neither receive a response nor an action from the vehicle. Interference and control by unauthorized signals is also minimized.

(3) Acknowledgment of the command is transmitted to the ground station prior to the execution of data/telemetry transmission. Verification of the acknowledgment is followed by the expected output from the vehicle or the execution of the command issued to the satellite.

e. *Channel Capacity.*

(1) The IDCSP is made up of satellites at quasi-synchronous altitudes. Each satellite has the capacity to handle a minimum of five multiplex channels (some satellites are multiplexed to 12 channels). As illustrated in figure 4-9, the satellite receives F_1 and F_2 from station A and translates these carriers to F_6 and F_7 for transmission to station B. At the same time the satellite receives F_8, F_9 and F_{10} from station B and translates them to F_3, F_4, and F_5 for transmission to station A.

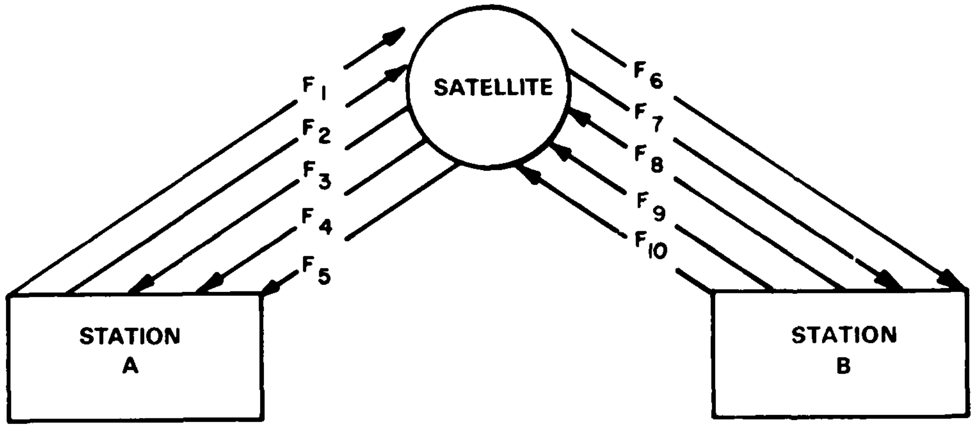


Figure 4-9. Channel depiction diagram.

(2) Channel configuration is flexible and can be altered to comply with operation requirements; spreading the spectrum to overcome noise, or jamming, would reduce the number of available channels.

4-6. EARTH STATION CONTROL AND SCHEDULING.

a. The satellite communications system requires a means of controlling and scheduling the particular satellites in orbit that will be used by pairs of earth stations (terminals) at any given time. With many satellites in orbit and many earth station pairs desiring to communicate with each other, the magnitude of this task can be appreciated. Also, at any time there may be several satellites in common view of earth station pairs. Almost all satellites will be in common view of more than one pair of earth terminals most of the time. Add to this already complicated situation, the fact that the satellites have multiple access capability and it becomes clear that we must also consider which channels of which satellites will be used by which earth stations. Further, malfunctions in equipment do occur, requiring the operational condition of satellite and earth-based equipment to be considered.

b. The SCCF at DCA Headquarters maintains overall system control responsibility. It assigns link terminals and satellites for test operations and assigns satellites to special users. Operational schedules are generated only by the SCCF or, in appropriate cases, by the subordinate area communications control function (ACCF). The weekly earth station communications status reports are effective for periods during which operational communications are handled (including emergency operations).

(1) *Earth station control function.*

(a) The functions required to control the communications satellites and the ground network associated with them are defined as the control subsystem (fig. 4-10). In terms of facilities, the control subsystem includes the SCCF and the link terminals. The control subsystem interfaces with the Defense Communication Agency Operations Center (DCAOC), satellite communications test operations center (SCTOC), satellite test center (STC), area communications operations center (ACOC), and the DCS technical control (DCSTC) facilities. In terms of functions, the subsystem includes the SCCF, the ACCF, and the earth station control function (ESCF). These functions are supplemented by the orbit and station visibility generation functions of the STC.

(b) Overall control is the responsibility of the SCCF as implemented through the control subsystem. In keeping with the concept of moderate decentralization, local control of an area is delegated to an ACCF, who is responsible for the direction and monitoring of activities of all ESCF's within the area.

(c) The ESCF coordinates and controls the operation and assignments of link terminals at each earth station. It is responsive to requirements from the SCCF and the ACCF and performs local coordination and control functions with the local DCSTC.

(d) The ESCF is not a facility and does not contain equipment as such. It is housed in the facility provided for the senior, or lead link, terminal at the earth station. Personnel normally performing the ESCF are assigned to the link terminal, and the function and duties are considered collateral to other link terminal routine duties. No electrical interfaces are associated with the ESCF. Communication equipment requirements are satisfied by the link terminal in which the ESCF is located.

(e) The ESCF functions are:

1. Accept and coordinate schedules received from the SCCF and the ACCF and distribute them to appropriate link terminals at the earth station.

2. Accept, coordinate, and distribute antenna pointing data (STC/SCCF data base).

3. Accept and process routine and exception status information from the link terminal(s).

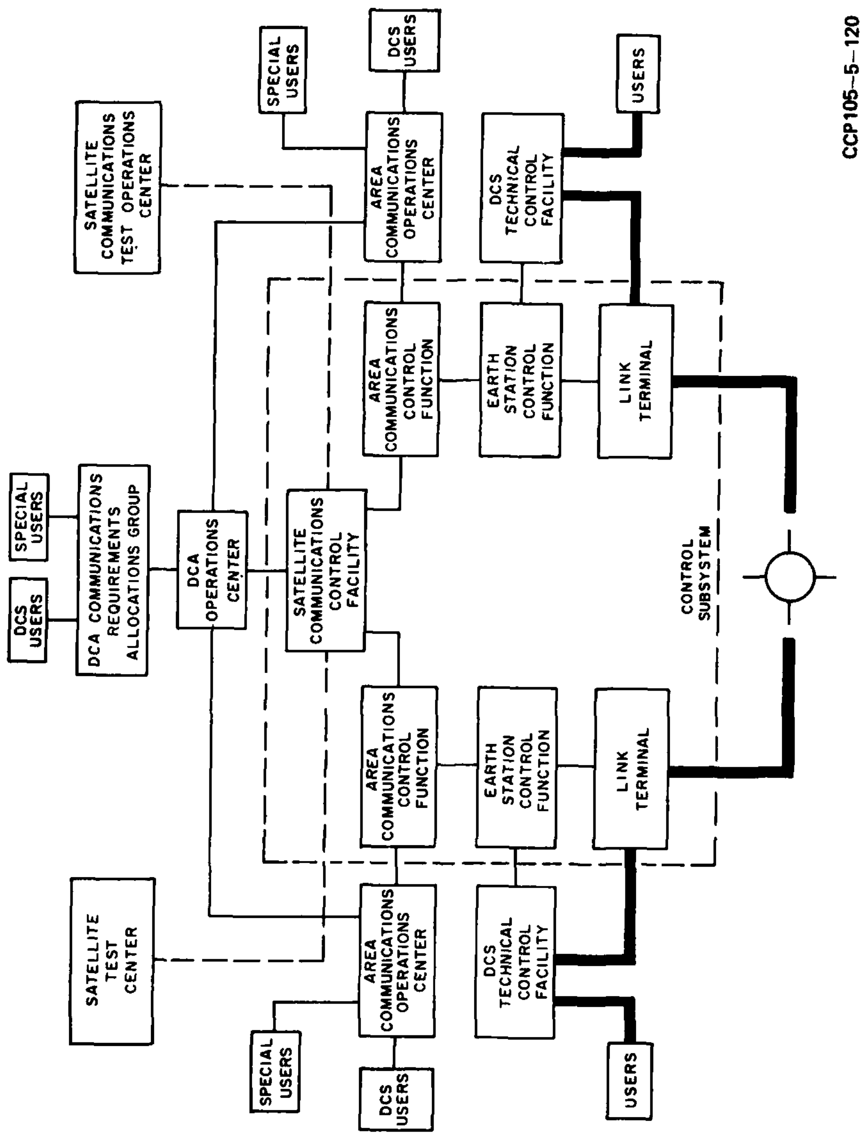


Figure 4-10. Control subsystem block diagram.

4. Provide the SCCF and the ACCF with routine and exception status reports.

5. Coordinate the gathering and distribution of link terminal collected track data to SCCF.

6. Conduct local coordination between link terminals and the local DCSTC.

(f) The ESCF is divided into three basic activities; operations control, status, and communications.

(2) *Control doctrine.*

(a) The SCCF is operated as an integral element of the DCS and is semidependent on the local DCS stations for communications support and the Headquarters, DCA, and also on local military headquarters for administrative technical and logistical support. Also, support is provided by the US Army Satellite Communications Agency, when applicable. ACCF's are control elements subordinate to the SCCF and collocated with an ACOC, and usually an existing DCS station. ESCF's are control elements subordinate to an ACCF and the SCCF. Each ESCF is dependent on the assigned technical control facility (TCF) and subordinate link terminals for communications, maintenance, and technical support. In fulfilling its functions, each element:

1. Extends the span of control to facilitate responsiveness to the requirements of senior military commanders and other authorized users.

2. Provides the capability requisite to establishing and reestablishing the assigned portion of the DCS, or other assigned communication links rapidly in event of a catastrophe.

3. Provides the capability for improving status reporting and accumulating basic inventory, historical, and analytical data.

4. Provides increased flow of control information for improved management direction.

(b) The locations of the control elements are listed in table 4-1.

(c) Because of the number and variety of control actions required to ensure proper performance, complete centralized control is not practical. Therefore, a moderate decentralization of the system control is established; to be performed in varying degrees at different locations where it is most practical and economical.

Table 4-1. Location of control elements

Control element	Location
SCCF	DCAOC, Bldg. 12, Arlington, Virginia
ACCF/Western	DCA-CONUS, Fort Carson, Colorado
ACCF/EUROPE	DCA-EUROPE, Vaihingen, Germany
ACCF/PACIFIC	DCA-PACIFIC, Kunia Hawaii

(d) To ensure optimum continuity and quality of communications, each level of control exercises some degree of monitorship over the action of subordinate levels of control. Using information derived from this activity, the SCCF supervisory control point can coordinate and support the activitites of subordinates; thus, controlling and solving problems in areas controlled or monitored by several subordinates.

(e) Each earth station operates assigned facilities and takes all corrective action possible to ensure that service is provided in accordance with assigned communications requirements, specified engineering requirements and priorities. When resolution cannot be accomplished by the earth station, the responsibility for solution is assumed by the lowest echelon authorized to coordinate resources (ACCF) and passes to a higher echelon (SCCF or DCAOC) if this lowest echelon also lacks resolution capability or authority. If the communications trunks involved emanate from and terminate outside one ACCF area, unless otherwise directed, responsibility for coordination rests with the SCCF and requests for assistance should be promptly relayed to the SCCF by, or through, the ACCF.

(f) When functioning as an element of the DCS, link terminals will operate in accordance with DCAC 310-55-1, "Operational Direction and Status for the Defense Communications System (DCS)," where applicable.

(g) At those link terminals where the satellite communications trunk does not interface with the DCS, the functions of the technical controller are performed by the control link terminal operator.

(h) In addition to the activities outlined above, the ESCF is also responsible for establishing coordination procedures for conducting operations with the local DCSTC and between link terminal facilities of the earth station. They must clearly indicate areas of responsibility, the requirements necessary for good working relations, and conformity with the expressed, or implied, requirements of the system.

CHAPTER 5

MULTIPLE ACCESS OF SATELLITES

5-1. INTRODUCTION.

a. For economic and technical reasons a separate satellite can not be made available for use by each pair of earth terminals that requires a satellite communications link. Also, an individual transponder can not be provided for every possible satellite link. Each satellite transponder must relay several communications links. That is to say, each transponder must simultaneously receive, translate, amplify, and transmit signals from several earth terminals. Having signals from several earth terminals pass through a satellite is referred to as multiple access.

b. Virtually all of the problems associated with providing a satisfactory multiple access capability stem in one way or another from the fact that present-day satellite transponders are hard-limiting repeaters. The TWT used in the power output stage of communications satellites must be driven to near saturation for most efficient operation. But operating in the nonlinear region of the TWT creates signal distortion problems. When more than one signal is being received, intermodulation products are generated within the transponder and small-signal suppression occurs.

c. Most of the newly designed satellites, such as the INTELSAT IV series, have not been hard-limiting satellites. That is to say, signals above a certain level are not abruptly cut off, but instead there is a rather gradual reduction in gain near the saturation level. Nevertheless, the transponders are not truly linear and intermodulation products are still generated, although their amplitude has been somewhat reduced. The problem of small-signal suppression also remains in some degree.

d. Linear transponders may be available some day. At present, the loss of efficiency in linear amplifiers more than offsets their advantages.

e. There are five different techniques for providing multiple access. They are as follows:

(1) Frequency division multiple access (FDMA).

(2) Time division multiple access (TDMA).

(3) Spread spectrum multiple access (SSMA).

(4) Single sideband uplink and frequency division multiplex downlink (SSB/FDM).

(5) Pulse address multiple access (PADMA).

5-2. FREQUENCY DIVISION MULTIPLE ACCESS.

a. FDMA has been the most frequently used technique primarily because it is the simplest to implement. With FDMA, each earth terminal accessing a satellite merely transmits on a different frequency, as depicted in figure 5-1. The satellite transponder acts as a common amplifier translator.

b. The shortcomings of FDMA are derived from the hard-limiting operation of the satellite repeater. The intermodulation products generated from the hard-limiting action waste satellite output power and add noise to the signal. Furthermore, the frequencies of the carriers accessing the satellite must be selected carefully to avoid having large amplitude intermodulation products appear at the same frequency as that of an output communication signal of the satellite.

c. FDMA also requires careful uplink power control and coordination to handle the problem of small-signal suppression and to ensure that a given pair of users does not use excessive amounts of the available satellite output power.

d. The advantages and disadvantages of FDMA are summarized as follows:

(1) *Advantages.*

(a) No special circuitry or equipment is required.

(b) No network timing is required.

(c) Voice can be transmitted in normal analog form.

(2) *Disadvantages.*

(a) The system is vulnerable to jamming.

(b) Intermodulation products waste power and add noise to the signal.

(c) Operating frequencies must be carefully selected to reduce interference from intermodulation product signals.

(d) Uplink power coordination is required to achieve efficient use of transponder power and to avoid large-signal capture.

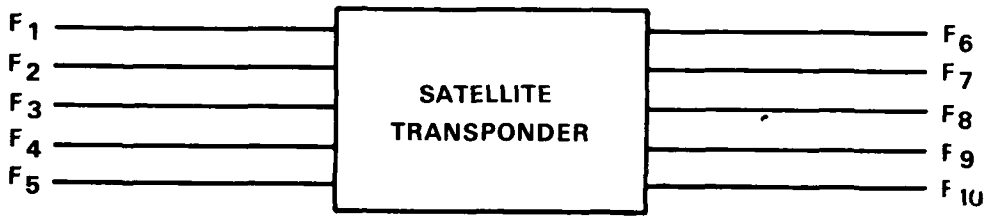


Figure 5-1. Frequency division multiple access diagram.

e. FDMA is the technique presently used in the INTELSAT, IDCSP and the DSCS Phase II satellites. In actual operation, the shortcomings of the FDMA technique have not proved to be a serious obstacle to satisfactory satellite communications.

5-3. TIME DIVISION MULTIPLE ACCESS.

a. With TDMA, each earth station is assigned exclusive use of the satellite repeater during specific time slots. Access is accomplished by designating a particular sequence of time slots to each earth station. The time slots for a channel occur periodically with the period of repetition called the "frame repetition rate" of the system. Figure 5-2 illustrates how a time frame might be divided into time slots.

b. Different length time slots can be used for the varying traffic requirements of the earth station. Figure 5-3 illustrates how a time frame could be divided into unequal time slots for handling different channel requirements.

c. Since each earth station has exclusive use of the satellite repeater during the period of its assigned time slot, the problem of large-signal capture is avoided and there is no need for careful coordination of output power for various accessing earth stations.

d. Theoretically, TDMA appears to be the most efficient multiple access technique that can be devised when working with hard-limiting satellites of present-day design. That is, TDMA can provide the highest information rate for a given repeater output power.

e. The main disadvantages of TDMA are that it requires careful network timing and synchronization and it is vulnerable to selective jamming of individual

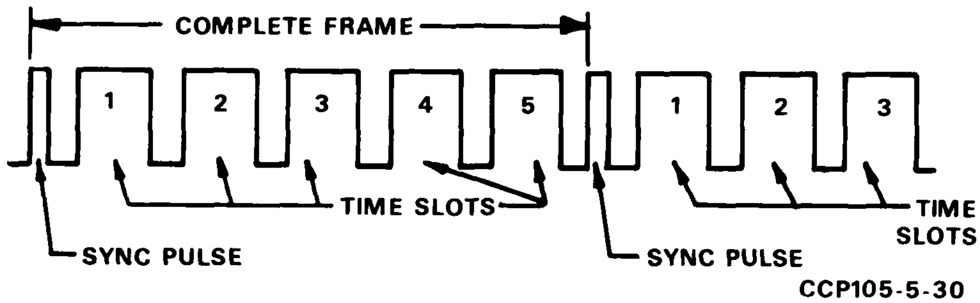


Figure 5-2. Equal time frame slot format waveform.

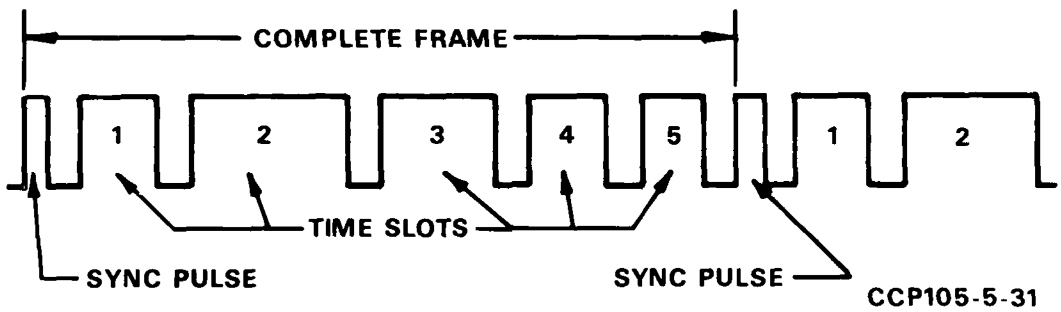


Figure 5-3. Variable time frame slot format waveform.

links. Network clock synchronization can be accomplished by timing marks originated in, or repeated by, the satellite. Recent tests, conducted by the Communications Satellite (COMSAT) Corporation, indicate that the timing synchronization required for TDMA may not be as difficult to implement as was previously assumed.

f. A further shortcoming of TDMA is that it can only handle information in digital form. Speech must, therefore, be digitized before it can be transmitted by a TDMA system. Speech digitizing is accomplished by pcm. The need for digitizing traffic may prove an advantage in the future when digital transmission becomes more common.

g. Another disadvantage of TDMA is that data buffering is required. The process whereby the difference in rate of flow of data from one device to another is compensated for is called "buffering." Normally data, which may be digitized speech, will be arriving at the earth station in a continuous stream. TDMA can only be transmitted during the assigned time slot periods. At the receiving end, the data will be received in bursts during the time slots. A reverse buffering process is required to smooth out the data at the receiving earth station before relaying it on to the user.

h. The advantages and disadvantages of TDMA are summarized as follows:

(1) *Advantages.*

(a) It is a very efficient system in terms of information capacity.

(b) There is no problem of interaction between signals; therefore, intermodulation is not a problem.

(c) There is no problem of large-signal capture of the satellite; hence, earth station power coordination is not required.

(d) It can accommodate earth stations of different sizes and stations having widely different characteristics.

(e) It can accommodate earth stations having widely different channel requirements.

(2) *Disadvantages.*

(a) Network timing is required.

(b) All information must be converted to digital form.

(c) The system is vulnerable to selective jamming.

(d) Buffering of data is required in the TDMA modem.

5-4. SPREAD SPECTRUM MULTIPLE ACCESS.

a. With SSMA, each carrier signal simultaneously occupies the same wide portion of the spectrum. The bandwidth of the signal will usually be at least 10 MHz wide and may occupy the whole bandpass of the satellite.

b. SSMA is usually accomplished by adding the digitized voice signal, or other digital signal, to a high rate pseudo-random noise code formed by a shift register generator. The pseudo-random sum signal modulates the earth terminal transmitter. At the receiving end, an identical shift register generator operating in synchronism with the transmitting-end generator produces the same noise code. The noise code is in effect canceled from the signal, leaving the intelligence signal.

c. Multiple access is achieved by the use of different pseudo-random noise codes for different pairs of earth stations accessing the satellite. Although there is a certain amount of mutual interference between different pseudo-random noise codes, the amount of interference between any two codes is actually very small. However, as more pseudo-random noise signals are added, the mutual interference increases; this limits the total number of earth stations that can occupy the same portion of the spectrum.

d. The advantages and disadvantages of SSMA may be summarized as follows:

(1) *Advantages.*

(a) It automatically provides the system with a high degree of jam resistance at low data rates.

(b) A passive monitor (such as an enemy might employ for intelligence purposes) can not tell how much traffic is being passed over the system, since the pseudo-random noise code looks the same regardless of whether or not it is carrying intelligence.

(2) *Disadvantages.*

(a) The spread spectrum modulation and demodulation equipment adds cost and complexity to the system.

(b) Wideband amplifiers are required; this adds cost and increases power requirements.

(c) Each earth terminal pair must keep their pseudo-random noise generators synchronized while communicating with each other.

(d) All information must be converted to digital form.

5-5. SINGLE SIDEBAND MODULATION UPLINK AND FREQUENCY MODULATION DOWNLINK.

a. Single sideband modulation uplink and frequency modulation downlink (SSB/FM) uses conventional single sideband for the uplink. Each voice channel is transmitted to the satellite via its own single sideband channel. The single sideband channels are multiplexed in the satellite to form a baseband signal for modulating the fm transmitter of the satellite to transmit back to earth. The fm downlink signal is demodulated at the earth station to recover the baseband signal. The baseband signal is demultiplexed into the individual voice channels. It should be noted that in order to recover any particular voice channel it is necessary to demodulate and demultiplex the entire fm downlink signal.

b. With SSB/FM, access to the satellite is achieved by merely checking to determine what channels are free and tuning the earth station transmitter to a clear channel. It is assumed that the receiving terminal would be constantly searching all channels for the appearance of a signal. This straightforward way of achieving access is the main advantage of the technique.

c. In theory SSB/FM is a very efficient system that would allow an almost unlimited number of earth terminals to access the satellite without system degradation. In practice the system has a number of disadvantages, particularly for a military system.

d. The advantages and disadvantages of SSB/FM are summarized as follows:

(1) *Advantages.*

(a) Provides the capability for many earth terminals to access one satellite.

(b) Power control is not critical for the earth terminals accessing the satellite.

(2) *Disadvantages.*

(a) The satellite is more complicated.

(b) Careful carrier frequency control must be maintained at each earth station.

(c) The entire system capacity is degraded to that of the smallest earth terminal accessing the satellite when different size terminals must receive and demodulate the same fm signal.

(d) The SSB/FM technique does not lend itself readily to the use of antijam modulation techniques.

e. It does not appear that SSB/FM will be used for military satellite communications systems in the foreseeable future. NASA has recently launched a satellite in its Applied Technological Satellite (ATS) series that uses the SSB/FM multiple access method. NASA will use this satellite to evaluate the advantages and disadvantages of this technique.

5-6. PULSE ADDRESS MULTIPLE ACCESS.

a. PADMA can take several forms. In general, multiple access with PADMA is achieved by assigning a unique combination or matrix of time and frequency slots to each user station. The time and frequency slots would occur periodically within the period of repetition being used by the earth station pair. Figure 5-4 illustrates a matrix of time and frequency slots. For example, the time-frequency slots t_1f_{15}, t_2f_3, and t_4f_9 might be assigned to a station to indicate the presence of a mark signal and t_5f_{11}, t_6f_5, t_7f_{12}, and t_8f_{16} might indicate the presence of a space signal. Each receiver will respond only to the unique time and frequency slot combination (pulse address) that it has been assigned.

b. There are various techniques for adding the message (either voice or data) to the pulse address signal. Using this concept of having one address code for a mark signal and a second address code for a space signal, information can be sent in binary form. This would require that speech be digitized using the method described in chapter 7.

c. The operational procedure for stations using PADMA is given below.

(1) Prior to transmitting a message the sending earth station operator adjusts his receiver to the addressee's pulse code address to verify that the channel is not busy.

(2) If the channel is not busy the message is transmitted in that channel.

(3) The receiving earth station responds to its own pulse address channel and rejects all other signals.

d. The advantages and disadvantages of PADMA are summarized as follows:

(1) *Advantages.*

(a) Rapid earth terminal pair synchronization can be accomplished.

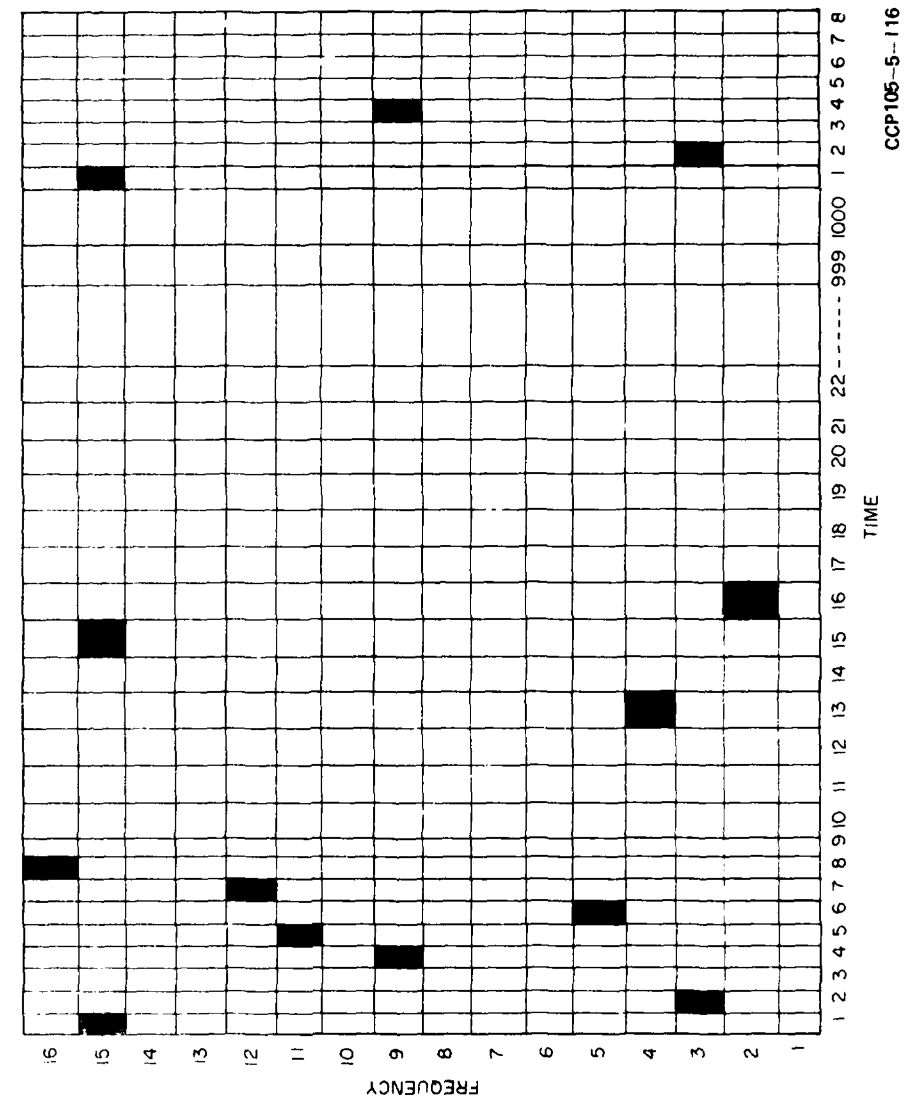


Figure 5-4. Time frequency matrix for pulse address system.

(b) Network timing is not required although the system will have greater capacity if network timing is employed.

(c) The system automatically provides some degree of jam resistance.

(2) *Disadvantages.*

(a) The technique has a limited capacity compared to other techniques.

(b) Stable frequency standards are needed.

(c) Stable time standards are needed to achieve maximum capacity.

CHAPTER 6

JAMMING OF SATELLITES

6-1. INTRODUCTION.

a. Jamming is defined as action taken against a communications system to prevent it from accomplishing its function. Antijamming (aj) is defined as action taken to nullify, or cancel, the effects of the jamming.

b. Jamming can be accomplished by physically disabling or destroying a portion of the communications system or subjecting the system to electromagnetic radiation to cause interference with, or loss of, communications. It is this electromagnetic radiation jamming that is of concern.

6-2. JAMMING TECHNIQUES.

a. At first thought it might appear that a satellite communications system could be jammed on either the downlink (satellite-to-earth receiving station) or on the uplink (earth transmitting station-to-satellite).

b. In considering jamming of the downlink it becomes obvious that the highly directional antenna of the earth receiving station will reject virtually all signals except those coming from the immediate direction of the satellite. Nevertheless, jamming of the downlink could still be accomplished by a satellite-based jammer placed in orbit so that it would appear (to the earth receiving station) to have tracking coordinates near those of the communications satellite. However, the time required to place such a jammer satellite in orbit would be too great to allow it to be of any tactical advantage. Also, the cost of such a project for jamming only one element of a communications link would be prohibitive. Therefore, only the jamming of the uplink by an earth-based jamming station will be considered.

c. It is safe to assume that shortly after a communications satellite is placed in operation, a potential enemy with any degree of sophistication in electronic warfare will have compiled data on the satellite. This data would probably include the following:

(1) Ephemeris data indicating the satellite's orbit characteristics.

(2) Information relative to frequency, bandwidth, and modulation of the satellite's communications equipment.

d. On the basis of this information, a jamming technique would be selected. Jamming operates on the premise that if enough jamming power is radiated into a receiver, the legitimate signal will be indistinguishable from the unwanted signal. Various jamming techniques are summarized, as follows:

(1) *Spot jamming.*

(a) Spot jamming consists of concentrating all of the jamming transmitter energy in a narrow band to achieve a high noise density in that band.

(b) For purposes of considering noise density in relation to jamming, noise density power is discussed in terms of watts/MHz. Therefore, a spot jammer concentrating a 1,000-watt noise signal in a 1-MHz bandwidth is said to have a noise density of 1,000 watts/MHz.

(2) *Broadband jamming.*

(a) Spot jamming is ineffective against a wideband signal or a signal that shifts in frequency faster than the spot jammer can follow it. Under these circumstances broadband jamming is used. Broadband jamming spreads energy out over a broad band of frequencies rather than on a spot frequency.

(b) However, the broadband jamming results in a lower noise density at any one particular frequency. Therefore, a spot jamming signal of 1,000 watts/MHz, when spread over a 100-MHz bandwidth, will have only a 10-watt/MHz noise density.

(3) *Swept-spot jamming.* Swept-spot jamming combines the high noise density of spot jamming with the wide bandwidth of broadband jamming. This is accomplished by sweeping the spot jammer, at a very high sweep rate, across the spectrum to be jammed.

6-3. ANTIJAMMING TECHNIQUES.

a. *Frequency Hopping.*

(1) A spot jammer can be easily defeated by changing frequency. This requires the jammer to determine the new frequency and move to it. During this period, the frequency can again be changed (frequency hopping). Since it takes approximately 0.25 second for the earth station-satellite-earth station trip, hopping frequency four times per second denies the spot jammer access to the communications link. However, it should be noted that frequency hopping uses a considerable amount of the available spectrum; thus, reducing the number of available communications channels.

(2) Frequency hopping forces the jammer to spread his energy (broadband jamming). This reduces the jammer noise density on any one channel.

b. *Spread Spectrum Modulation.*

(1) The most efficient aj technique is the processing game afforded by the spread spectrum modulation technique (with power). In this technique, the information to be transmitted is added to a hf pseudo-random noise code generated by a shift register and used to modulate the earth terminal transmitter. At the receiving end, an identical shift register noise generator, synchronized to the shift register at the transmitter, generates the same noise code to cancel itself from the incoming signal. Thus, only the transmitted information remains.

(2) The spread spectrum signal may occupy the entire bandwidth of the satellite simultaneously with several other spread spectrum signals (each having a different pseudo-random noise code). The pseudo-random noise code looks the same to the jammer whether or not it is carrying intelligence. This forces the jammer to spread his energy throughout the entire bandwidth of the random noise, resulting in a reduced jamming noise density and no knowledge of whether the jamming is effective.

c. *Effects of Hard-Limiting.*

(1) As the power input to a TWT (used in satellites as power output stages) increases, the power output increases. However, there is a saturation point where additional power input (overdriving) will produce no additional output; this is known as limiting (fig. 6-1). In a TWT, limiting occurs rather abruptly (hard-limiting). One result of overdriving a hard-limiting device is the production of large amplitude intermodulation products of the various input signals. This dissipates much of the available power throughout all of these intermodulation products, resulting in the output of the desired signals decreasing as the TWT is overdriven. Thus, a strong signal can effectively "capture" the satellite's communications channels by suppressing weak signals.

(2) Hard-limiting can be considered an aj technique since it becomes a power contest between the jammer and the legitimate ground station, with the legitimate ground station usually being in a more favorable geographic position to capture the satellite. However, it must be realized that in a situation such as this, legitimate small signals are suppressed and satellite channel capacity is reduced.

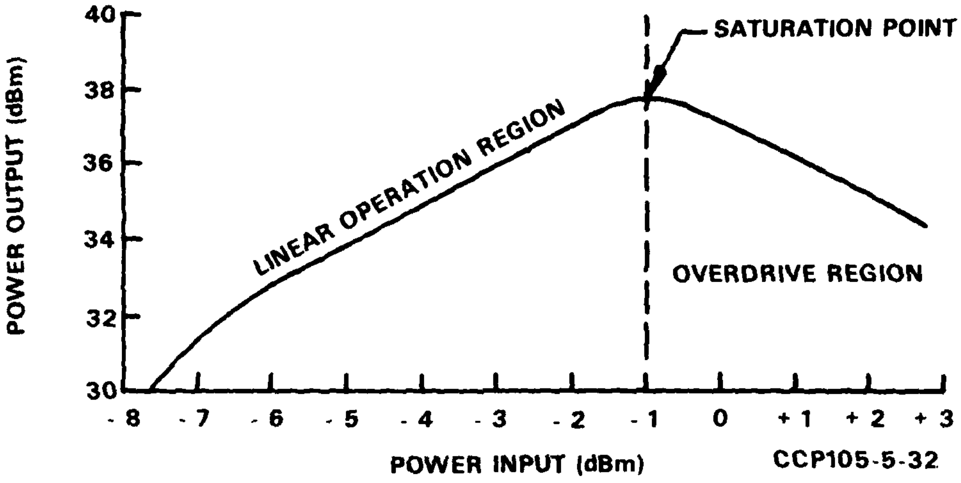


Figure 6-1. Typical TWT power curve.

d. *Satellite Directional Antennas.*

(1) Directional antennas on communications satellites provide an aj capability depending on the orientation of the satellite directional antenna. Jamming power received at the satellite receiver is represented by:

$$P_{rj} = \frac{P_j G_j G_{sj} \lambda^2}{(4\pi R_o)^2} \tag{6.1}$$

where

P_{rj} is received jamming power

P_j is the jammer power

G_j is the jammer antenna gain

G_{sj} is satellite antenna gain for the jamming signal

R_o is the distance between the satellite and the jammer

λ is the wavelength of the jamming signal.

(2) It can be seen from equation (6.1) that a low value of G_{sj} will have a large effect on the jammer power received at the satellite receiver. A 20 or 30 dB attenuation of the jammer signal due to off-axis loss will degrade the jammer signal to the extent that it will have no effect.

CHAPTER 7

SATELLITE COMMUNICATIONS LINK PARAMETERS

7-1. INTRODUCTION.

a. Active communications satellite systems are limited by satellite transmitter power on the downlinks and to a lesser extent by satellite receiver sensitivity on the uplinks. These limitations are constantly being reduced through research and development efforts. Until the new equipment and techniques, arising out of research and development have been operationally tested, it is necessary to plan operational systems based on proven state-of-the-art techniques. As a result, earth stations now being planned will have their characteristics largely determined by the limitations imposed by the current state-of-the-art in satellite design.

b. The communication link can be divided into two parts, the uplink and the downlink, as shown in figure 7-1. The uplink includes the earth station transmitter and antenna, earth station to satellite transmission path, and the satellite antenna and receiver. The downlink includes the satellite transmitter and antenna, the satellite to earth station transmission path, and the earth station receiving antenna and receiver. In both cases, the same general elements are discussed: transmitter, antenna, path, and receiver. Except for the transmission path characteristics, the characteristics of these elements vary considerably between the uplink and the downlink.

c. If the measure of performance is taken as the quality of the output signal at the receiving terminal of the link, it is possible to show the interrelationship of the parameters which determines the quality of the output signal. This quality is related to the input carrier-to-noise (C/N) ratio.

d. Paragraphs 7-2 through 7-9 discuss the interrelationships of the system parameters and describe each of these parameters in order to fully explain the trade offs necessary to complete a satellite communication link. The mathematical formula describing the involvement of each parameter is called the transmission equation. This basic formula applies to the uplink and the downlink.

7-2. TRANSMISSION EQUATION.

a. The basic transmission equation is developed in this text for the uplink (fig. 7-1). If the earth station transmitter power (P_T) is radiated by an isotropic antenna (fig. 7-7), the power per unit area at a distance R from the transmitting antenna is equal to the transmitter power divided by the surface area ($4\pi R^2$) of an imaginary sphere having a radius of (R). Since the transmitting antennas are

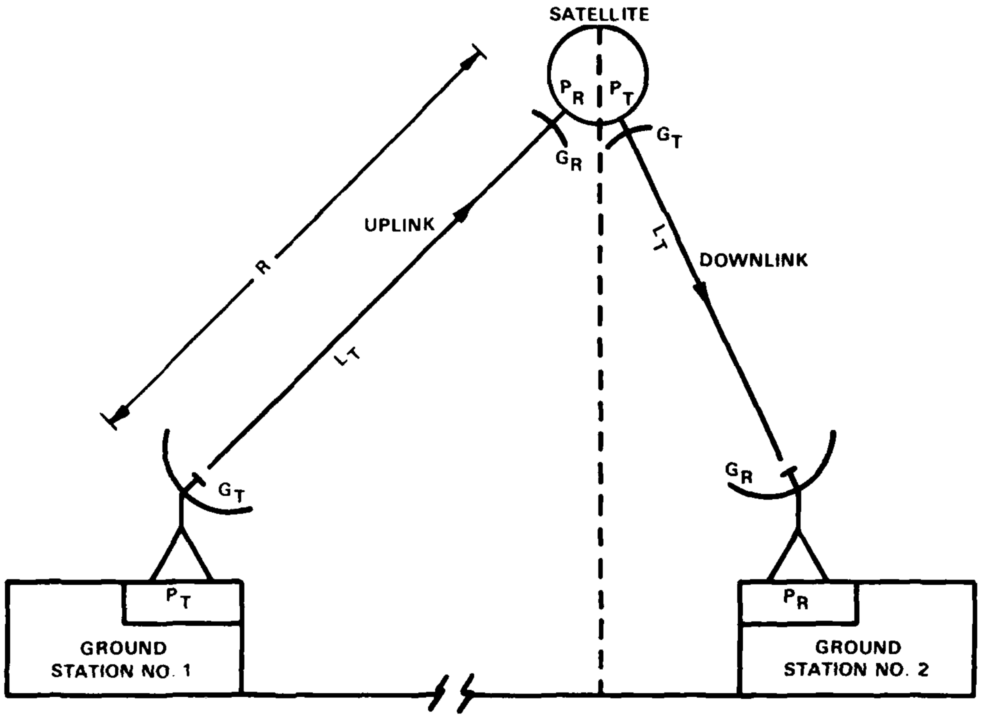


Figure 7-1. Typical satellite communications link functional diagram.

usually directional there will be a gain (G_T) associated with the transmitting antenna. Then the power per unit area or power density (W_L) at a distance R from a directional antenna is given by:

$$W_L = \frac{P_T G_T}{4\pi R^2} \tag{7.1}$$

where

P_T is transmitted power output of ground station

G_T is ground station transmitting antenna gain

The receiving antenna captures a portion of the transmitted power, giving a received power (P_R) based on an effective cross-sectional area of the receiving antenna equal to A_R.

$$P_R = \frac{P_T G_T A_R}{4\pi R^2} \tag{7.2}$$

Antenna theory gives the relationship between antenna gain (G) and the effective cross-sectional area (A) as:

$$G = \frac{4\pi A}{\lambda^2} \tag{7.3}$$

where λ is the transmitted wavelength.

Solving equation (7.3) for A and substituting in equation (7.2) gives:

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2} \tag{7.4}$$

Carrier-to-noise ratio (C/N) refers to the ratio of the energy in the desired signal to the energy in the undesired signal. To determine the C/N power ratio, both sides of equation (7.4) are divided by the effective noise power N, which will be shown to be equal to kT_eB_R.

where

k is Boltzmann's constant = 1.38×10^{-23} joules/kelvin

T_e is effective temperature in kelvin

B_R is receiver bandwidth in Hz

This gives:

$$\frac{C}{N} = \left(\frac{\lambda}{4\pi R}\right)^2 \frac{P_T G_T G_R}{kT_e B_R} \tag{7.5}$$

This is the standard form for the basic transmission equation. Normally a number of losses is associated with delivering the signal power to and from the antenna.

These losses (para 7-5) are usually lumped together as a single term L_R and inserted in equation (7.5). For convenience, the term $(\lambda/4\pi R)^2$ is usually plotted in the form of a nomograph and is called the free space loss (L_S). The free space loss L_S and the losses L_R are lumped together, giving a total fractional term L_T which, for the uplink, results in:

$$\frac{C}{N} = \frac{P_T G_T G_R L_T}{kT_e B_R} = \text{carrier-to-noise power ratio} \qquad (7.6)$$

The terms in equation (7.6), as they apply to the uplink, are summarized:

P_T is transmitted power output of ground station

G_T is ground station transmitting antenna gain expressed as a ratio

G_R is satellite receiving antenna gain expressed as a ratio

L_T is total power losses in the system expressed as a fraction

kT_eB_R is noise power in receiver bandwidth B_R at equivalent noise temperature T_e

b. The various parameters of a satellite communication link have been introduced in a form whereby the effect of each of the factors on the overall performance is easily visualized. To maximize C/N the terms P_T, G_T, G_R, and L_T should be as large as possible, and T_e and B_R should be as small as possible. This is the ideal case. In actual practice, performance specifications generally determine the system parameters which usually dictate the necessary trade offs.

c. It is necessary to understand each of the parameters in more detail before a discussion of the engineering advantages to be gained by these trade offs is possible. Paragraphs 7-3 through 7-6 discuss the major parameters: power limitations (P_T), antenna gains (G_T, G_R), losses (L_T), and noise (T_e, B_R), respectively.

7-3. POWER LIMITATIONS.

a. An active communications satellite is essentially a transponder. It translates the received signal to another (usually lower) frequency and amplifies the signal in an rf power amplifier. The rf power output, available from the power amplifier, is limited primarily by the amount of direct current power available from the prime power source of the satellite. The rf output amplifier of the satellite is the major consumer of satellite prime power.

b. The amount of prime power available is restricted by the size and weight limitations of the satellite. It is for this reason, when analyzing both up and down links, that the rf satellite output power available for the downlink usually limits the capacity of both links.

c. The quantity of prime power available within the satellite is determined by the power-to-weight ratios of state-of-the-art power sources and by the fact that only a limited number of sources are suitable for satellite use.

d. Many power sources have poor power-to-weight ratios or are too short-lived for satellite applications. However, solar cells in combination with storage batteries are suitable for use as prime power sources, although their power-to-weight ratio leaves something to be desired.

e. Solar cell-storage battery combinations provide continuous-duty service in satellites at the present time. They are limited to approximately 1 watt of deliverable power per pound of weight, owing to the relatively low conversion efficiency of the solar cells. Currently, about 10 percent of the energy in the form of sunlight converging on the solar cell is converted to electrical power. In addition to this low efficiency in conversion, solar cells degrade even further when subjected to bombardment by high energy particles such as those encountered in the Van Allen belt. Despite these limitations, the solar cell-storage battery combinations remain the most popular primary power source in satellites.

f. It is not economically feasible at the present time to build and orbit communication satellites weighing more than a few thousand pounds. This limits the total available power that can be obtained from the solar cells to less than a kilowatt.

g. When larger booster rockets become economically feasible, other sources such as nuclear power can be used. Nuclear power sources weighing 1,000 pounds with a capacity of 3 kW and a life of 27 years have already been constructed and operated. Nuclear power sources, which can be scaled upwards in power output with a sizeable reduction in the weight/power ratio, are well within the capabilities predicted for the near future. At present, these sources cannot be deactivated once the reaction has been initiated. This presents no handicap to their use since they have a more than adequate operating life.

h. Since the rf power amplifier in the satellite is the major consumer of prime power, maximum rf amplifier efficiency is required. Operational requirements generally dictate the degree of efficiency that is realized. For instance, certain types of modulation allow higher efficiency amplification than others, owing to the operating characteristics of the rf amplifier. Figure 7-2 illustrates the operating characteristics of a typical rf amplifier. Generally, the higher the degree of linearity

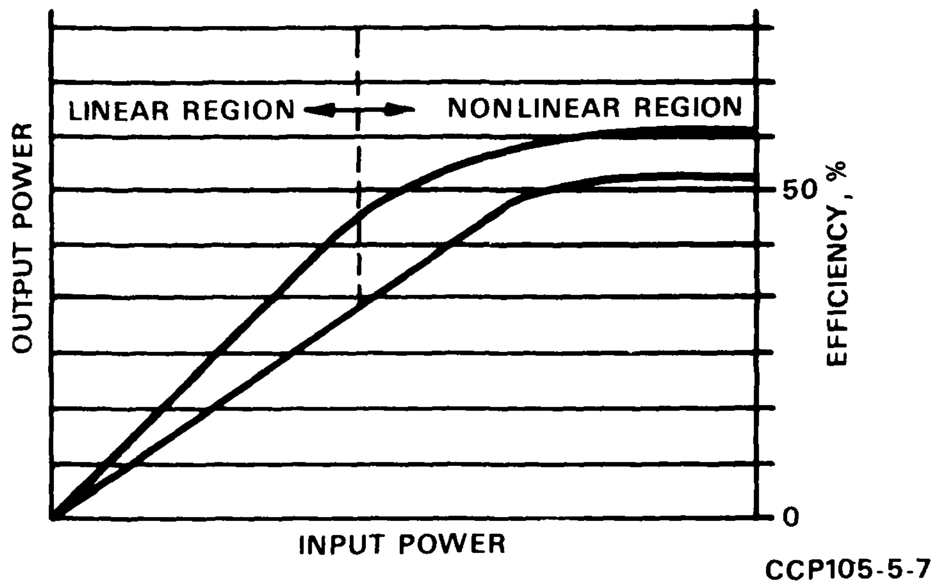


Figure /-2. Operating characteristics curve rf amplifier.

ıequired, the lower the level of efficiency that can be realized. A discussion of the efficiency of modulation is presented in paragraph 7-8.

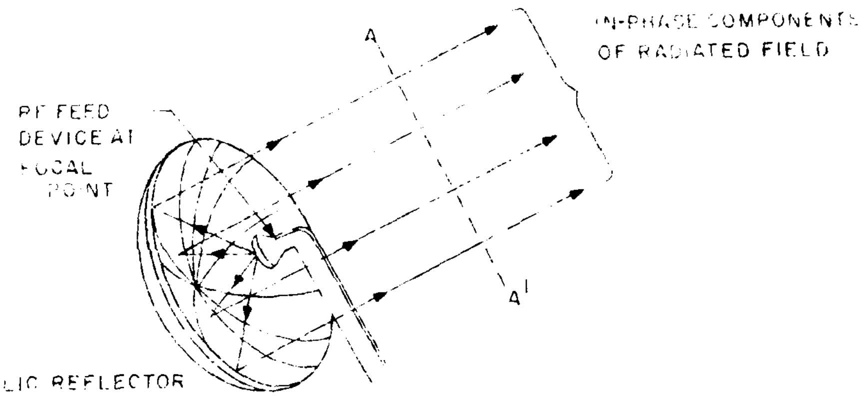
i. Earth station power sources do not present any major problems. Power requirements for an earth station, ranging from 10 to over 100 kW are easily obtained from commercial sources or by the use of portable or transportable power generating equipment.

7-4. **ANTENNAS.**

a. *Types of Antennas.*

(1) All rf power must be carefully conserved if a satellite communications system is to be successful. Antenna design and characteristics are extremely important. Usually some form of a parabolic antenna (fig. 7-3) is used at the earth station. The electric field is radiated from the feed device at the focal point of the parabolic reflector.

(2) As each component of the wavefront (traveling at the same velocity) is reflected from the parabolic reflector, it is shifted 180 degrees in phase. The



CCP105-5-105

Figure 7-3. Prime focus feed antenna functional diagram.

sum of the feed device-to-parabolic reflector, plus the parabolic reflector-to-line A-A1 path, is the same for all components of the wavefront. This provides a beam at line A-A1 with all components in phase.

(a) Figure 7-3 illustrates a prime focus feed parabolic antenna. It is so named because the feed device is at the prime focus point of the parabolic reflector.

(b) The Cassegrainian feed parabolic antenna (fig. 7-4) is usually fed from the vertex of the parabolic reflector and employs a hyperbolic subreflector usually located at the prime focus point of the parabolic reflector. Variations of the Cassegrainian feed system exist which have the feed at the rim of the parabolic reflector, with the subreflector offset in order to compensate for the offset feed. The Cassegrainian feed is preferable to the prime focus feed for the following reasons:

1 The Cassegrainian feed allows the receiver to be placed at the feed point whereas, the prime focus feed requires a long waveguide run with its associated losses.

2. The Cassegrainian antenna provides a better signal-to-noise (S/N) ratio, since spillover from the feed point in the Cassegrainian antenna goes past the subreflector to the sky; whereas, the prime focus feed spillover past the reflector goes to the relatively noisy earth.

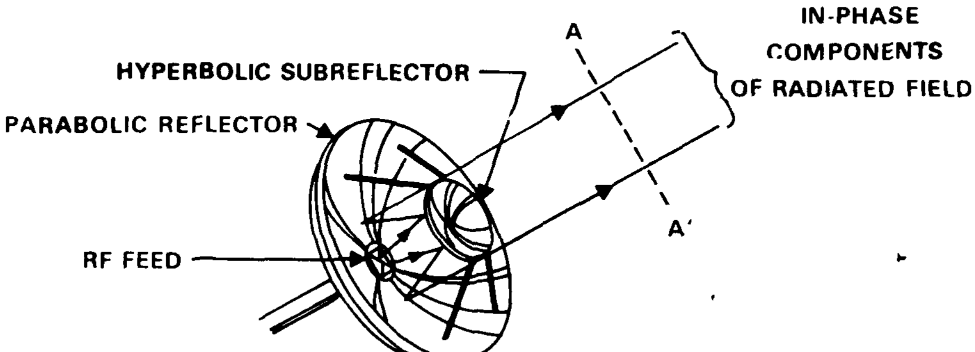


Figure 7-4. Cassegrainian feed antenna functional diagram.

3. Misalignment between the feed and the reflector, which occurs on prime focus feed antennas, is eliminated. The lightweight subreflector on the Cassegrainian feed system (as compared to the bulky, heavyweight waveguide and feed assembly on the prime focus feed) is mechanically more stable.

4. The main disadvantage of the Cassegrainian feed system is the aperture blockage by the subreflector and its support assembly. Also, critical alignment between the reflector and subreflector is required in order to prevent phase errors. However, these problems can be virtually eliminated in the design stage.

(c) The horn reflector antenna (fig. 7-5) uses a parabolic section as a reflector, with the horn offset from the reflector section. The horn feed is extended to meet the reflector. The COMSAT Corporation earth station at Andover, Maine, used a horn antenna; it proved to be highly efficient and had very little spillover from the reflector. However, it was over twice as large as a parabolic antenna of comparable gain (177 feet long) and weighed approximately 380 tons.

(d) In an attempt to reduce the size and weight of the horn reflector antenna, the Cassegrainian principle was employed in the casshorn antenna configuration illustrated in figure 7-6. The casshorn antenna is fed through the parabolic reflector, with the hyperbolic subreflector providing a virtual source similar to the full-size horn reflector.

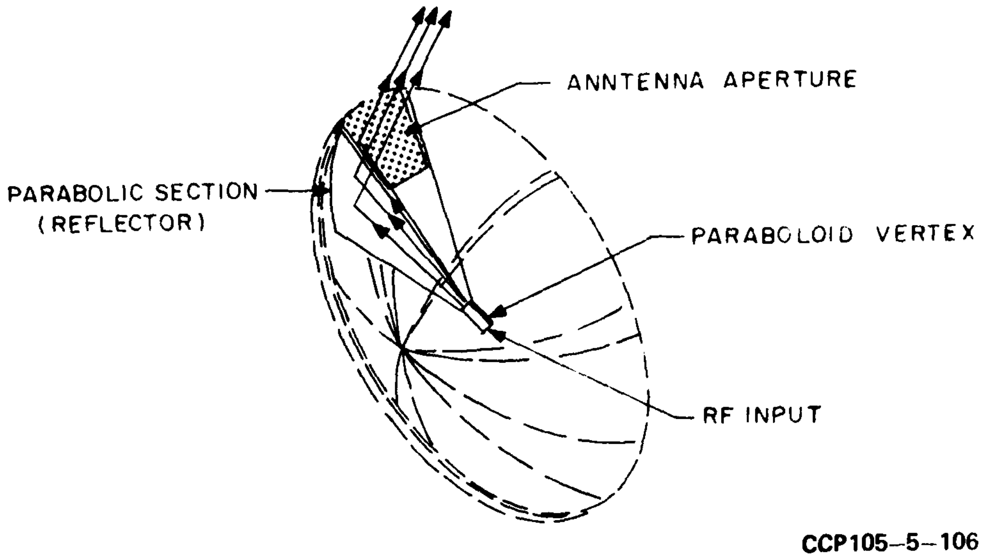


Figure 7-5. Horn reflector antenna functional diagram.

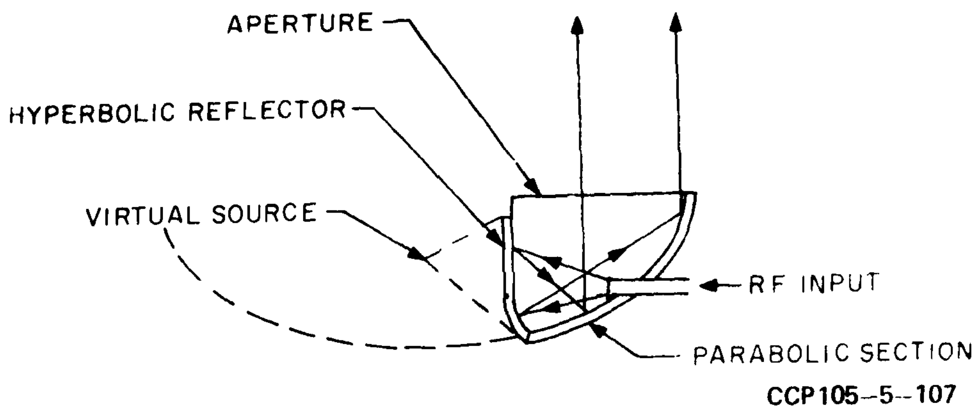


Figure 7-6. Casshorn antenna functional diagram.

b. *Isotropic Radiation Pattern.*

(1) When discussing antenna gain, think of an antenna that would radiate equally in all directions from a central point forming a sphere of radiation. Such a radiator is termed isotropic or is said to have an isotropic radiation pattern. Although such an antenna does not exist in practice, it is useful to consider as a standard or point of departure.

(2) As stated previously in this chapter, if P_T represents the available transmitter power, then at some distance R (in meters) from the antenna, the power density (W_L) in watts per square meter is given exactly by the formula:

$$W_L = \frac{P_T}{4\pi R^2} \qquad (7.7)$$

where

$4\pi R^2$ is the surface area of a sphere of radius R

Therefore, by definition, the power is radiated evenly over the entire surface. Figure 7-7 represents the field strength of an isotropic antenna.

c. *Parabolic Antenna Theory.*

(1) Antennas can be designed to radiate much more in one direction than in others. The radiation pattern for such an antenna is represented in figure 7-8. The ratio of the power density at the peak of the pattern W_P to what the power density would have been for an isotropic antenna W_L is called gain. The numerical value of an antenna gain can be expressed as a ratio of the powers or it can be expressed in dB by taking ten times the common logarithm of the power ratio. Figure 7-8 also shows that the power density W decreases rapidly as the direction angle θ increases. As this angle from the direction of maximum radiation intensity continues to increase, the emissivity experiences a series of maxima and minima. These secondary maxima are called the side lobes of the beam pattern. The intensity of the side lobes is expressed as a given number of dB down from the intensity of the peak of the main lobe. Good antenna designs have been achieved which have the side lobes 25 dB or more down (a power ratio of 320:1) from the main lobe.

(2) The angular width θ_B (beamwidth) of the main beam is the cross-sectional angle of a hypothetical conical beam, the edges of which are arbitrarily chosen to be that of the half-power points if power plots were made of the rf energy

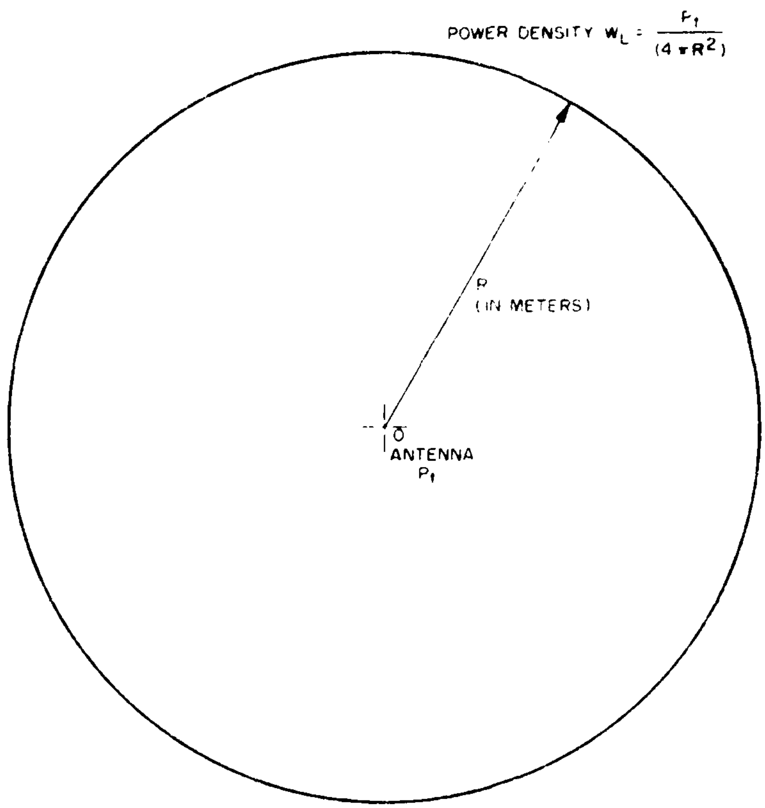


Figure 7-7. Isotropic antenna pattern.

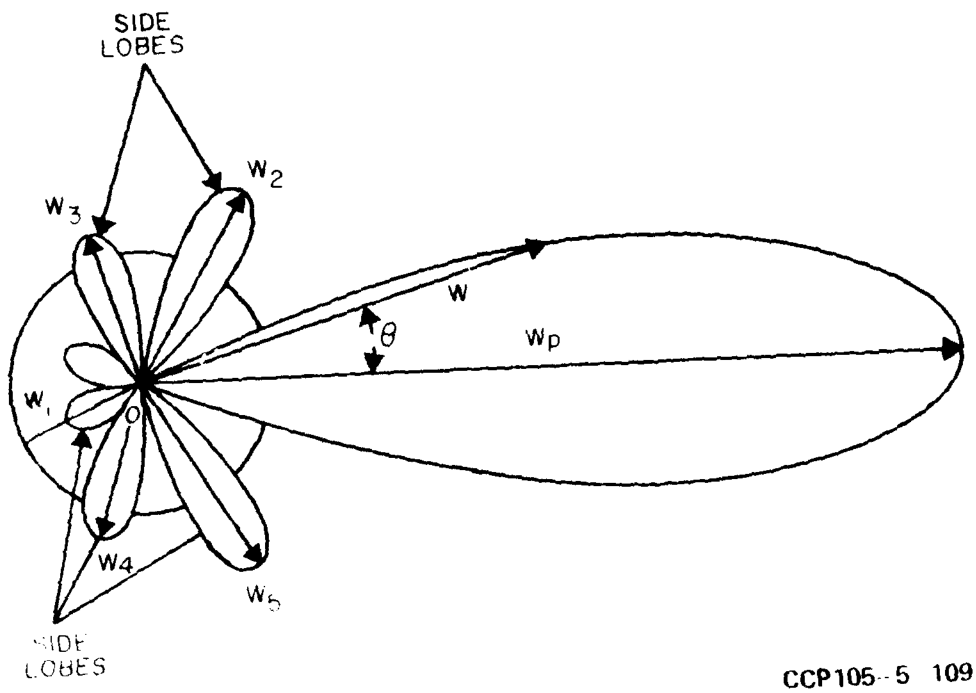


Figure 7-8. Directional antenna pattern.

in the beam pattern and a line drawn through the half-power points, the picture would be as shown in figure 7-9.

(3) Some very precise mathematical relationships exist between beamwidth, gain, wavelength, and antenna area in antenna design, as stated in equation (7.4).

where

A is antenna area (aperture area) in square meters

θ_B is half-power beamwidth in radians

G is gain (power ratio)

λ is wavelength, in meters $= \frac{c}{f}$

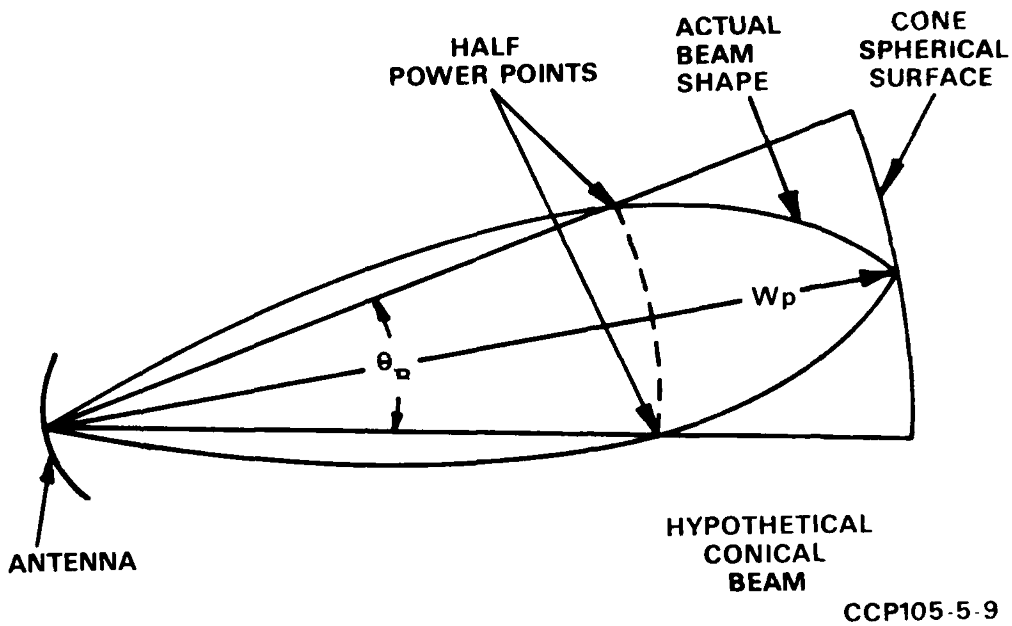


Figure 7-9. Antenna beamwidth pattern.

f is frequency, in Hz

c is speed of light (3 X 10^8 meters/sec)

d is parabola diameter in feet

Then for parabolic antennas common to microwave applications the following relationship exists:

$$G = \frac{4\pi A}{\lambda^2} = \frac{4\pi A f^2}{c^2} \cong \frac{4\pi}{\theta_B{}^2} \tag{7.8}$$

Typically, most parabolic antennas have an efficiency of 55 percent. This accounts for such factors as aperture blockage, antenna misalignment, and imperfections in the parabolic surface. The above gain formula then becomes:

$$G = \frac{4\pi}{\theta_B{}^2}\eta = \frac{(2.2)\pi}{\theta_B{}^2} \qquad (7.9)$$

η = antenna efficiency = 55 percent

Or in terms of decibels:

G (dB) = 20 log d + 20 log f − 172.5

The beamwidth can also be expressed as a function of frequency and antenna size:

$$\theta_B{}^2 = \frac{\lambda^2}{A} \text{ or } = \frac{(68{,}700)\,10^6}{fd} \qquad (7.10)$$

(4) The gain in figure 7-10 is shown in dB while the beam angle is shown in radians. The graph reveals that antenna gain increases rapidly with both an increase in diameter (or aperture area A) and frequency f. Also, gain (G) increases as beamwidth θ_B decreases

(5) Atmospheric disturbances, during bad weather, prevent beam angles from being less than 1 milliradian; hence, antenna gains of more than 72 dB for all weather use cannot be achieved. This is shown by the horizontal atmospheric limit line in figure 7-10. For a gain of 60 dB, the illustration shows that the dish size would approach 100 feet in diameter at S-band (3 GHz) frequencies. For the same gain, the size becomes proportionately larger as the frequency decreases. Conversely, for constant gain the dish size becomes smaller as the frequency increases.

(6) High gain antennas with their narrow beam angles must be accurately pointed. Ground station antennas must also maintain this pointing accuracy while tracking the satellite. Large antennas are prone to deform in shape to some extent as their elevation angle is changed. These deformations in shape, in turn, change the shape of the radiation pattern, reduce the antenna gain, and deflect the center of the beam away from its intended direction. Deformations of the antenna surface are allowable to only a small fraction of a wavelength, so that, although antennas can be made smaller to reduce the amount of deformation, as higher frequencies are used the allowable deformation also decreases. Practically speaking, high gain antennas used for tracking satellites in the 1 to 9 GHz frequency range all exhibit gains of 50 to 60 dB.

(7) The gain figures discussed are only valid for the far field (many wavelengths away from the antenna). Gain figures change with distance and are

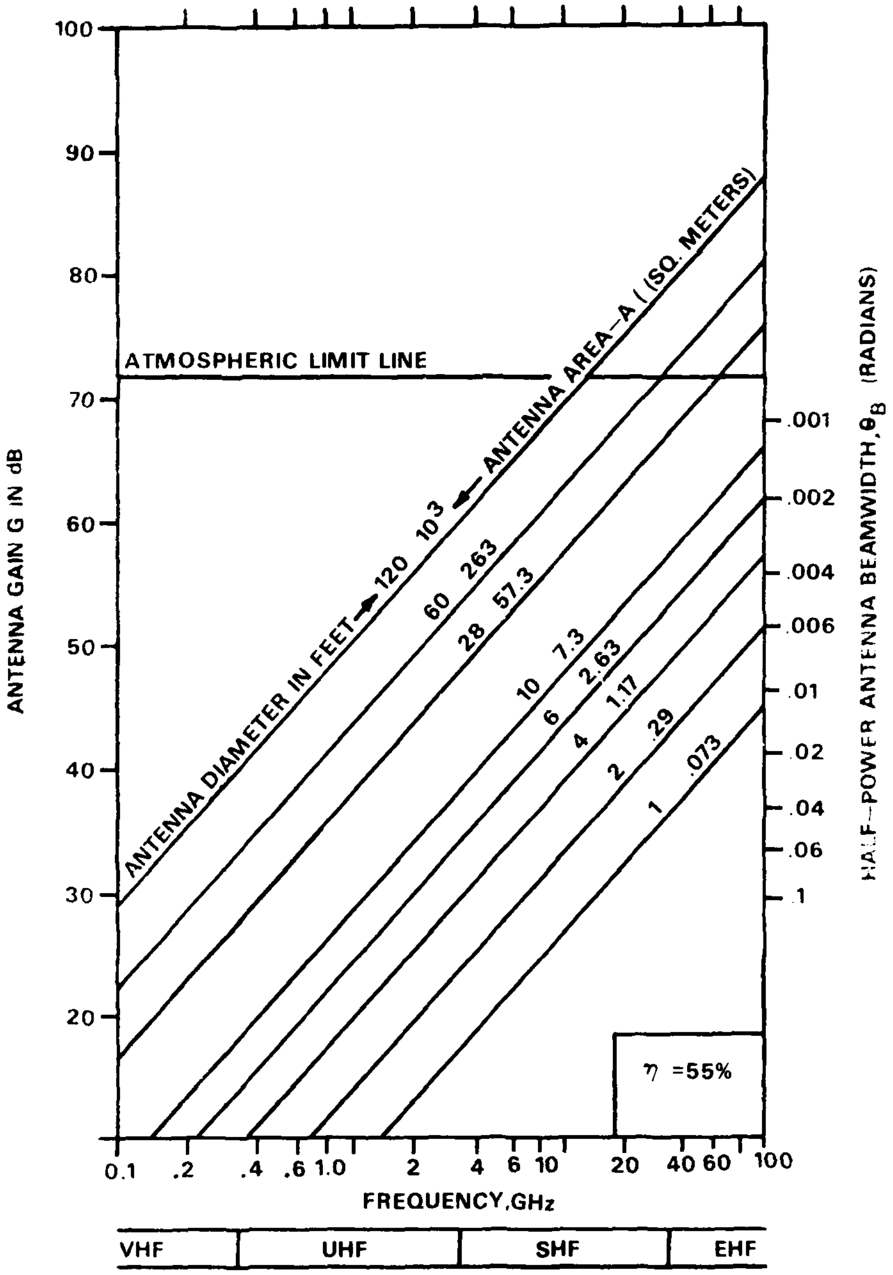


Figure 7-10. Antenna gain versus wavelength graph.

considerably different in the near field. Formulas for computing the distance beyond which the far field gain formulas apply have been developed but are of no practical value to this discussion.

(8) In the preceding material, antenna gain was described as a means of effectively increasing transmitter power (over isotropic) by the power gain ratio G. Directive receiving antennas also increase the power delivered to a receiver over that which would have been received by an isotropic receiving antenna by the same gain ratio G. Due to this reciprocity, a good transmitting antenna also makes a good receiving antenna.

(9) The size and weight necessary for high gain antennas restricts their use for the present to the earth station ends of a satellite communication link. Ground stations may use separate high gain antennas for the uplink transmitter and the downlink receiver. Normally, a duplexer is used with the antenna so that the antenna can be used for simultaneous transmitting and receiving.

(10) Because of the weight involved, principally in the complicated despinning or stabilizing mechanisms, high gain antennas for use on satellites had to await improvements in booster technology. For a simple satellite, the antenna is designed to be as nearly isotropic as possible so that no pointing mechanism need be carried in the satellite.

(11) As discussed in chapter 2, if the satellite is gyroscopically spin-stabilized on the proper axis, the antenna can be designed to produce a doughnut-shaped pattern, without requiring additional pointing mechanisms. The doughnut-shaped radiation patterns could theoretically achieve a gain of 10 dB, but normally a gain of 3 to 6 dB is obtainable. Using an antenna pattern that compensates for satellite spin is called electrically despinning.

(12) It is possible to provide oriented antennas for synchronous satellites with a beamwidth sufficiently large to illuminate the entire face of the earth. Such antennas have a gain as high as 20 dB. At present, most synchronous satellites use mechanically despun antennas. The antenna is rotated by a motor in the opposite direction of the spinning satellite body. The net result is that the antenna is stationary relative to the earth. Three-axis stabilization methods are under development. Using gravity gradient stabilization, or some other method, the satellite would be stationary in all three axes with respect to the earth. The antenna would then be mounted directly on the satellite.

7-5. LOSS FACTORS.

a. Major factors in rf considerations, as applied to satellite communications, are those of propagation and transmission loss.

b. With P_{TC} defined as the rf carrier power output delivered to the transmission line leading to the duplexer (either in the earth terminal or aboard the satellite) and P_{RC} defined as the carrier power delivered to the receiver amplifier input at either the earth terminal or satellite receiver, the ratio TC/P_{RC} can be called the transmission loss (L_T) of a single communications space link.

$$L_T = \frac{P_{TC}}{P_{RC}} \qquad L_T(\text{dB}) = 10 \log_{10} \left(\frac{P_{TC}}{P_{RC}}\right) \tag{7.11}$$

It is convenient to divide L_T into the sum of five losses for further study, thus:

$$L_T = L_S + L_a + L_r + L_t + L_{exr} \tag{7.12}$$

where

L_S is an isotropic space loss (dB)

L_t is the loss between the output of the transmitter power amplifier and the antenna (dB)

L_a is an absorption loss due to normal (clear) weather (dB)

L_{exr} is excess atmospheric absorption loss during periods of rain (dB)

L_r is a loss similar to L_t between the receiver antenna input and the receiver amplifier (dB)

(1) *Isotropic free space loss.*

(a) For any free space link, the ratio of the power radiated by the transmitter antenna P_T to the power absorbed by the receiving antenna P_R (if both antennas are assumed to be isotropic and in free space) is called the isotropic free space loss (L_S).

$$L_S = \frac{P_T}{P_R}$$

$$L_S(\text{dB}) = 10 \log_{10} \left(\frac{P_T}{P_R}\right) \tag{7.13}$$

(b) This isotropic free space loss is related quantitatively to the distance between the transmitter and receiver and to the frequency, or wavelength, of the signal. The free space loss formula (para 7-2) has been used to prepare the graph shown in figure 7-11. It can be seen from this graph that for a distance R of 10^4 km, which is typical of a ground-to-satellite link, and a frequency of 3 GHz (shf band), the isotropic free space loss is 190 dB. The isotropic loss L_s increases as the square of both distance and frequency. Expressed as a formula:

$$L_S = \left(\frac{4\pi Dr}{\lambda}\right)^2$$

$$L_S(\mathrm{dB}) = +79.3 + 20 \log f + 20 \log Dr \quad (7\text{-}14)$$

where

f is frequency in Hz

Dr is path length in km

(2) *Atmospheric and rain loss.*

(a) The atmosphere has a selective absorption of radiation in the microwave millimeter portions of the spectrum between 10^9 and 10^{11} Hz, as diagramed in figure 7-12. The right-hand curves represent absorption by oxygen and water vapor which occurs under even the best of conditions; i.e., clear weather. The left-hand curves represent excess rain loss in the atmosphere when rain or fog formations are dominant factors. The curves of figure 7-12 have been theoretically derived and experimentally verified.

(b) The amount of absorption due to water vapor and excess rain loss (L_{ex}) varies considerably depending on the weather. Both excess rain loss and clear weather atmospheric loss (L_a) are dependent to some extent on antenna elevation pointing angle. This is because, at low elevation angles, the radio signal traverses a longer path through the earth's atmosphere than at high elevation angles. At an antenna elevation angle of 5 degrees, the path length through the atmosphere is approximately 50 km in length. The constituents of atmospheric absorption and how they vary with frequency in the case of a 50-km path length are shown in figure 7-12. Figure 7-13 shows the total one-way clear weather absorption as a function of frequency for different antenna elevation angles.

(c) Since excess rain loss becomes a factor only during periods of rain or fog, it is expressed as a percentage of time that it exceeds a certain value. A curve plotted to show this type of relationship is known as a distribution curve. Figure 7-14 is a typical distribution curve of excess rainfall attenuation for a relatively dry climate such as San Diego, California.

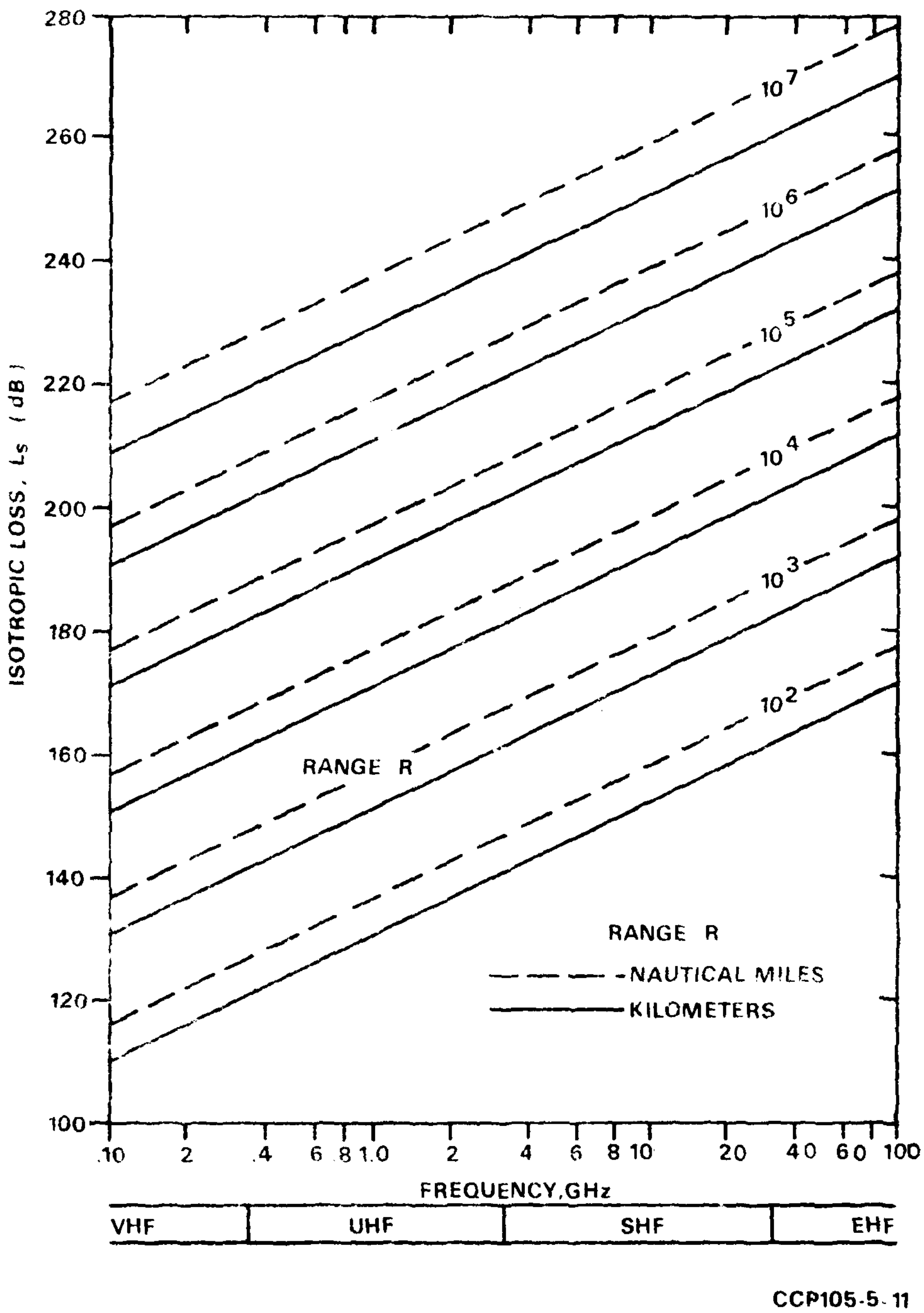


Figure 7-11. Isotropic free space loss graph.

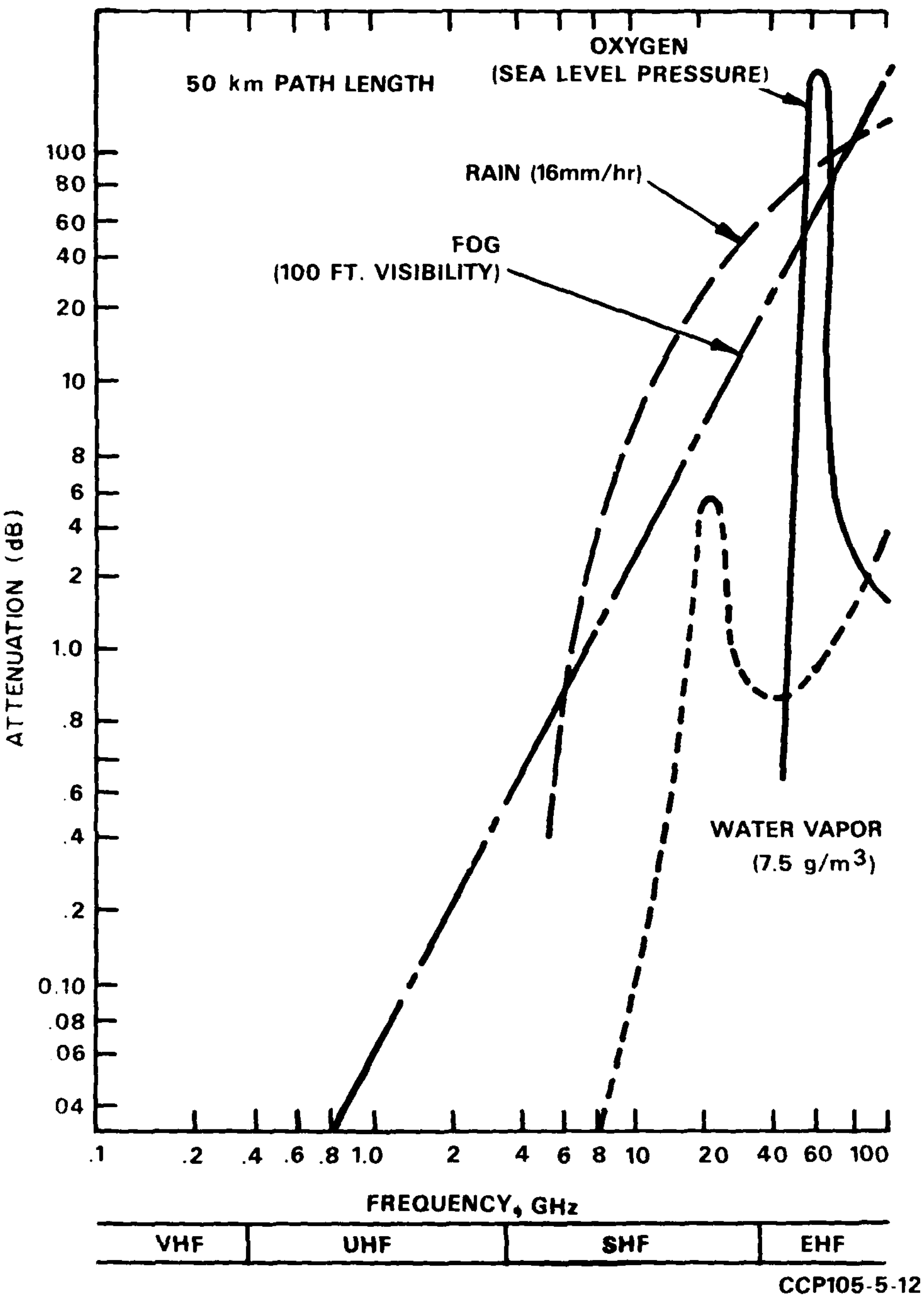


Figure 7-12. Atmospheric absorption graph.

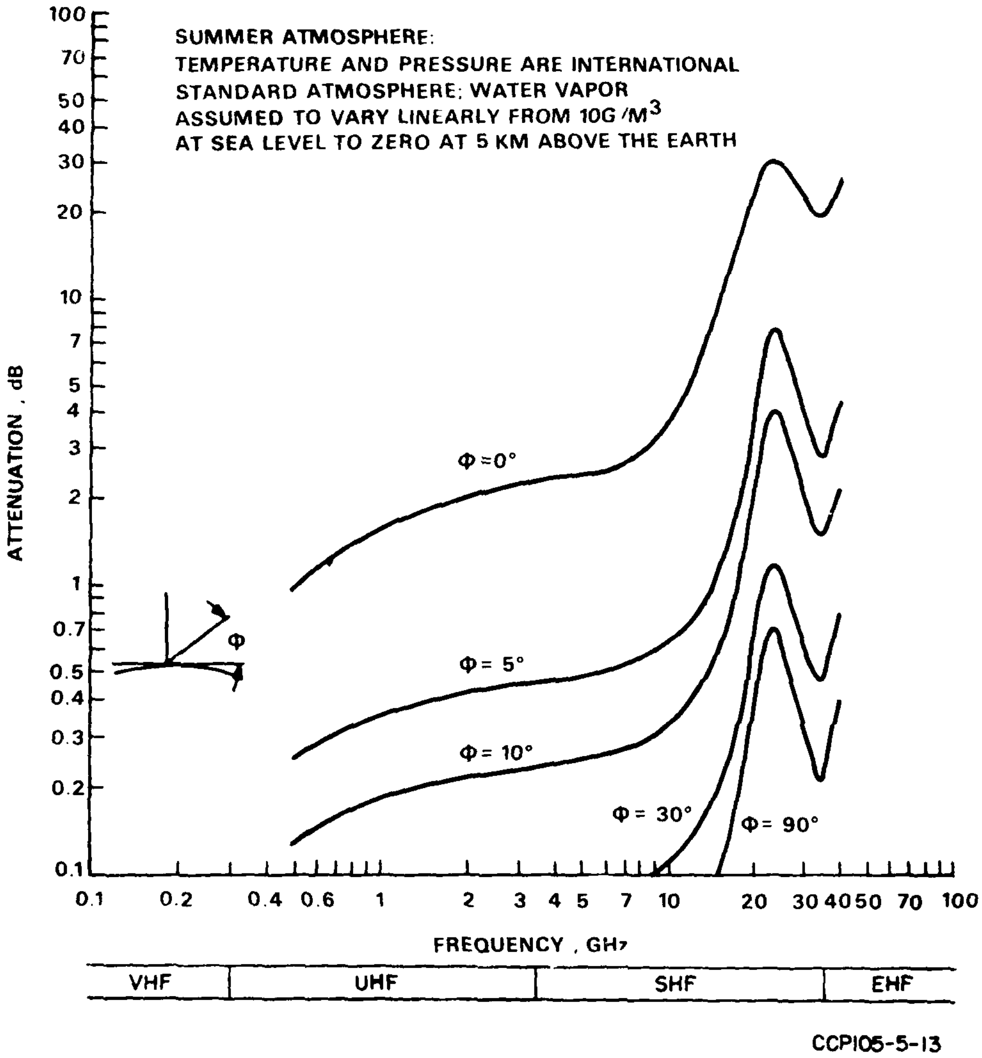


Figure 7-13. Clear weather atmospheric absorption graph.

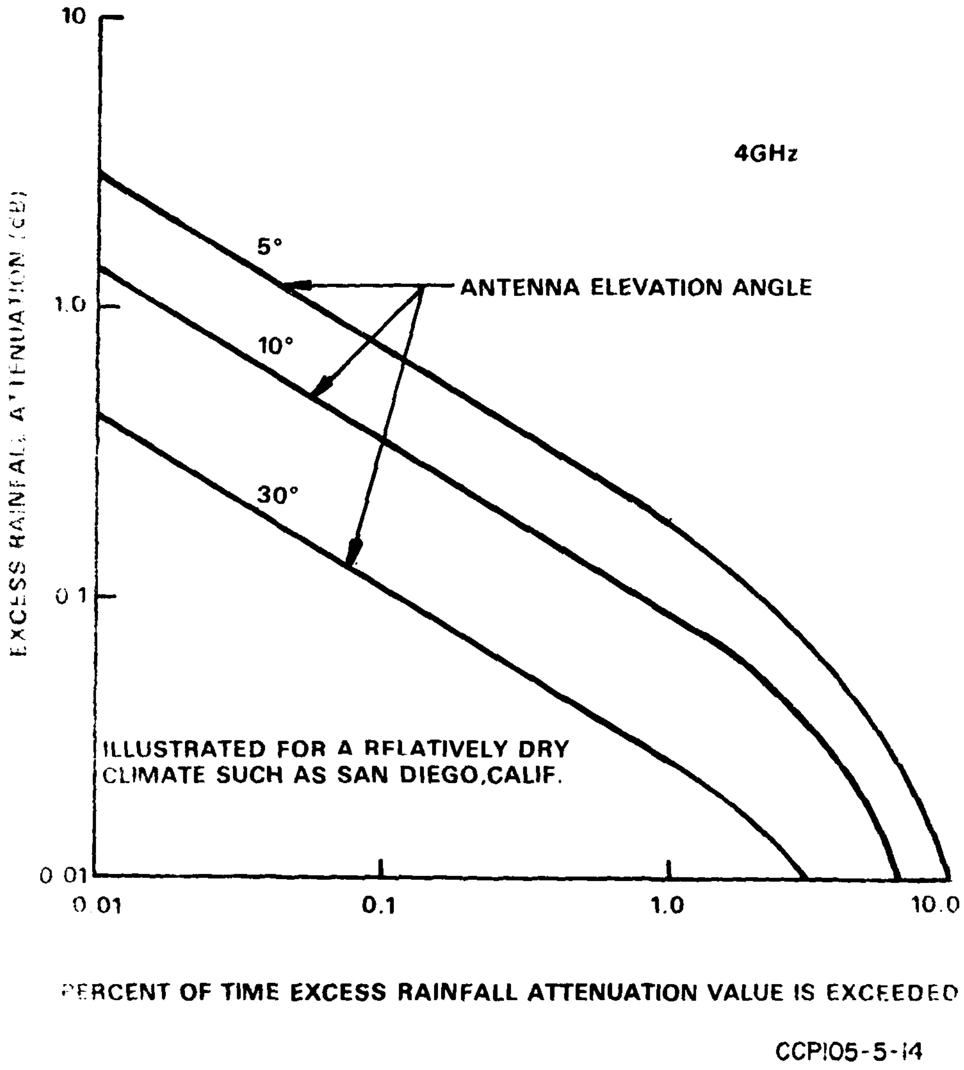


Figure 7-14. Graph of distribution of worst month excess rainfall attenuation.

(d) From figures 7-12 and 7-13 it can be seen that in ground-to-satellite links, frequencies above 10 GHz experience high atmospheric absorption.

(3) *Transmitter antenna line loss.*

(a) It is convenient to specify the rf power of a transmitter P_T as that delivered by the power amplifier to the transmission line leading to the duplexer. The rf power that is actually radiated is less than P_T. P_T is multiplied by a fractional transmitter loss (L_t) to give power radiated. This loss, composed of losses due to the duplexer, to network coupling, and to factors concerning the antenna beam, is practically unavoidable although every effort is made to minimize it as much as possible when designing the equipment.

(b) The loss in the network stems from the duplexer itself, from the transmission lines connecting the duplexer to the power amplifier, and to the antenna. The potential transmission line loss can be considerably reduced by the choice of high quality, low-loss transmission line material and arrangement of the antenna-duplexer-power amplifier locations to require the minimum length of line. The duplexer loss depends on its particular design, but can be reduced by using a greater separation between the transmitted and received frequencies. The same isolation is required but more freedom in design is allowed so that, generally, a lower loss duplexer (filter-isolator) results. Well designed duplexers exhibit a loss of the order of 1 dB, depending on rated power. It is more difficult to build a low-loss high-power duplexer than a low-power unit.

(c) The antenna beam loss is a composite group of loss factors which, when divided into the antenna gain G, reduces the theoretical gain to an effective operating antenna gain. This antenna beam loss group is composed of an illumination loss and a pointing error loss. The illumination loss occurs with beams formed by concave antenna dishes, which are illuminated by a point or a subreflector, in the case of a Cassegrainian feed system, at the dish's focal point. Due to practical considerations in design, not all of the energy from the feed point is intercepted by the dish surface. That which is not intercepted is lost. This illumination loss factor is usually on the order of 2 to 3 dB. The antenna gain derived from equation (7.8) applies only when the peak of the transmitted main beam lobe is in the direction of the receiving antenna. Inaccuracies in pointing the transmitting antenna can cause a pointing error (tracking error) loss ranging from a fraction of a dB to several dB.

(4) *Receiver antenna loss.*

(a) As in the case of the transmitter power ratios, the ratio of the single channel rf power intercepted by the receiver antenna (P_R) to the power delivered to the input of the receiver amplifier for that channel P_{RC} is the receiver antenna loss (L_r). This loss is also compounded by increments of network loss and antenna loss.

(b) Both the network and composite pointing error losses are similar in magnitude and occur for reasons similar to those described for the transmitter losses.

(5) *Polarization loss.*

(a) Electromagnetic radiation oscillates in a plane at right angles to its direction of propagation. By way of illustration, consider a rope tied to a post while the other end is moved up and down vertically. Vertical waves will be created which will travel down the rope from the source of vibration toward the fixed terminal. The waves in the rope will occur in only one plane and, since in this case the plane is vertical, the waves are termed vertically polarized waves. In a similar illustration, if the rope is vibrated from side to side and no gravity is present, the waves in the rope will occur in a horizontal plane and will be horizontally polarized. Polarized radiation is generated in a manner similar to this illustration. A receiving antenna and its transmission line coupling network can be designed to accept, with only negligible loss, radiation which is polarized in a plane and reject any radiation which is polarized in a plane at right angles to it. If the rope passed through a narrow slotted picket fence, horizontal oscillations (analogous to horizontally polarized waves) would be stopped abruptly at the slot. Varying degrees of hindrance would be observed if the angle of polarization were rotated from 0 (vertical) through 90 degrees (horizontal). Thus, a plane-polarized receiver antenna is similar to the picket fence and when aligned to the polarization plane of the transmitter, it can deliver the radiated power intercepted by the antenna to the antenna transmission line. As the angular misalignment (short of 90 degrees) between the transmitter and receiver antenna increases, the polarization loss will become greater. When the receiver polarization plane is at right angles to the transmitter polarization plane (a condition called cross polarization) the loss, theoretically, is infinite. In practice, cross polarization loss is not infinite, but it does approach the order of 30 dB or a factor of 1,000 to 1.

(b) A more sophisticated scheme of polarization is called circular polarization. This is a wave whose plane of polarization rotates through 360 degrees as it progresses forward. The rotation can be clockwise or counterclockwise. Figure 7-15 illustrates this effect, depicting the waves as traveling away from a viewer. Circular polarization is created by combining equal magnitudes of vertical and horizontal plane polarized waves, with a phase difference of 90 degrees. Depending on the phase relationship, this creates rotation either in one direction or the other. If the phase-time relation between the horizontal or vertical components is not exactly

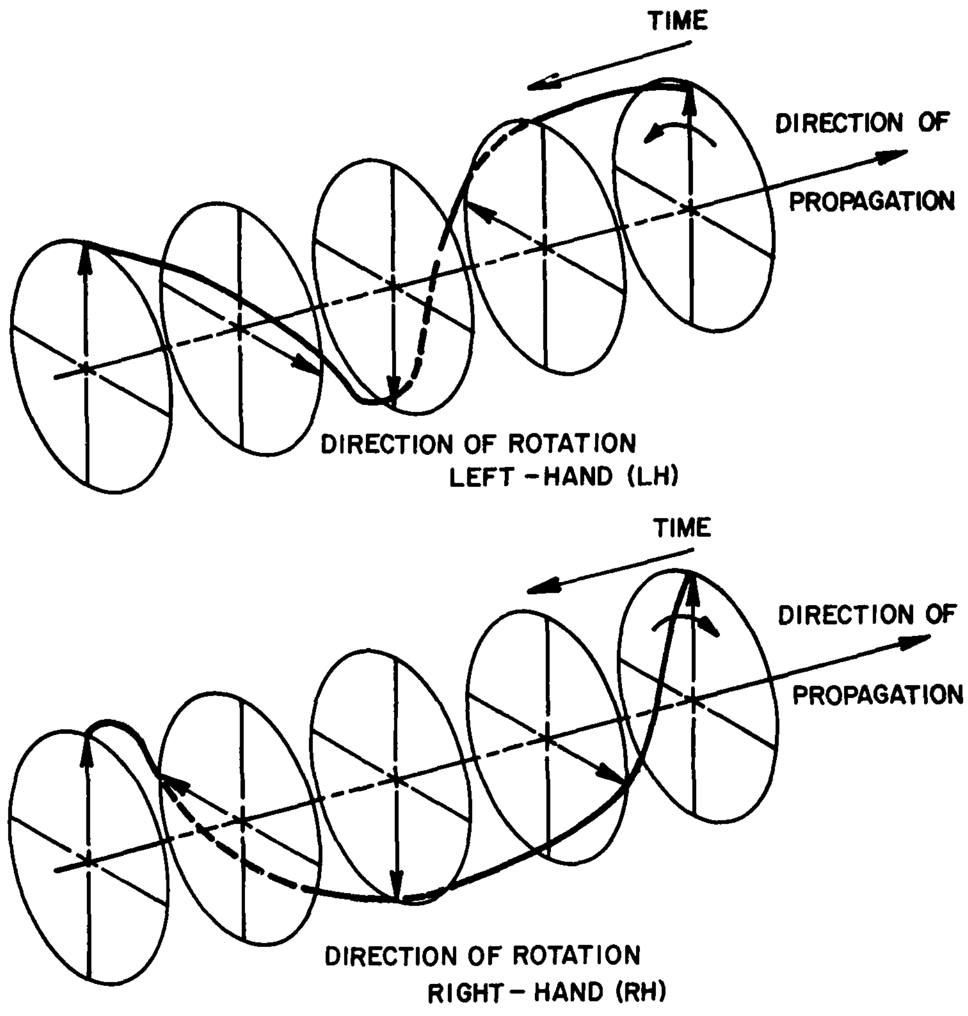


Figure 7-15. Representation of rotation of circular polarized waves.

90 degrees or if the magnitudes are not equal, a right-hand or left-hand elliptical polarization results, as illustrated in figure 7-16. The magnitude of the radiated wave varies in a cyclical manner.

(c) A receiving antenna designed for circular polarization in one rotational sense (right-hand, for example) will pass all intercepted right-hand circularly polarized radiations to the transmission line. Conversely, an antenna designed for one rotational sense will exhibit a high rejection of radiation polarized in the opposite rotational sense. Theoretically, this loss is infinite, but again in practice, it is more on the order of 30 dB.

(d) There are two outstanding reasons for using circular polarization rather than plane polarization in ground-to-satellite and satellite-to-ground communications. Ground station receivers can reject circularly polarized radiation, emitted from their own transmitters and back-scattered from reflecting layers in the atmosphere (rain, fog, and clouds) far more effectively than if the reflected radiation is plane polarized. Additionally, the use of circular polarization makes the use of a polarization alignment control unnecessary. For example, suppose a linearly polarized wave were rotated 90 degrees as the result of rotation of the satellite's transmitting antenna or because of atmospheric phenomena. A linearly polarized receiving antenna would then have to be rotated simultaneously through 90 degrees to avoid a large loss (as much as 30 dB) of signal. With circular polarization there would be, in theory, no loss of signal strength as the result of the 90 degrees rotation of the transmitted wave. In practice, there might be as much as a 3 dB loss resulting from a 90 degrees rotation of the wave with circular polarization. On the other hand, a circular polarized antenna is more complicated than a plane polarized antenna and the polarization will always have some ellipticity in practice.

(e) Even when employing circular polarized antennas, some polarization loss occurs. This is due to the fact that actual antennas are never exactly circularly polarized and polarization is affected by the ionosphere. These factors give rise to slightly elliptical rather than pure circular polarization. When the elliptical polarization seen by the receiving antenna is not exactly matched by the antenna polarization, a polarization loss occurs. This polarization loss is typically on the order of 1 or 2 dB.

7-6. NOISE.

a. One of the major factors determining the capacity of a satellite relay link is the ratio of the strength of the received signal to the noise present within the receiving system. As has been discussed, the received power is determined by the transmitted power, the transmitting and receiving antennas, and the associated system and path losses. The noise within the receiver originates from two principal

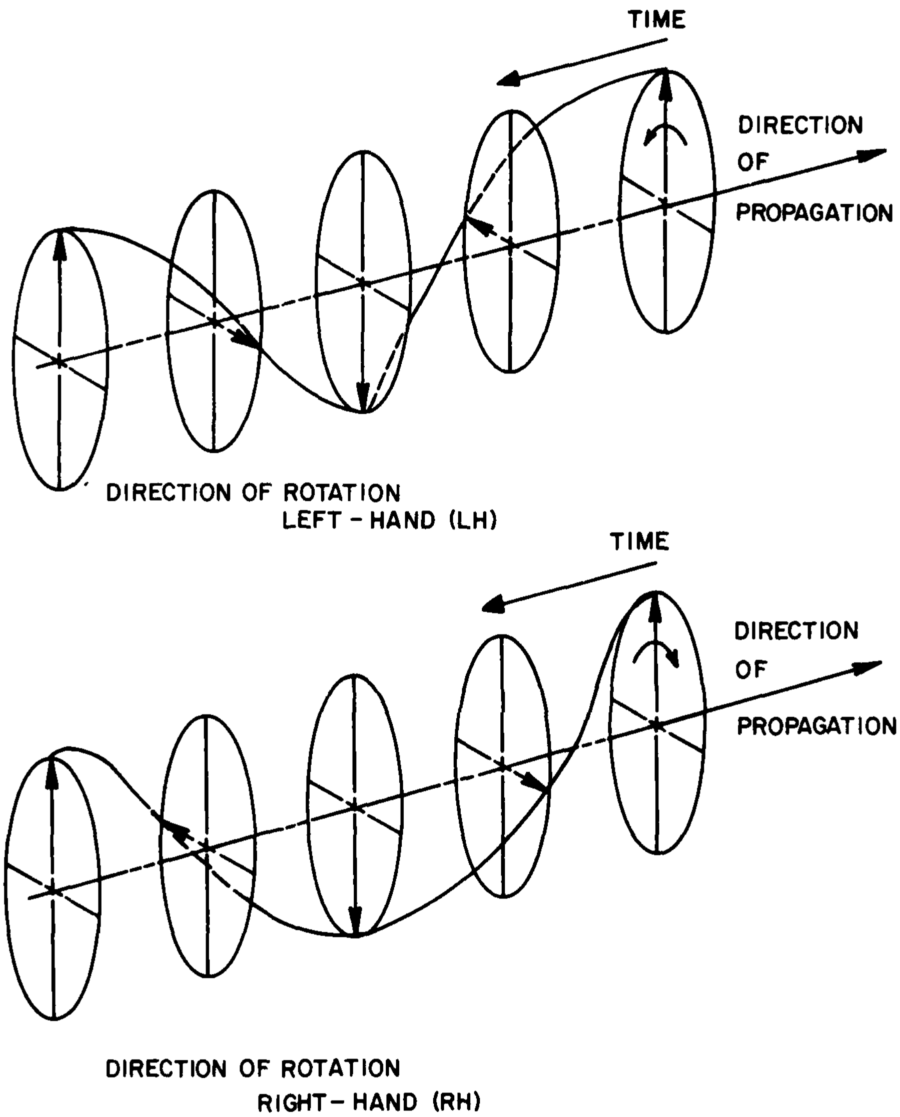


Figure 7-16. Representation of rotation of elliptical polarized waves.

sources. The internal noise associated with the electron flow in the amplifying devices and the external noise contributed by various sources.

b. Until quite recently, tube noise contribution at microwave frequencies was so large that it completely masked the external noise. Typical noise temperatures for amplifiers ranged from 600 K at 200 MHz to 3,000 K at 10 GHz. Two relatively new developments have completely altered this situation. They are parametric amplifiers (paramp) and masers (chapter 8). These devices can have such low noise temperatures that noise from the galaxies is now often the limiting factor in determining receiver performance.

c. Before developing the subject further, an introduction of two concepts, noise temperature and noise figure, will greatly simplify the presentation.

(1) *Noise temperature.*

(a) The earth terminal receiving system noise temperature is a measure of the quality of the receiving system (the lower the temperature, the higher the quality). The receiving system noise temperature indicates, more than anything else, how strong a received signal the earth terminal requires for satisfactory operation or, on the other hand, how many channels of communication a given earth terminal can provide when receiving a signal of a given strength.

(b) If a resistor is at a specific temperature, a specific noise power will appear across the terminals of the resistor. The noise power is due to thermal agitation; that is, motion caused by heating of the electrons in the structure of the resistor. If the resistor is heated to a higher temperature, the noise power increases; if the temperature is lowered, the power decreases. A useful measure of these powers is a quantity proportional to voltage squared. The rule that this noise voltage or power increases with temperature can be expressed in a more precise way if the noise is measured as a noise power and the temperature is measured on an absolute scale; that is, one whose reference is 0 K (-273 °C). Mean square noise voltage is defined as:

$$V^2 = 4RkTB \qquad (7.15)$$

where

V^2 is mean square voltage of the noise

R is equivalent resistance or impedance in ohms of the circuit under measurement

k is 1.38×10^{-23} joules/kelvin (Boltzman's constant)

T is temperature in kelvin of the resistor (or a receiver, as normally applied)

B is bandwidth of the device under consideration (the receiver or the instrument making the measurements) in Hz

(c) To find the maximum noise power (*N*) that the resistor will develop (for a given T and B), replace the resistor by an ideal resistor which develops no noise and a voltage generator of V's mean magnitude, as shown in figure 7-17, and terminate the theoretical equivalent circuit in a matched load $R_{load} = R_{ideal}$ (impedance match).

Then:

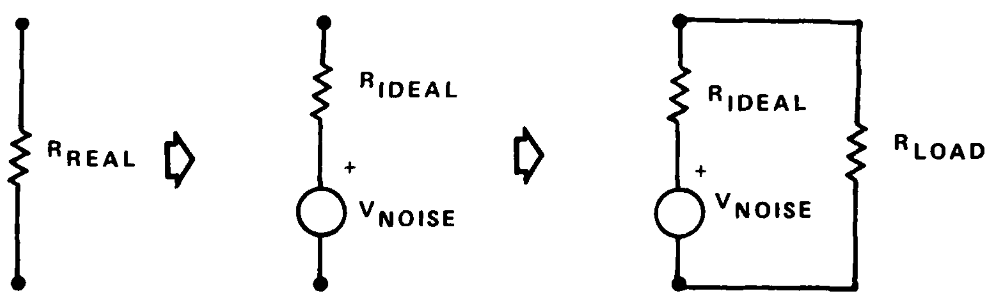
$$I = \frac{V}{2R} \text{ ; therefore, } (N) = I^2 R = \frac{V^2}{4R}$$

where

N is noise power

Since:

$$V^2 = 4RkTB \text{ therefore } N = kTB \qquad (7.16)$$



CCP 105-5--178

Figure 7-17. Resistive noise source schematic diagram.

(d) This indicates that the noise power (N) is directly proportional to the absolute temperature (T). Thus, a doubling of absolute temperature results in a doubling of noise power. This leads to the concept of noise temperature. Since a given resistor R generates a given amount of noise for a given temperature, it is possible to refer to that amount of power by an equivalent noise temperature. The normal reaction to this might be: "Why bother with noise temperature when a perfectly good term, power, is available?" The answer to this becomes apparent if more detailed measurements are made on the resistor. The measuring system, which includes amplifiers, has some bandwidth. If this measuring system is used to measure a signal of a fixed bandwidth that is less than the measuring system bandwidth, then a further increase in the bandwidth of the measuring system would not change the measured power. Such is not the case with the noisy resistor. Doubling the bandwidth of the system doubles the measured power; halving the bandwidth halves the power. This means that the power available from the resistor depends on bandwidth as well as temperature.

(e) It is more convenient to use dBm as a unit of comparison. When compared to a 1-Hz bandpass at 1 K, N, results in a constant of −198.6 dBm. This indicates that for the reference receiver considered (1-Hz bandpass, 1 K temperature), any increase in bandwidth or temperature will cause an increase in noise contributed by the receiver. Therefore, noise contributed (in dBm) for the actual temperature and bandwidth is expressed as:

$$N \text{ (dBm)} = -198.6 + 10 \log T_{actual} + 10 \log B_{actual} \qquad (7.17)$$

(f) For example, if the temperature of the resistor or receiver is 300 K and the power measurement configuration or receiver has a bandwidth of 10 kHz, the noise power is:

$$N = -198.6 + 10 \log 300 + 10 \log 10^4$$

$$= -198.6 + 24.7 + 40$$

$$= -133.9 \text{ dBm}$$

(g) If the receiver requires a C/N ratio of 10 dB for fm quieting to occur (i.e., to be above threshold) then the carrier level must be greater than -123.9 dBm.

(h) In a calculation of this type, the bandwidth generally used is that of the last IF strip and not of the paramp or first several IF strips, because it is from that last bandwidth that the detector or demodulator receives the noise and carrier.

(2) *Y-factor measurements of system noise temperature.*

(a) A common technique for measuring noise temperature is by the use of the Y-factor. This employs a noise source such as an argon lamp of which the noise temperature has been established both for cold (turned off) and hot (turned on) conditions. The Y-factor measurements are made using the figure 7-18 test configuration.

(b) With the noise source off, the equivalent cold temperature at the receiver input (T_{ce}) is composed of the cold temperature of the source (T_c) modified by the isolation and filter unit. With the noise source turned on, the equivalent hot temperature at the receiver input (T_{he}) is composed of the hot temperature of the source (T_h) modified by the isolation and filter unit.

(c) To determine the Y-factor a reference is set on the power meter with the noise source turned off and the attenuator set at zero dB. The noise source is turned on and the attenuator is adjusted so that the power meter is set on the reference point again. The attenuator setting is the Y-factor in dB.

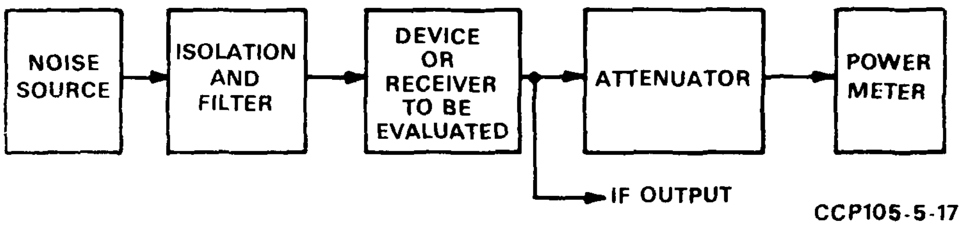


Figure 7-18. Y-factor noise measurement test configuration block diagram.

(d) The receiver noise temperature (T_r) can be calculated from the following equation since Y, T_{ce}, and T_{he} are known:

$$Y = \frac{T_{he} + T_r}{T_{ce} + T_r} \qquad (7.18)$$

(e) Note that Y is measured in dB but must be converted to a power ratio for insertion into the equation. Also, the noise contribution from the isolation and filter unit must be included in the calculation. A practical example for use of the Y-factor equation is given below:

Assume:

T_h = 9000 K, T_c = 270 K, $T_{ambient}$ = 290 K, isolation and filter loss (L) at $T_{ambient}$ = 1 dB, Y = 12.2 dB (a power ratio of 16.6:1)

Then:

$$T_{he} = \frac{T_h}{L} + \frac{(L-1)}{(L)} \text{ at } T_{ambient}$$

$$T_{he} = \frac{9000\ K}{1.26} + \frac{(1.26-1)}{(1.26)}\ 290\ K = 7140\ K + 60\ K = 7200\ K$$

$$T_{ce} = \frac{T_c}{L} + \frac{(L-1)}{(L)} \text{ at } T_{ambient} = 274\ K$$

Solving equation 7.18 for T_r:

$$T_r = \frac{T_{he} - YT_{ce}}{Y-1} = \frac{7200 - 16.6(274)}{15.6} = 170\ K$$

and using $T_r = (F-1)\ 290\ K$, the noise figure (F) is obtained:

$F = 1.586$ (power ratio)

$F = 2.00$ dB

(3) *System noise temperature.*

(a) The receiving system noise temperature is the equivalent noise temperature of the system as applied to the input of the paramp. It is composed of the equivalent noise temperature of the total of all components preceding and including the paramp, noise contributed by the waveguide, and the noise temperature of the antenna. The system noise temperature can be computed using the following procedure.

(b) Determine the receiver noise temperature (including the paramp) using the Y-factor technique or other system of noise temperature measurement. Label the result T_r for receiver noise temperature.

(c) Connect the input of the paramp to a 50-ohm load (cold source) and obtain a reference level in the receiver's IF. Then connect the input of the paramp to the antenna (hot source) and, as before for the measurement of T_r, obtain the Y-factor for this situation. Label this second Y-factor Y_2 and in a power ratio form Y_{2pr}. To keep things straight, label the first Y-factor Y_1 or Y_{1pr}. At this point the effective noise temperature of the antenna and feed combination T_e can be computed by using:

$$T_e = \frac{290 + T_r(1 - Y_{2pr})}{Y_{2pr}} \tag{7.19}$$

where ambient room temperature is 290 K

(d) The antenna noise temperature is extracted as T_a (antenna) $= L_{fpr}T_e - 290\ (L_{fpr}-1)$ where L_{fpr} is the antenna feed line loss in a power ratio form.

(e) Finally, the system noise temperature T_s is found using:

$$T_s \text{ (system)} = T_a + (L_{fpr} - 1)290 + L_{fpr}T_r \tag{7.20}$$

(4) *Noise figure.*

(a) Some criterion is needed to rate receivers and receiving systems. The noise figure provides such a criterion as far as the noise performance is concerned. The noise figure, however, does not completely specify a receiver performance, since it says nothing about gain, bandwidth, or distortion, all of which must be satisfactory as well.

(b) The concept of noise figure has gone through many stages of development, and many slightly different types of noise figures (e.g., spot noise figure and average noise figure) have been defined. This section treats only one type; average noise figure (the noise figure normally used in measuring system performance). The noise figure to be considered is a single number characterizing the receiver. It is, in a sense, an average noise figure over the passband of the system. Separate noise figures (spot noise figures) could be quoted at each frequency within the band, much as different gains can be quoted at various frequencies for a simple amplifier. Just as it is common to refer to an amplifier as a 20 dB amplifier, meaning that its maximum gain is 20 dB, so it is also common to quote a single noise figure that will be of most interest as a criterion for rating system performance. More specifically, it will be the average standard noise figure.

(c) The average standard noise figure gives a measure of the amount of noise that the amplifier itself contributes to its output. The noise figure is defined as the total noise power output of the amplifier divided by the output noise contributed by a matched source impedance at a temperature of 290 K at the amplifier input (290 K is approximately room temperature). The noise figure can be written as:

$$F = \frac{N_{out}}{N_o G} \tag{7.21}$$

(d) The noise figure concept is shown in figure 7-19. A matched source at the standard reference temperature of 290 K drives an amplifier with gain G. The noise power N_o generated by the matched source is multiplied by the gain of the amplifier G and appears in the matched load as a noise power $N_o G$. If the amplifier were a perfect amplifier and contributed no noise of its own, this would be the total noise output of the amplifier. However, any practical amplifier will contribute some noise and this is designated N_A. The total noise output of an amplifier is the sum of $N_o G$ and N_A. Since the output noise power N_{out} is just $N_A + N_o G$, an alternate expression for noise figure is:

$$F = \frac{N_o G + N_A}{N_o G} = 1 + \frac{N_A}{N_o G} \tag{7.22}$$

where

F is average standard noise figure

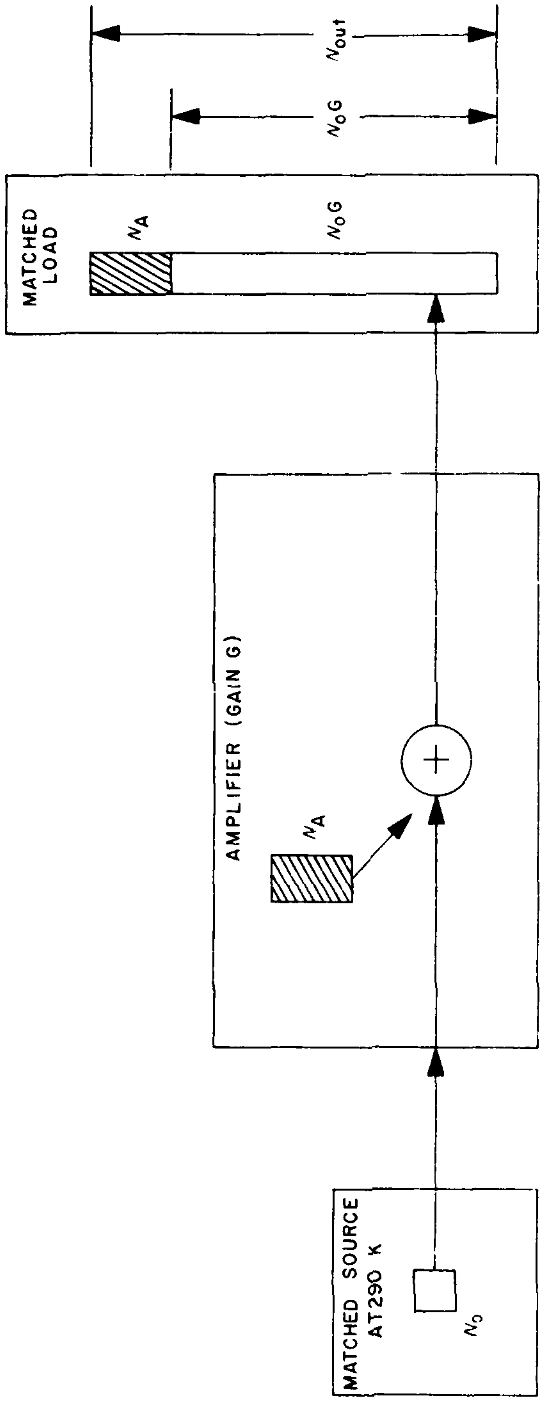
N_A is output noise power contributed by the amplifier

N_{out} is total output power

$N_o G$ is output noise power due to source noise

(e) The noise figure of a receiver may be measured with the setup in figure 7-19. This technique is not used in practice to measure the noise figure since it requires an accurate measurement of the gain of the amplifier and other quantities. Other techniques are available which do not require an accurate measurement of amplifier gain.

(f) Noise figure (F) may be quoted as a number, a ratio, or in dB as is common with power ratios. Thus, a noise figure of 2 and a noise figure of 3 dB are equivalent. Both indicate that the amount of noise added by the amplifier is equal to the noise caused by a 290 K matched source.



CCP105-5-110

$$F = \frac{N_{out}}{N_o G} = \frac{N_o G + N_A}{N_o G} = 1 + \frac{N_A}{N_o G}$$

Figure 7-19. Noise figure concept block diagram

(g) If an amplifier added no noise to the output it would be considered ideal. Its noise figure would be 1 or 0 dB. This is the minimum noise figure possible. Strictly speaking, in accordance with the definition, a negative or fractional noise figure is not possible. When dealing with modern very low noise amplifiers, such as cooled parametrics and masers, this fact has often led to confusion. The confusion results because the definition of noise figure assumes that every receiver will have the noise input of a matched source impedance at room temperature (290 K). Using equation (7.17), this amounts to a noise power input of 290 KB watts. Therefore, by this definition even a perfectly noiseless amplifier would have an output noise of 290 GKB watts. When analyzing a radio link for the purpose of predicting its S/N ratio performance, it has in the past been very convenient to estimate the noise contributed by a receiver as being 290 KB multiplied by the noise figure of the receiver, or F (290 KB) watts. So long as the noise contributed by a receiver amplifier was large in comparison to 290 KB and so long as the receiver was actually operated at something near a room temperature of 290 K, this gave a perfectly good engineering result. With modern low-noise cooled receiver amplifiers, this is no longer the case. A good cooled microwave receiver may actually have a noise output under operating conditions that is less than the 290 GKB watts of the ideal receiver having a noise figure of 1 (0 dB).

(h) If one were to use noise figure as a basis for comparison of such a receiver with a conventional receiver, it would be necessary to assign the low noise receiver a negative noise figure based on how many times better it is than the supposedly 0 dB ideal receiver. For very low noise figures, and in cases where it is necessary to add the performance of additional items to the receiver performance, it is more convenient to express noise performance in terms of an equivalent receiver noise temperature. This receiver noise temperature is related to N_A, where an ideal amplifier with no noise contribution is assumed. The question is then asked: "How much noise contribution would have to be made at the input to give the observed noise at the output?" The answer is $N_A/G + N_o$. The N_o part of the noise is normally due to the source and the N_A/G is that part which will give the same result as the noisy amplifier. If we express this N_A/G part as an equivalent noise temperature, we obtain the receiver noise temperature noise figure relation:

$$T_e = (F - 1)\ 290\ K \tag{7.23}$$

where

T_e is equivalent receiver noise temperature (K)

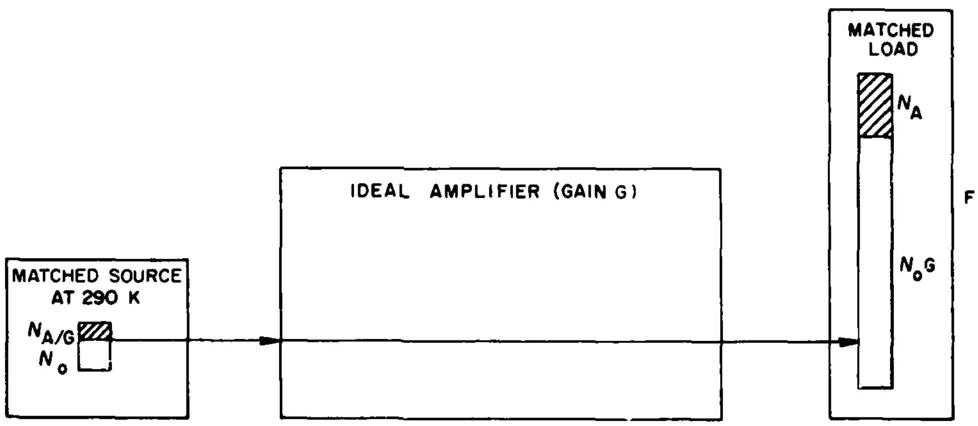
F is noise figure (a ratio, not in dB)

290 K is temperature factor for the matched source

(i) The relationship between noise figure and receiver noise temperature is shown in figure 7-20. If an amplifier is known to have a certain noise figure, a ratio for amplifier noise and source noise in the output is specified. This is depicted in figure 7-20 for the same case shown in figure 7-19. Instead of the noisy amplifier with gain G and noise contribution N_A, an ideal amplifier with gain G is assumed. Noise power N_A/G is added to the source noise N_o to give the same N_A at the load. Thus, the noise temperature of a receiver represents the amount of noise that would be needed at the input of an ideal amplifier to give the same performance as the actual amplifier under discussion. It could be calculated from $T_e = (F-1)$ 290 K if G, B, and N_A are all known.

$$N = kTB \text{ or } \frac{N_A}{G} = kTB \tag{7.24}$$

The advantage of measuring noise figure directly is that conversion can be made to noise temperature without knowing G, B, and N_A separately.



$$F = 1 + \frac{N_A}{N_o G} = 1 + \frac{(N_A/G)}{N_o}$$

$$T_e = (F-1)\,290K$$

CCP105-5-111

Figure 7-20. Noise figure - ideal amplifier block diagram.

(j) The noise temperature T_e of the best types of present-day receivers ranges between 10 K for the frequency range 1 to 100 GHz as presented in figure 7-21. The maser has the lowest noise temperature T_e (about 10 K), but because it is relatively large in size, heavy in weight, and is complicated (requires liquid helium cooling), it is presently limited to ground station receivers. The tunnel diode amplifier, crystal mixer, front end amplifier, transistor amplifier, and paramp are solid state amplifiers featuring small size, lightweight, and low power consumption characteristics, as needed for satellite receivers. TWT designs, though larger, heavier, and not having as low noise as tunnel diode and paramp, are often used in satellite receivers because they have wider bandwidth and greater immunity to radiation damage. However, tunnel diodes and paramps are also used.

7-7. SIGNAL-TO-NOISE RATIO, CARRIER-TO-NOISE RATIO.

a. As described in paragraph 7-6, noise in a satellite communication system is made up largely of thermal noise, sky noise, excess rain noise, and satellite intermodulation noise.

b. Although, ultimately, the S/*N* ratio of the output signal from the earth station receiving system is of prime importance, the C/*N* at several intermediate points between the transmitting earth station and the receiving earth station should be considered. The power ratios at the following points are of particular interest:

(1) At the input to the satellite transponder (C/*N*).

(2) At the input to the earth station receiving system (C/*N*).

(3) At the output of the earth station receiving system, as measured in the individual voice or data channels (S/*N*, C + *N*/*N*).

c. This paragraph describes the various factors that must be taken into consideration in computing S/*N* at the points of interest.

(1) *Effective radiated power.*

(a) Effective isotropic radiated power (EIRP) is a term that has been found convenient to use in describing the radiated power from the satellite. EIRP is the product of the actual rf transmitter power output and the antenna gain. For instance, a satellite transponder with a 10-watt final amplifier and an antenna with a gain of 10 would have an EIRP of 100 watts or 20 dBW. The EIRP of a satellite can be calculated providing that certain parameters (readily available at the earth station) are obtained. These parameters are antenna elevation, slant range to the satellite, weather, ground antenna net gain, system noise temperature, C/*N* (or

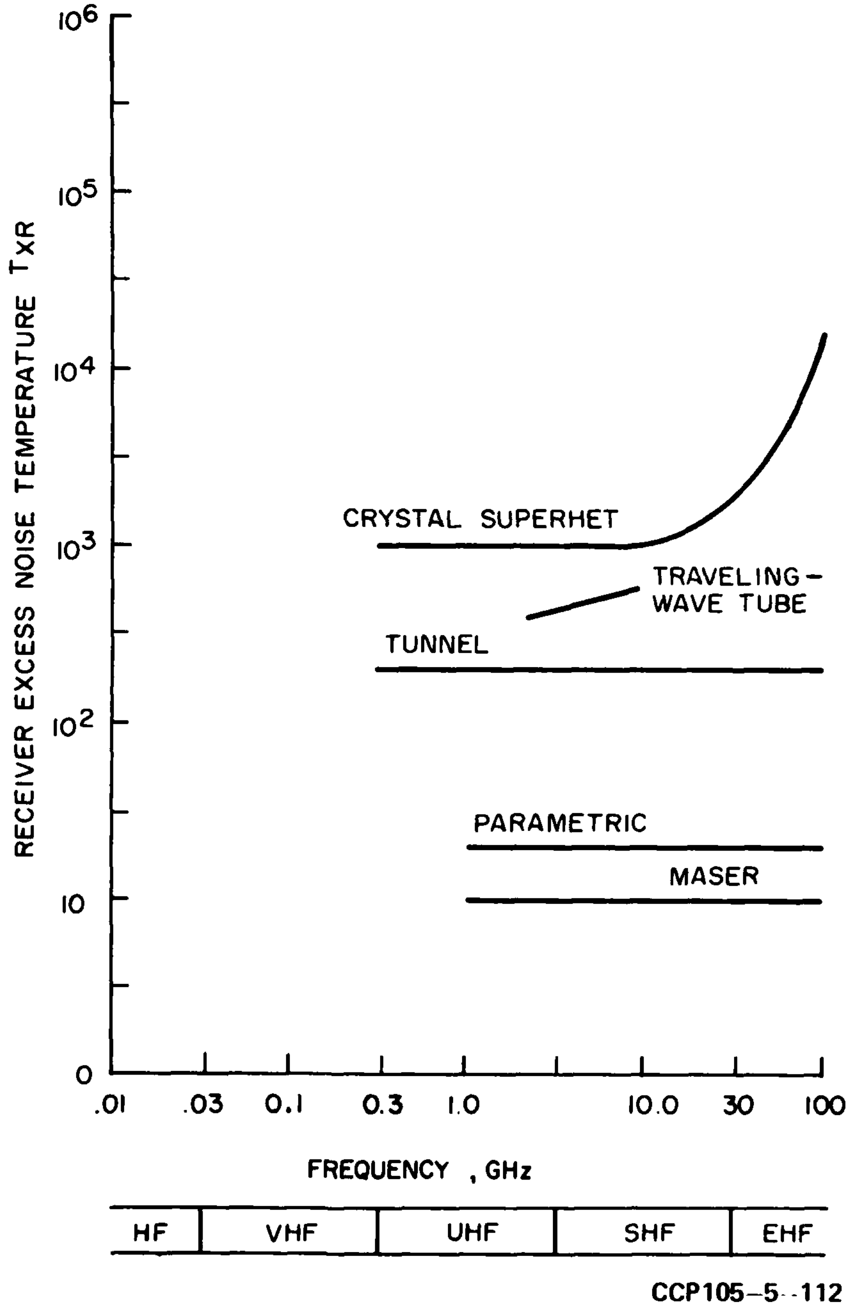


Figure 7-21. Receiver excess noise temperature graph.

C/*k*TB) of the receiver, and bandwidth of the receiver or measuring device (whichever is narrower). The received signal strength (S_r) at the ground station (referred to the paramp input) is given by:

$$S_r = EIRP - L_{fs} - L_a + G_{ga(net)} \quad (7.25)$$

where

S_r is received signal strength (referred to paramp input)

L_{fs} is free space loss

L_a is atmospheric attenuation

$G_{ga(net)}$ is antenna net gain - to include radome loss (when applicable), tracking (crossover) loss, polarization (ellipticity) loss, and feed loss (including waveguide lines) to paramp input.

Based on the above it can be seen that by transposing equation (7.25), EIRP is expressed by:

$$EIRP = S_r + L_{fs} + L_a - G_{ga(net)} \quad (7.26)$$

(b) Once the major parameters are stated, it is simply a matter of addition and subtraction to determine EIRP.

(2) *Multiple access backoff.*

(a) Multiple access backoff is a term that has come into use in connection with FDMA. The techniques of FDMA are described in chapter 5.

(b) Multiple access backoff refers to the satellite output power that is lost due to the necessity for backing off on the earth station radiated power to avoid generating excessively high intermodulation products in the satellite. Multiple access backoff will normally amount to between 1 and 2 dB.

(3) *Noise bandwidth.*

(a) A figure for *C/N* is meaningless unless the noise bandwidth associated with the *C/N* is known. The noise bandwidth of interest varies, depending on which *C/N* is being determined.

(b) The only noise of concern in the earth terminal receiving output signal is the noise that falls within the output voice or data channel. In the case of a voice channel, this noise bandwidth will normally be 4 kHz.

(c) At the input to the earth station receiving system we are concerned with the C/N in the IF bandwidth. The C/N in this bandwidth determines whether or not the signal is strong enough to be above receiver threshold. Knowing the C/N in the IF bandwidth, the C/N in the output channel can be calculated.

(d) At the input to the satellite receiver there are two noise bandwidths of interest; the C/N in the entire transponder passband, and C/N in the rf signal bandwidth. The ratio of the signal to the noise power in the entire transponder passband largely determines how the output power of the satellite will divide between the desired signal and the noise signal. Also, the satellite power lost to intermodulation products must be considered. The C/N in the rf signal bandwidth determines how much noise power will be contained in the output signal that will ultimately fall within the IF passband of the earth station receiving system.

(4) *Carrier-to-noise density.*

(a) In the C/N density expression (C/kT), C refers to the ratio of the rf signal level, or carrier level, of the received signal at the earth terminal receiving system. Boltzmann's constant is k and T is the receiving system noise temperature. From equation (7.15), kT is the noise power in a bandwidth of 1 Hz; hence the term, noise density.

(b) The significance of this expression comes from the fact that formulas for computing the capacity of satellite communication links show that this factor is basic to determining the channel capacity of the system. It suffices to say, however, that once the arbitrary factors such as desired channel C/N and modulation index have been determined, then channel capacity can be determined from C/kT. Figure 7-22 shows how the number of voice channels varies as a function of C/kT under set conditions.

(c) Because C/kT is fundamental to so many link performance calculations, manufacturers of satellite receiving equipment generally describe the performance of their equipment based on the C/kT of the received signal.

(5) *Antenna gain-to-noise temperature.*

(a) The antenna gain-to-noise temperature (G/T) ratio may be thought of as a figure of merit for an earth receiving station. G refers to the gain of the earth station antenna in the receive mode, and T is the equivalent noise temperature of the receiving system.

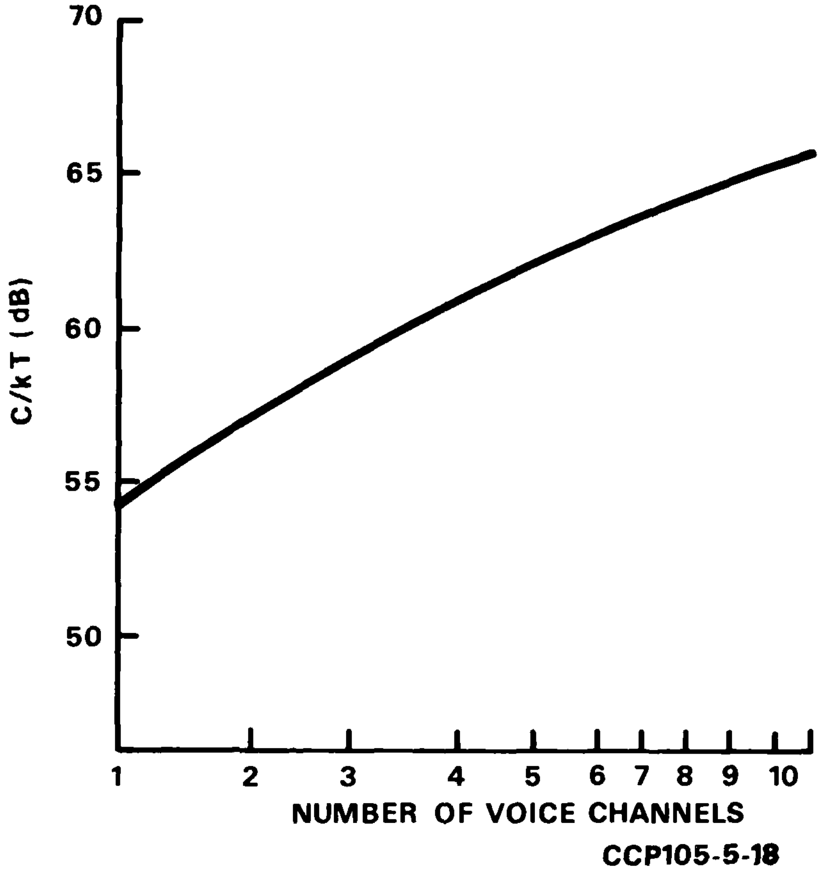


Figure 7-22. Channel capacity graph in relation to C/kT.

(b) To understand the importance of G/T, recall that C/kT determines the capacity of a satellite communication link. C, in the expression C/kT, corresponds to P_R and is expressed as:

$$C = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2} \tag{7.27}$$

Equation (7.27) indicates that any increase in the receive antenna gain G_R will bring about a corresponding increase in the value of C/kT.

(c) The equivalent noise temperature T in the expression C/*k*T is based on the total amount of noise in the received signal. This includes sky noise, excess rain noise, satellite intermodulation noise, as well as the earth terminal receiving system noise temperature. However, the earth terminal receiving system noise temperature is usually the major component of the total noise in the receive system and reducing it will automatically reduce T in the C/*k*T expression; thereby increasing the value of C/*k*T. There is a direct relationship between an increasing value of G/T and an improvement in C/*k*T. Satellite earth equipment designers always strive for as high a ratio of G/T as possible.

(d) The measurement of G/T is a difficult procedure because of the problem of finding a suitable reference and measurement procedure. The National Bureau of Standards (NBS) is working on a procedure that offers the most promise in terms of accuracy.

1. Tests were conducted at Camp Roberts, California, using the satellite communications ground station. The measurements employed the flux from two radio stars, Cassiopeia A(3C461) and Cygnus A(3C405), and assumed them to be two known noise sources. The parenthetical designations for the stars are references to the Revised Third Cambridge Catalogue. The values of star flux, assumed for the measurements performed, were extracted from several sources of published data available in the literature.

2. The measurements were performed by comparing the signals received (antenna first directed at selected radio stars and then at cold sky in the immediate vicinity of the radio stars). The ratio of the two noise power signal levels was compared using the NBS computer controlled Y-factor system. The results were read out directly in G/T value. A satisfactory analysis of the measurements' error has not been completed, but it is estimated to be less than ±0.8 dB.

(6) *Signal-to-noise ratio computation.*

(a) Four sets of computations must be performed to determine the S/*N* ratio of the earth receiving station output signals to the ultimate user of the link. The four sets of computations are:

1. An uplink power budget analysis.

2. A computation of how the signal power divides in the satellite between signal power and noise power.

3. A downlink power budget analysis to determine C/*k*T.

4. A computation of the number and quality of voice channels based on C/*k*T.

(b) The uplink power budget analysis may be performed by solving equation (7.13) for the free space loss. To the free space loss must be added the atmospheric absorption loss, which can be obtained from figure 7-12. The total loss consisting of free space loss, plus atmospheric absorption, determines the amount of power received by the satellite from the earth station. Excess rain noise and sky noise can be ignored in the uplink analysis because they will be very small in relation to the thermal noise introduced by the satellite receiver.

(c) The computation to determine how the signal power in the satellite divides when FDMA is used is best explained by going through a simple example calculation. Let us assume that the satellite has a rated power output of 25 watts and that there are four stations all of equal power accessing the satellite. Let us further assume that the accessing stations are small tactical stations so that the signal strength from each station reaching the satellite, as computed in equation (7.13), is equal to the thermal noise power of the satellite receiver. The 25 watts of output would, therefore, be divided equally into five parts, one part to each of the four signals it is repeating and one part to noise, so that each output carrier would be 5 watts.

(d) The 5-watt carriers, however, must be reduced by the amount of multiple access backoff that is required. In a typical satellite repeater this will amount to -1.5 dB, or a factor of 1/1.41, resulting in 3.5 watts per carrier.

(e) In computing the split between satellite signal power and thermal noise power, the noise power in the entire passband of the satellite has to be considered. Of the noise power leaving the satellite, however, we need be concerned only with the noise in the rf carrier bandwidth. For ease of computation, assume that the total satellite bandwidth is 200 MHz and that the rf carrier bandwidth is 2 MHz. The thermal noise power within each carrier passband will be one-hundredth of 5 watts or 0.05 watts.

(f) The satellite intermodulation noise must be added to the thermal noise power. In a typical satellite repeater, the intermodulation noise will be about the same magnitude as the thermal noise. Therefore, to complete this example of how the power divides within the satellite, we can assume that the output from the satellite would consist of 3.5 watts of signal power and 0.1 watt of noise power.

(g) An accurate downlink power budget analysis is somewhat more complicated than uplink analysis. Free space loss (equation (7.13)) and atmospheric absorption loss (fig. 7-12) must be applied to both the satellite signal power output and the satellite noise power output to determine how much signal power and how much noise power reach the earth station antenna. The sky noise (fig. 7-23) must be added to the noise power received from the satellite. The total noise

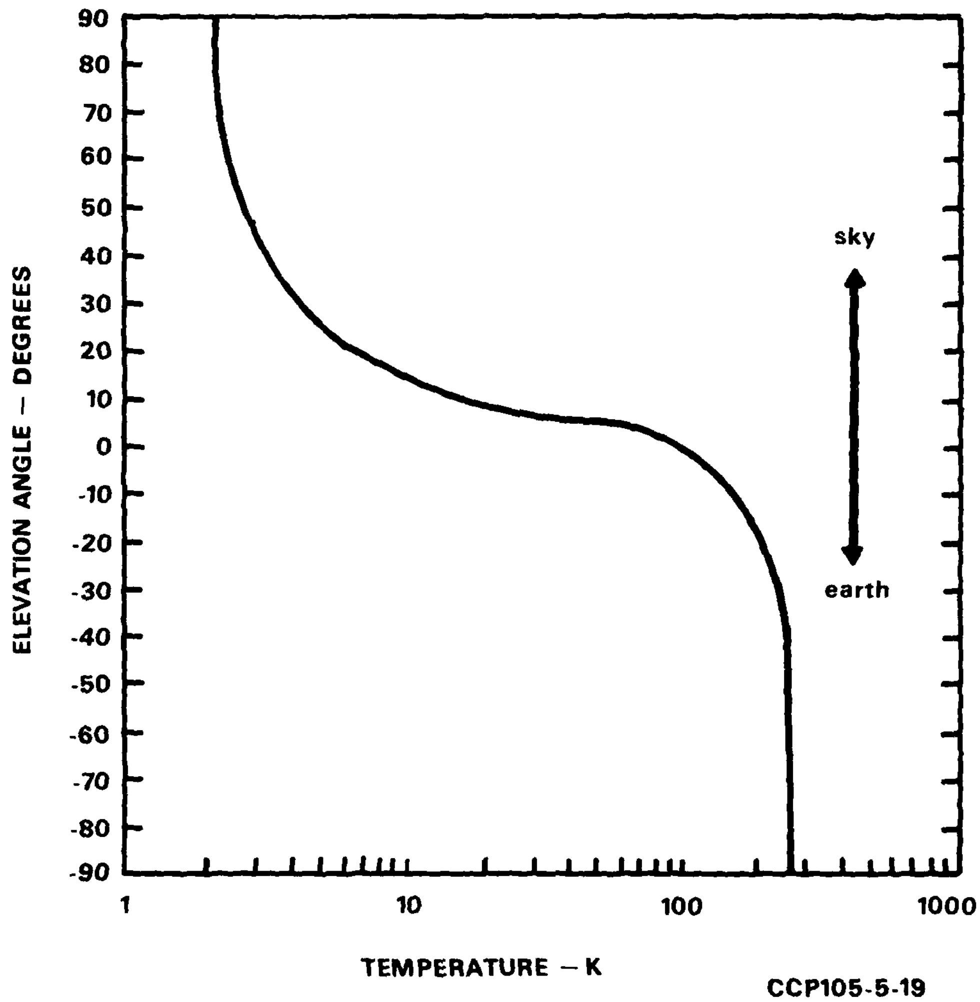


Figure 7-23. Space noise temperature graph as a function of elevation angle at 4 GHz.

in the receiving system may be calculated by adding the receiving system noise temperature to the noise from the satellite. The ratio of the signal power reaching the earth to the total noise power is C/*N*. Multiplying by 1/B gives the carrier to noise density C/*k*T.

(h) To determine the C/*k*T during high precipitation conditions, apply the excess rain loss to both the satellite signal and noise power. Also, add the excess rain noise to the noise from the satellite and sky noise. This will give a new figure for total received signal and noise. By adding the receiving system equivalent noise temperature to the received equivalent noise temperature of the total received noise, C/*k*T under rain conditions can be obtained.

(i) The number and quality of voice and data channels that a given satellite communications link can support can be determined from C/*k*T by using equipment curves such as shown in figure 7-23.

(7) *Required signal-to-noise ratio.*

(a) While the S/*N* ratio gives a measure of the performance of a communications link, the decision as to whether a particular S/*N* ratio is good or bad is still somewhat subjective, although various standards have been established. An S/*N* ratio that would be intolerable on a military system might be perfectly acceptable on a commercial system, or vice versa.

(b) Because of the difficulties involved in achieving high C/*N* ratios on satellite communication links, initial military systems will operate with moderate S/*N* ratios (probably in the vicinity of 25 dB). Such magnitudes will be achievable with military equipment that is ruggedized and transportable.

(8) *Other noise considerations.*

(a) The major galactic sources of noise (sky noise) are the sun, the Milky Way, and the scattered hydrogen clouds. The sun is the most intense source, having a noise temperature T_{sun} as high as 1,000,000 K at 300 MHz, although this decreases to 6,000 K for frequencies higher than 10 GHz, as shown by the line labeled T_{sun} in figure 7-24. If an antenna beam has a beam angle width θ_B, which is identical to the angle subtended by the sun from the earth, then when the beam is pointed at the center of the sun the antenna beam temperature T_B will be exactly the sun temperature as shown in figure 7-24. If the antenna beamwidth θ_B is larger than the angle subtended by the sun θ_S, then with the beam again pointed at the center of the sun, the beam temperature T_B is less than the sun temperature. During the most active years in the sunspot activity cycle, the temperature T_{sun} can be as much as 10,000 times that shown in figure 7-24.

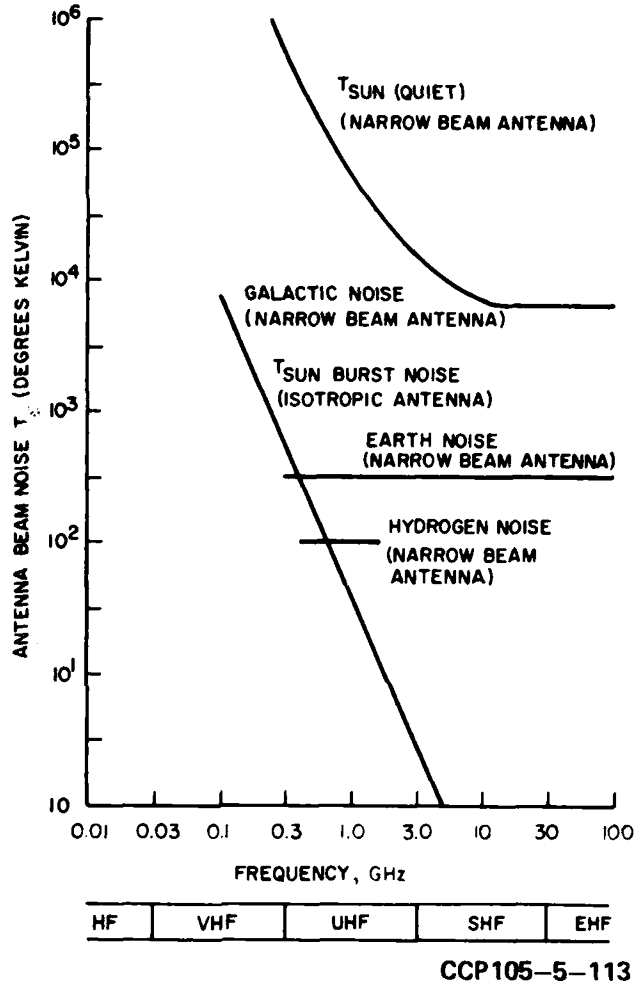


Figure 7-24. Antenna beam noise graph.

However, by ensuring that a high gain antenna does not point within ten beamwidths of the sun, the beam temperature T_B of an isotropic antenna resulting from the sun's radiation at its high burst intensity will not exceed that given by the line marked sun burst noise in figure 7-24.

1. The most intense galactic noise comes from our own galaxy, the Milky Way, in the direction of the constellation Sagittarius. For antenna beams sufficiently

narrow (fraction of a degree) T_B is given approximately by the galactic noise line in figure 7-24. There are isolated clusters of hydrogen in space that also radiate noise. High gain antennas with narrower beamwidths, when pointed at these sources, have beam temperatures of about 100 K, as shown by the hydrogen noise line in figure 7-24. For larger-than-fractional-degree antenna beamwidths, the beam antenna noise temperature would be less than 100 K.

2. Satellite antennas with a beamwidth narrow enough to be completely subtended by the earth would have an antenna beam temperature equal to the earth or earth's atmospheric thermal temperature (approximately 290 K), as shown by the line marked earth noise. The antenna beam noise would be less than 290 K for antennas with beam angle widths larger than the angle subtended by the earth.

3. In general, the antenna beam background noise temperature can be anything from 1,000,000 K to a few K, depending on narrowness of beamwidth angle, frequency, and direction of pointing. Both isotropic and high gain antenna beam noise are minimized by choosing frequencies higher than 1 GHz. High gain antennas must avoid pointing at discrete noise sources such as the sun, Milky Way, hydrogen clouds, and the earth.

4. Despite the high sky noise temperatures that can be encountered, the amount of sky noise entering the antenna under normal conditions does not present a severe problem. For an antenna pointed slightly downward so that it is actually looking at the earth, the noise temperature will be approximately the temperature of the earth, or 290 K; as the antenna pointing elevation is raised, the temperature rapidly drops off to something less than 10 K at an elevation angle of 15 degrees and continues to drop off as the antenna is raised. This is shown in figure 7-22.

(b) During periods of heavy rainfall, the antenna is in effect pointed at a large mass of material in the form of countless raindrops. Excessive rainfall noise energy in the form of blackbody radiation will be received by the antenna. Since the rain drops do not form a solid mass, the noise equivalent temperatures of the rainfall will be somewhat lower than the actual temperature of the rain.

1. If the rain is occuring some distance away from the antenna, the rainfall probably will not extend through the entire cross section area of the antenna beamwidth. This will have the effect of reducing the noise equivalent temperature further.

2. Since excess rain temperature is present only during periods of rain, it is expressed as a percentage of time that it exceeds a certain value in the same manner that excess rain loss is expressed. A typical distribution curve of excess rainfall attenuation is shown in figure 7-25.

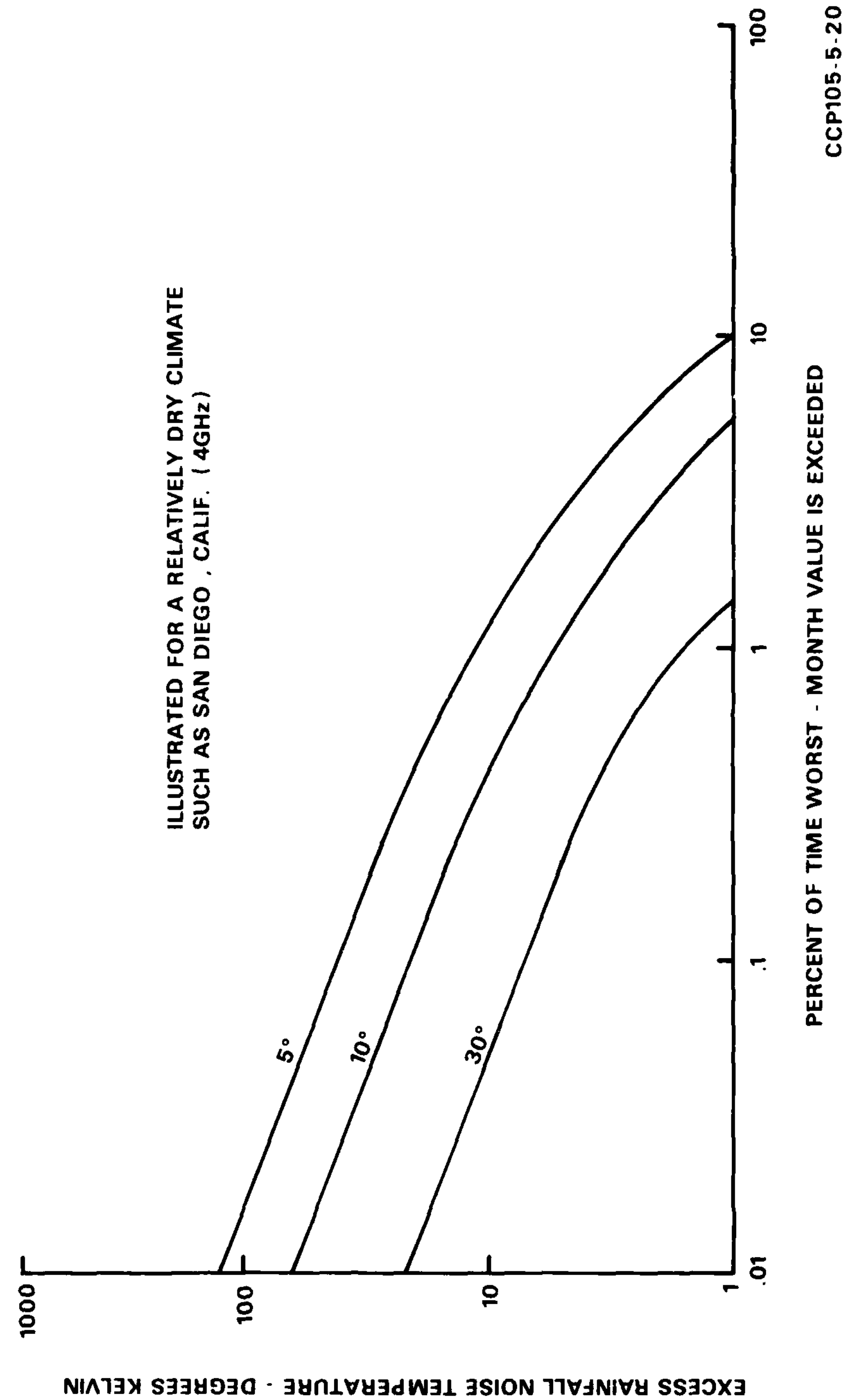


Figure 7-25. Distribution graph of worst month excess rainfall and temperature.

(c) The receiver portion of the satellite transponder generates a certain amount of thermal noise in the same manner as any other receiver. The amount of thermal noise generated is dependent upon the noise figure or equivalent noise temperature of the satellite receiver.

1. This noise is added to the signal received from the earth station and is rebroadcast by the transmitter portion of the satellite transponder with the desired signal. The amount of satellite thermal noise in the rebroadcast signal will depend on the ratio of received signal strength to satellite thermal noise.

2. A typical satellite transponder receiver will have an equivalent noise temperature of 3,000 K. The amount of noise power generated by a 3,000-degree receiver having a 50-MHz bandpass can be calculated from equation (7.16) to be:

$$N = k\mathrm{TB}$$

$$= 1.38 \times 10^{-23} \times 3 \times 10^{3} \times 5 \times 10^{7}$$

$$= 2.07 \times 10^{-12} \text{ watts} \qquad (7.28)$$

3. Most communications satellites employ automatic gain control so that as the received signal decreases in strength, the gain of the satellite increases. This ensures that, regardless of receiver signal strength, the satellite transmitter is driven to saturation or near saturation. In the event that no signal is received, the noise generated by the receiver will constitute the signal that drives the satellite transmitter.

4. The amount of satellite thermal noise in the output signal can vary anywhere from constituting the entire output signal (no-received-signal condition) to a small portion of the output signal (strong-received-signal condition).

(d) Since all satellite repeater transponders built to date have been nonlinear devices, they generate intermodulation noise. Intermodulation noise consists of the sum and difference signals that are produced whenever two or more signals are amplified in any device that is not exactly linear (output signal directly proportional to input signal) in its operation. The amount of intermodulation noise that is added to the signal by the satellite transponder is dependent to some extent on the type of multiple access that is employed. There is no accurate way of calculating the amount of intermodulation noise contributed by the satellite.

7-8. MODULATION.

a. Information-bearing signals are transmitted between the transmitter and receiver in satellite communication systems by means of radiated energy. For efficient transmission, the information is superimposed on a radio carrier wave. This is called the modulation process; likewise, the extraction of this information from the radio carrier is called the demodulation process. Sine wave modulation (emphasizing amplitude and angle modulation) and pulse modulation are discussed in this paragraph. Occasionally, a communication system will use combinations of the various forms of modulation. Various modulation-demodulation techniques are available which allow improvement in the output S/N ratio over the input S/N ratio. The theory involved is basic to the understanding of a communication link and also will be discussed in detail.

b. Sine wave, or continuous wave modulation, allows frequency multiplexing, or stacking, of many baseband signals (channels) into a specified frequency band. This allows many telephone messages to be transmitted simultaneously over a single pair of wires, or a number of television channels to operate via a wideband communication satellite. Systems have been developed whereby messages numbering in the thousands have been successfully multiplexed simultaneously into a single wideband channel.

c. The angle modulation (frequency or phase modulation) technique provides the advantage of a better S/N ratio as compared to amplitude modulation for a given input S/N ratio. In the case of amplitude modulated signals, noise superimposed on the carrier affects the amplitude of the carrier, and these disturbances appear at the output of the receiver. In angle modulation, two effects must be considered; the noise produces an amplitude disturbance in the carrier and the noise produces a frequency disturbance in the carrier. The first effect is normally eliminated by the limiting action in the receiver. The second effect in many circumstances can be reduced to an insignificant value by employing an angle deviation that is as large as possible.

d. In pulse modulation, the samples may be coded into pulses of uniform height and reshaped, when necessary. This minimizes the effect of noise and interference in many cases. Pulse modulation also results in an efficient use of transmission facilities. Messages are time multiplexed, or staggered, in time sequence so that each channel uses the total system bandwidth. Further, with the message in pulse form, coding and error correction can be applied for even more efficient channel usage. Pulse modulation has been widely used for radio, telegraph, and telemetry.

(1) *Carrier.*

(a) Sine wave modulation is normally applied to a carrier in which the amplitude, phase, or frequency of the carrier is caused to vary in accordance with the message. Figure 7-26 represents an unmodulated sine wave carrier of frequency

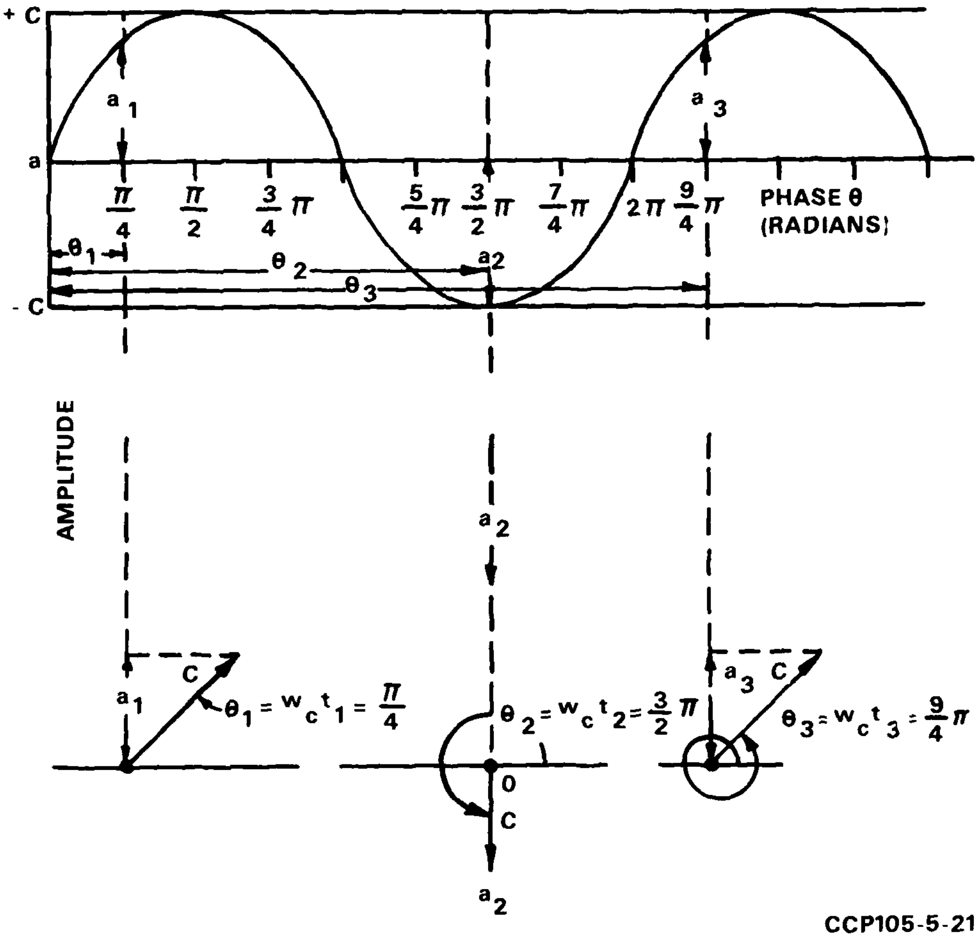


Figure 7-26. Sine wave modulation representation.

f_c whose amplitude (a) varies between peak values of +C and -C as the phase angle θ increases. Phase angle θ is measured in radians where 2π radians is 360 degrees. The sine wave carrier may be written in a variety of forms, such as:

$$a = C \sin(\theta_c + \phi) = C \sin(\omega_c t + \phi)$$

$$= C \sin(2\pi f_c t + \phi) = C \sin\left(\frac{2\pi t}{T_c} + \phi\right) \qquad (7.29)$$

where

θ is phase angle in radians

ϕ is a phase angle

ω is radians per second

$2\pi f_c t$ is carrier wave

T_c is period of one cycle of f_c

(b) Any of the three quantities, C, θ_c, or ϕ may be varied in accordance with the message, or modulating signal. The phase θ_c increases uniformly with time t at a rate ω_c (radians per second). The rate of change of phase can be also measured in Hz.

(c) The sine wave carrier can also be represented as the projection of a spinning vector of length C and rate ω_c on a line through the origin, 0 (fig. 7-26). As a matter of convention, the line is taken as the vertical axis and the direction of rotation as counterclockwise. The length of projection on the vertical axis for each of these vectors is θ_1, θ_2, and θ_3.

(2) *Amplitude modulation.*

(a) In amplitude modulation (am) the amplitude of the carrier is varied in accordance with the message; i.e., C varies. For simplicity, a single sine wave with frequency f_m, is considered as the modulation. The resulting modulated wave C_m can be represented mathematically as:

$$C_m = C_o \left[1 + M_a \cos 2\pi f_m t\right] \sin 2\pi f_c t \qquad (7.30)$$

where

C_o is carrier amplitude

M_a is modulation factor

$\cos 2\pi f_m t$ is the modulating wave

$\sin 2\pi f_c t$ is the carrier wave

(b) This is called double sideband amplitude modulation (dsbam). The unmodulated carrier with constant peak amplitude of C_o is represented in part A of figure 7-27. The message (modulating) wave of peak amplitude A and frequency f_m is represented in part B of figure 7-27. The modulated wave C_m is represented in part C of figure 7-27. Characteristics of dsbam are as follows:

1. When the baseband, or modulating wave has a zero value (points a, c, and e of part B of fig. 7-27) the modulated wave has the same magnitude C_o as when unmodulated.

2. When the baseband wave is at its maximum positive value (points b and f), the modulated wave is increased beyond C_o by an amount $(1 + M_a)$.

(c) The modulating factor M_a is a number between 0 and 1 representing the amount of modulation and is proportional to the magnitude of A. The value of M_a, shown in part C of figure 7-27, is about 50 percent.

(d) When the baseband wave is at its maximum negative value (point d) the modulated wave is decreased below C_o by an amount $(1 - M_a)$. For 100 percent amplitude modulation $(M_a - 1)$, the maximum possible amplitude of the modulated wave (part C of fig. 7-27) would be twice C_o at points b and f and zero at point d.

(e) Using the vector concept (fig. 7-26), a vector representation of the carrier wave, rotating at a rate $2\pi f_c$, is shown. Take as a frame of reference, the rotating carrier (assume the carrier is stationary so that the axis is now rotating oppositely at the carrier rate, $2\pi f_c$) and add to the carrier the upper sideband and lower sideband at their amplitudes $C_o M_a/2$ and respective instantaneous phase angles $2\pi f_m t$ and $-2\pi f_m t$. The resulting vector is a representation of amplitude modulation.

(f) If equation (7.29) is multiplied out, $C_m = C_o \sin 2\pi f_c t + C_o M_a (\sin 2\pi f_c t)(\cos 2\pi f_m t)$ and, after expanding the last term, using trignometric identity, to give:

$$\sin x \cos y = 1/2 \left[\sin (x + y) + \sin (x - y)\right] \qquad (7.31)$$

we arrive at:

$$C_m = C_o \sin 2\pi f_c t + \frac{C_o M_a}{2} \sin 2\pi(f_c + f_m)t + \frac{C_o M_a}{2} \sin 2\pi(f_c - f_m)t$$

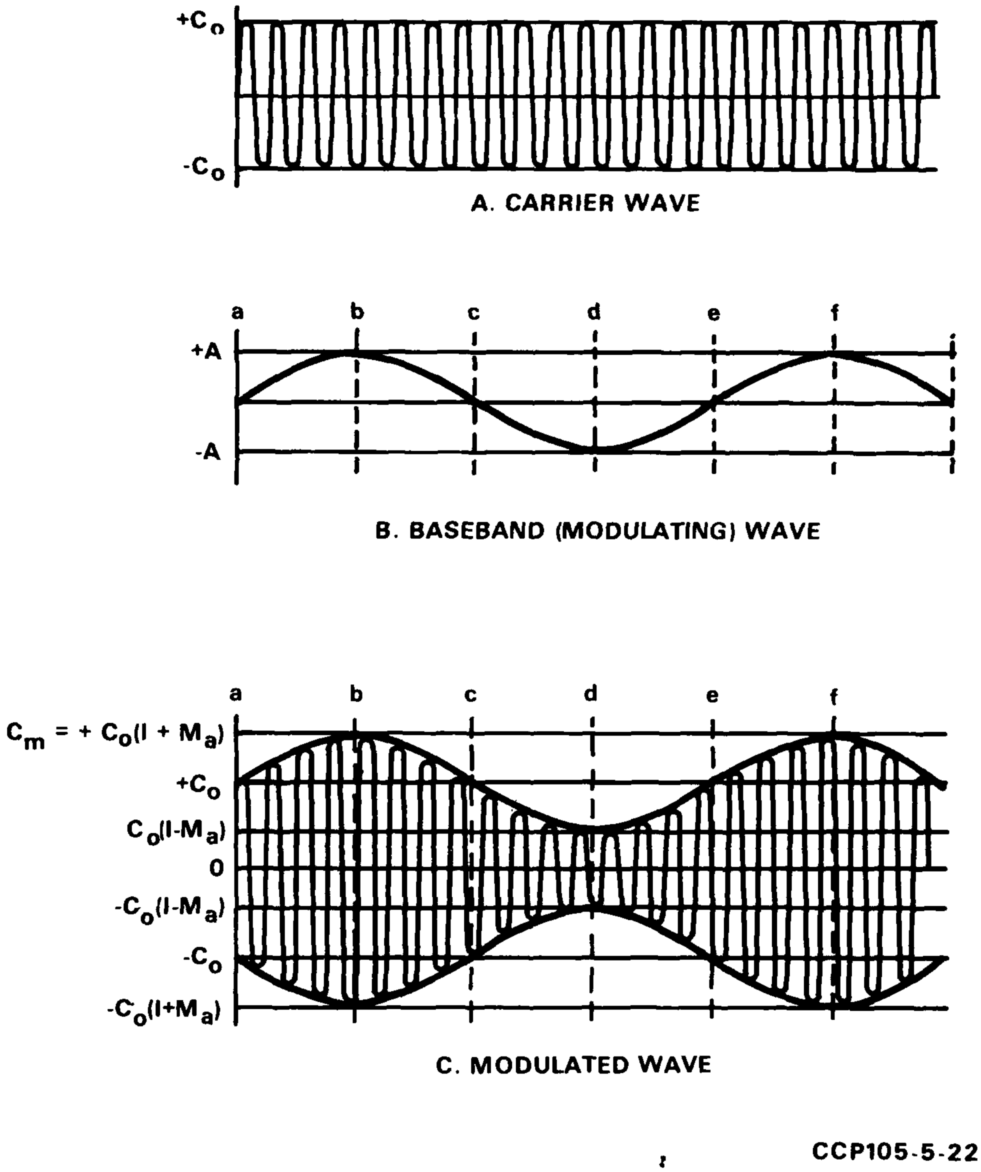


Figure 7-27. Amplitude modulation representation.

where

$\sin 2\pi f_c t$ is carrier

$\sin 2\pi(f_c t + f_m)t$ is upper sideband

$\sin 2\pi(f_c - f_m)t$ is lower sideband

(g) A frequency plot of the amplitude modulated wave would contain three components; the upper sideband displaced above the carrier frequency by the modulating frequency (f_m), the carrier (f_c), and the lower sideband displaced below the carrier frequency by the modulating frequency (f_m) with relative amplitudes ($C_o M_a/2$), (C_o), and ($C_o M_a/2$), respectively. Notice that the sidebands are equal and can be at most one-half of the amplitude of the carrier which occurs when M = 1 (100 percent modulation). A plot of the frequency spectrum of am is shown in part A of figure 7-28. Of special interest is that the bandwidth at the carrier frequency band is twice the highest modulating frequency $2f_m$. Since each term in equation (7.30) can be thought of as a voltage, the power in each sideband is proportional to $(M_a C_o/2)^2$, which can at most be one-fourth of the carrier power. As all of the information is carried in the sidebands, a simple calculation will show that at least one-half of the total power is wasted in the carrier. This results in a relatively inefficient modulation scheme

(h) Consider the effect of a complicated modulating signal such as voice or multiple tones. So as not to exceed 100 percent modulation (which would produce distortion), the sum of the individual modulation factors from each tone must be less than unity. In practice, this would be too restrictive because it is quite unlikely that each tone in a large number of tones would reach its peak value at the same instant. In practice, overmodulation is allowed for a certain percentage of time (normally 1 percent).

(i) Equation (7.31) is repeated so that the various types of am can be easily explained.

$$C_m = C_o \sin 2\pi f_c t + \frac{C_o M_a}{2} \sin 2\pi(f_c + f_m)t + \frac{C_o M_a}{2} \sin 2\pi(f_c - f_m)t$$

(j) This equation is a mathematical representation of dsbam or, more often, double sideband (dsb). To generate double sideband suppressed carrier (dsbsc), the carrier term is simply removed, leaving the two sidebands (table 7-1). Usually a balanced modulator is used to generate dsbsc. By using dsbsc, the percentage of information-carrying power in the total output signal is at least doubled, since no power is wasted in transmitting a carrier. Notice, though, that the width of

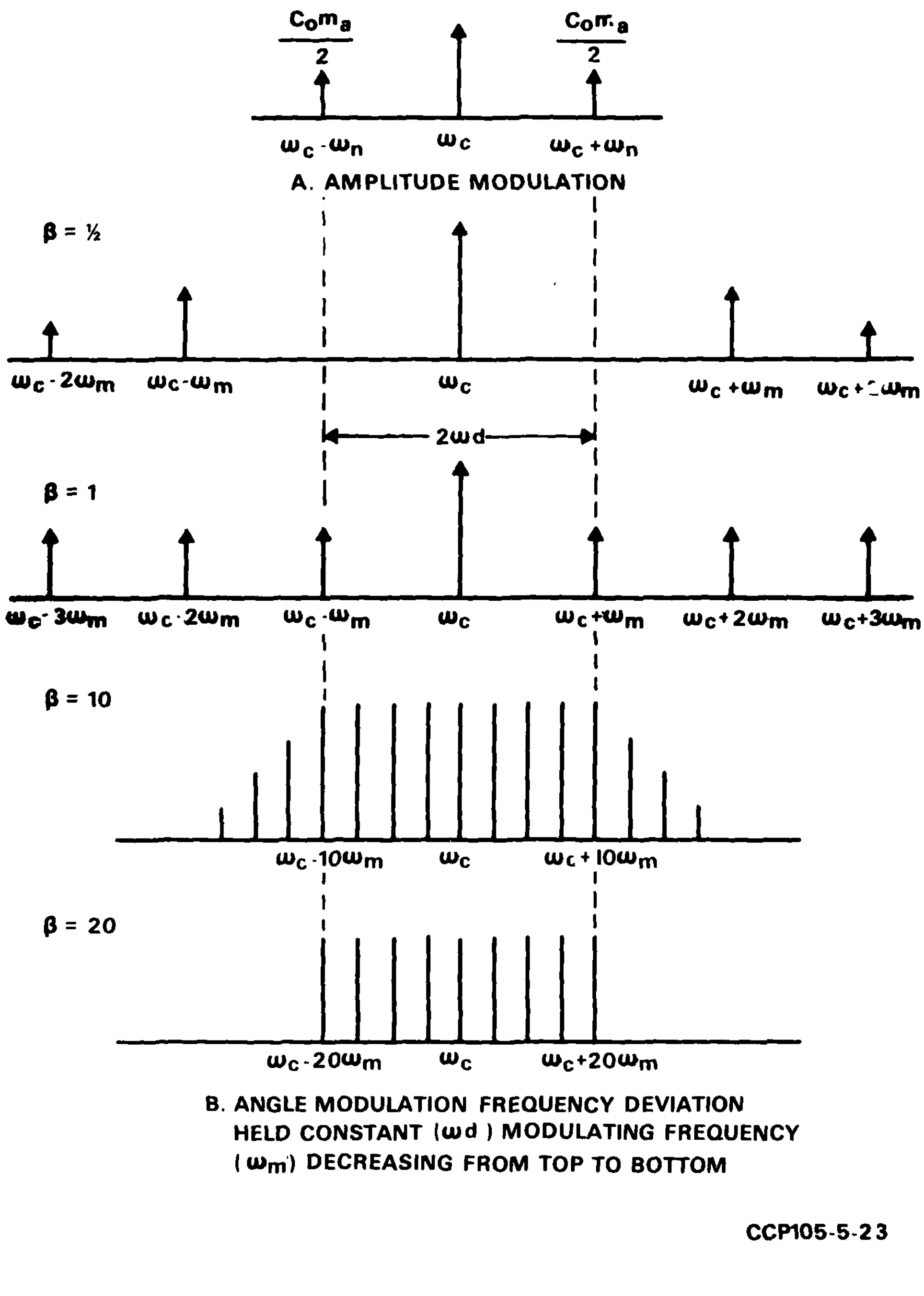


Figure 7-28. Representation of spectrum of modulated signals.

Table 7-1. Sine wave modulation

Type of modulation	Equation	Frequency components	Bandwidth	Improvement factor fm (maximum)
Double sideband amplitude modulation	$C_o \sin 2\pi f_c t +$	f_c		
	$\frac{C_o M_a}{2} \sin 2\pi(f_c + f_m)t +$	$f_c + f_m$	$2f_m$	1/2
	$\frac{C_o M_a}{2} \sin 2\pi(f_c - f_m)t$	$f_c - f_m$		
Double sideband suppressed carrier amplitude modulation	$C \sin 2\pi(f_c - f_m)t +$	$f_c + f_m$	$2f_m$	1/2
	$C \sin 2\pi(f_c - f_m)t$	$f_c - f_m$		
Single sideband amplitude modulation	$C \sin 2\pi(f_c + f_m)t$ or	$f_c + f_m$ or	f_m	1
	$C \sin 2\pi(f_c - f_m)t$	$f_c - f_m$		
Frequency, phase modulation	$^*\sum_{n=o}^{N} J_n(\beta)\cos \left[(2\pi)(f_c \pm nf_m)t\right]$	$f_c \pm \sum_{N=1}^{N} nf_m$	$2f_m(M + 1)$	$\frac{3}{2}M^2$

*where $J_n(\beta)$ are Bessel functions of the first kind and n is the highest significant sideband.

the frequency spectrum will be the same as dsb; i.e., $f_c - f_m$ to $f_c + f_m$, which is $2f_m$ wide. By removing either sideband from the dsbsc, not only is the percentage of information carrying power again doubled, but also the frequency spectrum will be halved (the same information is being transmitted in half the bandwidth of either dsb or dsbsc). This modulation technique is called single sideband amplitude modulation (ssbam or ssb). The generation of ssb is accomplished by two methods; either a sharp cutoff filter is used to eliminate the carrier and the unwanted sideband, or the carrier and one sideband are canceled out using a pair of balanced modulators and two phasing networks. Sometimes a combination of both methods is used. Table 7-1 contains the basic equations for the various modulation techniques. The use of ssb is almost universal in high -frequency, long-distance radio communications, mainly because of the complete concentration of the signal power in a minimum bandwidth. The main limitations of ssb are that a carrier must be reinserted at the receiver with a reinsertion error of less than 25 Hz and linear amplification must be used throughout the transmitter and receiver to hold distortion to a low level. Because of the effect of Doppler shift ssb is not suited for satellite communications, unless at least a partial carrier is also transmitted. Angle modulation does not require linear amplifications or accurate carrier reinsertion.

(3) *Angle modulation.*

(a) Modulation in which the frequency or phase of a carrier is caused to vary with the modulating signal is called angle modulation. The general angle modulated wave can be expressed as:

$$C_m = C_o \cos \left[2\pi f_c t + \phi(t)\right] \quad (7.32)$$

where

C_m is angle modulated carrier in volts

C_o is a constant equal to peak amplitudes of carrier in volts

f_c is carrier frequency in Hz

$\phi(t)$ is time varying modulation angle

(b) If angle modulation is used to transmit information, ϕ (t) must be related to the modulating signal to be transmitted. Phase modulation (pm) is angle modulation in which the instantaneous phase deviation ϕ (t) is proportional to the modulating signal. Similarly, frequency modulation (fm) is angle modulation in which the instantaneous frequency deviation, $d\phi(t)/dt$ (i.e., the rate of change of instantaneous phase deviation is proportional to the frequency of the modulating signal).

(c) Frequency modulation differs from phase modulation in that the peak frequency increase of the modulated wave relative to that of the carrier f_c, occurs at the time of peak value of the modulating wave. In phase modulation the peak value occurs when the modulating wave passes through the zero level (fig. 7-29). This maximum value of frequency increase or decrease in the modulated wave is called the frequency deviation f_d. Frequency deviation is directly proportional to the amplitude of the modulating wave.

(d) For the simplified example of the modulating wave having only a single frequency f_m, a modulation index β is given by:

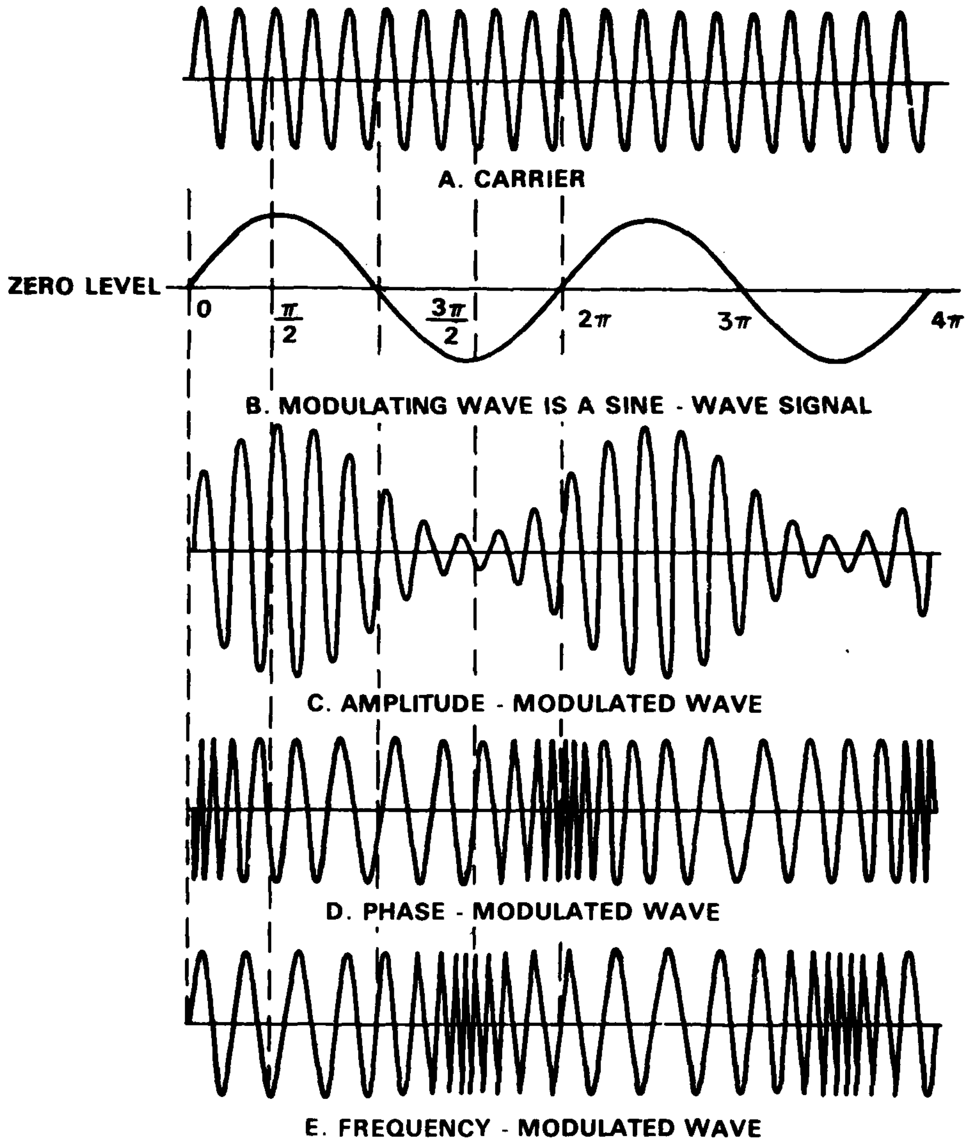
$$\beta = \frac{f_d}{f_m} \qquad (7.33)$$

and the basic equation for single tone fm is:

$$C_m = C_o \cos \left[2\pi f_c t + \beta \cos 2\pi f_m t\right] \qquad (7.34)$$

(e) Figure 7-29 presents a comparison of both sine wave amplitude and angle modulated signals. Part D of figure 7-29 (phase modulation) shows that the modulated wave frequency is highest at 0, 2π, 4π * * * of the modulating wave, lowest at π, 3π * * * , and the same at $\pi/2$, $3\pi/2$, $5\pi/2$ * * *. Similarly, in part E of figure 7-29 (frequency modulation), the identical events occur but advanced in phase by $\pi/2$ radians. Whether a particular angle modulated wave is phase or frequency modulated can not be determined unless the modulating signal is also known. There is no essential difference between the phase and the frequency modulated waves. Hence, the remainder of this discussion will be limited to frequency modulation.

(f) The frequency spectrum for either pm or fm angle modulation for a single modulating sine wave (f_m) with a small modulating index; i.e., M less than unity, is similar to am in that the carrier (f_c) is approximately the same amplitude with and without modulation, and a pair of sidebands spaced f_m on either side of the carrier appear (fig. 7-28). As the modulation index (β) is increased above unity, more pairs of sidebands appear in equal numbers and symmetric magnitudes on either side of the carrier (fig. 7-28). The greater β is, the greater the number of sideband pairs, each spaced f_m apart. Both the sideband and carrier magnitudes change in a complex manner as the modulation index is changed (parts B, C, D, and E of fig. 7-28). For large β, the carrier and sidebands are comparable in magnitude and difficult to distinguish. For am with large modulation index ($M > 1$) an infinitely large number of sidebands exist with some quantity of energy in a large frequency band. More than 99 percent of the total sideband energy lies within a spectrum bandwidth B_s centered at f_c, given by Carson's Rule:



CCP105-5-24

Figure 7-29. Representation of am, pm, and fm of sine wave carrier by sine wave signal.

$$B_s = 2f_m(\beta + 1) = 2(f_m + f_d) \qquad \text{For } \beta >> 1 \qquad (7.35)$$

(g) As shown in figure 7-29, the magnitude of an angle modulated wave, independent of the modulation, is constant and equal to the carrier magnitude C_o. Therefore, the total power of the sidebands plus carrier of an angle modulated wave must equal the power of the unmodulated carrier. As the modulation index increases, the number of sidebands increases, but the average magnitude of the individual sidebands and; thus, of the carrier, must decrease in such a way as to maintain a constant total power. In contrast, an amplitude modulated wave always contains more power than the carrier alone.

(h) The preceding discussion on fm has been based on the simplest form of a single sine wave modulating signal. Unfortunately, the extension to more complicated and more realistic modulating signals is difficult. The nonlinearity of the fm process precludes the use of superposition, so that even the computation of the frequency spectrum of an fm signal with two modulating sine waves becomes a tedious task. For the two-tone case, there are sideband components displaced from the carrier by all possible multiples of the individual modulating frequencies. Also, there are components displaced by all possible sums and differences of multiples of the modulating frequencies. When higher numbers of waves are considered, the task of computation becomes extremely time consuming unless accomplished with the aid of a computer.

(i) When the baseband contains a large number of frequencies, the complex angle modulated frequency spectrum will be a jumbled collection of the spectra due to each baseband component, each of which has a different frequency f_m and modulation index M. The spectrum width B_s which would contain all but a negligible (less than 1 percent) portion of the constant sideband power, would be as before:

$$B_s = 2(f_m + f_d) \qquad \text{For } \beta >> 1 \qquad (7.36)$$

where f_d is the highest single frequency of the modulating signal.

(j) As the amplitude of both phase and frequency modulated waves is constant at the unmodulated carrier level, the power transmitted is a constant equal to the unmodulated carrier power level. Further, since the amplitude is constant, a nonlinear device such as a class C amplifier, which has very high efficiency, can be used in the transmitter stages to amplify the signal after the modulation has been applied to it in some low level stage. Since all of the message information is contained in the instantaneous frequency or, equivalently, in the zero crossings of the modulated carrier wave, any device, such as a limiter, which maintains these zero crossings retains all of the message. These features make angle modulation attractive for satellite communications.

(4) *Pulse modulation.*

(a) Pulse modulation includes the types of modulation wherein the amplitude, width, frequency or phase, position, or quantity of a set of carrier pulses is altered in a definite pattern corresponding to the message to be transmitted. The number of variations that may be employed and the complexities of some systems make a complete discussion on pm outside the scope of this publication. However, two principal types of pm, those of pulse amplitude modulation (pam) and pulse code modulation (pcm) are discussed.

(b) Both pam and pcm are based on the regularity of the sine wave. In examining a typical sine wave pattern, it becomes apparent that it is only necessary to send a number of samples in each cycle in order to describe the wave.

(c) On still closer examination it is apparent that these samples need not both be amplitude. One might be amplitude and the other, for example, might be phase information. So long as two independent samples are sent for each cycle, the original sine wave can be reconstructed.

(d) It is possible to send more than one sine wave using this technique as long as two independent samples per cycle are transmitted. Thus, for two sine waves, the samples would be two samples for each cycle from each wave, or four samples. This may be continued for any number of sine waves as long as the ratio of at least 2n samples per n waves is maintained for each cycle.

(e) Figure 7-30 illustrates the timing and amplitudes used in a typical pam transmission of two waves. The period of lowest frequency, which could be divided into at least four time intervals, is in this case divided into eight. During the first time interval, a pulse equal to the height of the first sine wave at the beginning of the time interval (at t_1) is transmitted as a burst of carrier, as shown in part C of figure 7-30. During the second time interval, a pulse equal to the height of the second sine wave at t_2 is transmitted. At t_3 a pulse is transmitted equal in height to the first sine wave; at t_4 the second sine wave and so on for the transmission. These pulses indicate the shape of the waves, as shown in part C of figure 7-30.

(f) This method may be extended to include any number of sine waves and may incorporate different methods, such as sending the width, frequency, or phase, or varying the position of the samples.

(g) Transmitting n sine waves during a time interval when only one sine wave could be sent with sine wave modulation techniques demands an increase in bandwidth. The bandwidth is needed in order to distinguish the individual pulses.

AD-A114 385 ARMY COMMUNICATIONS COMMAND FORT HUACHUCA AZ F/G 17/2
INTRODUCTION TO SATELLITE COMMUNICATIONS.(U)
APR 74
UNCLASSIFIED ACC-CCP-105-5 NL

3 OF 4
AD
A114385

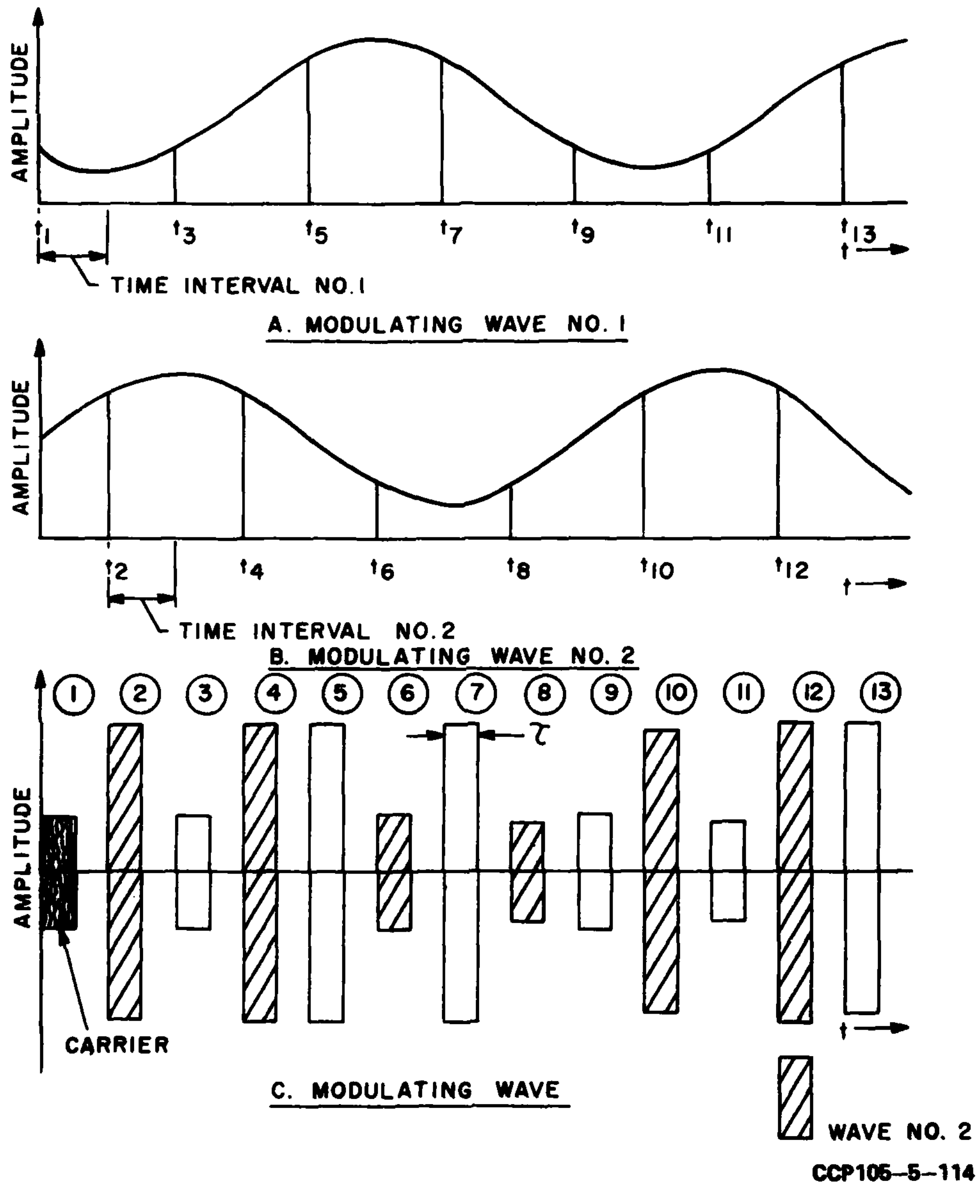


Figure 7-30. Representation of pulse amplitude modulation time-division multiplexed.

It can be shown that the minimum bandwidth necessary to resolve two successive baseband pulses is equal to one-half of the reciprocal of the pulse width; i.e.,

$$\text{Minimum bandwidth} = \frac{1}{2\tau}, \text{ where } \tau = \text{pulse width} \tag{7.37}$$

For double sideband pulse amplitude modulation (dsbpam) the transmitted bandwidth will be twice this minimum bandwidth.

(h) Since at least two pulses per cycle of the signal to be modulated must be sent to completely specify the signal, the widest any one of a set of equal width pulses can be is one quarter of the period of the modulating frequency f_m. If n signals are to be transmitted, the maximum pulse width per signal must be 1/n of the single frequency case. Since the minimum bandwidth is related to the pulse width in equation (7.37), the bandwidth (Bw) per sideband is:

$$Bw = \frac{1}{\frac{2\tau}{n}} = 2nf_m \tag{7.38}$$

(i) For dsbpam, the minimum bandwidth would be $4nf_m$, which is just 2n times that of dsbam. The same result could be accomplished by frequency stacking of the individual channels; i.e., frequency multiplexing and time division multiplexing use the same basic bandwidth.

(j) An important point is that, although theory gives a lower limit on the number of pulses (samples) and minimum required bandwidth, these limits are rarely attained in practice. For instance, three pulses per cycle give about a 1 percent error when reasonably good data filters are used; i.e., 60 dB per octave cutoff rates. The factor limiting performance is normally the data itself, in that it has no definite bandwidth and, hence, must be restricted by filters. The minimum bandwidth follows the pulse width criteria stated in equation (7.37) with a negligible error. In practice, the bandwidth is as given in equation (7.37) but usually the number of samples per cycle is increased from two to at least three, giving an increase in bandwidth of at least 50 percent.

(k) This detailed discussion of dsbpam has been given because of the simplicity of this type of pm. An obvious extension, more applicable to space communications, would be to code the modulating wave amplitudes into either carrier pulse widths or frequencies. This form of pm, wherein the signal is transmitted as a coded group of pulses corresponding to their sample amplitudes, is called pulse code modulation (pcm). The pcm method takes advantage of the limitations in the ability of the observer to distinguish the fine detail of a signal by transmitting only a finite number of levels. A specific pulse code group is assigned to each quanitzed sample level, and this pulse code group is transmitted during a time interval equal to the time between samples (part C of fig. 7-31).

(l) Take a particular interval, for instance, the interval beginning with the time 6 as shown in part A of figure 7-31. At this time the signal is between 3 and 4 volts, but is closer to 3 volts, so this particular sample is quantized into a 3-volt sample as shown in part B of figure 7-31. Each voltage level is assigned a pulse, or code grouping, with 3 volts being 011 (part C of fig. 7-31). Part C of figure 7-31 also shows the coding for each interval corresponding to the quantized samples. If a 0 is assigned as a 0-volt pulse and a 1 as a 1-volt pulse, the pulse pattern for a 3-volt quantized sample will be three pulses, with the first pulse being 0 volt and the next two pulses being 1 volt each (part D of fig. 7-31).

(m) The quantization process introduces some error in the eventual reproduction of the signal. It will be noticed that the quantized samples (part B of fig. 7-31) can be in error from the actual samples by as much as 0.5 volt (one-half the distance between the levels). This error is the quantization error and is called quantization noise. Since quantization noise is dependent on the distance between the levels, it can be reduced to any level desired by reducing this distance, or equivalently increasing the number of quantization levels. Unfortunately, by increasing the number of levels, a longer code is needed to specify each level uniquely; however, a longer code requires more pulses and means wider bandwidths. A complete discussion of coding is beyond the scope of this document, but an introduction to the subject is given in chapter 8.

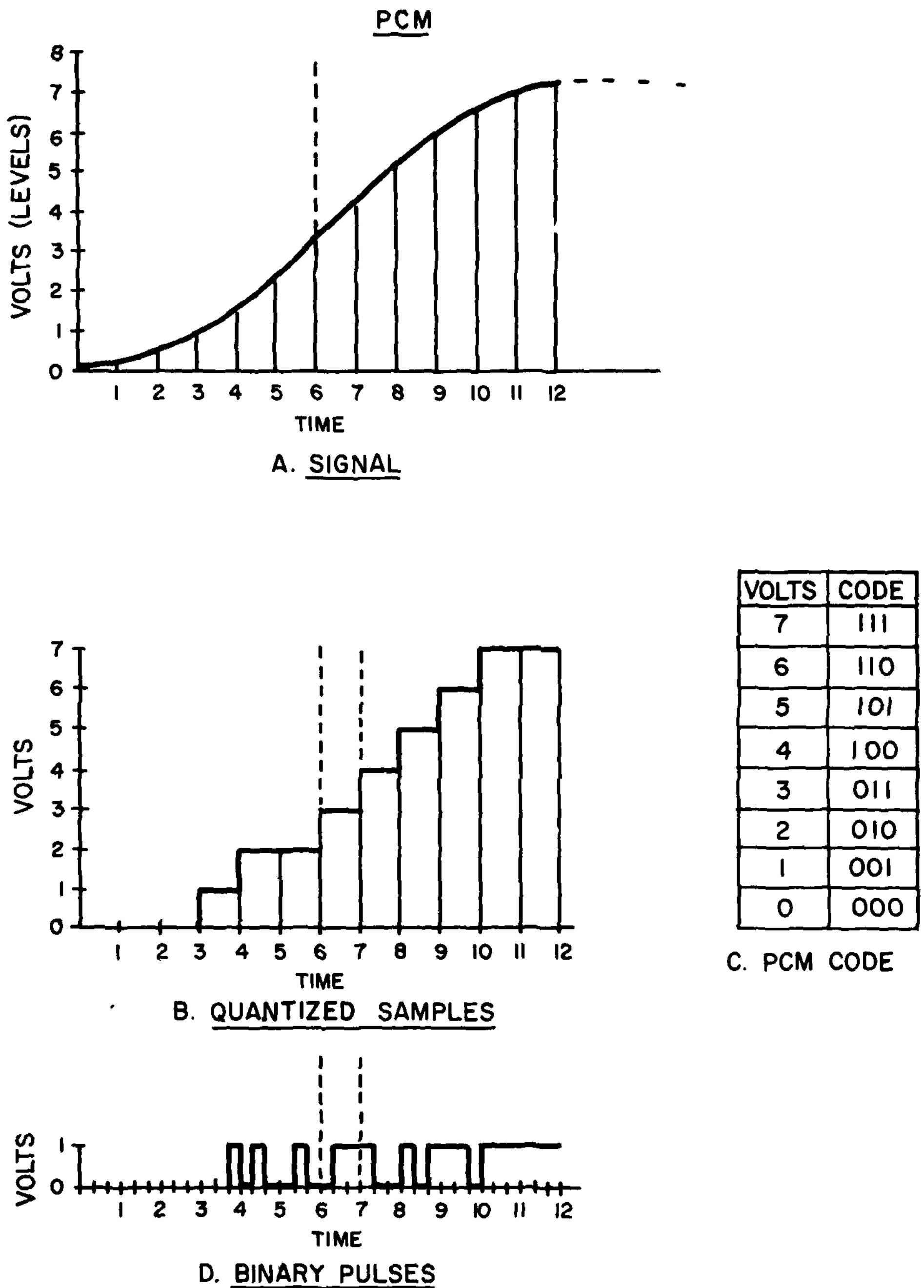
(n) As has been pointed out earlier, fluctuation noise introduced during transmission and at the receiver may cause errors in recognition of the coded pulses. If the received pulses are above a certain threshold; however, the average number of errors can be kept quite low. In this case it can be shown that the output S/N ratio depends on quantization noise alone and in dB is directly proportional to the number of pulses used for each quantization, that is:

$$(S/N)_O = 20 \log M = 6n \tag{7.39}$$

where

M is the number of quantizing levels, if binary coding is used

n is the number of pulses for each sample $M = 2^n$



VOLTS	CODE
7	111
6	110
5	101
4	100
3	011
2	010
1	001
0	000

C. PCM CODE

CCP105–5–115

Figure 7-31. Pulse coding representation.

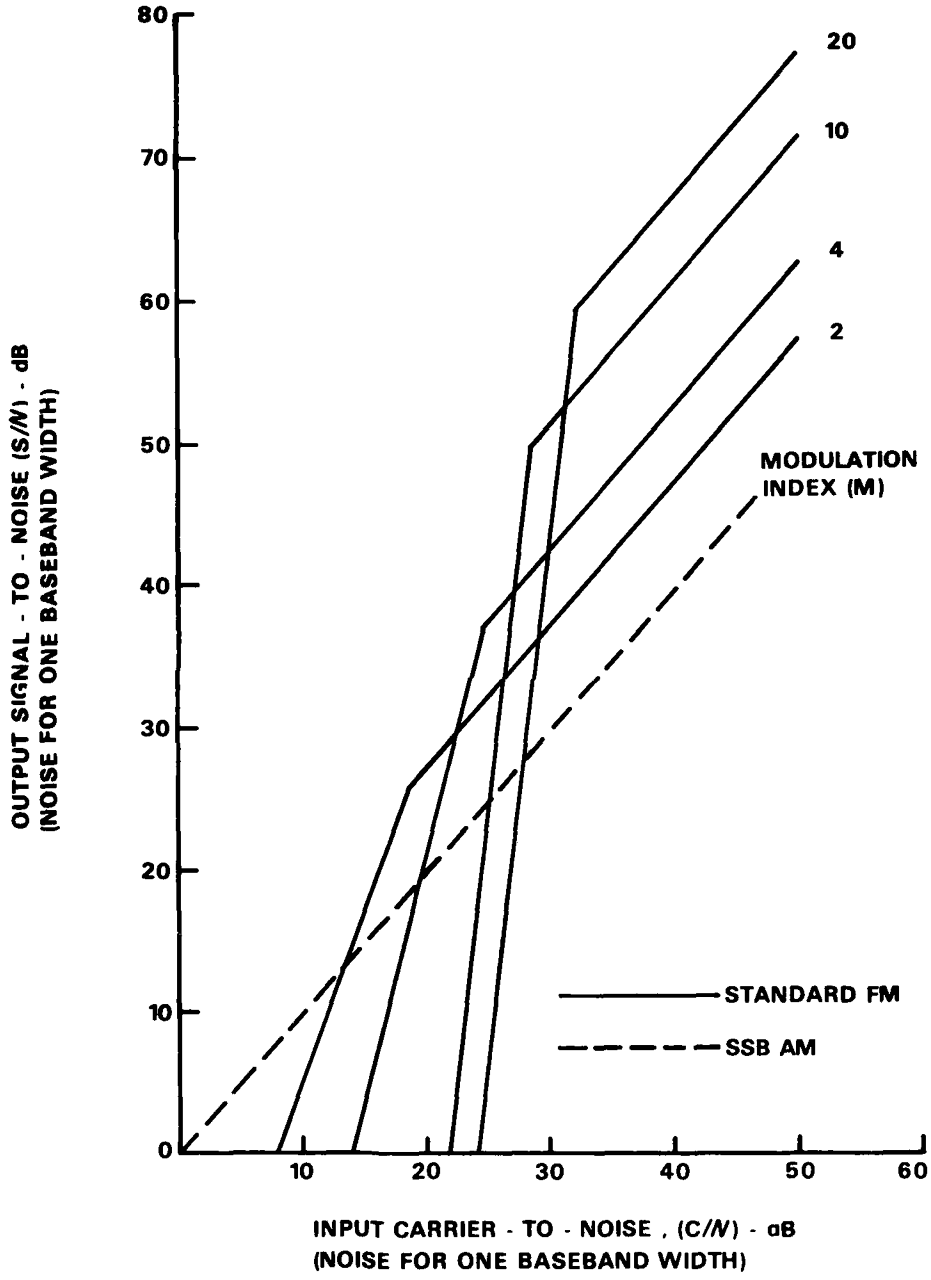
(o) Since bandwidth was shown previously to be related to the number of pulses, the S/*N* ratio in dB is directly proportional to bandwidth. This is a very important discovery, since we can trade bandwidth for S/*N* on an exponential scale. The best we can obtain using any other modulating method is a direct one-for-one trade, which is for angle modulation. For fm, doubling the bandwidth improves S/*N* ratio by a factor of two. The pcm method is used in cases of repeated pulse regeneration and retransmission, such as data transmission and telemetry.

7-9. DEMODULATION.

a. The ground receiver-detector combination must deliver to the baseband output (fig. 7-27 and 7-29) a recovered baseband wave of best possible quality. The receiver amplifier and demodulator must be designed not only for the type of modulation (i.e., am or fm), but also for the exact parameters of the modulation chosen. The satellite receiver function is only to amplify and frequency translate. If the modulation is am, the receiver must have a sufficiently accurate automatic gain control (agc) to avoid overload, nonlinearity, or limiting. Any nonlinear amplifying on an am wave will cause distortion in the demodulated baseband wave. On the other hand, angle modulated receivers in both satellites and ground stations usually limit the amplitude intentionally to provide a better immunity against noise. As can be appreciated from the previous discussion on angle modulation, amplitude limiting has no effect on the frequency or phase deviations in the modulated wave.

b. The receiver bandwidth (B_r) must be wide enough to pass the modulated spectrum bandwidth (B_s) to avoid distortion in the demodulated baseband signal. This means that for ssb the receiver bandwidth need be only as wide as the baseband width (f_m), but for angle modulation the necessary receiver bandwidth can be many times this, depending on the chosen magnitude of the modulation index β. The receiver bandwidth should not be wider than the minimum necessary to pass the modulated wave spectrum bandwidth in order to minimize receiver noise power (N_r), which has been shown to be proportional to receiver bandwidth.

c. Modulation-demodulation theory has developed the best possible single baseband modulation improvement factor (F_m) achievable from amplitude and angle modulation. F_m has been defined as the ratio of the baseband S/*N* output divided by C/*N* input where the noise power in C/*N* is normalized to single baseband width. It was also pointed out that achieving the desired S/*N* output quality with the highest possible F_m is extremely desirable. The maximum possible F_m can be considerably larger than unity, as shown by the solid lines in figure 7-32. For comparison, the dotted line shows the S/*N* output to be exactly equal to the C/*N* input (i.e., $F_m = 1$), which is for ssb. The angle modulation improvement ratio F_m (in dB) is the S/*N* difference (dB) between the angle modulation line and



CCP105-5-25

Figure 7-32. Curves for S/N versus C/N for fm.

ssb line. It is seen (fig. 7-32) that the larger the modulation index (M), the greater the improvement factor for angle modulation. However, angle modulation has a carrier-to-noise threshold ($(C/N)_T$) below which the F_m decreases rapidly. The larger the modulation index the larger the $(C/N)_T$, which must be exceeded for the full modulation improvement factor to be in effect. For example, an M of 2 has an F_m of 6 (8 dB) but the C/N must exceed an 18 dB threshold (fig. 7-32). For the higher M of 10, the F_m is 150 (22 dB) and the threshold is 28 dB. For C/N ratios greater than the $(C/N)_T$, the F_m for angle modulation is:

$$F_m = \frac{3M^2}{2} \tag{7.40}$$

which converts to:

$$S/N = \frac{C}{N} \times \frac{3M^2}{2} = AB_s \text{ for } M = 1$$

which shows that S/N is directly proportional to bandwidth where A is a constant for any particular system.

d. As the C/N level decreases below threshold, in addition to the modulation improvement factor degrading rapidly, the noise character in the baseband output changes from a fine-grain hiss to an erratic sputter or pop. In practice, the threshold demarcation is gradual, rather than the sharp discontinuity shown in figure 7-32. A summary of the important equations and results is presented in table 7-1.

7-10. COMPANDER ACTION.

a. A compander actually consists of two separate items of hardware; a compressor at the transmitter and an expander at the receiver. The compressor is usually located at the baseband input of the transmitter. It compresses the amplitude range of the baseband signal by imparting more gain to the weaker components and less gain to the stronger components of the baseband signal. The circuit accomplishing compression operates by sensing baseband signal level and adjusting amplifier gain so that weaker signals are amplified more than stronger signals.

b. At the receiver, the expander circuit senses the baseband signal level and adjusts amplifier gain so that stronger signals are amplified more than weaker signals. This restores the original levels of amplitude variation to the baseband signal (provided the compressor and expander have identical characteristics).

c. The prime advantage of compander action is that it raises the average modulation level of the baseband; thus, increasing the S/*N* ratio of the received signal.

7-11. PREEMPHASIS AND DEEMPHASIS.

a. *Preemphasis.*

(1) One characteristic of human speech is that the higher audio frequencies make the greatest contribution to the intelligibility of the speech pattern. Unfortunately, another characteristic is that the higher audio frequencies do not have the intensity that the lower frequencies possess. Therefore, the lower voice frequencies (even though they contribute less to message intelligibility) cause more deviation of an fm signal; hence, a higher S/*N* ratio than the higher voice frequencies. To avoid degrading the higher voice frequencies by means of a poor S/*N* ratio, more amplification is provided for the higher frequency voice components than for the lower frequency voice components. This is known as preemphasis.

(2) A simplified preemphasis network is shown in part A of figure 7-33. A preemphasis curve showing relative response to voice frequencies of a preemphasis network is shown in figure 7-34.

b. *Deemphasis.*

(1) At the receiver, the reverse characteristic of preemphasis occurs, so that the normal high-frequency-to-low-frequency ratio of the voice is restored.

(2) This is accomplished by a deemphasis network which attenuates the higher frequencies (part B of fig. 7-33).

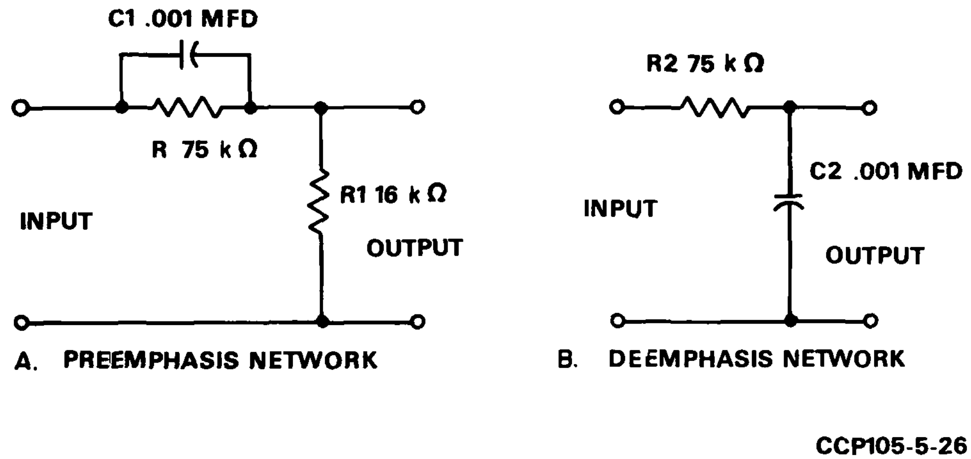
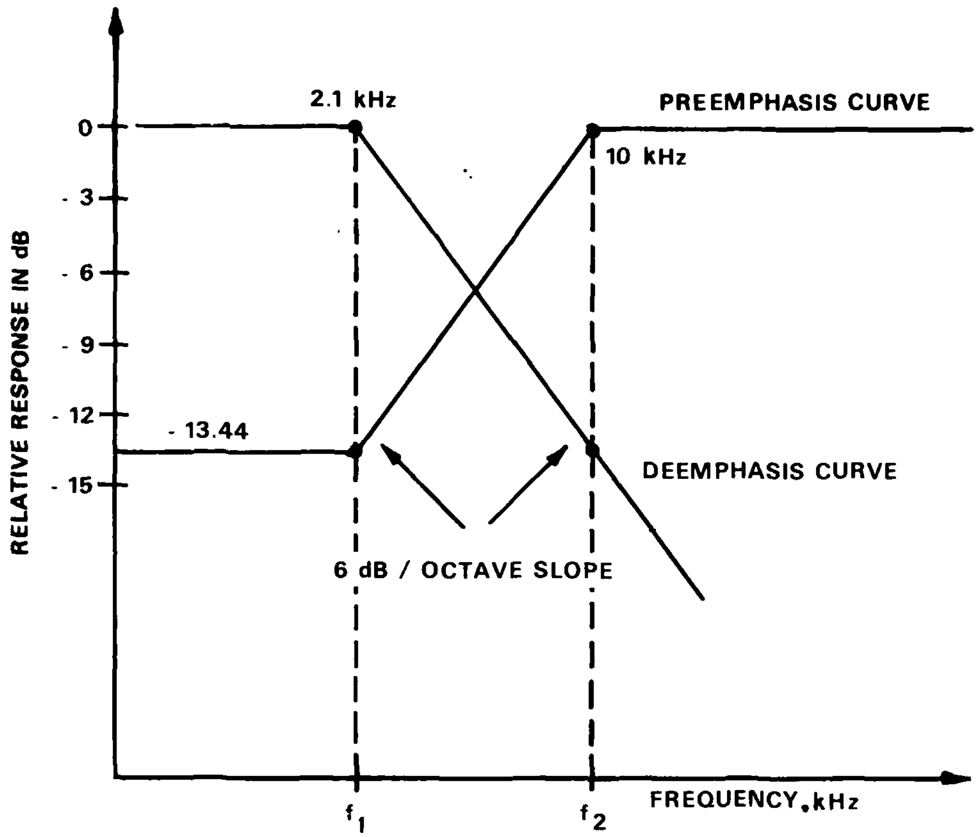


Figure 7-33. Preemphasis and deemphasis networks schematic diagram.



$$f_1 = \frac{1}{2\pi\gamma} = \frac{1}{2\pi R C_1} = \frac{1}{2\pi R_2 C_2} = 2.1 \text{ kHz}$$

$$f_2 = \frac{1}{2\pi\gamma_2} = \frac{1}{2\pi R_2 C_2} = 10.0 \text{ kHz}$$

CCP105-5-27

Figure 7-34. Preemphasis and deemphasis curves.

CHAPTER 8

EARTH TERMINAL COMPONENTS, DEVICES, AND SYSTEMS

8-1. GENERAL.

a. Chapter 4 described a satellite earth terminal and its use in the DCS. The explanations of the specialized techniques and components were given in a very abbreviated form because a full description would have distracted from the overall view of the ground station.

b. This chapter presents a more detailed explanation of some of these techniques and components grouped under the subsystem in which they are most likely to be employed. Some techniques and components may be used in more than one subsystem; as an example, a TWT might be found in either receiver or transmitter subsystem, but generally the grouping will provide accurate reference to the components.

8-2. RECEIVER SUBSYSTEM.

a. *Feedback Demodulation.*

(1) As described in paragraph 7-8, the angle modulation technique achieves improvement factors greater than one. This means that the demodulated or baseband S/*N* ratio is higher than the C/*N* ratio at the receiver input. For ordinary demodulation (without feedback) of an angle modulated signal, the C/*N* ratio at the input must be greater than moderately high thresholds if the improvement ratio is to be realized. A space communication link from the satellite to the ground station must achieve a satisfactory baseband S/*N* ratio quality at the lowest possible C/*N* power. Any carrier power larger than the minimum necessary requires the satellite transmitter to be larger, heavier, and more costly. The relatively high threshold of carrier-signal-to-noise ratio that an angle modulated signal must exceed for satisfactory demodulation of the signal with conventional fm detectors is an obvious disadvantage for space link communications. However, in recent years two demodulation techniques have been developed that can demodulate satisfactorily down to considerably lower threshold levels and still retain the improvement factor of angle modulation at signal levels above threshold. One demodulation technique used frequency modulation feedback (fmfb), and the detector used is called an fmfb detector. The second technique employs a phase-lock loop (phll) detector. The circuitry of fmfb and phll differ considerably, but the performance is essentially the same for both. Some of the functional mechanisms of fmfb and phll are qualitatively described.

(2) When frequency or phase modulated, the carrier deviates higher and lower in frequency or ahead and behind in phase, according to the waveshape of the baseband signal. However, it also contains noise. The greater the deviation, the greater the bandwidth occupied by the spectrum of the modulated carrier.

(3) Refer to figure 8-1 and assume that switch S is in the OPEN LOOP position and that a large deviation fm wave with modulation index β_1 is applied to the mixer at point A. At the same time an identical fm wave with a slightly reduced deviation (modulation index β_2), is applied to the other terminal of the mixer at point B. The mixer output will be the sum and difference frequencies of the two waves. The difference frequency is selected by an IF filter. This difference frequency at point C will have a deviation with index β_3 which will be the difference of the modulation indexes of the two fm waves ($\beta_3 = \beta_1 - \beta_2$). This resulting reduced deviation wave may be passed through a filter whose bandwidth is approximately β_3/β_1 times that required of the large deviation wave. The signal is frequency detected in a circuit such as a discriminator. The second fm wave (β_2) can actually be derived by feeding the output signal of the frequency discriminator through a low-pass filter to frequency deviate a voltage controlled oscillator (vco). The larger the gain in the feedback loop, the more the input deviation is reduced in the IF.

(4) To explain the threshold improvement gained by using fmfb, the threshold mechanism of conventional fm will be detailed. The threshold occurs in a conventional fm receiver when the random noise peaks exceed the carrier amplitude, prior to the frequency modulation detector (discriminator), for a sufficient percentage of time. Each time a noise peak exceeds the carrier amplitude, an impulse in amplitude (spike) appears at the frequency discriminator output. This noise appears as a random sequence of spikes, which are heard as sharp pops in an audio system or seen as spots on a television screen. For a voice, data, or television channel, operation below threshold is generally unsatisfactory.

(5) These noise spikes are reduced by fmfb and, therefore, the threshold is reduced. This is accomplished by feeding the detected signal and noise back to the vco. The noise feedback will reduce the incoming noise and reduce the threshold of the system. At the same time the improvement factor for high level *S/N* ratios will remain that of the transmitted wave. To demonstrate this action, consider two examples. First, a 50-dB *S/N* ratio is desired at the output or baseband signal. By using conventional fm, a minimum *C/N* ratio of 27.5 dB is required (fig. 8-2). Only 18.5 dB is required for fmfb, and the carrier power can be reduced by 9 dB, which is a factor of 8. As the second example, take a more typical baseband *S/N* of 35 dB. Referring to figure 8-2, it can be seen that the carrier power can be reduced by 6.6 dB, or a factor of 4.6. The modulation indexes necessary for an optimum system are also shown in figure 8-2. Since bandwidth

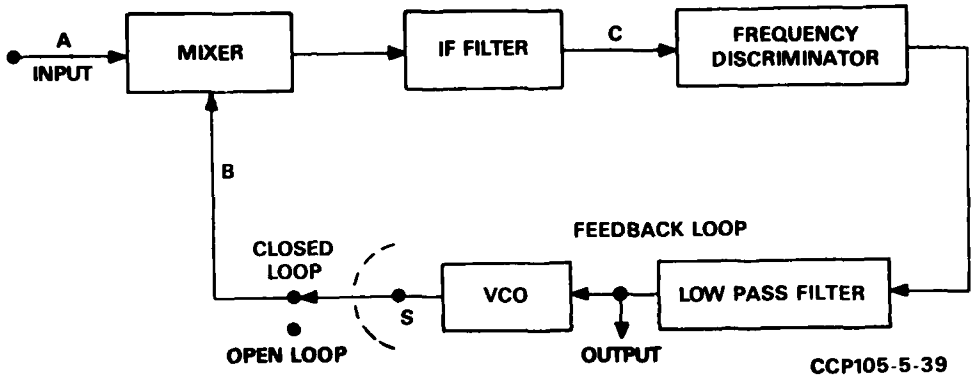


Figure 8-1. Block diagram of fm feedback demodulation system.

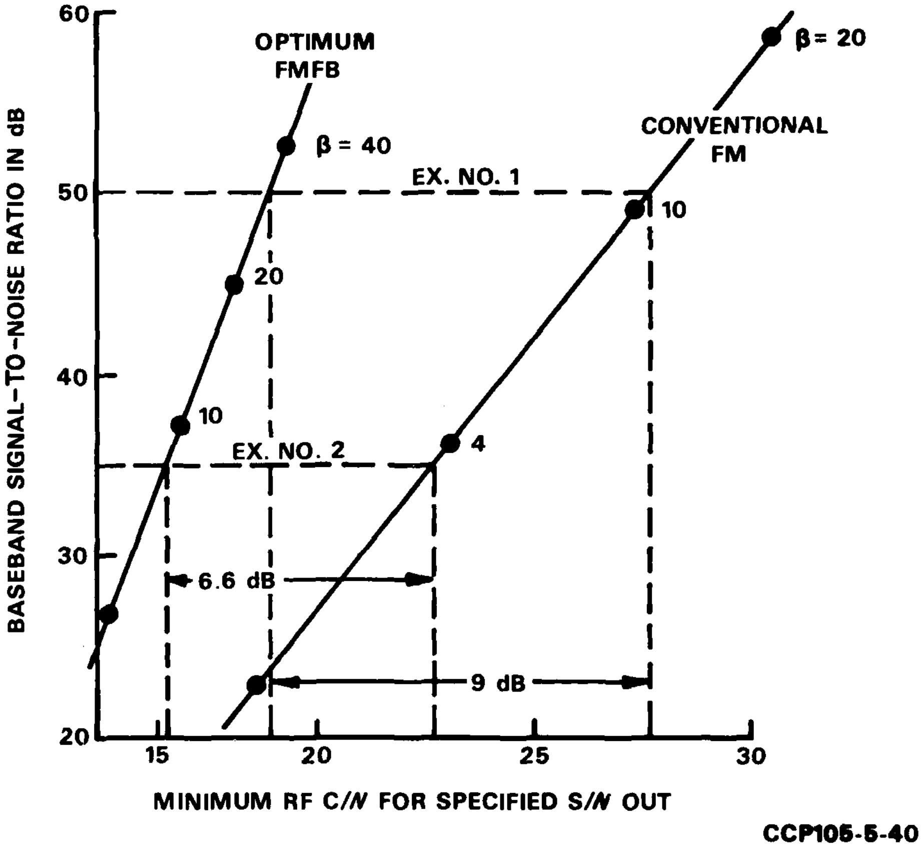


Figure 8-2. Graph of rf C/N threshold in dB $(C/N)_T$.

is directly proportional to the modulation index, the rf bandwidth can be computed. For these two examples, the modulation indexes are between 2 and 3 times larger for fmfb; hence, the rf bandwidth for optimum fmfb is 2 to 3 times larger than for conventional fm.

(6) Noise operates on an angle modulated signal in such a way that only some of the noise energy angle modulates the carrier and only the angle modulation portion of the noise is reduced by the feedback. For moderately high input angle modulation indexes, requiring only a moderate amount of loop feedback gain, networks which ignore the amplitude modulation portion of noise are used. As the modulation index of the input signal is increased, the amplitude portion of the noise modulation causes the threshold to rise to some intermediate value between the 6 dB lower angle modulation limit and the high threshold it would have had without feedback.

b. *Phase-Lock Loop Demodulation.*

(1) The phll system, like the fmfb system, uses a signal-tracking filter and responds only to a narrow band centered about the instantaneous carrier frequency as it sweeps through its deviation. Since at any instant the incoming signal is but a single frequency, the detection bandwidth required is limited to that needed for modulation information, as opposed to the wide bandwidth required by conventional fm circuits. The phll-system generates a replica of the incoming carrier, as opposed to the fmfb system which provides only a narrow-band IF signal. Since the phase-lock demodulator detects signals that would normally be below the threshold level of standard demodulators, it is extremely useful in satellite earth station receiving equipment.

(2) Figure 8-3 illustrates a typical phll threshold extension system. The circuits are conventional until the signal reaches the phase-lock demodulator. The phase-lock demodulator consists of a phase detector, low-pass filter, and a vco. These circuits form a feedback loop which continuously adjusts the frequency of the vco to that of the incoming fm carrier. In operation, the phase detector compares the instantaneous phase of the incoming frequency-modulated carrier with the phase of the locally generated replica of the modulated signal obtained from the vco. The output of the phase detector is an error voltage proportional to the phase difference of the two signals. After passing through the low-pass filter the error voltage is applied as feedback control to the vco. Thus, the vco is locked in phase with the incoming carrier. Since the output of the phase detector must adjust the instantaneous frequency of the vco to the deviating incoming carrier to maintain phase-lock, the vco control voltage is a replica of the modulation and, thus, represents the demodulated receiver output.

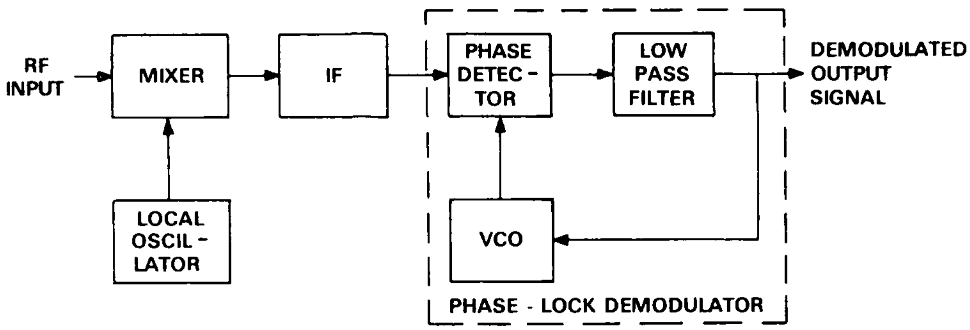


Figure 8-3. Phll threshold extension system block diagram.

(3) A minor disadvantage is that the narrow feedback loop bandwidth leads to sluggishness. In the phase-lock demodulator the loop must control the phase difference (error) to less than nonambiguous limits (±90 degrees) of the phase detector or the output signal is subject to severe distortion. This requires an increase in the loop tracking bandwidth and a cost in the threshold extension range.

(4) To analyze the effects of the narrow loop bandwidth in a little more detail, consider the error-sensing device (phase detector). The phase detector is simply a mixer, a nonlinear device driven by the sum of two signals, the vco output and the received signal. The output is passed through the low-pass filter and provides both the feedback voltage and the baseband output. The important point is that the phase detector output is unambiguous only if the phase difference between the vco signal and the carrier is known to be within the limits of ±90 degrees; it is impossible to deduce from the output if this is, in effect, a reality or if the vco has "slipped" the carrier input by an integral multiple of 180 degrees. In operation, if the response is too sluggish, tracking will be imperfect. Even in the absence of noise, instances will occur when the vco phase differs from the carrier signal by more than ±90 degrees and the two signals slip phase by an integral multiple of 180 degrees. When this occurs the output waveform becomes a "folded" version of the modulation and the intermodulation distortion becomes enormous because of the gross nonlinearity introduced.

(5) To describe the nature of the problem a little further, consider the following: The phase detectors in many phlls measure the phase difference between the input wave and that of the vco in a linear range somewhat less than ±90 degrees. This is also the range of phase difference that provides negative feedback control. If the phase of one of the waves changes abruptly, causing the phase

difference to fall somewhat outside the range of ±90 degrees, the loop becomes regenerative. The vco will sweep through the regenerative range as rapidly as the loop time constant permits until it has reached a new degenerative region of phase difference into the regenerative (and cycle skipping) range; whereupon, a peak noise impulse appears abruptly in the phase detector output.

(6) In summary, when phase reference is lost momentarily, further complications may arise in some phll systems. Lock is apt to be lost and not recovered properly. In particular, a state of anomalous lock (false lock) can occur and though it is popularly thought to be caused by a trivial malfunction, it is in fact, an inherent characteristic of some phll systems. Figure 8-4 illustrates a method of correcting this condition. The principle of operation is simple. The output of a conventional discriminator is applied with the vco control voltage to a control circuit. In normal operation, the vco control voltage keeps the error phase difference within the limits of ±90 degrees. However, if the vco phase slips, the conventional discriminator, which is affected only by the incoming signal, momentarily assumes control and forces the vco to within the ±90-degree phase-difference boundary, the vco is then controlled by the feedback loop in a normal manner.

(7) However, a trade off between threshold extension range and stability is required. The loop natural frequency may be made large relative to the top channel frequency and ensure a peak phase difference that is small relative to ±90 degrees. This provides relatively low intermodulation distortion but high threshold. In the opposite direction, the loop natural frequency may be picked near or even below the top channel to produce a low threshold, but at the expense of increased intermodulation distortion. Demodulators currently in use appear to use a natural frequency approximately 2.5 to 3 times the top baseband frequency, a penalty of about 5 dB in noise.

(8) As shown in figure 8-3, typical phase-lock fm demodulators use the oscillator control voltage as the baseband source. This leads to an increase in intermodulation distortion, as compared with conventional receivers. Here, however, the problem is more severe since the phase detector characteristic generally is far from linear and is dependent upon the signal and S/N ratio. Far better results (as much as 30 dB improvement) in intermodulation performance have been measured when the vco is discriminated to provide the output signal.

(9) Mention should be made of the impulse-noise effects that are present in the output when the input signal drops below the extended threshold of the phll system. In essence, the phase-lock demodulator appears to involve two thresholds. First, the IF limiter threshold (a soft or gradual threshold), the point at which the IF input to the phase detector begins to drop in amplitude and,

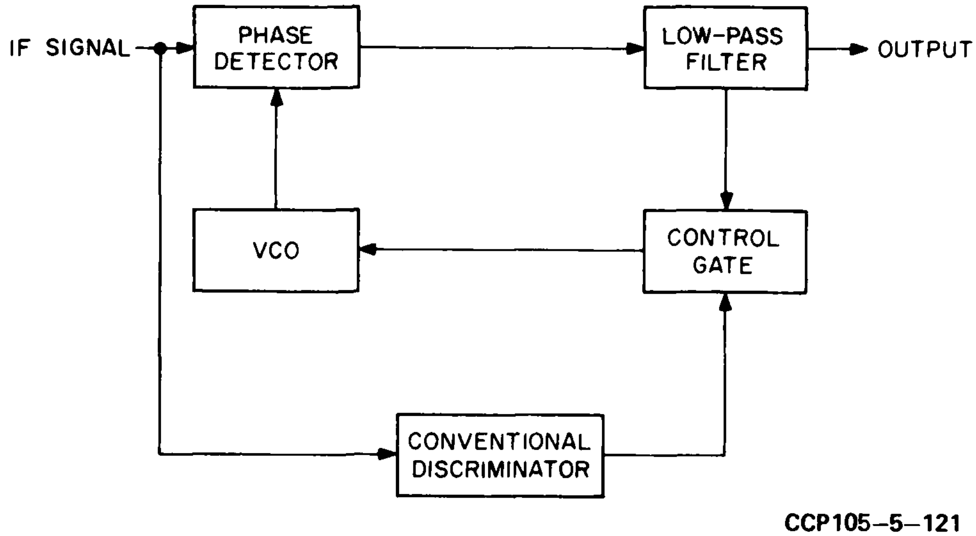


Figure 8-4. Phase detector with control circuit to prevent false lock block diagram.

in consequence, to induce changes in the loop parameters; and second, the point at which phase-jump effects are observable in the output point of figure 8-3. Phase-jump effects will not be observed in the output of the vco when the output of the vco is demodulated, except perhaps for an attenuated version of those observed at the output of figure 8-3.

c. *Parametric Amplifier.*

(1) A paramp works through modulation of energy storage devices (reactances), the reactance of which is made to vary as a function of time. In electrical systems, the reactance can be a capacitance, which stores potential energy; it can be an inductance, which stores kinetic energy. Or it can be a beam of electrons, which has both potential and kinetic energy. The following paragraph explains how this time-varying reactance can amplify an electrical signal.

(2) The classic example of a paramp is the LC-resonant circuit in which the capacitor acts as a storage device. To explain the amplifying action, first consider part A of figure 8-5. Assume that at some instant the capacitor C is charged to some specific value Q and the switch is closed. If there is no resistance in the circuit, the circuit will oscillate, causing the voltage to swing about the zero point, as shown in part B of figure 8-5, at a frequency that depends on the resonant

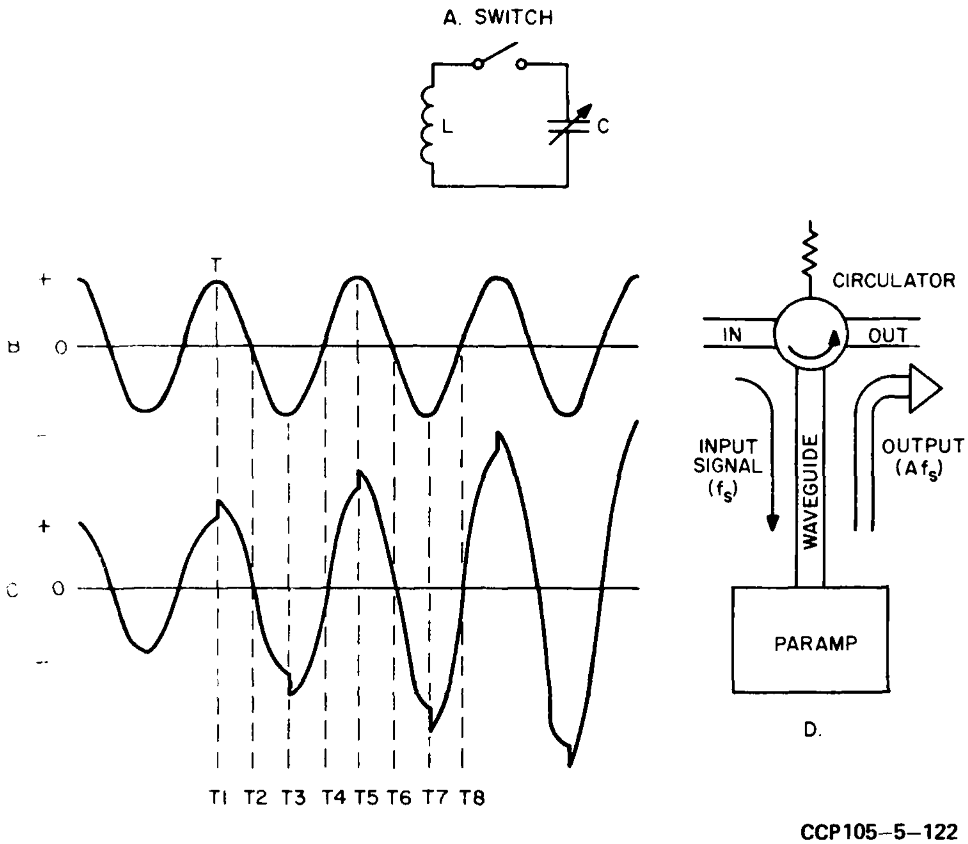


Figure 8-5. Paramp, principles of operation, functional diagram.

frequency of the LC circuit. If the plates of the capacitor could be separated at time T1 (part B of fig. 8-5) and pushed together at T2, separated again at T3; etc., voltage amplification would occur as illustrated in part C of figure 8-5.

(3) This action can also be expressed mathematically as V = Q/C. If the denominator (C) of the right side of the equation is decreased (plates pulled apart), V will increase, since Q is assumed to remain constant during the pulling action. If C is increased (plates pushed back together) when V equals zero, the charge (Q) on the capacitor will be zero and no change in the voltage will occur; i.e., power will neither be added to or taken from the LC tank circuit. Thus, the overall result and the pumping action is to increase the voltage or power.

(4) The paramp described is a one-port negative resistance amplifier. The paramp is connected through a short piece of waveguide to a circulator, as shown in part D of figure 8-5. Briefly, the paramp can be thought of as a device having a reflection coefficient greater than one. Thus, a signal entering the paramp is immediately reflected, but at a greater amplitude. The circulator is a ferrite device by which, due to certain properties of ferrites, the input signal (f_s) is separated from the amplified output signal (Af_s). A more detailed description of the mechanism of a paramp is presented below.

d. *Theory of Parametric Amplification.*

(1) It was explained briefly in the previous paragraphs how a single frequency signal could be amplified, using a time-varying capacitance, once it is injected into the resonant circuit. However, a number of problems must be solved before this principle can be applied in a practical manner. These are as follows:

(a) The signal to be amplified is complex instead of a single sinusoidal signal as depicted. In this case, it is a very high frequency (7.5 GHz) varying over a 50-MHz range. Thus, instead of the capacitor plates opening and closing at a steady rate, the change must take place at some varying frequency and keep a precise instantaneous phase relationship with a varying frequency.

(b) The amplifier to be described is a one-port device; i.e., the output is taken from the same terminals to which the input is applied. These two signals, both appearing at the same terminals, must be separated.

(c) The frequency at which the capacitor plates are opened and closed must be at least twice the signal frequency, or 15 GHz; this obviously can not be done with any mechanical device. In the actual case, the capacitance is varied at a much higher rate than twice the signal frequency or approximately 26.5 GHz. The pump frequency chosen affects the gain, bandwidth, and noise characteristics. This is important but need not be discussed here. Referring for a moment to the simplified explanation (para 8-2c(2)), it appears that, for the device to work, the pump frequency must be exactly twice the signal frequency and properly phased with the signal frequency. Now, we seem to have destroyed this important consideration by using a pump frequency that has no apparent relationship with the signal frequency. How is this relationship attained and how is the capacitance varied at the very high rate?

(d) Other problems such as the filtering out of unwanted signals, amplifier stability, and noise also arise, but an understanding of these problems is unnecessary to the general explanation and; hence, will be ignored.

(2) A schematic representation showing the principal elements of an operating paramp (figure 8-6) will illustrate how these elements combine to solve these problems. Referring to figure 8-6, the paramp consists of the following elements:

(a) The circulator (paragraph 8-2e) is a ferrite device. The laws governing the conductivity and dielectric constant of ferrites are not well understood. For the present; therefore, it suffices to say that as a result of certain properties of ferrites, the input signal (f_s) is separated from the output signal (Af_s).

(b) The signal tank circuit, L_1C_1, idler tank circuit L_2C_2, and pump tank circuit, $L_3C_3C_4$, are representative of the hi-Q resonant cavities of the actual paramp. The idler tank is tunable by a knob (part A of fig. 8-6) and is the primary means for setting the paramp frequency. The pump frequency is made much higher than twice the signal frequency f_s; actually f_p is made equal to $f_s + f_1$. These frequencies are chosen to provide the best gain-bandwidth product and noise factor.

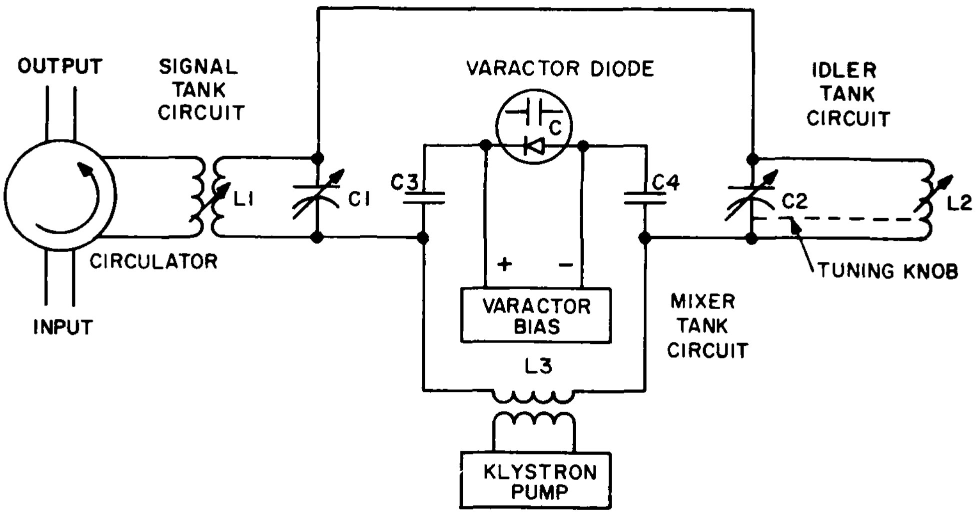
(c) The kystron pump provides pump power to cause the cyclic change in the capacitance of the varactor diode. This effect is equivalent to the push-pull action on the plates of a capacitor.

(d) The varactor diode is a low-loss silicon diode, whose junction capacity is a function of the voltages applied to its terminals. Because of the low LC product associated with the junction and terminals, the capacitance can be made to vary at a super high frequency. Because of the loss in the diode, it can be represented by a small resistor (which accounts for the loss) in series with a variable capacitor. This resistive component introduces noise as a function of its temperature. The magnitude of this noise is decreased by cooling the diode to cryogenic temperatures (paragraph 8-2f).

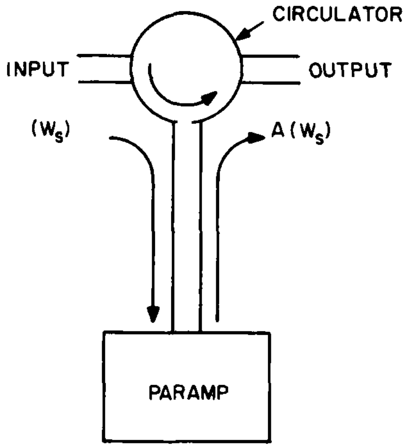
(e) The varactor diode is biased to a specified point on its operating curve to decrease the amount of diode current, which would produce noise due to "shot effect" and also to provide the proper operating point for pump operations.

(3) If the above statements regarding the functions performed by the elements are acceptable, then two of the listed problems may be considered as solved; namely, how both the input and output signals, appearing at the same terminals, can be separated; and how the capacitor can be varied at the very high frequency required. The major problem remaining is to show how this variation in the capacitor can be kept in phase and synchrony with the varying incoming signal. This will be accomplished with the aid of part A of figure 8-6.

(4) The incoming signal from the feed horn is passed through the circulator to the paramp, where it is mixed with the pump signal. This mixing action is



A. PARAMP, SCHEMATIC DIAGRAM



B. PARAMP, BLOCK DIAGRAM

CCP105–5–123

Figure 8-6. Paramp schematic and functional diagrams.

somewhat analogous to that of a superheterodyne receiver in which a local oscillator signal is mixed with the incoming signal to produce both their primary signals in addition to their sum and differences, and harmonics.

(5) It will be helpful to discuss one point, which may cause confusion. This results from the necessity of talking about mixing frequencies to produce additional frequencies, and the apparent similarity of this action to that of a superheterodyne receiver. In a superheterodyne receiver, the incoming signal is mixed with a local oscillator to produce an IF. The IF signal is passed through a number of relatively narrow-band amplifier stages. In this case, it is relatively unimportant what the conversion gain of the mixing stage is since most of the amplification takes place in the IF stages. Confusion may arise from the fact that in the superheterodyne receiver amplification takes place stage-by-stage as the signal progresses; amplification in a paramp, on the other hand, takes place in a more simultaneous manner. In a paramp, the action is more like that shown in part B of figure 8-6. Here the paramp appears as a "black box" connected through a short piece of waveguide to a circulator. As far as the results that could be seen at the circulator are concerned, the paramp acts like a reflector having a reflection coefficient greater than one. A very weak signal going toward the paramp and the same signal being reflected back (highly amplified) will appear at the circulator output.

(6) In the actual paramp, the sum frequency is eliminated so that there will exist in the paramp only:

Input signal	f_s	$=$ 7.5 GHz
Pump signal	f_p	$=$ 26.5 GHz
Difference frequency	f_i	$= f_p - f_s =$ 19.0 GHz

(7) These signals which appear in the signal tank, pump tank, and idler tank circuits are represented graphically in parts A, B, and C, respectively, of figure 8-7. The fact that the idler frequency f_i exists is substantiated in paragraph (8).

(8) Under the influence of the pump signal, the capacitance of the time varying varactor (capacitor) is given at any moment by $C = C_o + C_1 \cos \omega_p t$ and the voltage appearing across the varactor by $V + V_o \cos (\omega_s t + \phi)$. The instantaneous charge across the capacitor is the product of the varying voltage and capacitance. By expanding this product in a Fourier series, it can be shown that the composite periodic waveform is made up of the two primary frequencies and various harmonics (each at some specified amplitude) of the sum and difference of the primary frequencies. Among these will be the difference frequency, $\omega_p - \omega_s = \omega_1$, which is the desired idler frequency.

(9) The interaction between two signals of different frequencies can be of two possible types: (1) Direct superposition of the two signals (which amounts

to adding the ordinates point-by-point) followed by rectification of the combined signals and (2) systems in which one of the signals is modulated by the other.

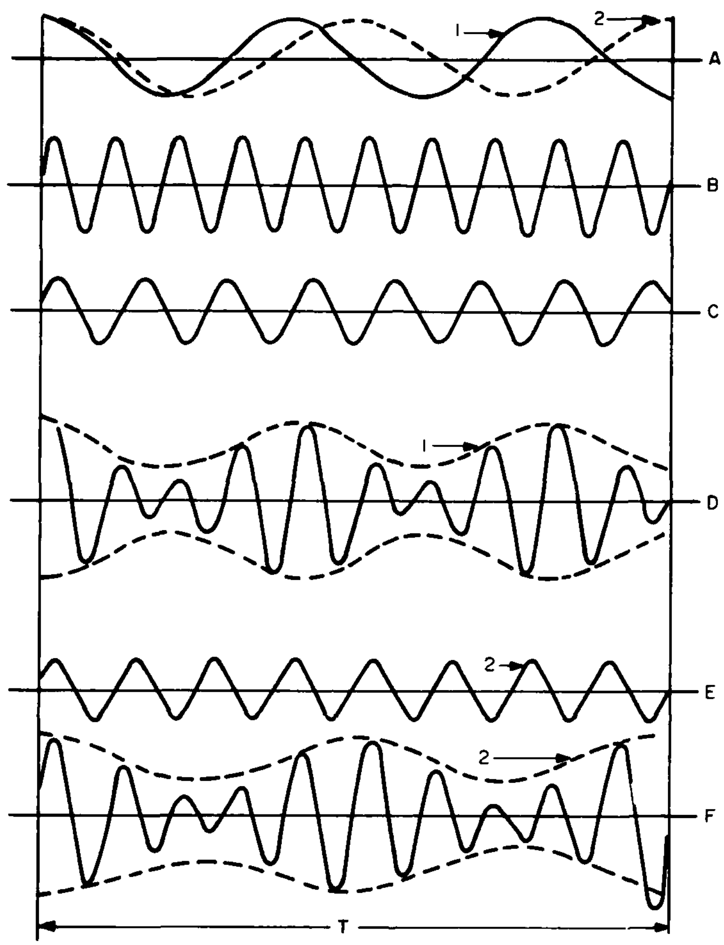
(10) In the first case, when two alternating voltages of different frequencies (shown in parts B and C of fig. 8-7) are superimposed, the result is as shown in part D. As can be seen, the two waves alternately add and subtract as the higher frequency wave advances (relative phase advance). A study of figure 8-7 shows that the resulting envelope goes through one cycle of amplitude variation in a time interval required for the higher frequency to gain one cycle. The frequency of the envelope is, therefore, the difference between the higher and lower superimposed signals.

(11) In the second case, if one of the waves in figure 8-7 is used to modulate the other, the result is the production of sidebands. One of these sidebands is the difference frequency, so that the end result is exactly as described.

(12) Thus far it has been shown that, under the given conditions, three frequencies of interest appear in the paramp. A closer look will show that the voltage across the varactor (capacitor) is not simply that of the input signal, but is a combination of input signal and the idler signal. The same approach could be used as before to show that the idler signal combines with the pump signal to produce signals at other frequencies. Among these new signals will be a difference frequency signal equal in frequency and phase to the input signal but greater in amplitude. Thus, for all practical purposes, the input signal and an amplified replica of the input signal appear across the varactor. Generally, what appears to happen is that as the incoming signal starts positive, the varactor sees the combined results of the signal and a signal at the same frequency produced by the idler (and pump), which causes the voltage across the varactor to vary as the instantaneous sum of the two. Hence, the output appears as an amplified replica of the input.

(13) From the above discussion, it might be appropriately asked, "How does the idler signal become synchronized with the time varying input signal to maintain the proper relationship between the input signal and the pump signal?" To answer this, refer again to waveforms A, B, C, and D of figure 8-7.

(14) In the time period T input signal A goes through two and one-half cycles, pump signal B goes through 10 cycles and idler frequency C through 7.5 cycles. Thus, the idler frequency equals 10 − 2.5 = 7.5 cycles for the time period T. Now suppose the frequency of the input signals changes as indicated by the dashed lines in A. This results in the new idler frequency shown in E. When the pump signal B is combined with the new idler signal E the result is as shown in F. When the input signal decreased to two cycles for the time period T, the idler signal increased to eight cycles, maintaining the equality between input signal, pump



CCP105–5–124

Note: For the difference frequency to show clearly when diagrammed in this manner, the ratio between the two frequencies must be in the order of 3 to 4. (The lower the ratio, the more apparent the result.) Thus, in an attempt to show the production of the difference frequency using the paramp input signal (7.5 GHz) and the pump frequency (26.5 GHz), the result will be disappointing because of the high (approximately 3.5 to 1) ratio. This relationship between input signal and pump signal does exist, however, so that waveform C can be considered as having been developed from waveforms A and B.

Figure 8-7. Paramp waveforms.

signal, and idler signal. In general, as the input frequency changes causing a shift in phase as compared to some reference, the idler also shifts in phase (frequency) a corresponding amount. Thus, phase coherence always exists between the idler, pump, and signal frequencies.

(15) One other important problem remains to be solved regarding phase and synchrony. The previous paragraph explains how synchrony is achieved, but it does not show how the proper phase relationship is attained. As mentioned previously, with regard to phase and synchrony, in order to have energy transferred from the pump to the signal frequency, the capacitance must increase when the applied voltage is maximum and must decrease when the voltage is zero.

(16) There is no simple way of showing this except by mathematics. As mentioned previously, both the input signal f_s and idler signal f_i appear across the varactor. The instantaneous value of these two voltages separately can be represented as $v_s = \sin 2\pi f_s t$ and $v_i = \sin 2\pi f_i t$. The instantaneous value of their sum is $v_t = \sin 2\pi f_i t + \sin 2\pi f_s t$. By referring to a table of trigonometric identities, one obtains an equivalent expression $v_t = 2 \sin \pi(f_i + f_s)t \cos \pi(f_i - f_s)t$. The term $(f_i + f_s)$ is exactly equal to the pump frequency. When the term $\sin \pi(f_i + f_s)t$ goes through zero, the voltage across the varactor also goes to zero. This occurs at just one-half the pump frequency. Stated in another way, the varactor is being pumped at just twice the rate at which the voltage across the capacitor goes through zero, which is the correct relationship between varactor voltage and variation in capacitance to produce energy transfer from the pump to the signal.

e. *Circulator/Isolator Analysis.*

(1) Physically, a circulator (fig. 8-8) consists of three Y-connected strip-line conductors, sandwiched between two layers of ferrite material, in a steady magnetic field. The rf energy entering the signal input port of the circulator, travels in both clockwise and counterclockwise directions around the center of the circulator to arrive at the output signal port. Due to the unilateral characteristics of the ferrite material, the lengths of the paths (in wavelengths) traveled around the circulator differ for energy traveling in the two possible directions.

(2) As shown in figure 8-8, energy traveling the path from point A to point B experiences a 120-degree phase shift, as does the energy traveling from points A to C to B. Since the phase shifts are identical, energy from the two paths will add at point B, and the resulting signal becomes the output signal from the circulator. Following the paths energy would take going from point A to point C, and adding the indicated phase shifts, it can be seen that energy arriving at point C will be 180 degrees out of phase; consequently, practically none of the energy entering the input port is dissipated in the 50-ohm termination at point C, and all of the input signal is fed out of the circulator at the output port.

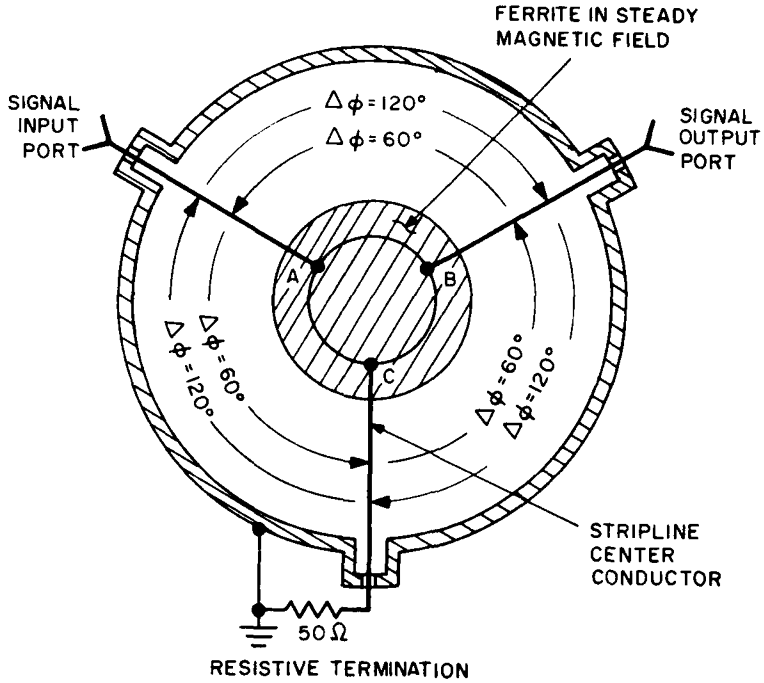


Figure 8-8. Circulator/isolator signal paths.

(3) Suppose now that the reflected energy feeds back into the circulator at the output port and tries to follow the two paths around the circulator to the input port. In this case, energy from point B to point A receives a 60-degree phase shift by the counterclockwise path and a 240-degree phase shift by the clockwise path. Hence, energy arriving at point A is 180 degrees out of phase and cancels. At point C, reflected energy (that is, energy entering the circulator at the output port) is in phase and is dissipated in the termination at point C.

(4) In general, a circulator has very low resistance in the forward direction; this is on the order of 0.02 to 0.05 dB. A very high impedance appears in the reverse direction.

f. *Parametric Amplifier Cooling.*

(1) The paramp is cooled in order to keep thermal noise to a minimum (cooling is accomplished by a cryogenic system using helium (the Cryodyne helium

refrigerator)). Operation is based on a principle similar to that used in a household refrigerator - namely, that most gases cool when they are expanded. The low temperatures that are obtainable with the Cryodyne refrigerator result from the use of: (1) An efficient thermodynamic cycle, (2) special heat exchangers, (3) two stages of refrigeration, and (4) a refrigerant (helium) that remains fluid even at temperatures approaching absolute zero.

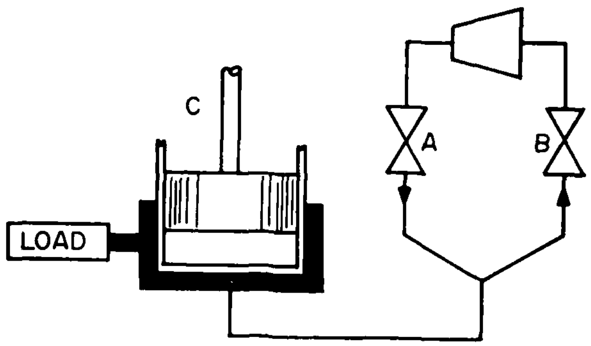
(2) The flow of helium in the refrigerator is cyclic. The sequence of operations is best explained by beginning with an elementary cooling circuit consisting of a single cylinder and piston as shown in figure 8-9. A source of compressed gas (a compressor, in this case) is connected to the botton of cylinder C through inlet valve A. Valve B is in the exhaust line leading to the low-pressure side of the compressor. Assume that the piston is at the bottom of the cylinder; valve B is closed and valve A is opened. The piston is caused to move upward and the cylinder fills with compressed gas. When valve A is closed and valve B is opened, the gas expands into the low-pressure discharge line and cools. The resulting temperature gradient across the cylinder wall causes heat to flow from the load into the cylinder. As a result, the gas warms to its original temperature. The piston is lowered, displacing the remaining gas into the exhaust line, and the cycle is complete.

(3) Such an elementary system, while workable, would be very limited in the minimum temperatures that it could achieve. To operate the cycle in a lower range of temperatures we must cool the incoming gas with the exhaust gas before the incoming gas reaches the cylinder. This is accomplished in the Cryodyne helium refrigerator by a device that extracts heat from the incoming gas, stores it, and then releases it to the exhaust stream. The device (fig. 8-10) is a periodic-flow heat exchanger, or regenerator. The regenerator contains a single flow path, packed with a matrix of material that readily accepts heat from a warmer fluid and yields it to a cooler one. Once steady-state operation is established, the matrix maintains a temperature between that of the inlet and the exhaust gases at each point along its length.

(4) Operating under steady-state conditions, a system of this type exhibits the characteristic temperature profile shown in figure 8-11. The steps of the cycle are as follows (numbers in parentheses refer to the applicable portions of figure 8-11):

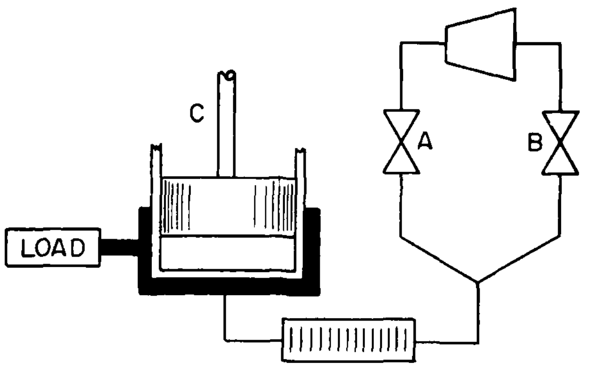
(a) With the piston at the bottom of its stroke, compressed gas enters through valve A (fig. 8-12) at room temperature (1).

(b) As the piston rises, the gas passes through the regenerator. The matrix absorbs heat from the gas (warming from 3 to 4), and the gas cools (2).



CCP105–5–126

Figure 8-9. Elementary cooling circuit functional diagram.



CCP105–5–127

Figure 8-10. Cooling circuit with regenerator functional diagram.

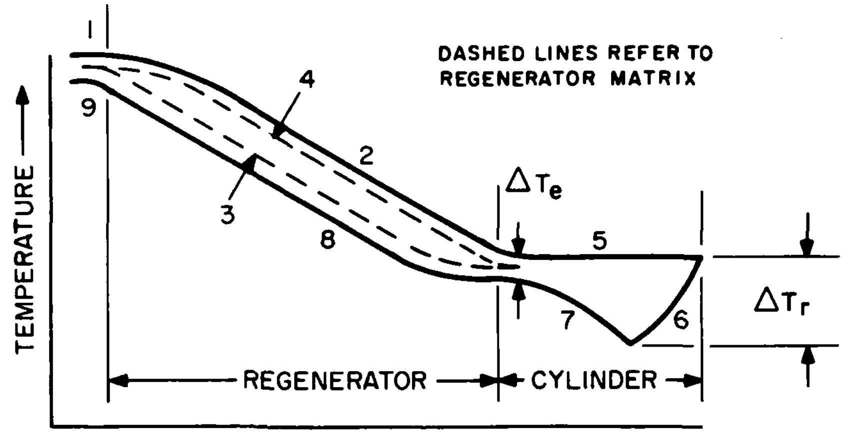


Figure 8-11. Temperature profile of single-stage cryodyne refrigerator.

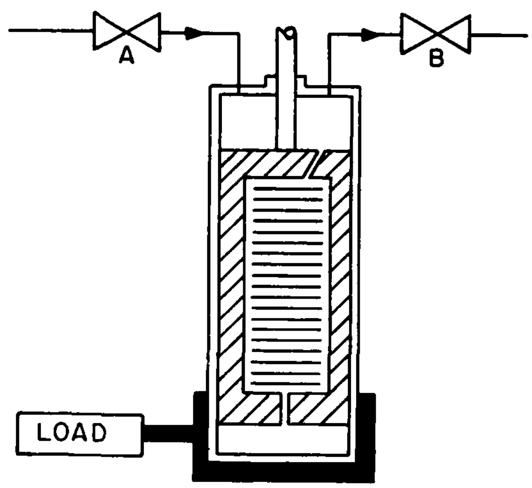


Figure 8-12. Improved single-stage regenerator functional diagram.

(c) Still at inlet pressure, the cooled gas fills the space beneath the piston. The gas temperature at this point (5) is about the same as that of the load.

(d) Valve A closes and exhaust valve B opens (fig. 8-12). The gas now expands into the low-pressure discharge line and cools further (6). This temperature drop (ΔT_r) is responsible for the refrigerating effect.

(e) Heat flows from the load through the cylinder walls, warming the gas to a temperature slightly (ΔT_e) below that at which it entered the cylinder (7).

(f) As the gas passes through the regenerator it warms up (8) as it receives heat from the matrix and the matrix is cooled (4 to 3).

(g) The piston descends, pushing the remaining cold gas out of the cylinder and through the regenerator. However, because the regenerator is not 100 percent efficient, there is always a temperature difference between the gas and the matrix; thus, at any point shown in the diagram, the exhaust gas remains slightly cooler than the inlet gas.

(h) The low-pressure gas leaves through valve B (fig. 8-12) at approximately room temperature (9).

(5) In figure 8-10, the piston would require a pressure seal and would have to be designed to withstand unbalanced forces. A more practical adaptation of this cycle is illustrated in figure 8-12. This system uses a double-ended cylinder and an elongated piston made from a material of low thermal conductivity. Since the pressures above and below the piston are substantially equal, the piston needs no pressure seal. The piston is more correctly called a displacer, because it merely displaces gas from one end of the cylinder to the other; no mechanical work is introduced (except in overcoming friction) and, thus, the system is said to use a "no-work" cycle. The regenerator is placed inside the displacer to avoid unnecessary piping and heat losses.

(6) The regenerator matrix material must have considerable heat capacity to enable it to perform its function in the system. Unfortunately, solids lose heat capacity as the refrigeration temperature is lowered. This phenomenon not only impairs the efficiency of the regenerator, but also imposes a lower limit on the temperature that can be attained. The improved single-stage regenerator (fig. 8-12) may achieve a temperature as low as 30 K, but useful refrigeration may be limited to 60 K.

(7) The refrigeration available from a given work input; e.g., a given size of compressor, is a function of the refrigeration temperature. As this temperature

is lowered the gross refrigeration available falls. The net refrigeration falls more sharply because the rate of heat loss increases. Therefore, to obtain useful refrigeration in the region below 60 K, it is advantageous from the standpoint of the work input or compressor size to use a cascaded, two-stage system as shown in figure 8-13. In this arrangement, refrigeration (80 K) is provided at the first stage to intercept most of the conduction and radiation leaks from room temperature. This stage also provides for the principal regenerator loss. Refrigeration can be provided in the second stage at temperatures down to 15 K.

g. *Varactor Diode.*

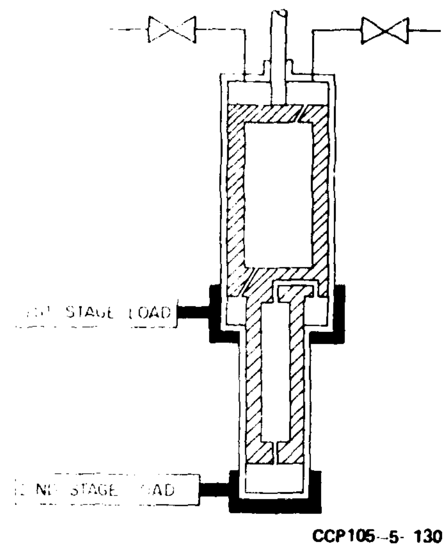
(1) The explanation of parametric amplification has assumed a capacitance that is, in some way, varied in amount at the pump frequency. It is obviously impractical to vary the capacitance at the pump frequency by any mechanical means. However, the varactor diode exhibits a quality that enables this to be accomplished electronically. When biased in the reversed direction, all semiconductor diodes exhibit a variation in capacitance as the reverse voltage is increased and decreased. This variation can be explained as follows. The junction of the diode has P type semiconductor material on one side and N type semiconductor material on the other side. The current carriers in the N type material are electrons, while the current carriers in the P type material are holes. These holes are the equivalent of mobile positive charges and are discussed in paragraph 9-4.

(2) The application of a reverse voltage repels the electrons and the holes from the junction, leaving a depleted region in the neighborhood of the junction. The layer of electrons on one side of the junction and of holes on the other side can be considered the plates of a capacitor. Since variation of the reverse voltage causes the electrons and holes to move closer and farther away from the junction, the equivalent capacitance changes accordingly.

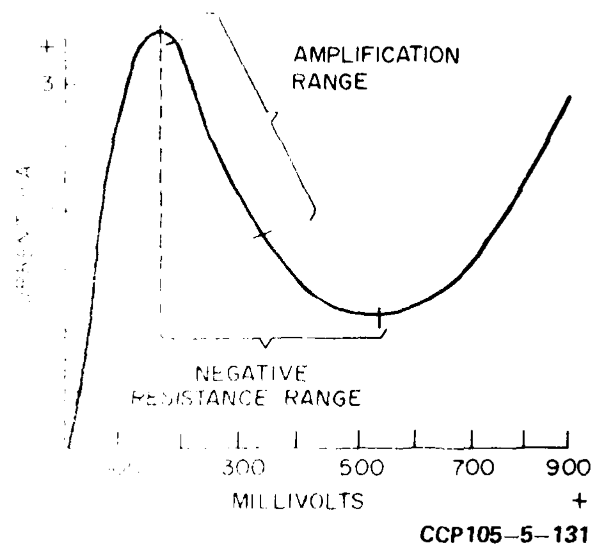
h. *Tunnel Diode.*

(1) The tunnel diode is ideally adapted for use in satellite communications since it can amplify at microwave frequencies (tunnel diodes operate at the speed of light), is resistant to nuclear radiation without additional hardening, and can operate in a wide temperature range.

(2) The tunnel diode amplifies by using the negative resistance characteristic (fig. 8-14) peculiar to the tunnel diode. Note the linear increase in current with application of positive voltage until the negative resistance range is reached. At this point, increasing voltage causes a current decrease; likewise a decrease in voltage causes an increase in current. Only the linear portion of the negative resistance



Two stage cryodyne refrigerator functional diagram.



Tunnel diode negative resistance characteristics curve.

range is used for amplification. A further increase in voltage (beyond the negative resistance range) causes a further increase in current.

8-3. TRANSMITTER SUBSYSTEM. The wideband, low-noise amplifiers presently used in satellite communications links are either klystrons or TWT's. These amplifiers are explained below.

a. *Kylstron.*

(1) The klystron tube consists of a cathode and heater, one or more resonant cavities, a drift tube, a collector, and some means of producing a magnetic field around the device. For the purpose of this discussion a three-cavity klystron is considered. The cross-sectional view of such a tube is depicted in figure 8-15.

(2) The electrons produced at the cathode by the heater (not shown) are drawn down the drift tube toward the collector by the difference in potential between the cathode and collector (anode). (The drift tube, cavities, and collector are all at ground potential in order to eliminate the hazard associated with tuning the cavities.) A large potential difference between the cathode and collector causes the electrons to be accelerated down the axis of the tube toward the anode. These electrons are concentrated into a thin pencil beam by the electromagnetic focusing coil surrounding the tube. The signal to be amplified is fed into the resonant cavity and, during the transit time of the electrons in the beam, acts upon the rapidly moving electrons to further accelerate or decelerate them depending on the polarity of the rf field in the cavity. The decelerated electrons are overtaken by the accelerated electrons, and bunching of electrons occurs. This bunching takes place at the same rate as the signal voltage, but at a much higher power level. The purpose of the drift tube is to allow the electrons time to complete the bunching process. Introduction of more cavities and a longer drift tube causes the electron beam to be acted upon by further bunching processes, thereby extracting still more power from the beam. Finally, the greatly amplified signal is extracted from the final cavity by a suitable stub.

(3) From the foregoing it might be supposed that, if the drift tube were made long enough, maximum bunching or velocity modulation would occur. This is true only for a certain length limit, for eventually the electrons would strike the walls of the drift tube and serve no useful purpose. In fact, keeping the electrons focused down the drift tube is an extremely difficult task. One hundred percent effectiveness has not been achieved and the electrons which do strike the walls exit through the body of the tube to ground. The current which results is called body current.

(4) For specific applications, klystrons differ in the size of the cavity, the transit time through the drift tube, the voltage difference from cathode to anode, and the frequency of operation.

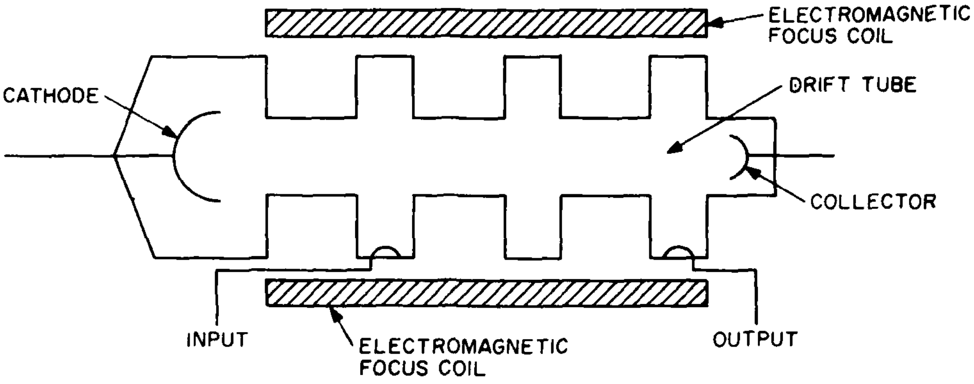
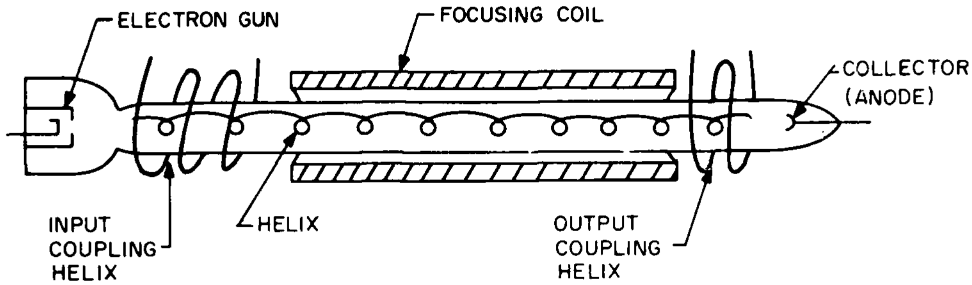


Figure 8-15. Klystron, cross sectional diagram.

b. *Traveling-Wave Tube.*

(1) The TWT has an electron gun (cathode) and a collector (anode) as shown in figure 8-16. A helix, whose axis is the path of the electron beam, is placed about the path of the electron stream. The electron stream is caused to traverse this path from the gun down the axis of the surrounding helix to the collector anode. The entire assembly is sealed inside an evacuated glass envelope. Another helix, which is used to couple the signal voltage into the inside helix, is placed outside the glass envelope at the cathode end. A similar helix, which is used to extract the greatly amplified signal, is placed at the anode end. A focusing coil is fitted closely around the inside helix, but outside the tube envelope. This coil is similar to the focusing coil used on a klystron and performs the same function of keeping the electrons packed tightly in a beam.

(2) The beam of electrons produced in the gun is projected toward the anode or collector. The speed of electrons is controlled by the difference in potential between the cathode and the collector. The electrons are kept in a tight beam shape by the focusing coil. The signal voltage is introduced into the coupling helix nearest the cathode and proceeds down the helix at the speed of light. These are termed fast waves. However, since the conductor is wound into a helix, the speed of propagation along the axis of the helix is some value less that the speed of light. Depending on the diameter of the helix and the number of turns per unit distance, the speed of propagation now becomes something on the order of 1/10 to 1/30 the speed of light. These are termed slow waves. The slow waves produce an electrostatic field which varies according to the signal voltage and interacts with the electrons in the beam in a manner similar to the klystron.



CCP105–5–133

Figure 8-16. TWT, cross sectional diagram.

(3) Acceleration and deceleration of the electrons take place and bunching occurs. The bunching occurs at an exponential rate as the process is continued down the length of the tube. Simultaneously, the signal on the helix increases in magnitude as it proceeds down the helix. The increase in signal energy is balanced exactly by the loss of kinetic energy in the electron beam. The basic difference, and a very important one, between the klystron and the TWT is that the klystron has only a very short time to act on the electron beam and; therefore, high electric fields are necessary for efficient operation. These are produced in a resonant cavity which is sharply tuned. Specifically, the interaction time is only that transit time in the cavities and, to an indirect degree, in the drift tube. In the TWT; however, the interaction time is that time required for the rf field to travel down the helix proper, which is a much longer physical distance than in the klystron.

(4) Due to the fact that no high impedance or resonant devices are employed in the TWT, it operates over a very wideband of frequencies. TWT's have been designed and built which are essentially linear over a range of 7 GHz. The only critical requirement is that the electron speed be approximately equal to the rate of propagation of the slow wave. The requirement for this is self-evident, for if the speed of the electrons were either much greater or much slower than the slow wave, no interaction between the two would take place. Once the speed of the electrons has been synchronized by adjusting the voltage difference between the cathode and anode little, if any, further adjustment is necessary. Amplification factors of 10^4 or 40 dB are common with a single TWT.

(5) TWT's exhibit a variety of characteristics, some of which are discussed briefly as follows:

(a) *Narrow band amplifiers.* TWT's have been constructed to operate over a narrow band. In these cases larger gains are exhibited; i.e., 60-70 dB.

(b) *Low-noise amplifiers.* Where exceptional gain was not a factor, TWT's have been constructed which have a noise figure as low as 6 dB. However, these operate at a relative low power level.

(c) *Special purpose amplifiers.* Sometimes an extremely flat (linear) amplifier is required, or one which has increasing gain with increasing frequency. These, too, have been constructed at the expense of some of the other features of the TWT.

8-4. MODEM TECHNIQUES.

a. Multiplexing is the process of simultaneously transmitting more than one independent information stream over a single link. An important part of many of these multiplexing systems is the modulation and demodulation section, the theory of which was discussed in paragraphs 7-8 and 7-9. This modulation and demodulation equipment is often packaged in common units and the term modem is often applied to these single packages.

b. Multiplexing can take many forms; however, frequency-division and time-division are the most common types. Because of this popularity, only these types will be discussed.

(1) *Time-division multiplex.*

(a) Time-division multiplex, because of its very nature, is most commonly used to multiplex digital information. Time-division multiplexing of analog information, such as a voice signal, is not impossible but is quite uncommon. Therefore, only time-division multiplexing of digital information will be discussed. In many digital transmissions, the information to be transmitted is contained in the amplitude or frequency of the waveform describing the intelligence to be transmitted, and the duration in time of the waveforms is of no particular significance. This offers the possibility of reducing the duration of the individual symbols transmitted, and transmitting additional information in the time intervals vacated. An example of this process is shown in figure 8-17.

(b) Part A of figure 8-17 represents a section of a digital message. Positive values represent ones and negative levels represent zeros. Ones and zeros are the symbols given to the two possible states in a binary transmission.

(c) Part B of figure 8-17 shows the same sequence in which the length of symbols has been cut in half. Part C represents a second independent digital message; part D, the same message with halved symbol length. If the message, represented by part D, is moved one unit in time and combined with the message in part B, the sequence in part E is the result. This message represents a composite of

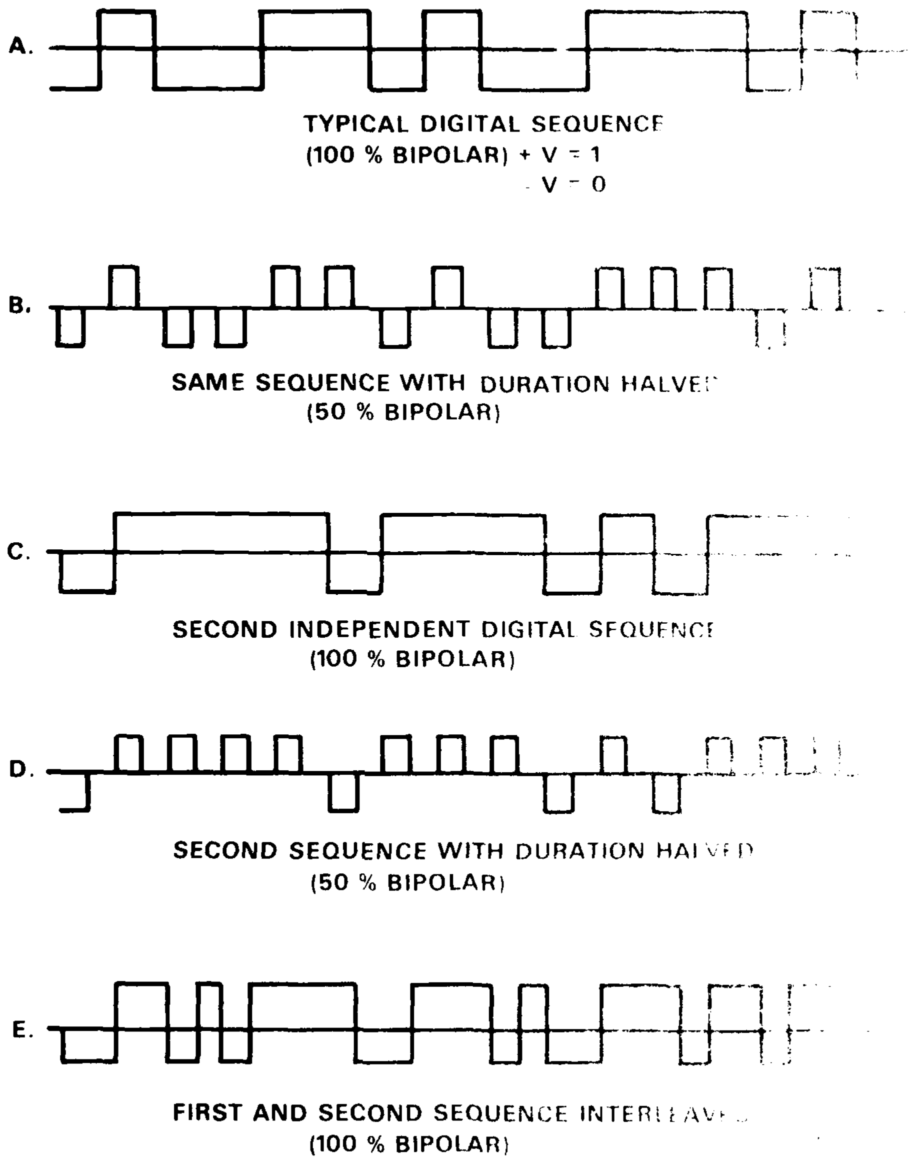


Figure 8-17. Time-division multiplex

all the information contained in both messages. This is the principle behind time-division multiplex. By taking smaller and smaller duration samples of the sequences, any number of digital signals can be multiplexed. One result of the multiplexing is that the reduction in individual symbol length requires a wider bandwidth for transmission. This expanding bandwidth requirement becomes a limit on the number of channels that can be multiplexed in this fashion.

(d) While the sequence in part E contains all the information in the sequences of both parts A and C, it is necessary to know which portion of the sequence in part E belongs to each message. In order to know this, the receiving station must be able to divide the message into time intervals corresponding to those used for transmission. For this reason, synchronization between the transmitter and receiver end of a link is required. As the individual symbols become shorter and shorter, synchronization becomes more difficult. This increasing demand also tends to limit the number of channels that can be time-division multiplexed. Figure 8-18 is a block diagram of a system that can be used to accomplish the interleaving of the individual symbols and the equipment required to interpret the composite sequence received.

(e) Figure 8-19 shows the waveforms involved in the operation of such a system. Parts A and B show two inputs which may be regarded as typical waveforms of the input to the four channels. These unsynchronized inputs must be given some definite time relationship before the time multiplexing scheme can be applied. This synchronizing is achieved by the channel synchronizer and input storage unit of each individual channel. The synchronizer, in turn, gets its timing information from the system clock. The action of the input storage and synchronizer can be observed in the input to gates shown in figure 8-19, parts C and D. The beginning space and all subsequent symbols in the inputs, parts C and D, are held in storage since they occurred at a random time relative to the clock pulse. The clock pulse will trigger the channel synchronizer and release the symbols from the input storage. The beginnings of marks - spaces in both channels will now be simultaneous at the input to the gates. The gate in this circuit is known as an "and" gate and gives a positive output if both the input and clock pulse are positive, and a negative output if either is negative.

(f) Parts G and H of figure 8-19 show the output of gates 1 and 2. The output of all such gates is fed to a summing network and the composite multiplexed signal results.

(g) This summed output is ready for transmission on a wire line, or it may be used to modulate a carrier for transmission. In the case of communication via a satellite, this waveform would be used to modulate a microwave carrier.

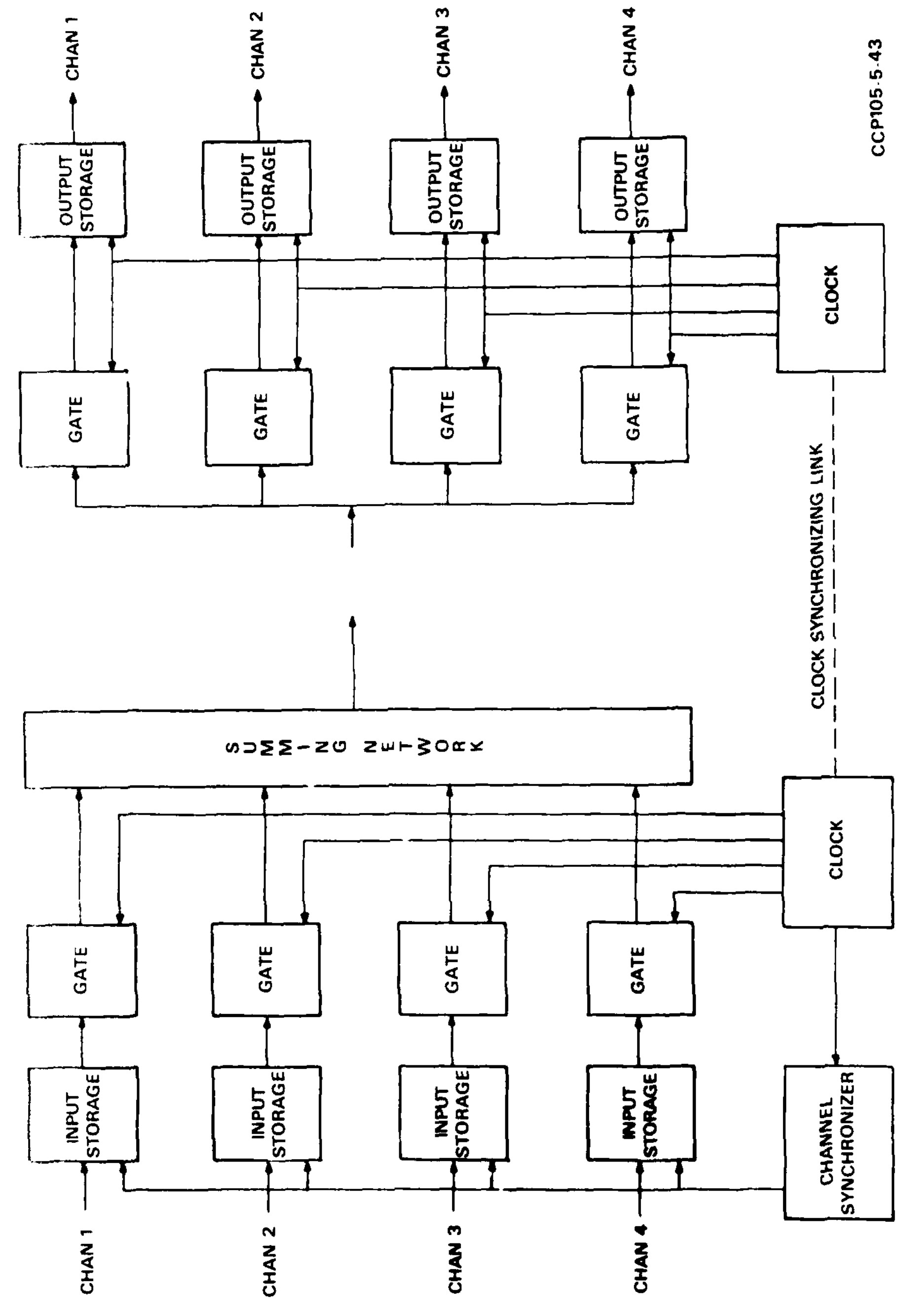


Figure 8-18. Clock synchronization system block diagram.

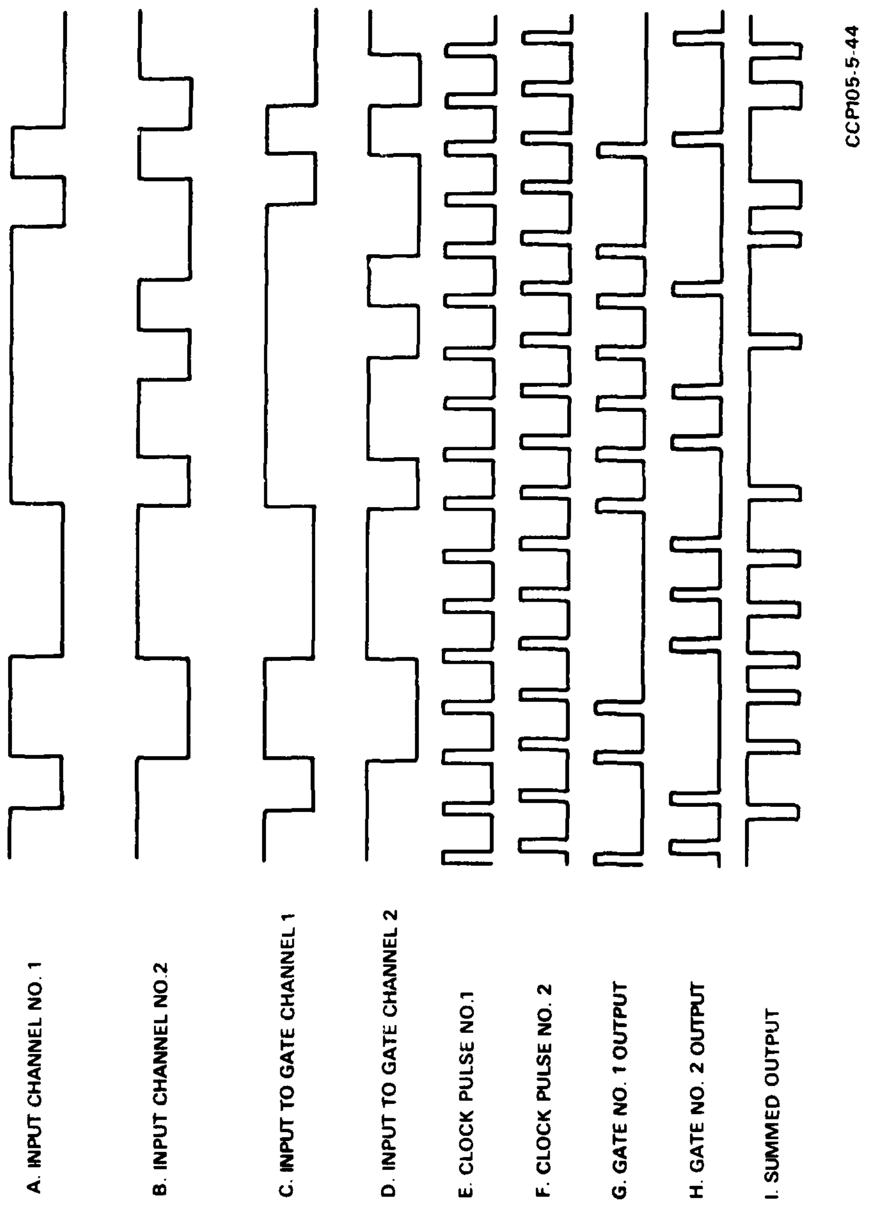


Figure 8-19. Time-division multiplexing waveforms (transmit).

(h) Returning to figure 8-18, the right half of the block diagram shows the equipment needed to recover the original four-channel information from the composite signal. Figure 8-20 gives the waveforms involved. Part A shows the input waveform (identical to part I of figure 8-19) after its detection in the receiver. The receiving demultiplexer requires a clock, which generates pulses similar to the multiplexing equipment, and which must be synchronized with those of the multiplexing equipment. The dashed line in the block diagram represents the signal, which is sent between clocks when necessary to ensure synchronization. The clock pulses for reception are shown in parts B and C of figure 8-20. These do not occur at exactly the same time as those shown previously, but occur at a slightly later time to compensate for any delay in transmission.

(i) The gates in the receiving system act like those in the transmitting system; that is, they give a positive output when both inputs are positive and a negative output when either input is negative. The outputs of the gates for channels 1 and 2 are given in parts D and E of figure 8-20. These are identical to the individual gate outputs of the transmitting equipment. The function of separating the individual signals is now complete, but the original pulse length must be restored. This is accomplished by a flip-flop which receives the gate outputs and clock pulses. The flip-flop is designed to give a positive output only if a pulse from the gate and clock are applied to the inputs simultaneously. Once put into this state, the flip-flop remains in this condition until triggered by additional pulses. If a clock pulse alone arrives, the flip-flop switches to the negative state and remains there until again triggered simultaneously by a gate and a clock pulse. When the gate outputs (parts D and E) and the clock pulses (parts B and C) are applied to the flip-flop, the outputs shown in parts F and G result. These are the final outputs of the system. They are similar to the input signal. The exceptions are that the output of channel 2 is delayed by the length of a clock pulse and the interval between characters (letters) is often changed by a small amount.

(2) *Frequency-division multiplex.*

(a) Frequency-division multiplexing is used primarily to multiplex voice channels, although teletype channels are often multiplexed in a similar fashion into the individual voice channels. As the name implies, frequency-division multiplexing depends on translating the various voice channels to different frequencies and transmitting this composite signal. When this composite signal is received, the individual voice channels are separated and demodulated. This process results in a repeat of the original voice channel information.

(b) A simple example of a four-channel frequency-division multiplex system is given in figure 8-21. In this example four individual voice channels are to be multiplexed. The four channels are applied to four individual input circuits. The output of each of these circuits is fed to a balanced modulator. At the same time,

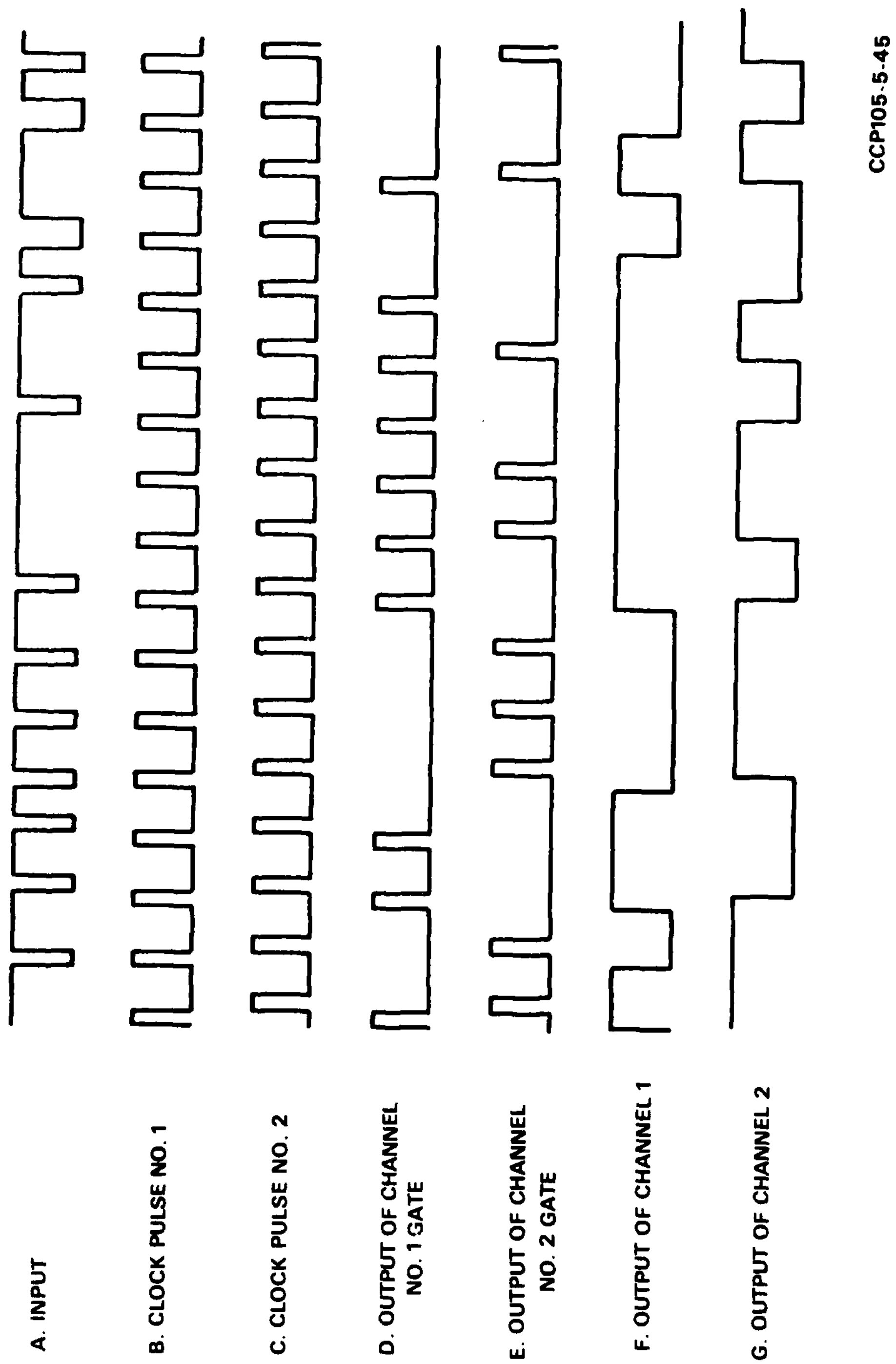


Figure 8-20. Time-division multiplexing waveforms (receive).

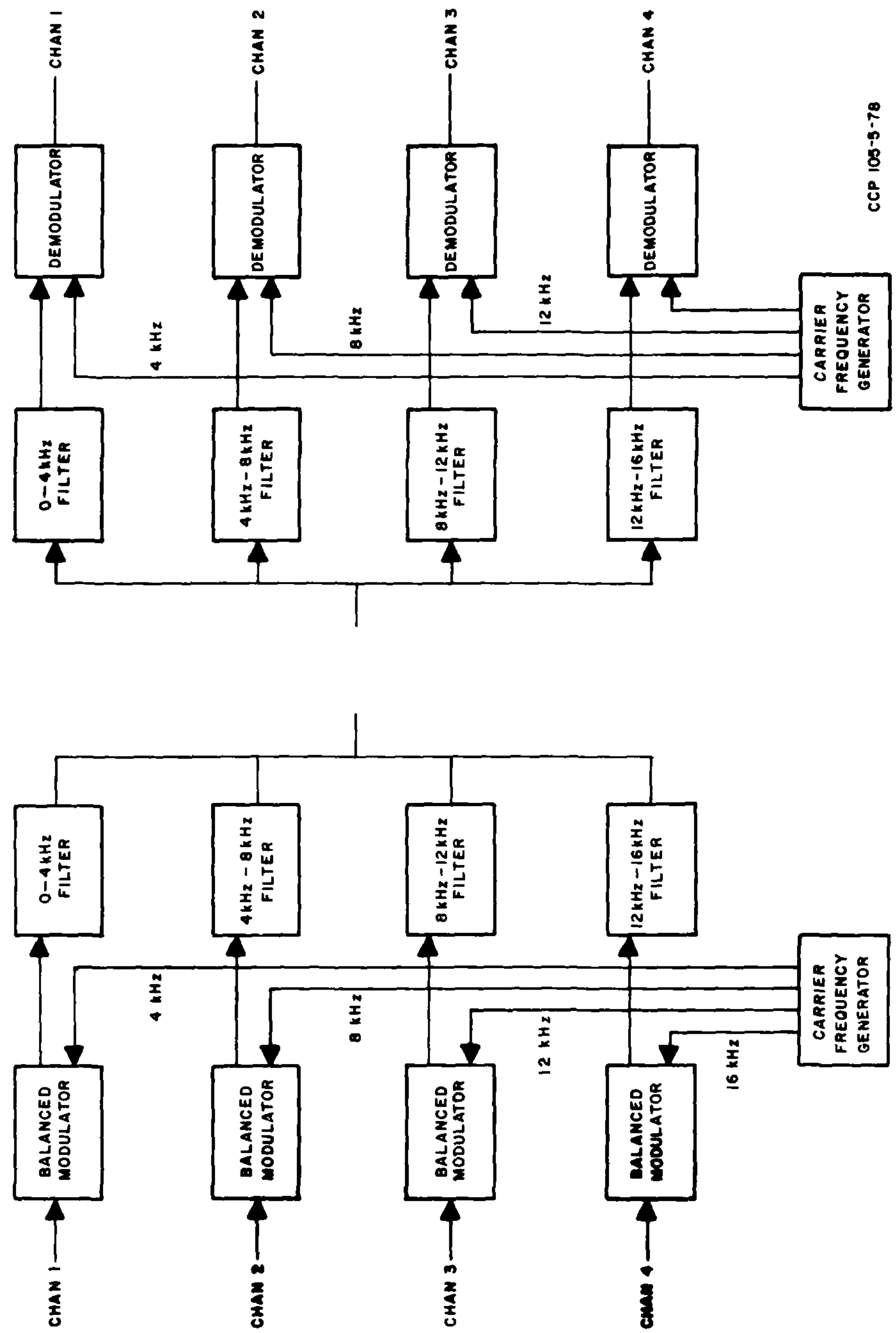


Figure 8-21. Four-channel frequency-division multiplex system block diagram.

each balanced modulator also receives a carrier frequency from the carrier frequency generator. The modulator in channel 1 receives a 4 kHz carrier, channel 2 receives an 8-kHz carrier, channel 3 a 12-kHz carrier, and channel 4 a 16-kHz carrier. The output of each balanced modulator is a double sideband suppressed carrier signal centered around the particular carrier frequency employed in that channel. The carrier frequency is suppressed (reduced) to a small amplitude by the balanced modulator, carrier suppression being one of the characteristics of a balanced modulator output. The double sideband output of each modulator is passed through a filter, which passes the lower sidebands and attenuates the upper sideband. Thus, the output of each channel is a single sideband signal, which is a replica of the original signal's spectrum, shifted to another frequency. This signal has the same shape spectrum and contains all the information needed to reconstruct the original signal. The envelope of this output, however, does not resemble the input, this being one of the characteristics of a single sideband signal. The output of these four filters is summed and the composite signal is ready for transmission.

(c) The carrier frequency generator which furnished the carrier frequency to the balanced modulators must generate accurate, properly related frequencies which can be reproduced at the other end of the link for proper demodulation. This is usually achieved by establishing a 4-kHz output derived by dividing down from some higher frequency standard and then providing the 8-, 12-, and 16-kHz signals by multiplying the base 4-kHz signal. This gives uniform spacing to the various carriers. The composite signal output of the summing network can then be used directly on cable or used to modulate an rf carrier as would be the case in satellite communications.

(d) At the receiving end, a bank of four 4-kHz wide bandpass filters are used to separate the composite, multiplexed signal into individual channels. Each channel filter output is demodulated by mixing with the proper carrier frequency. In this case, 4-, 8-, 12-, and 16-kHz carriers for channels 1 through 4, respectively, are used. To supply these carriers, a carrier frequency generator similar to that employed in the transmitting equipment is used.

(e) Since a single sideband technique is used in multiplexing and, as mentioned, the envelope of the single sideband signal does not resemble the original information, the carrier frequency supplied to the demodulator must be very nearly the same as that originally used for modulation. Demodulation of a single sideband signal with a carrier which is even a few cycles different from the original carrier frequency results in severe distortion of the output. A change in carrier frequency as small as 50 Hz is usually great enough to make a message unintelligible. The outputs of the channel demodulator are replicas of the original channel inputs and the multiplexing-demultiplexing process is complete.

(f) Although it will be some time before we can expect high-channel-capacity military satellite communications systems, when the time arrives, techniques will be available to multiplex large numbers of channels. Frequency-division multiplex units now multiplex as many as 1,800 voice channels on a single carrier frequency over conventional links, and systems providing even more channels are being developed.

(g) When multiplexing large numbers of channels, techniques similar to those described for a four-channel system are used, with the exception that not all of the voice channels are translated to their final transmission frequency in a single step. A typical arrangement for multiplexing 600 channels is shown in figures 8-22, 8-23, 8-24, and 8-25. The following standards have been established: 12 voice channels are usually mutiplexed to form a group which has a bandwidth of 48 kHz; five of these groups are combined to form a supergroup which has a bandwidth of 240 kHz and will handle 60 voice channels; and ten of these supergroups are combined to form the 600 channel baseband, which is 2.6 MHz wide. This combination (groups, supergroups, and baseband) is shown in figure 8-22. The details of combining 12 voice channels into a group are shown in figure 8-23. Channels 1 to 12 are mixed with carriers and filtered to provide a composite signal between 60 and 108 kHz. This composite signal is passed through a 60- to 108-kHz group filter and is ready for further combination into supergroups.

(h) An additional feature, available on many multiplexing equipments, is an input directly into the group filter which allows transmission and reception of a single wideband 48-kHz channel in place of the 12 voice channels. Usually any number, from 0 to 50, of the groups might handle 48-kHz channels. For example, a 600-channel multiplex might be transmitting twenty 48-kHz channels and 360 voice channels. In figure 8-24, the method of combining five groups into a supergroup is shown. The same modulation and filtering technique is repeated with the final filter passband being from 312 to 552 kHz. A wideband input is provided so that a 240-kHz channel is also available. Any number of supergroups can be devoted to 240-kHz channels, with the remaining channels used for 48-kHz or voice channels. As an example, there might be three 240-kHz channels, 22 48-kHz channels and 156 voice channels in the 2.6-MHz baseband. Figure 8-25 depicts the method of combining 10 supergroups into the total baseband.

(i) The technique is similar for both the channel and grouping systems with some slight exceptions. Supergroup 2 is passed through a 312- to 552-kHz filter directly without modulation. In addition to the combined channels, two pilot tones at 60 and 2604 kHz are added to the transmitted signal. As in the four channel system, the modulating and demodulating carriers must be identical in frequency. This condition is especially hard to maintain when carriers at frequencies as high as 2852 kHz are used in supergroup 10. In order to ensure that these frequencies will be identical, pilots derived from the carrier generators at one end of the link

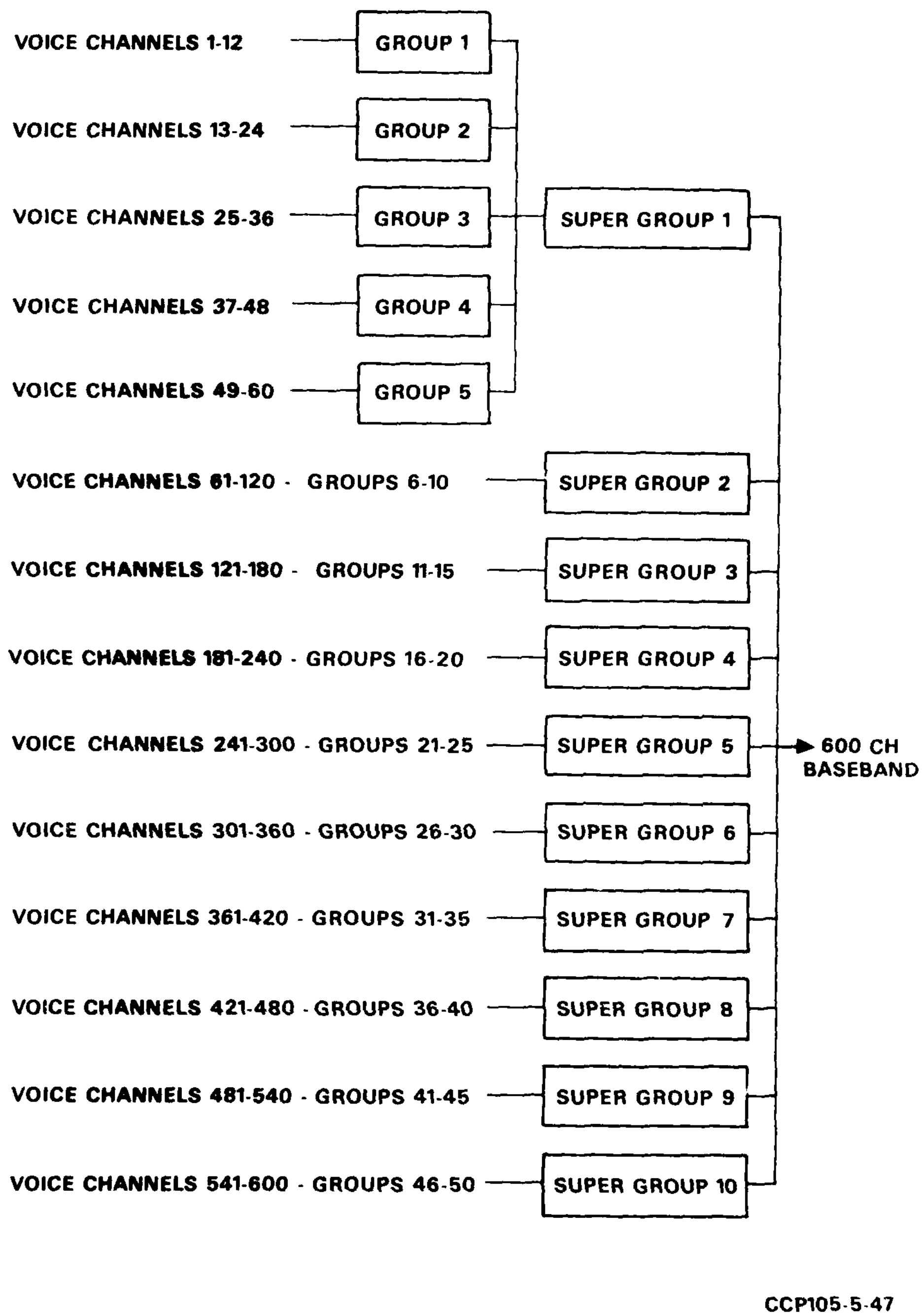


Figure 8-22. Frequency-division multiplex channels, groups, and supergroups block diagram.

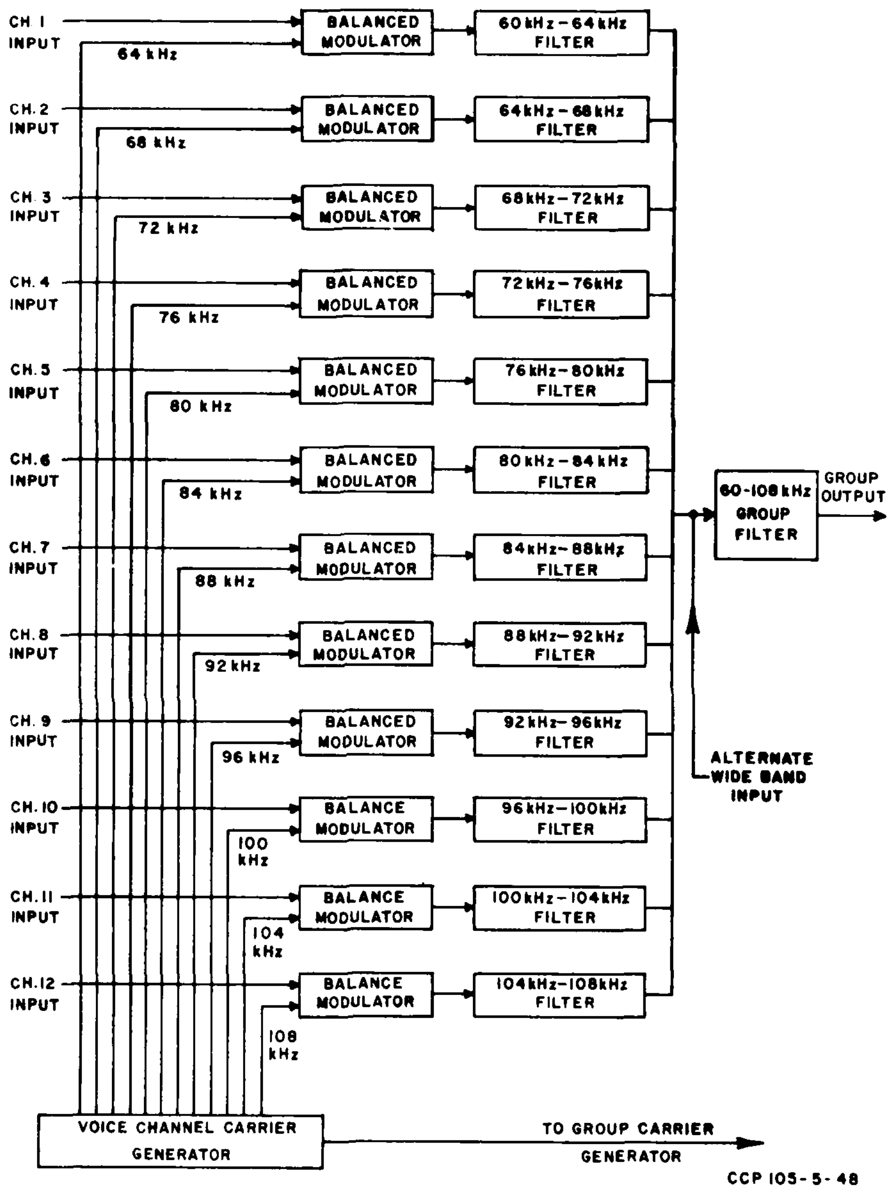


Figure 8-23. Combination of 12 voice channels into one group block diagram.

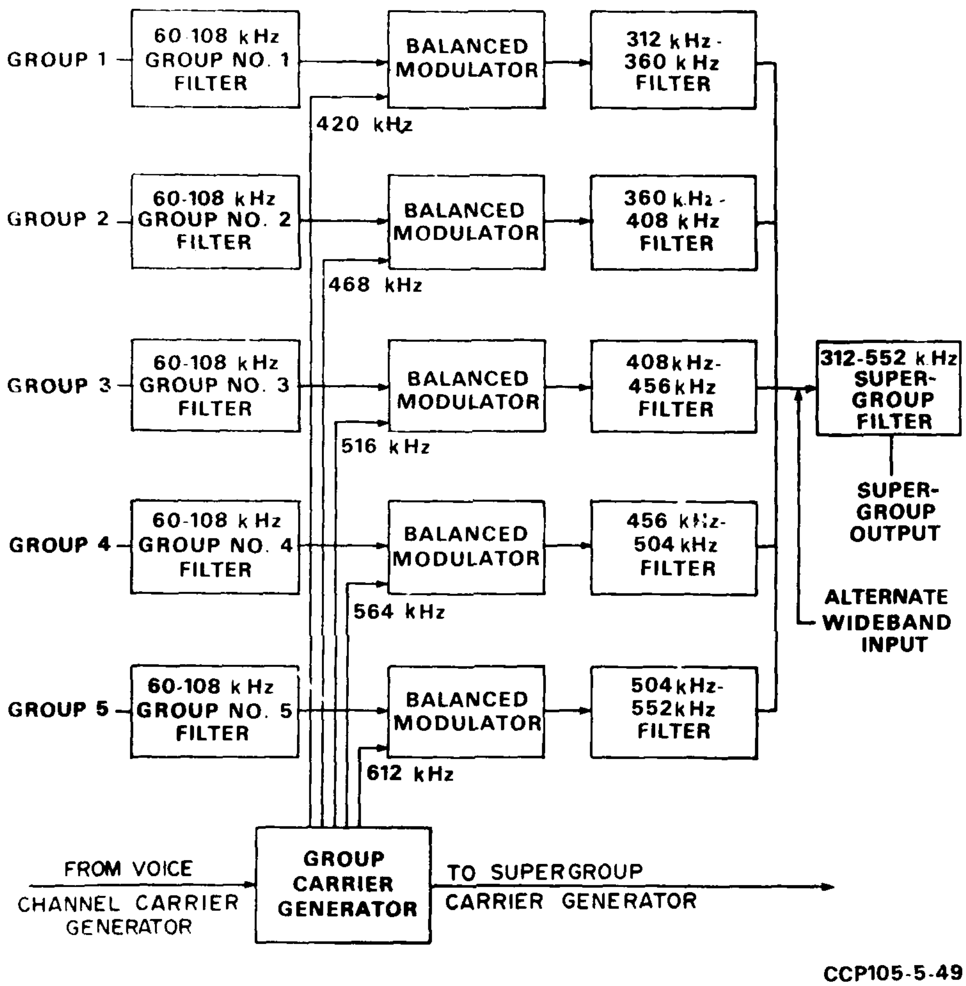


Figure 8-24. Combination of five groups into one supergroup.

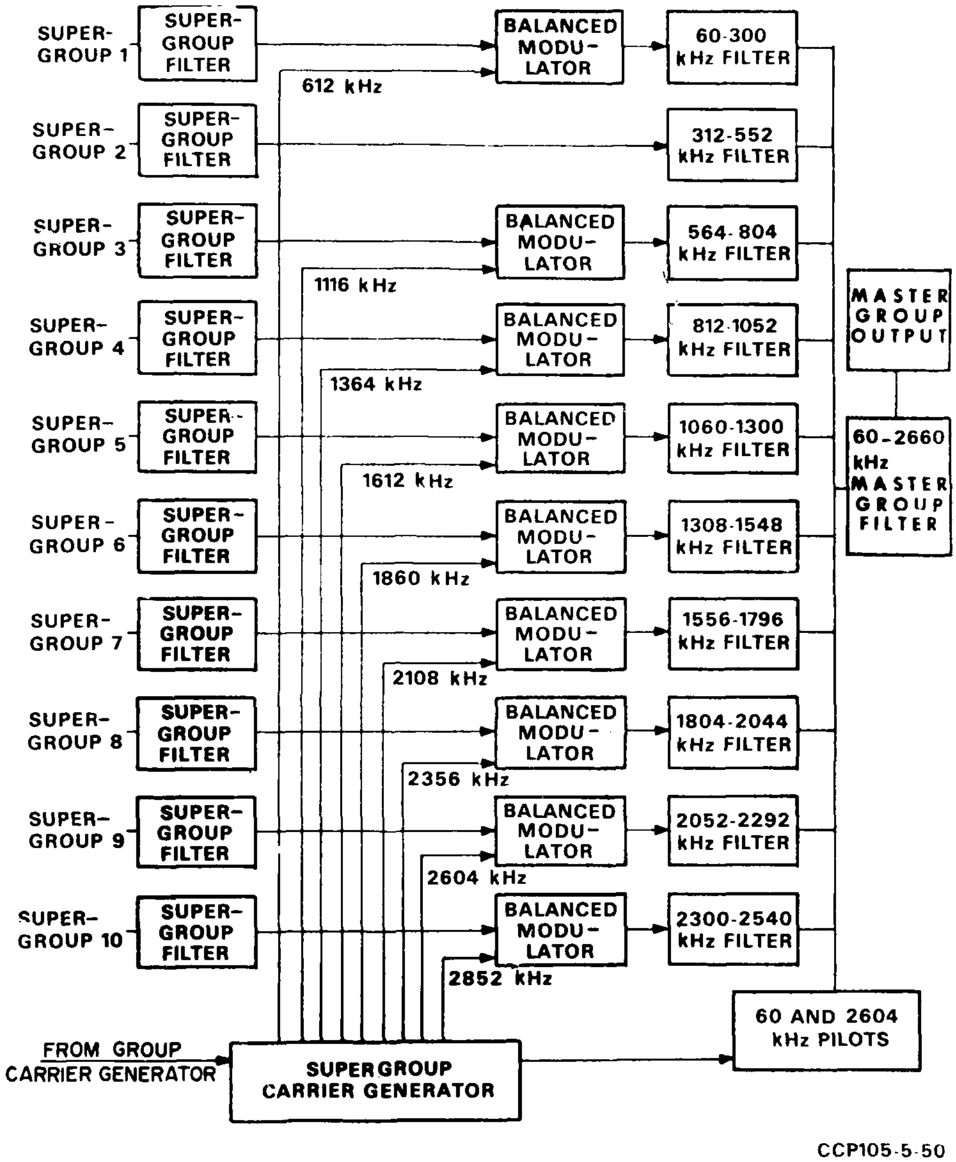


Figure 8-25. Combination of 10 supergroups into one baseband block diagram.

are received and used to derive carriers at the other end. As a result, any drift in frequency at one end of the link is followed at the other end of the link and no distortion is introduced in demodulation.

8-5. TIMING AND FREQUENCY STANDARDS.

a. The satellite communications system depends on maintaining time and frequency accuracy, without which some of the systems functions will not operate properly. Some of the systems functions which could be degraded because of inaccurate time and incorrect frequency control are as follows:

(1) Satellite acquisition time (the day, hour, minute, and second that a satellite will appear over the horizon to a particular ground station).

(2) The precise channel frequency so that proper Doppler shift correction can be computed for each satellite communication link (receiver and transmitter).

(3) The precise beacon receiver frequency so that proper Doppler shift correction can be computed to permit quick acquisition of the satellite.

(4) Encrypted data synchronization.

(5) Time reference to the local computer, which provides pointing angles to the antenna system for each satellite communication link.

(6) Frequency standards for test equipment used to calibrate the many frequency determining devices found within a ground station.

b. Special broadcast stations transmitting in the very low frequency (vlf), low frequency (lf), and hf bands transmit national time and frequency standards to all interested users who have need of accurate time and frequency reference standards. These transmissions can generally be received at any location throughout the world. In the United States, the National Bureau of Standards and the Naval Observatory operate time and frequency standards stations whose frequencies and time are maintained as constantly as possible, with respect to the United States Frequency Standard.

c. Table 8-1 is a list of lf and vlf transmissions of frequency and time standards. Not listed in the table are the hf WWV and WWVH (U. S. National Bureau of Standards radio stations) time and frequency standards broadcasts on 2.5, 5.0, 10, 15, 20, and 25 MHz.

d. The standard of time is related to the standard of frequency because one determines the other. The passage of time is measured by counting the

Table 8-1. Lf and vlf frequency and time standards

Station	WWVB	WWVL	NBA	GBR	MSF
Location	Boulder, CO	Sunset, CO	Summit, Canal Zone	Rugby, England	Rugby, England
Frequency	60 kHz	20 kHz	18 kHz	16 kHz	60 kHz

repetitions of some regular event, such as the rotations of the earth, the vibrations of a piece of quartz crystal (as in a conventional crystal oscillator), or the waves emitted by an atom. When the regular event occurs many times in a second it is more convenient to speak of the number per second or its frequency rather than the time between two events.

e. A typical basic time and frequency standards system, which is generally referred to as the local clock, is portrayed in figure 8-26. It consists basically of a local oscillator of ultra high stability delivering a sine wave signal of high purity. High purity means that when the fundamental frequency is multiplied many times the signal will remain free of any unwanted spurious noise or disturbing frequencies, which are slightly different from the desired one.

f. The heart of the local oscillator frequency standard is a highly stable quartz crystal oscillator. The oscillator quartz crystal is of especially high quality and is mounted in such a manner that it is held firmly without damping its oscillations. In order to accomplish this, overtone vibrations having nodal points are chosen. The quartz crystal assembly is contained in a oven which is fitted with precise temperature control, since temperature affects the natural frequency of the crystal. In some cases the oven with the crystal is enclosed in an outer oven to improve the degree of temperature stability.

g. Of nearly equal importance to oscillator stability is the requirement for a very stable and well regulated power supply. The oscillator power supply must not reflect any of the fluctuations that occur in the primary power source. A reliable battery power source is always associated with the basic oscillator and clock portion of the system, to provide a continuous source of power in the event of an inadvertent main power outage. The batteries will normally supply power for at least 6 hours, which is generally quite adequate.

h. Additionally, high quality and high reliability components (capacitors, resistors, inductors, tubes, and transistors) must be used throughout the oscillator unit.

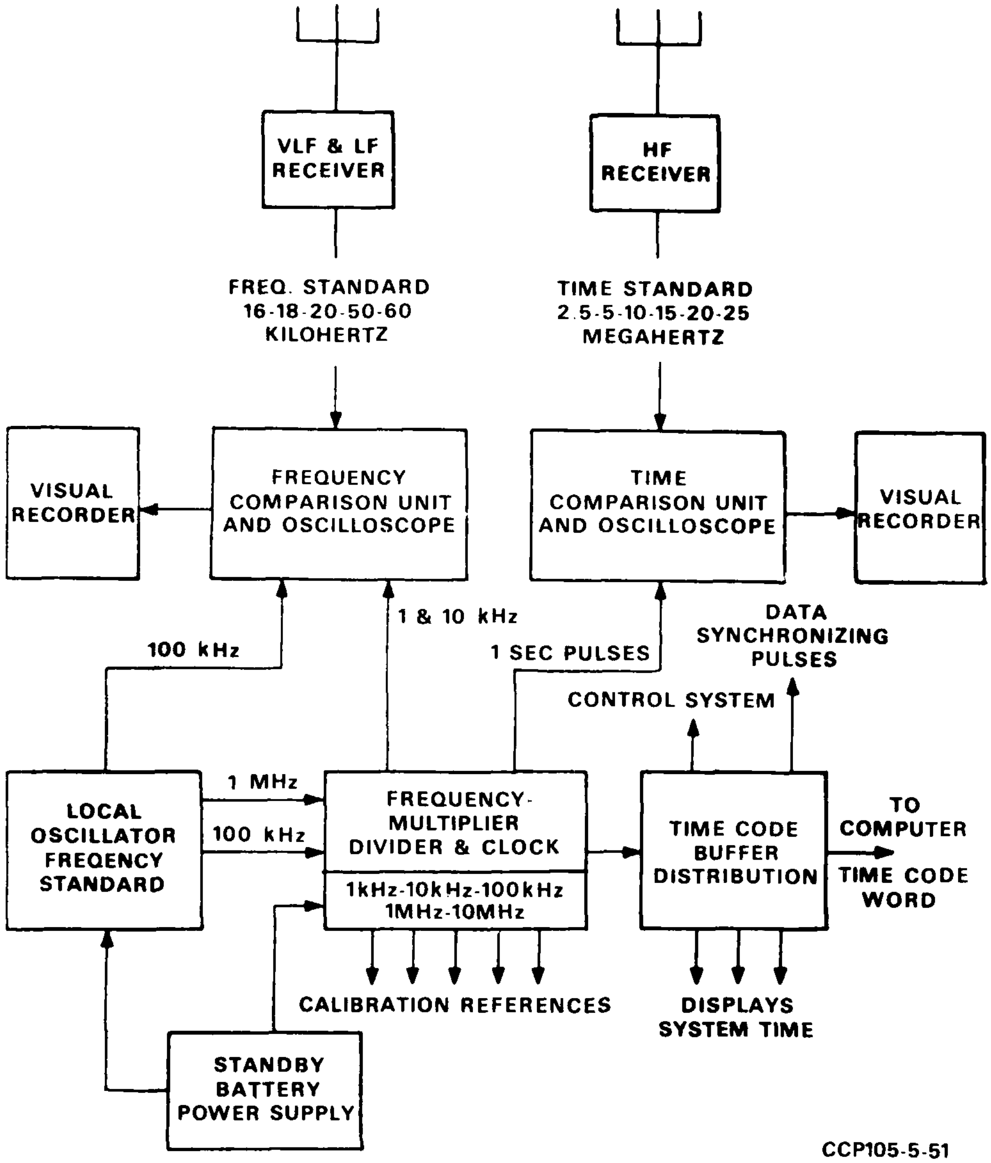


Figure 8-26. Typical local clock and frequency standard block diagram.

i. The oscillator furnishes power to a synchronous clock, which is very similar to the familiar household electric clock. The accuracy of the time kept by these clocks is dependent on the accuracy of the power source frequency.

j. If the oscillator is low in frequency, the clock will lose time; if the frequency is high, the clock will gain time accordingly. The rate at which the clock gains or loses time is extremely important, and extended observations and calibrations to determine this are required. These observations are accomplished by use of the frequency and time comparison units. These units contain a type of visual recorder that displays the difference in time and/or frequency between the time and frequency standard being received (vlf, lf, or hf radio time and frequency standards) and the local clock and frequency standard. If large errors exist they will usually be quite apparent to the operator observer, however, when the local clock settles down after initial start-up and has been operating for at least 3 weeks (usual time required to stabilize), the errors can only be detected from observations and recordings over hours, days, and frequently weeks.

k. Adjustment of the basic oscillator should not be attempted without considerable prior thought, data evaluation, and reference to the manufacturer's instruction manual. It is generally permissible; however, to advance or retard the time code generator as required, as this will not adversely affect the basic standard. Timing and frequency comparison is made using the procedure below.

l. Local clock time is compared with the hf stations, WWV or WWVH, during either full daytime or nighttime hours. The twilight hours are generally avoided because of the rapid fluctuations that normally occur to the rf propagation characteristics associated with natural changes in the ionosphere. The frequency standard, or local oscillator, is compared to the vlf and lf stations as their propagation characteristics are considerably more stable and are virtually unaffected by ionospheric changes. In both cases, the appropriate delay charts and equipment procedures, peculiar to each geographic location and station equipment, must be consulted.

m. The foregoing procedure assumes that the time and frequency accuracy requirements will be such that changes in hf propagation time will be too great to permit use of the hf WWV transmission for checking the local oscillator. In actuality, this will often not be the case and it may be found that the hf WWV transmission is used exclusively.

n. The minimum oscillator accuracy or acceptable stability is one that will not drift or deviate more than 1 part in 10^8 (one cycle out of 100 million cycles). This is, of course, an exceptionally high requirement. However, ground terminals may well operate at a frequency of 8 GHz, which is 8 x 10^9 Hz. A drift of one part in 10^8; therefore, would result in a drift of 80 Hz which, though tolerable, is a significant drift. With the rapid advances in technology, it is not uncommon for frequency standards in field use to have stabilities of 1 part in 10^{10}, and the near future may find an atomic frequency standard and clock at ground stations rather than the quartz crystal types now in use.

o. These atomic clocks can provide an accuracy of two parts in 10^{11}. Their operation depends upon emission or absorption of spectral lines that arise from the difference in the energy states of an atom. To understand what is meant by the two energy states of an atom, visualize a case where light is given off from a phosphorescent material in a dark room. Phosphorous atoms are capable of being excited to a higher energy state by the absorption of some of the light energy of ordinary white light, such as provided by the sun or an incandescent bulb. The phosphorous atoms that have been excited to a higher state will gradually, and in a rather random fashion, decay back to their original unexcited state. As this occurs, the light energy that was previously absorbed now is emitted in the form of light energy that has a very specific wavelength or frequency. Some atoms have energy states which give rise to the emission or absorption of wavelengths sufficiently low in frequency for electronic counters to count the number of cycles emitted in a given period of time. Since the emitted wavelength (spectral line) is at a very precise frequency, this provides, in effect, an extremely stable source of rf energy.

p. Atomic clocks have been constructed that use a beam of caesium atoms. The beam is separated into two components. The component made up of atoms in the lower of two possible states is passed into a resonant chamber fed by an oscillator. These atoms absorb energy from the oscillator and are raised to a higher state. The atoms which have absorbed energy are deflected into a detector cavity Upon detection they can be counted very precisely. If the microwave oscillator varies slightly in frequency, less atoms are deflected into the detection cavity and; hence, less are counted. Since the output lacks direction of error, the count is less whether the frequency of the oscillator is high or low. A small amount of phase modulation is added to the microwave oscillator's output. The phase modulation can be detected in the output; thus, giving sense to the output. This enables the frequency of the oscillator to be controlled precisely.

q. As ground stations become more sophisticated and higher and higher data rates are encountered, these areas of timing and frequency standards will assume an increasing importance.

8-6. ANTENNA SUBSYSTEM.

a. Satellite communications systems require antenna systems which do not contribute any additional noise to that of the receiver in the overall system. The Cassegrainian-fed parabolic antenna is one type of antenna design which, because of negligible side and back lobes, yields good performance. Other types of antenna feed systems are also used and are discussed in paragraph 7-4.

b. The Cassegrainian system consists of a hyperbolic surface (subreflector) mounted between the focus of the parabola and the surface of the parabola. A typical feed system and antenna are shown in figure 8-27. The reflector is part of a Cassegrainian system. The feed is mounted near the surface of the parabolic dish. It radiates energy which is reflected off the hyperbolic surface onto the parabolic surface and back into a parallel beam. Typical rays are F' (the feed position), A, A', A", and F', B, B', B". Having the feed at F' allows the receiving components, such as the paramp, to be mounted near the dishes; this lowers the losses in the waveguide and the attendant degradation of system performance. Spurious rays are not reflected off the hyperbolic surface to the side of the dish, and side lobes and back lobes are kept low.

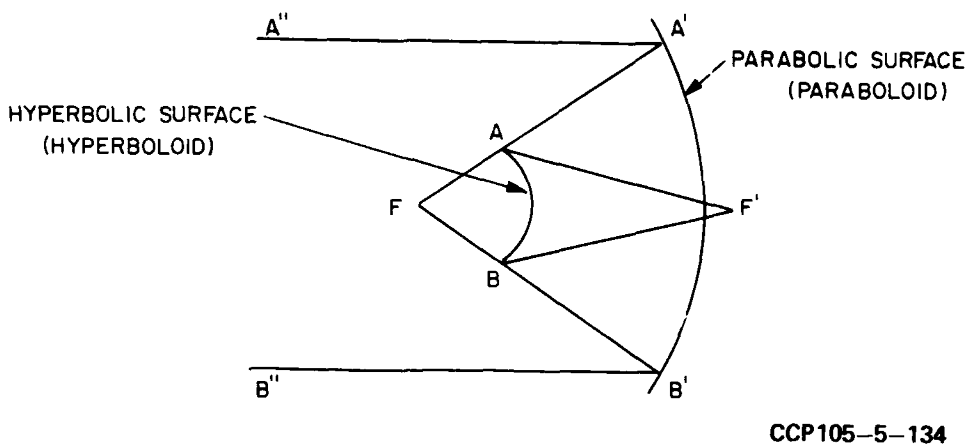


Figure 8-27. Typical Cassegrainian parabolic dish and feed system functional diagram.

8-7. TRACKING SUBSYSTEM.

a. Whenever reliable communications must be maintained between two fixed sites via an orbiting satellite, whose relative angular position is changing, some form of automatic or manual tracking must be employed.

b. Tracking is accomplished by positioning the antenna beam with a servomechanism so the error signal is minimized.

c. An error signal is generated by causing the antenna beam to bracket or encircle the target, either in a continuous or discrete pattern, by offsetting the beam at a slight angle. For a target centered exactly in the pattern, a signal is produced which will be reduced slightly due to the beam offset angle, but will be equal for all beam positions and will cause no error signal. If, on the other hand, the target is not centered an error signal will be produced and will indicate the direction of the error. The error signal is amplified and applied to the antenna servomechanism. The servomechanism points the antenna in a direction to reduce the error signal.

d. In satellite tracking applications, the beams (which are called lobes) are those of the ground station receiving antenna. The lobes are generated either by offsetting the receiving horn in a discrete or continuous pattern, or by using three or more separate horns simultaneously. An example of the single offset horn is a sequential, or conical-scan system, and an example of simultaneous lobing is the multiplex horn arrangement called monopulse. Both of these methods, detailed below, are techniques which have been developed principally for use by tracking radars, but are almost universally used at satellite tracking stations.

(1) *Simultaneous lobing (monopulse and pseudomonopulse).*

(a) In the monopulse method, two offset antenna lobes are combined to measure elevation error and two additional lobes are combined to measure azimuth error. These lobes are formed by the feed system.

(b) The receive feed system consists of four horns feeding four hybrids through four receive filters. The transmit feed consists of a feed horn fed through a cooled transmit filter and a harmonic filter.

(c) The four receive feed horns are arranged around the transmit feed horn to form the monopluse tracking feed. The arrangement of the four feeds is shown in figure 8-28. The rf energy picked up by feeds A and B is channeled through waveguides to hybrid number 1. This hybrid gives two outputs, a sum and a difference. The sum output is represented by (A + B), the difference by (A - B). The signals arriving at horns C and D are processed in the same way in hybrid

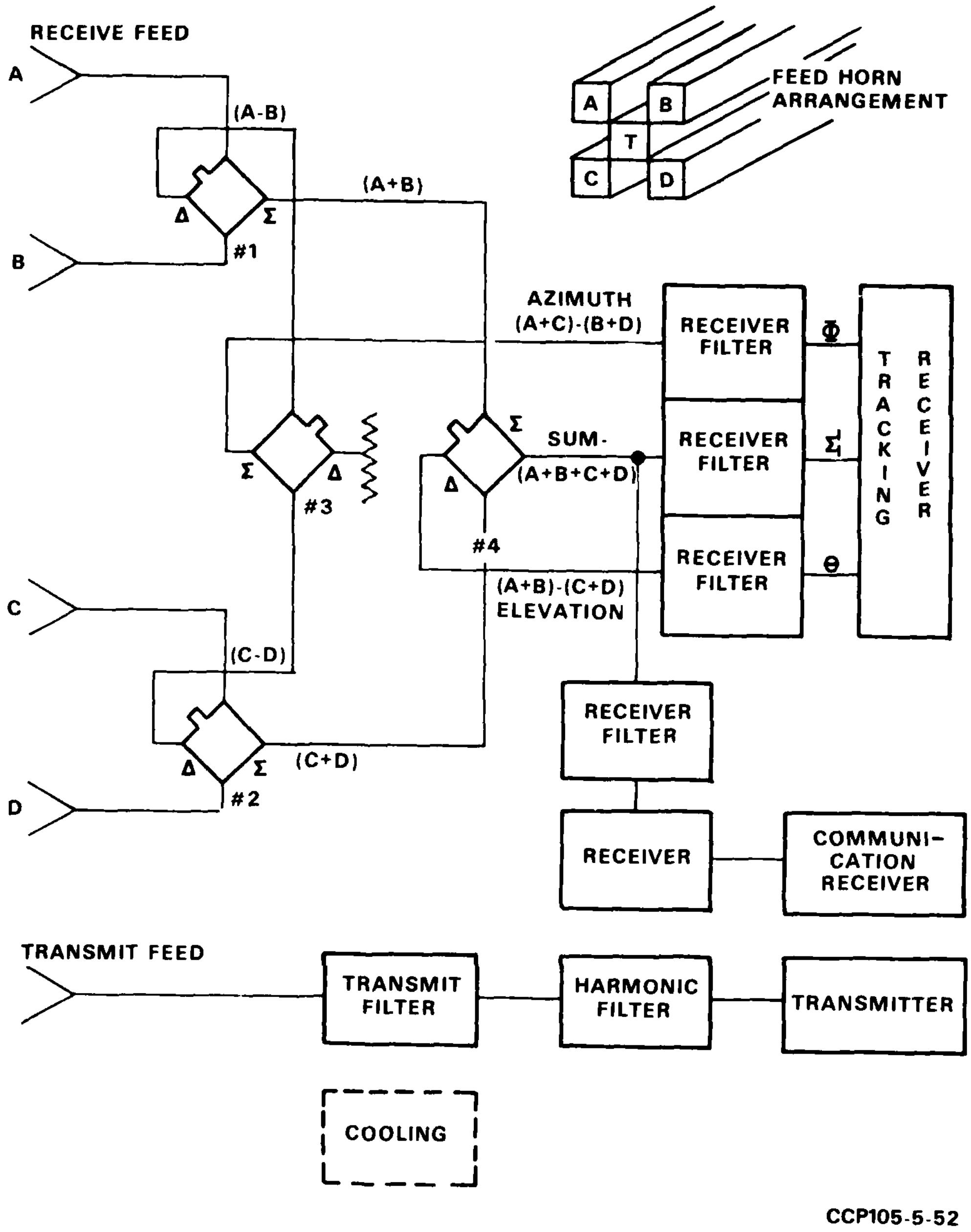


Figure 8-28. Four-horn antenna feed system block diagram.

number 2, giving a sum output (C + D) and a difference output (C − D). The sum outputs of hybrids 1 plus 2 is processed in a similar manner in hybrid number 4. The difference outputs of hybrids 1 plus 2 is processed in hybrid number 3. The difference output of hybrid number 3 is not used, and this output is terminated in a matched load resistor. The sum signal from hybrid number 3 represents (A − B) + (C − D) or (A + C) − (B + D), and is used as the azimuth, or X, signal. This signal is zero when the rf energy source is midway between lines drawn through AC and BD. The difference output of hybrid number 4 is similarly (A + B) − (C + D) and is used as an elevation, or Z, tracking signal. The sum output of hybrid number 4 is the sum signal A + B + C + D, and is used as a reference for tracking and as the communications signal. If plots are made of the voltages at the outputs of the hybrids, two types of pattern occur. These are shown in figure 8-29. They are commonly called sum and difference patterns. The sum output of hybrid number 4 will yield a sum pattern. The difference output of hybrid number 4, the elevation output, will yield a sum pattern if the antenna is swung in azimuth, but a difference pattern if swung in elevation. The sharp changes in amplitude, of the difference signals near the axis of the pattern, are used to generate the error signals in the amplitude monopluse tracker. These error signals are used to drive the antenna servomechanism to a zero error position; the desired antenna pointing direction.

(2) *Sequential lobing (conical scan).*

(a) In contrast to monopulse, where multiple stationary lobes are used, the conical scan technique uses a single offset lobe, which is caused to rotate symmetrically about a centerline in space as shown in figure 8-30. One of the simplest methods for generating the conical scan pattern is to use a parabolic antenna with an offset feed, which can be rotated about the centerline of the reflector. If the feed maintains the same plane of polarization as it rotates it uses a flexible coupler. Otherwise, it is called a rotating feed and requires a rotary coupler.

(b) The error signal in the conical scan method results from the apparent amplitude modulation of the signal as the lobe is rotated about the centerline. This can be visualized by extending the monopulse pattern from four to an infinite number of lobes. The resulting pattern viewed from the satellite would be a circle with its center on the centerline (fig. 8-31). The instantaneous lobe pattern would appear as a single lobe, rotating inside of the larger circle at a rate equal to the conical scan frequency. The beacon signal will be modulated at the conical scan frequency. The amplitude of the modulation would depend on the position of the beacon in the pattern for a given offset angle.

(c) If the beacon signal appeared on the centerline, point A of figure 8-31 would be zero modulation, since the centerline appears at a point of constant

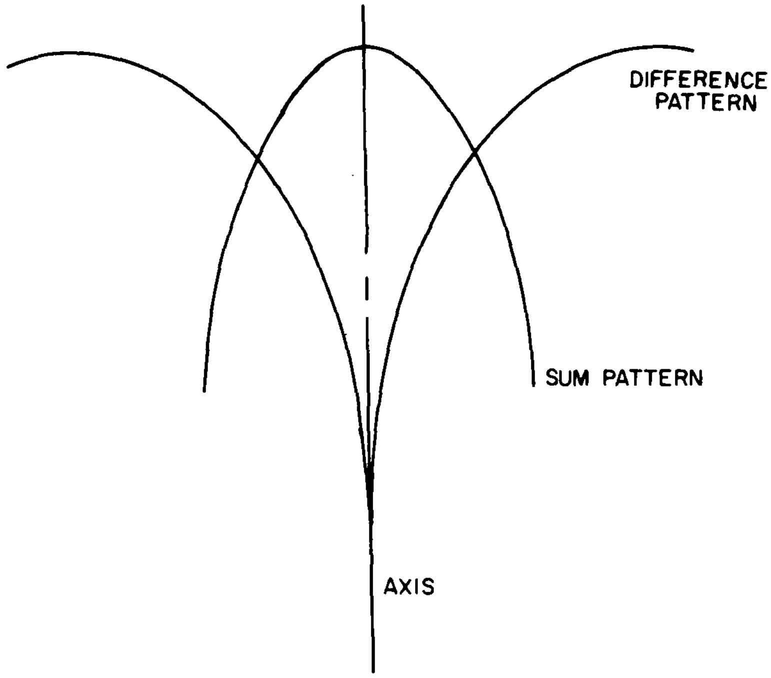


Figure 8-29. Typical monopulse system beam pattern.

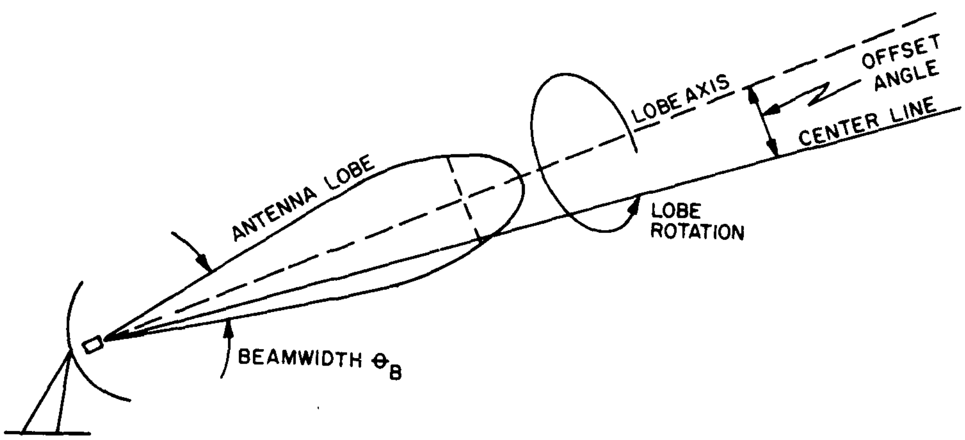
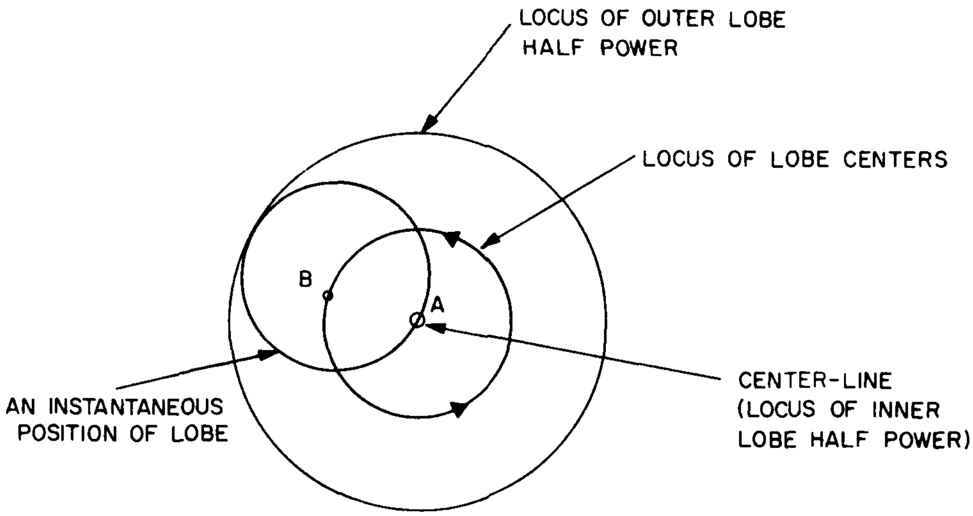


Figure 8-30. Conical scan technique representation.



CCP105–5–137

Figure 8-31. Conical scan lobe pattern.

signal. That is, the rotating lobe always has its circumference touching the centerline. If the beacon signal appeared elsewhere; for example at point B, the received beacon signal would increase and then decrease in a sine wave manner at the scan-frequency rate as the lobe swept past the beacon (this generates the modulation). The phase of the modulation generated depends on the direction of the angle from the beacon to the centerline. A typical conical scan modulation would appear as in figure 8-32.

(d) A block diagram of a typical conical scan tracking system is shown in figure 8-33. The conical scan modulation is extracted from the beacon signal in a conventional receiver; that is, a mixer, an IF amplifier, and an amplitude detector. After filtering to remove all but the scan frequency, the modulation is applied to two phase-sensitive error detectors. The error detectors have as their reference a phase-shifted version of the scan frequency, with the azimuth channel 90 degrees out of phase with the elevation channel (fig. 8-32). The output of these detectors will be a dc voltage depending on the phase relationship between the reference and the modulation.

(e) To be specific, the output of the azimuth error detector is a dc voltage proportional to the magnitude of the modulation times the cosine of the phase difference angle, θ_D, between the modulation and the azimuth reference generator.

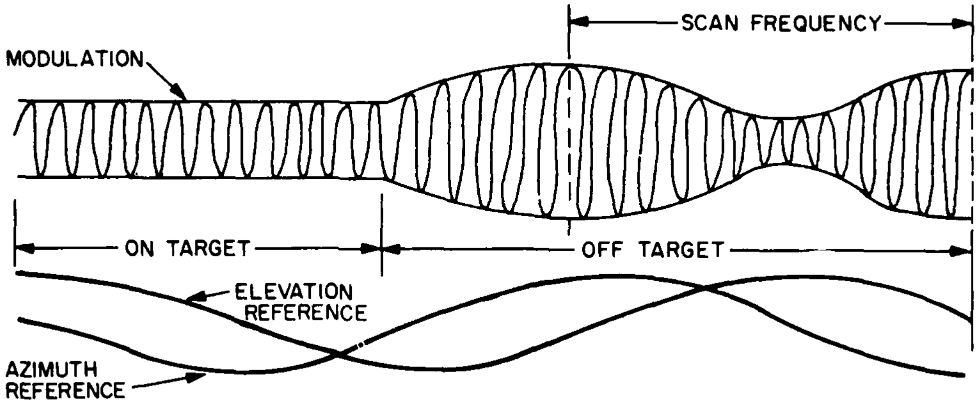


Figure 8-32. Conical scan modulation waveforms.

Since the magnitude of the modulation is proportional to the angle, θ_T, that the beacon makes with the centerline:

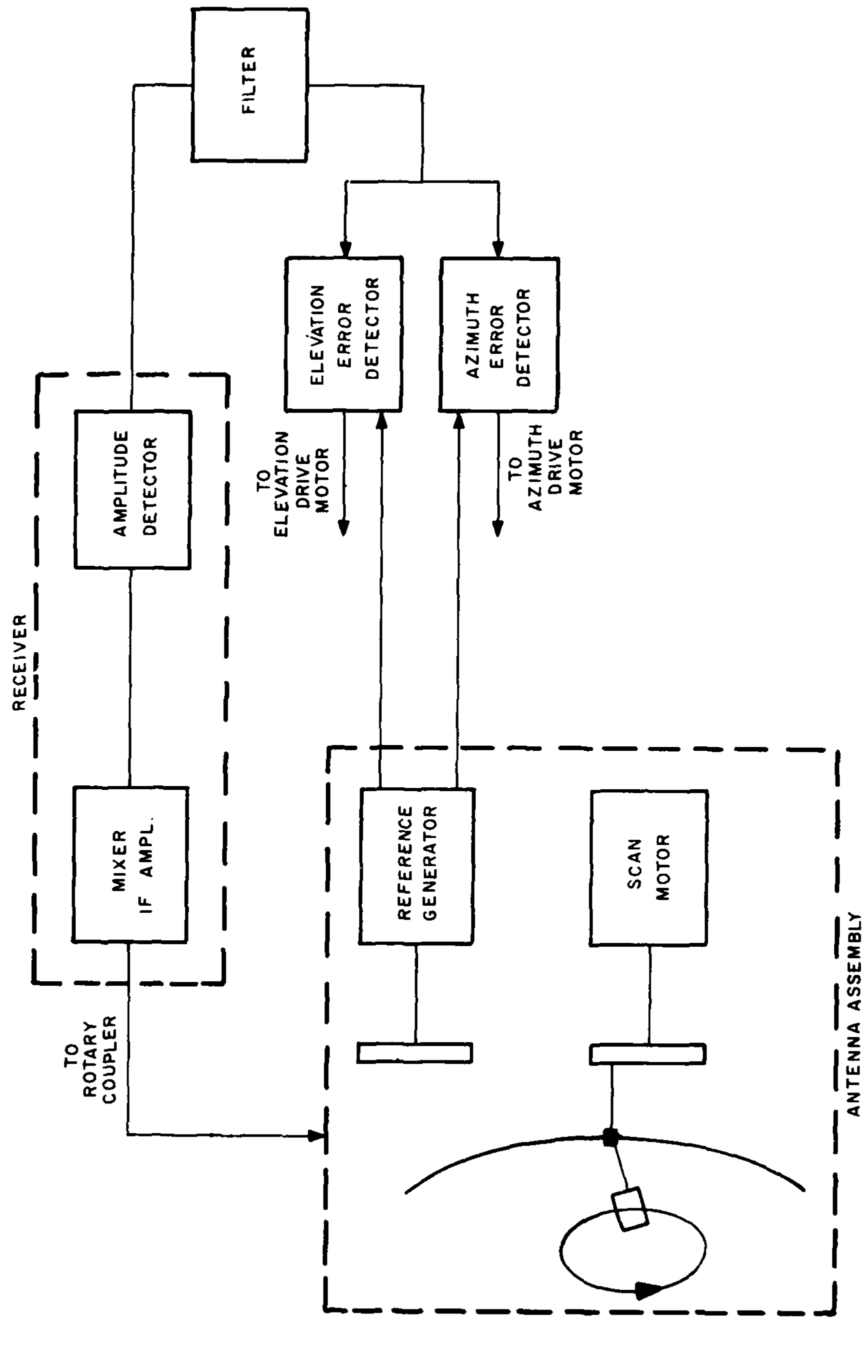
$$\text{azimuth-error signal} \approx K\theta_T \cos\theta_D \tag{8.1}$$

where K is the proportional constant

Likewise:

$$\text{elevation-error signal} = K\theta_T \sin\theta_D$$

These azimuth and elevation error signals are applied to their respective drive motors to position the antenna for zero error signal.



CCP105–5–139

Figure 8-33. Typical conical scan tracking system block diagram.

8-8. CONTROL SYSTEMS.

a. *General.*

(1) Throughout a communication satellite ground station there are many functions and devices that must be remotely controlled and many indicating devices that must be remotely displayed. The most significant and important of the remotely controlled devices is the large antenna associated with each communication satellite terminal. The size of the dish (as large as 60 feet), coupled with the high frequencies used, results in an antenna beamwidth on the order of one-tenth of a degree. To minimize any loss of signal due to pointing error, the antenna must be remotely controlled to point with an error of even less than the beamwidth of one-tenth of a degree.

(2) A control system is an arrangement of physical components connected or related in such a manner as to command, direct, or regulate itself or another system.

(3) Automatic control devices of one kind or another have been used by man for hundreds of years, and descriptions of early control systems can be found in literature at least as far back as the time of Leonardo da Vinci. The accumulated knowledge and experience that comprise the present-day science of control system design received a great initial impulse from the work of Nicholas Minorsky on *Directional Stability of Automatically Steered Bodies,* in 1922, and the work of H. L. Hagen, in 1934, on the *Theory of Servomechanisms.* These papers contained mathematical analyses based on the direct study of solutions to the differential equation. This approach to the design problem was the only one available for many years, and was used to great advantage by designers of control systems. In 1932, Harry Nyquist published a procedure for studying the stability of feedback amplifiers by the use of steady-state techniques.

(4) World War II created a great demand for high performance control systems. The demands of military security, however, restricted the dissemination of this valuable information for general use in the field. It was during this time that A. C. Hall published (for restricted use) *The Analysis and Synthesis of Linear Servomechanisms* (1943). This article gave a comprehensive approach to the steady-state analysis of stability and popularized the name "transfer-locus method" for this approach.

(5) Y. W. Lee and Norbert Wiener jointly described the fundamental relationships of the transfer functions for a large class of physical systems. H. W. Bode applied, in great detail, these basic relationships to the design of electrical networks and feedback amplifiers. All of these contributions have made possible

the advanced state-of-the-art that exists in the design and implementation of today's control systems.

b. *Control System Basics.*

(1) Examples of control systems exist everywhere in our daily life. A common control system would be a thermostatically controlled home heating system. A person sets the input to the system, which is the desired temperature setting of the thermostat. The thermostat compares this reference temperature with the actual temperature in the room. When the room temperature drops below the reference setting a heating unit is turned on. When the room temperature goes above the reference the heating unit is turned off.

(2) The two types of control systems are open loop and closed loop (fig. 8-34). In a closed loop system the control action is dependent in some way on the output. An open loop system has a control action independent of the output. The thermostat is an example of a closed loop system. The control action of turning on the heating system is dependent on the system output, which is the actual temperature of the room.

(3) Closed loop systems are also called feedback systems. Some of the output is fed back to affect the control action. Feedback can be defined as a property which permits the system output to be compared with the system input, or some internal parameter of the system, so that an appropriate control action may be performed.

c. *Control System Functional Elements.* Any control system, whether open loop or closed loop, can be generalized into four fundamental elements. Generalized open and closed loop systems are shown in figure 8-34.

(1) *Reference input.* The reference input is an external signal applied to the central system. It is used to command a specific action from the plant. In many cases, the input must be converted from one form of energy to another. The thermostat setting is the reference input.

(2) *Control elements.* These components take the input reference and the feedback signal and generate an appropriate signal to be applied to the plant. Going to our example, the bimetallic strip in the thermostat generates the on or off signal which is applied to the heating system.

(3) *Plant.*

(a) The plant, or the system to be controlled, is a process, machine, or other physical body. Some quantity or condition of the plant is to be controlled. For

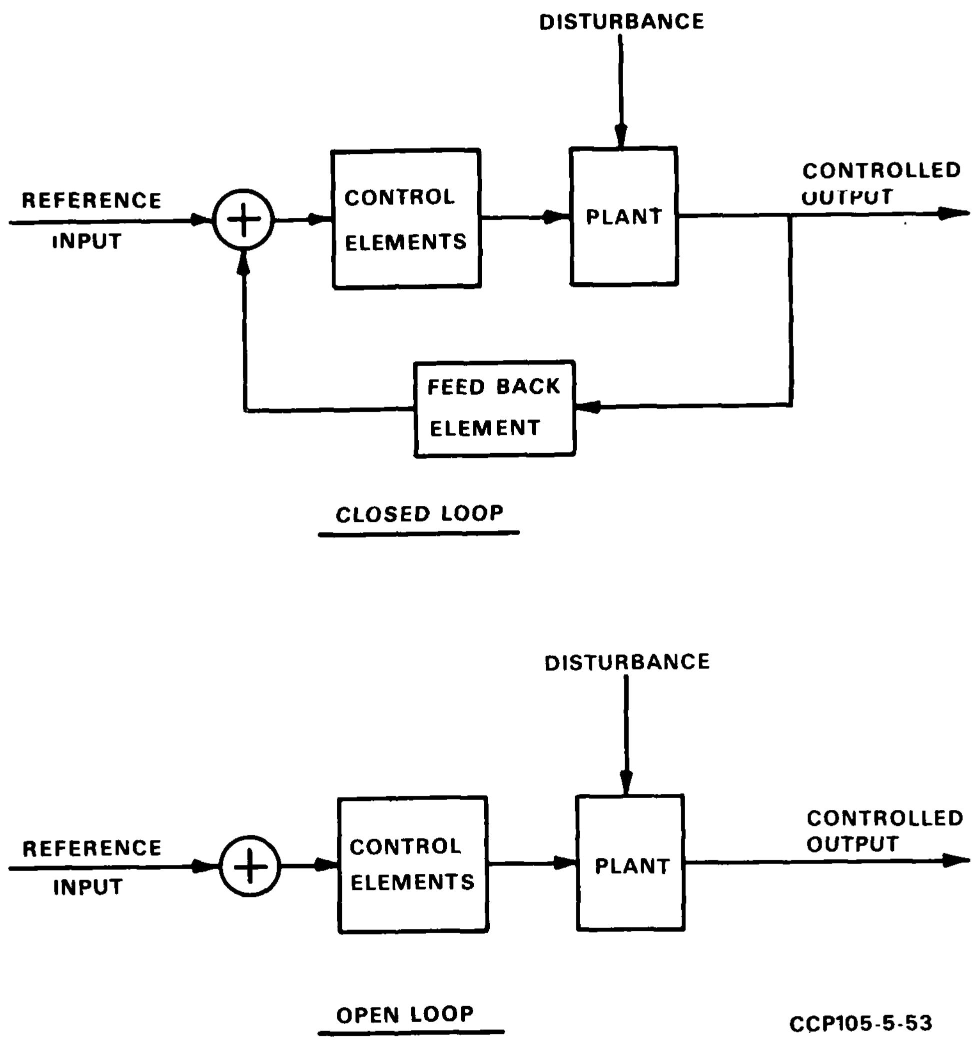


Figure 8-34. Generalized control systems block diagram.

this example the plant is the room to be heated. The temperature is the particular condition of the plant to be controlled.

(b) A disturbance is an undesired input to the plant. It could be anything that causes the room temperature to change. An example would be an open window on a cold day. This disturbance would lower room temperature and the thermostat would compensate for it.

(4) *Feedback elements.*

(a) These are the components that take the output and modify it for the input side of the control system. Again, the feedback elements may have to convert the output to another form of energy. In the thermostat example, the feedback is accomplished by the room temperature acting on the bimetallic strip. The strip converts the room temperature (thermal energy) back to mechanical position, so that the room temperature may be compared with the reference temperature set by the thermostat.

(b) The design of the feedback elements is the most critical in the system. The feedback should be a stabilizing factor, so that the controlled output always reaches the desired value. Much analysis goes into the design of a control system to ensure system stability under all conditions.

d. *Servomechanisms.*

(1) A servomechanism is a special form of control system. It can be defined as a power amplifying feedback control system in which the controlled output is mechanical position, velocity, or acceleration.

(2) In satellite terminals, an important servomechanism is the antenna pointing system.

e. *Positional and Velocity Tracking.*

(1) In the application of servomechanisms to tracking units a number of techniques are in common usage. Among these are two basic techniques which reflect the type of servosystem being used. These are manual tracking and aided tracking. In manual tracking, the operator can be considered as a part of a servoloop. The operator observes the misalignment (serves as the error detector) between a telescope and the target, and turns a handwheel in the direction that tends to reduce this misalignment. The handwheel drives the tracking unit which in turn positions the telescope and closes the feedback loop.

(2) In the simplest type of manual tracking system the operator is the only power source in the closed loop. In more complex systems, where torque amplification requires the use of a driving motor, the operator becomes analogous to a secondary servomotor. Limitation on the speed and torque available apply to both the power motor and the human operator; in addition, the performance of the operator is governed by conditions of fatigue, and mental and physical comfort. It is characteristic of operators that there is a time lag between the instant when an error is observed and the instant when corrective action is started.

(3) As in any servosystem, in order to obtain good performance the loop must have high gain and stability. The equalization is achieved by the judicious choice of the available parameters; the handwheel ratio, the optical magnification, and the tracking time constants.

(4) Aided tracking is a combination of positional and velocity tracking. In pure positional (or displacement) tracking, the operator has a direct connection, either mechanically or electrically, with the controlled member. In tracking a target, moving at a constant angular rate, the operator must turn his handwheel at a constant rate. If he is lagging the target, he will turn faster until the error is corrected; if he is leading the target, he will turn more slowly.

(5) In pure velocity tracking, it is the speed of the output that is determined by the position of the input handwheel. In tracking a target, moving at a constant rate, the handwheel need not be turned after the proper adjustment has been made.

(6) When these two types of tracking are combined an error in rate and the resulting displacement error are corrected simultaneously; a change in the handwheel position changes the rate of motion of the output, at the same time that the displacement error is corrected. Thus, aided tracking makes use of velocity servos and positional servos.

f. *Basic Position Servosystem.*

(1) To demonstrate the principles of a basic position servosystem (fig. 8-35), consider the following example. Assume that the control device is a crank, or a handwheel, and the load is an antenna. The handwheel is connected to a synchro generator and the antenna to a control transformer. The stator windings of the generator and control transformer are connected electrically. The output of the control transformer is an ac voltage that depends on the relative shaft position for magnitude and phase. There is no output when the shafts are in the same position and; accordingly, no output from the amplifying device or control amplifier. Hence, the servomotor does not turn and the antenna remains stationary. To change the position of the antenna merely turn the handwheel to the desired position. This produces an error voltage in the control transformer. This voltage

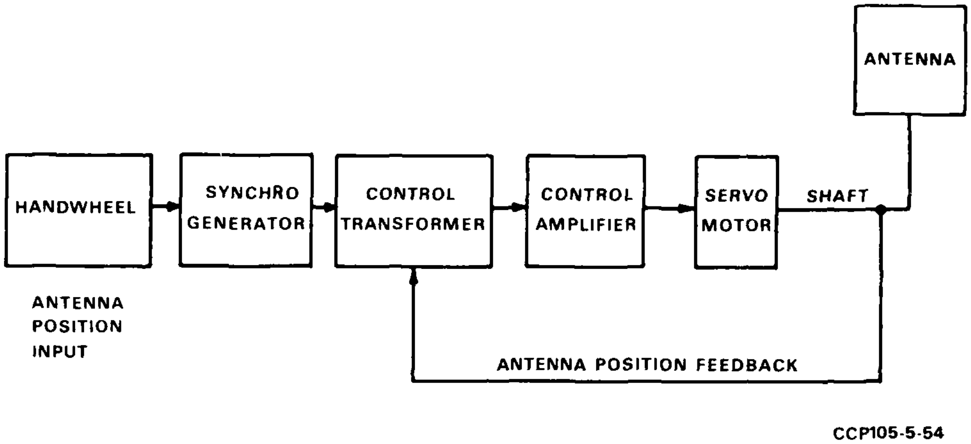


Figure 8-35. Basic position servosystem block diagram.

is rectified and increased in magnitude by the control amplifier and applied to the servomotor, causing it to turn the antenna to the desired position.

(2) The synchro generator and control transformer may be interchanged in position; however, such interchanging necessitates switching five electrical leads from one position to the other and gives no advantages. If the control device rotates continuously in one direction, the servomotor will rotate in the same direction at the same rate but with a small angle of lag. Remember that an error voltage is necessary before torque is developed. The greater the torque required, the greater the angle of lag.

g. *Control Amplifier.*

(1) The control amplifier takes the error signal, developed by the control transformer, and amplifies its power level to the high power levels necessary to drive the servomotor.

(2) The amplifier can be a mechanical device, such as the amplidyne, or an electronic device using thyristors or silicon controlled rectifiers.

h. *Servomotor.*

(1) An ac motor, being inherently a constant speed device, is not suitable for use with most servomechanisms and; therefore, a dc motor is usually preferred.

There are two methods of controlling the speed and direction of rotation of the dc motor; controlling the voltage applied to the field coils and controlling the armature voltage. Controlling the field voltage permits control of speed from a certain minimum speed upward only. However, reduction of the motor speed to zero is necessary at times; hence, the control in the servosystem is always accomplished by use of the armature current.

(2) Servomotors operate with constant fields. The field is supplied either by permanent magnets or by dc applied to the field coils. The method of applying dc voltage to the field coils is wasteful of power, since power is consumed even when the motor is not turning. In addition, the design of the equipment is much more complicated than the permanent magnet arrangement. For this reason, the permanent magnet is more widely used.

(3) One problem in the use of permanent magnets is the danger of their demagnetization by the field, set up by current flow in the armature. This is overcome by the inclusion of a special compensating winding that is connected in series with the armature and wound in the opposite direction. A compensating field is set up by the flow of current through this winding. Since the current in the compensating winding varies with the armature current, the compensating field balances out the armature current and protects the permanent magnets.

i. *Synchro Generator.*

(1) This unit, sometimes referred to as a synchro transmitter, consists of a rotor which carries a single winding and a stator made up of three windings displaced 120 degrees. The rotor is supplied from an ac source and is usually coupled, either directly or through gears, to the controlling shaft.

(2) The voltages induced in the stator winding, as a result of the alternating field set up by the rotor windings, are representative of the angular position of the rotor at any time.

j. *Control Transformer.*

(1) Synchros cause two shafts to rotate in synchronism and produce an error voltage that indicates the difference in position of two shafts. The synchronous generator and motor are used for the express purpose of establishing synchronism. For the purpose of producing error voltages, a type of synchro, known as the control transformer, is used in conjunction with the synchro generator. The generator and the transformer have their shafts connected to a load. The voltage, induced in the rotor of the control transformer, results in a voltage whose magnitude and polarity depends on the relative position of the two shafts.

(2) The stator of the transformer is similar to that of the generator, with the stator current being determined by the impedance of the stator's windings, and being practically independent of the rotor's position. All of the rotor windings are in series, with their external connections completed through two slip rings and brushes. The shaft is always connected to a load making an oscillation damper unnecessary.

k. *Measurement of Error.*

(1) Among the components necessary to a closed loop system are those devices which measure the difference between the actual output and the desired output. It is important that this difference, or error, be presented in a manner compatible with the other components in the control system. Thus, the location of equipment, as well as the physical type of error signal and its transmission are important. For example, a mechanical differential is rarely used as a device to obtain the difference between the actual output and the desired output, because it is often impractical to realize these two quantities as physically adjacent shaft rotations.

(2) Any error-measuring system selected must possess an inherent accuracy greater than that required of the overall loop. Often, the static accuracy and the dynamic accuracy of the system must be examined. Many error-measuring devices produce noise of such a nature that the component frequencies are proportional to the rate of change of the input and output quantities. The response of the loop to this noise must be considered.

(3) In addition to such noise, error-measuring systems using ac voltages commonly have present in the final voltage not only a voltage proportional to the error (with a fixed phase shift from the excitation voltage), but also a voltage proportional to the rate of error. Harmonics of both voltages are usually present. The phase shift of the error voltage, if constant for the class of device, is not a serious problem, but must be considered in the design of the other components.

(4) The components of the error-measuring system must be strong enough to fulfill their functions for a sufficient length of time to withstand the velocities and accelerations of the variables being transmitted and operate at the temperatures expected. Because of the ease with which their signals may be transmitted and the resultant freedom in the placement of the equipment, electrical synchro devices have had a wide application as error-measuring devices.

(5) In the satellite tracking system the positioning of the antenna is accomplished by using the azimuth and elevation error signals, which are derived from the tracking receivers. The error-signal voltage from the output of the receiver is fed into an antenna-error servoloop. The signal is demodulated and amplified

to activate a motor, which drives a differential control transmitter and a synchro control transmitter (via a power servo). These control transmitters, in turn, control the positioning of the antenna through another servoloop. The measurement and use of the azimuth and elevation error signals is one of the primary examples of the application of servo theory and servomechanisms in the satellite communications system.

(6) The error-measuring system selected must possess an inherent accuracy greater than that required of the overall loop. This is one of the reasons why ground station antennas will frequently use a digital readout device as a part of the servoloop during the acquisition phase. A digital readout is obtained by attaching a code wheel to one of the antenna shafts. A code wheel is a device that is marked, or engraved, in some sort of a sequence. If one looks at the code wheel through a narrow slit, a different arrangement of the markings will appear for different positions of the code wheel. It is possible; therefore, to scan along the slit with some optically sensitive device such as a photocell and produce a series of pulses arranged in a code form which will vary with the shaft position. This gives a highly accurate digital readout which can be fed into a digital computer and compared with the signal that directed the antenna. If they are different, the computer can generate an error signal to turn the antenna in the direction necessary to reduce the error. During the acquisition phase the signals to direct the antenna will normally be derived from a tape that contains the computed path of the satellite in digital form. Once the satellite is acquired, the tracking receivers will generate the error signals and the acquisition loop is disengaged.

CHAPTER 9

SOLID STATE TECHNIQUES

9-1. GENERAL.

a. Solid state electronics is the name applied to a general class of electronic devices that are characterized by the use of materials which are neither conductors nor insulators, but fall into a group somewhere between the two extremes and are called semiconductors. Semiconductors are not a new development and have been used for years as rectifiers and signal detectors in crystal radios. With the birth of the atomic age in the early forties, increased interest in the physical properties of elements brought about a union between solid state physics and the field of electronics. One outcome of this union was the invention of the transistor, a solid state device that could be made to amplify signals.

b. In 1948, Bardeen and Brattain, of the Bell Telephone Laboratories, announced the discovery of the transistor and revolutionized the electronics industry. The supremacy of the vacuum tube was destined to be toppled by the growth of the transistor's popularity. Since the transistor is smaller, simpler, more efficient, longer lived, and more rugged than a vacuum tube, a whole new field of electronics was opened. This is the field of solid state electronics.

c. The transistor and other solid state devices have made possible the present-day electronic age, with its computers, space vehicles, and high-speed data communications. There are examples on every hand of how the state-of-the-art has been advanced with the advent of solid state electronics. Being relatively small, semiconductor devices lend themselves to miniaturization techniques that result in compact and lighter equipment. This is a primary requirement for components designed for use in satellites.

d. One of the few drawbacks to the use of semiconductor devices has been their low power outputs. For many applications low power levels are sufficient, but for many rf applications high powers are required. In the high power area thermionic devices are still widely used. The heat and frequency limitations of solid state devices have been virtually eliminated. The use of silicon in place of germanium has increased the temperature range of devices, and new fabrication techniques have pushed the operational range of transistors into the uhf range. Bulk effect devices, such as the Gunn diode, have been used up into the gigahertz range.

9-2. INTRODUCTION TO THE ATOM.

a. In order to understand the workings of solid state devices, one must be familiar with the basic concept of atomic structures. It is generally accepted as fact that all matter is made up of many atoms, which are joined in a complex structure. The atom is made up of a nucleus and one or more electrons, which revolve or orbit around the nucleus. The number of particles in this nucleus, the atom's weight, the number of electrons in orbit, and the electrical charge in the nucleus, vary in different elements and are the basic differences which distinguish one element from another.

b. The nucleus of an atom consists of protons and neutrons surrounded by a sufficient number of planetary electrons to maintain the atom's electrical neutrality. A proton has a charge of $+q$, an electron a charge of $-q$, and a neutron has a zero charge ($q = (1.6)10^{-19}$ coulomb). The atom also contains subatomic particles such as positrons, neutrinos, and mesons. For an analysis of semiconductor electrical properties, considering only the protons, electrons, and neutrons is sufficient.

c. A diagram of a normal hydrogen atom which consists of a nucleus containing one proton with a charge of $+q$, and a planetary electron with a charge of $-q$ is shown in part A of figure 9-1. Notice that the normal hydrogen atom does not contain any neutrons. Part B of figure 9-1 shows a helium atom. A helium atom has a nucleus containing two protons; hence, an electrical charge totaling $+2q$, two neutrons with zero charge and two planetary electrons with an electrical charge totaling $-2q$. Since these two sets of electrical charges are of opposite polarity, the total electrical charge of the atom is zero making the atom electrically neutral.

d. The atoms of different elements have characteristics which make each atom different from the atoms of every other element. In addition to the diagrams representing the hydrogen and helium atoms, figure 9-1 contains some additional data which describes these differences.

e. First, there is the atomic weight. The atomic weight of an atom is defined as the weight of the atom relative to the weight of an oxygen atom. Referring to a table of atomic weights, hydrogen has an atomic weight of 1.008. The atomic weight of an oxygen atom has been taken as 16. Using this as a basis, and since the atomic weight of hydrogen is approximately one, the hydrogen atom is approximately 1/16 the weight of oxygen. As a further example, the atomic weight of sulphur is given as 32.06 which is approximately twice the weight of oxygen.

f. The second item of descriptive data is the atomic number. The atomic number of an atom is defined as the number of positive charges in the nucleus.

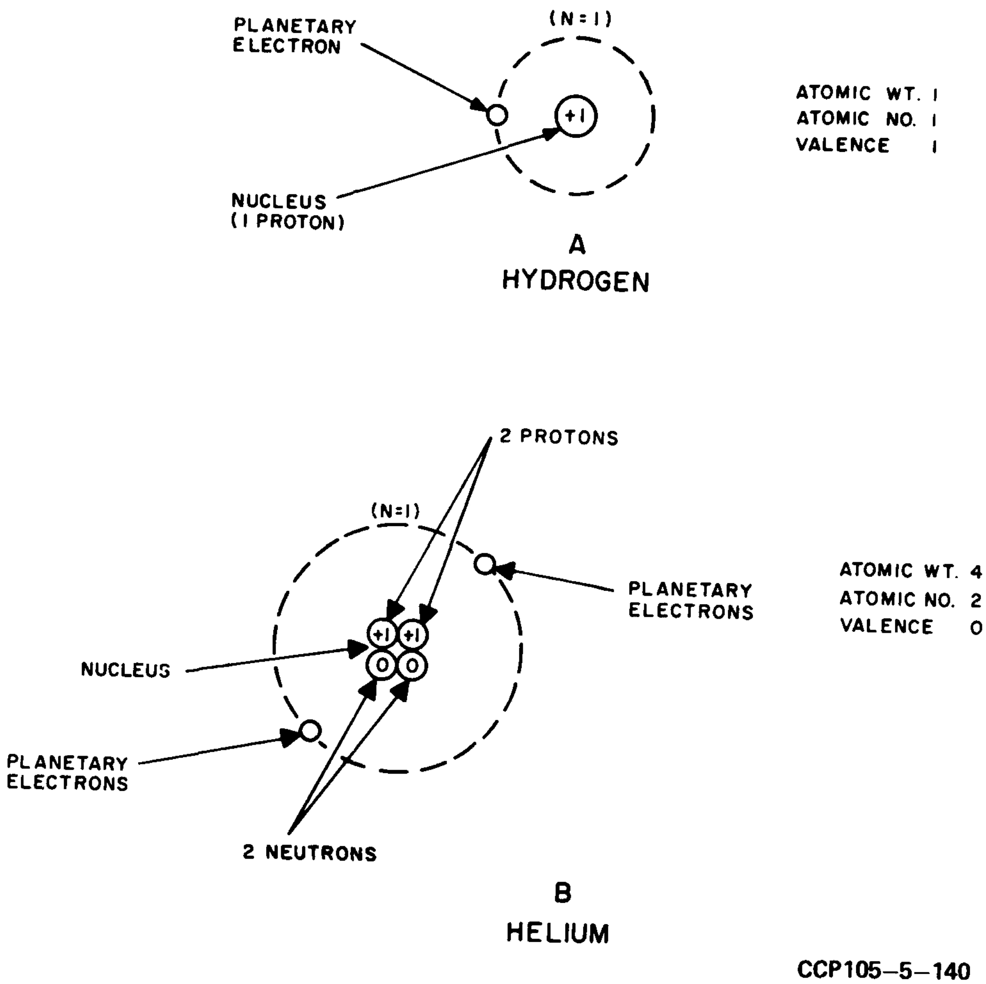


Figure 9-1. Schematic diagrams of hydrogen and helium atoms.

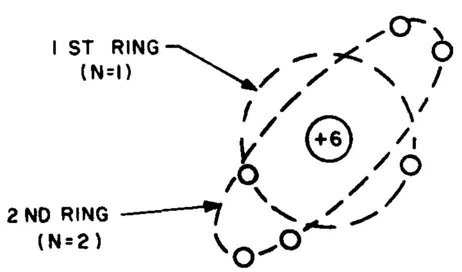
In the case of hydrogen the nucleus has one proton; therefore, the atomic number is 1. Similarly, helium with two protons has an atomic number of 2.

g. The third descriptive characteristic is the valence of the atom. Valence can be shown to be the number of electrons in the atom that are available for bonding with other atoms. Two common types of bonding are ionic, where electrons are transferred from one atom to another, and covalent, where electrons are shared in common by the atoms.

h. A schematic diagram of a carbon atom is illustrated in figure 9-2. It can be seen that this is a more complex atom than the hydrogen atom. Note that the electrons are distributed in two rings around the nucleus. This is an important observation because the distribution of electrons in each of the rings is not a random distribution, but governed by the principal energy state, which is designated by the letter N and is called the principal quantum number. For a given principal energy state there is a maximum number of electrons which can exist. For example, in the state specified as N = 1, as in helium, the first ring can contain a maximum of two electrons. For N = 2, as in carbon, the second ring can contain a maximum of eight electrons.

i. An example of how quantum states affect valence can be seen by looking at the hydrogen and carbon atoms. The principal quantum number for hydrogen is N = 1, permitting a maximum of two electrons in the first ring. But looking at part A of figure 9-1 we see that the hydrogen atom has only one planetary electron. Since two are permitted the hydrogen atom can combine with one more hydrogen atom making a hydrogen molecule and this ability is called the atom's valence. The carbon atom's (fig. 9-2) valence is 4. Since only four electrons occupy the second ring, there is room for four more and the carbon atom readily combines with other atomic structures. Other elements (such as helium) which have a valence of zero and contain the maximum number of electrons stated by the principal quantum number, do not react under normal conditions. Elements of this nature are called inert.

j. Atoms which do not have complete outer rings can combine chemically. All elements try to complete their outer rings. As an example, an atom of hydrogen will strive to combine with other atoms in order to complete the rings. In molecular hydrogen the atoms are bonded together each sharing the other's electrons, which satisfies the principal energy state. The hydrogen molecule (H_2) is a stable gas.



ATOMIC WT.	12.01
ATOMIC NO.	6
VALENCE	4
ELECTRONS	6
PROTONS	6
NEUTRONS	6

CCP105–5–141

Figure 9-2. Carbon atom schematic diagram.

9-3. INSULATORS, SEMICONDUCTORS, AND CONDUCTORS.

a. Under the proper conditions the carbon atom will join other carbon atoms in a covalent bond arrangement to form a crystal structure. In the covalent bond each carbon atom will share its electrons with four neighbors so that the N = 2 energy state is filled. A diamond crystal arrangement results in strong bonds between atoms that leaves no electrons available for conduction. In all insulators most of the electrons remain bound to the constituent atoms. This means that insulators will have no appreciable electrical conductivity since very few charges are able to move through the material.

b. For good conductors, the number of free electrons is very high. In most metals each atom supplies one free electron for bonding. This puts the number of free electrons per cubic centimeter at about 10^{23}. Since these electrons are shared by all the atoms, they are free to wander throughout the substance. This allows the material to conduct easily when an electric field is applied.

c. Semiconductors fall somewhere between the extremes of insulator and conductor. Typically, the resistivity of a semiconductor of standard cross-section and length ranges from 10^{-4} to 10^{9} ohm/cm. In comparison an insulator will have a resistivity range of 10^{9} to 10^{25} ohm/cm and a conductor a resistivity of 10^{-5} ohm/cm. The structure of a germanium crystal (resembles silicon crystal) is the same as the diamond structure of crystaline carbon. Both have four valence electrons in the outer shell, but these electrons are not bound as tightly as in the carbon crystal. In carbon the valence electrons are in the second ring (shell); in silicon the valence electrons are in the third shell, and in germanium the valence electrons are in the fourth shell. Because the valence electrons in silicon and germanium are farther from the nucleus, they are not bound as tightly to the atom. This allows the semiconductor to have a moderate amount of electrons available for conduction.

9-4. ELECTRONS AND HOLES.

a. The study of semiconductors reveals two types of carriers, electrons and holes. Carriers are charged particles that can be moved to change the electrical state of a material. The movement of electrons and holes imparts heat through thermal energy to a valence electron and causes break-away from the valence band. The free electron is now available for conduction. The motion of the electron creates a vacancy in the valence band which is called a hole and has a positive charge. The process of generating electron-hole pairs is continuous because electrons and holes recombine by a mechanism similar to their generation.

b. When an electron is raised to the conduction band (valence band) it leaves a vacancy which can be filled with an electron from an adjacent atom, which develops a hole in its structure. This motion has the effect of the hole moving from one atom to another. The resultant effect is as though a positive charge moves across the semiconductor to the external lead. One electron flows from the lead to neutralize the hole. Thus, a hole can move across a valence band by a replacement process. When holes are present in the valence band, electrons can change their energy state and conduction is possible by hole movement. Also, electrons in the conduction band can change their energy state and conduction is also possible by electron movement. Therefore, conduction within a semiconductor is caused by movement of positive (hole) and negative (electron) carriers.

c. To summarize this concept, when an electron moves from one point to another in order to fill a hole in the valence structure, the net effect is the movement of a positive charge in the opposite direction. However, when an electron moves from one point to another without filling a hole, the net effect is the movement of a negative charge in the same direction as the electron motion.

9-5. SEMICONDUCTORS WITH IMPURITIES ADDED.

a. The crystal structure of certain elements can be modified to the extent that their characteristics can be changed for some particular application. In the case of semiconductors, we do this by adding impurities to the crystal structure. An intrinsic semiconductor is one to which no impurity has been added.

b. When the semiconductor has been doped with impurities, conduction takes place primarily by one type of carrier. When there are more holes, the holes are called the majority carrier and the electrons the minority carrier. If there are more electrons, the electrons are the majority carrier and the holes are the minority carrier.

c. Two common types of impurities are arsenic and gallium. They are added to the semiconductor crystal during manufacture to produce an abundance of either electron or hole carriers.

d. Germanium and arsenic have atomic numbers of 32 and 33, respectively. When arsenic goes into a covalent bond with germanium, one electron of each arsenic atom is not included in the covalent bond. This group of loosely bound electrons is available for conduction; thereby, providing a large number of electron carriers.

e. When gallium, with an atomic number of 31, goes into a covalent bond with germanium, there is one covalent bond which is not complete. This provides a vacancy into which an electron from the valence band can move; thus, providing an excess of holes available for conduction. Electrons try to fill the valence band of the gallium atoms and leave a large group of hole carriers.

f. The impurity material that is added to semiconductors to create a majority of electron carriers is called a donor impurity, and the extrinsic material which results is called N type. The impurity material that is added to semiconductors to create a majority of hole carriers is called an acceptor impurity, and the extrinsic material which results is called P type. Materials suitable as acceptors are boron, aluminum, gallium, and indium. Materials suitable as donors are phosphorus, arsenic, antimony, and bismuth.

9-6. PN JUNCTION.

a. A PN junction is made up of P and N type semiconductor material. The manufacture of a junction is a complex procedure; however, for simplicity this discussion will assume that the P and N materials can be brought together to form a PN junction. The width of the semiconductors is the same in each case. The number of electrons in the N type material is equal to the number of holes in the P type material. Each material is electrically neutral.

b. When the materials are brought into contact to form a PN junction, a concentration gradient of electrons and holes exists. More holes are located at one end and more electrons at the other end. Since the semiconductor will strive for a neutral condition, a process of diffusion starts. This is much like the action which occurs when ink is poured into a glass of water. The holes diffuse from the P material into the N material, and the electrons diffuse from the N material into the P material. Therefore, an initial diffusion current is flowing as a result of unequal concentrations between the P and N materials.

c. This process continues until the donors and acceptors near the junction have lost their compensating carriers and a potential barrier is set up, which tends to oppose any further diffusion currents. Thus, an equilibrium condition is reached wherein current across the junction is zero. Since some of the majority carriers (electrons in N type and holes in the P type) have been effectively cancelled, the semiconductor materials at the junction assume a positive charge in the N type and a negative charge in the P type.

d. It has been stated that each material is electrically neutral and the majority carriers were equally distributed throughout the material. This is no longer true as a junction has been formed. The electrons in the N material are repelled from the junction by the negative charge in the P material and the holes are repelled by the positive charge of the N material. These majority carriers maintain a position away from the junction. In order for further conduction to take place an electron or hole carrier must acquire sufficient energy to climb this barrier. This energy may come from radiation in the form of heat, light, X-ray, battery, or power supply.

e. Assume that a battery has been connected across a PN junction in such a manner that the voltage aids the movement of carriers across the potential barrier. This is called forward bias. To obtain forward bias connect the negative side of the battery to the N type material and the positive side of the battery to the P type material.

f. Figure 9-3 illustrates a PN junction to which a battery has been connected to apply forward bias. Since the battery is connected with the negative pole to the N terminal, and positive to the P terminal, it adds energy to the majority carriers and assists them across the barrier. Thus, we have a current flow across the junction which will result in a current flow in the external circuit.

g. If we change the battery connections, so the negative side of the battery is connected to the P terminal and the positive side is connected to the N terminal, we will increase the barrier; thus, pulling the carriers further away from the junction. This is called a reverse bias and no current flows in the circuit.

h. If the battery is replaced with an ac signal, rectifier action results. That is, during the positive half cycle, when the voltage forward biases the junction current will flow in the circuit. During the negative half of the cycle the voltage polarity is reversed and no current flows in the circuit. Therefore, we have current flow in one direction and have effectively converted ac to a pulsating dc voltage. This process is called rectification.

i. Figure 9-4 shows the input and output waveforms of a PN junction diode circuit.

9-7. JUNCTION TRANSISTOR.

a. The action of a simple PN junction has been discussed and this concept shall be used to explain the transistor action which makes this solid state device useful.

b. A junction transistor is made up of two PN junctions. In order to make a transistor, one element of each junction must be shared in common. To further explain this phenomenon assume that there are two PN junctions as shown in figure 9-5.

c. If connected back-to-back, they will not yield any transistor action. If constructed in the manner shown in figure 9-6 (a common element), we have a

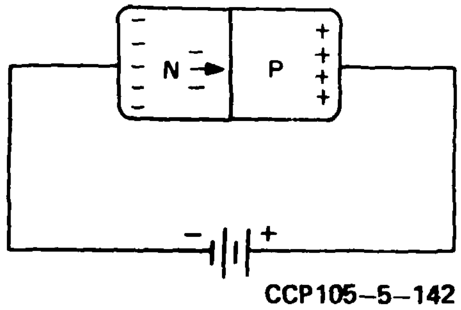


Figure 9-3. Forward biased PN junction functional diagram.

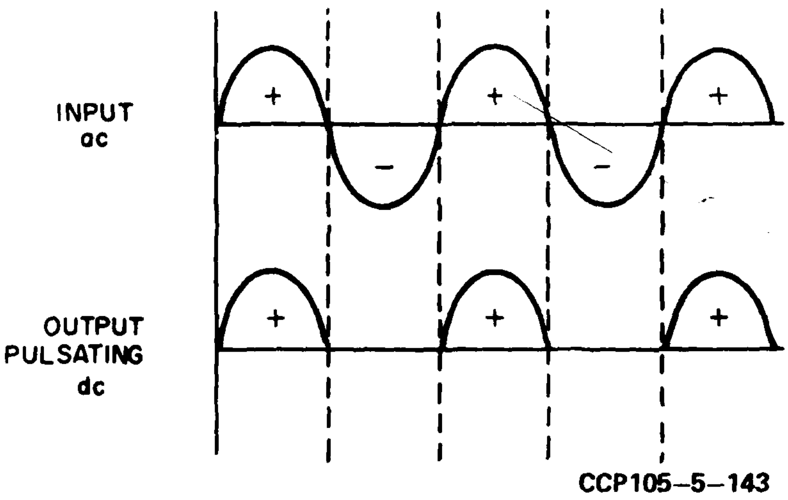


Figure 9-4. Input and output voltage waveforms half-wave rectification.

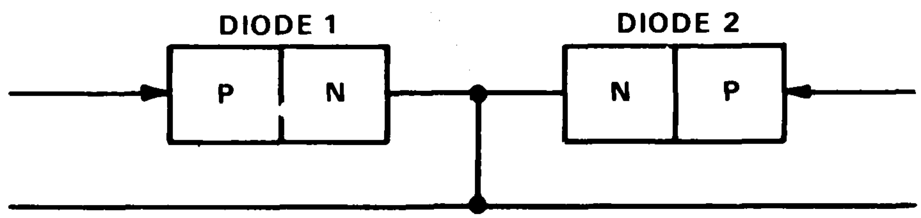


Figure 9-5. Two diodes connected back-to-back functional diagram.

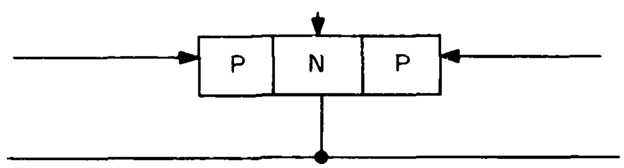


Figure 9-6. Two diodes with common elements functional diagram.

transistor which is capable of amplifying a signal. In practice, the common element can be either P or N type material.

d. Figure 9-7 shows the schematic representation for a PNP transistor. The common element (base electrode) is shown as a vertical line, intersected by a horizontal line. The emitter electrode, shown on the left, is for P type material and is represented by a diagonal with an arrowhead into the base.

e. The collector electrode (second P type material) is shown as a diagonal line which intersects the base at an angle, and in figure 9-7 appears on the right side of the schematic. Regardless of the attitude in which it is drawn, the diagonal with the arrowhead is always the emitter electrode and diagonal without the arrowhead is always the collector electrode.

f. Voltage can be measured as a difference of potential between two points. To signify these points, a sign convention has been adopted for specifying the

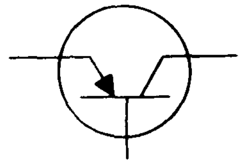


Figure 9-7. PNP transistor symbol.

polarity of bias, signal potentials, and current at the electrodes of a transistor. Figure 9-8 illustrates these symbols and the areas they refer to as related to the example.

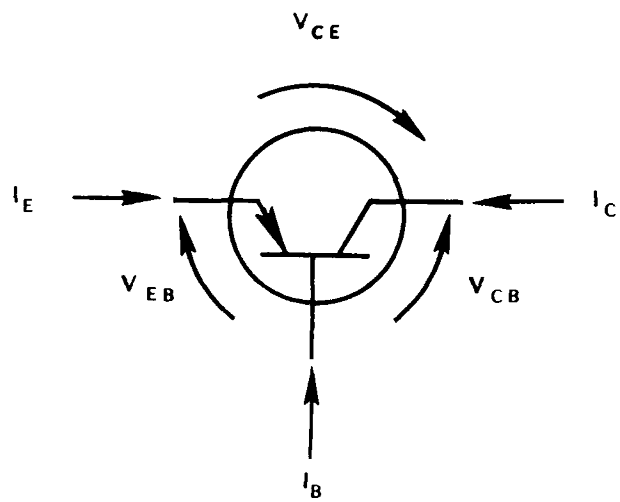
g. Note that uppercase I and V are used for dc values, and lower case i and v are used for time varying (TVNG) currents and voltages. To obtain amplification, the transistor must be connected in a circuit and biased. Forward bias of the emitter-base junction and reverse bias of the collector-base junction is shown in figure 9-9.

h. In this typical transistor circuit, holes are diffused through the base from emitter to collector. The emitter-to-base resistance is very small in comparison to the 3-kilohm series resistor; therefore, the emitter current I_E can be calculated by Ohm's law: $I_E = E/R = 3/3000 = 0.001$ ampere, or a 1-milliampere (mA) current. Notice that the collector current is approximately the same value. The difference is the base current, which is caused by a number of holes being filled by electrons in the N material. The collector circuit contains a 10-kilohm load resistor.

i. Using Ohm's law the voltage across the 10-kilohm resistor is 0.00099 x 10,000 = 9.9 volts. Since the battery is connected opposing this voltage V_{CB} is −10.1 volts.

j. When an ac signal source is added to the emitter input and varies the emitter current plus and minus 0.5 mA, the following action occurs. As I_E is increased to 1.5 mA, I_C will increase to 1.48 mA, and as I_E is decreased to a 0.5 mA, I_C will follow to 0.5 mA.

k. Since the input circuit (emitter) is of low impedance and the output circuit (collector) is a high impedance circuit, a voltage gain and a power gain is achieved by the transistor circuit. This gain is termed transistor action and can be defined as the ability of the emitter current to control the current in the collector.



DC	DC TVNG	
I	i	CURRENT
I_E	i_E	EMITTER CURRENT
I_C	i_C	COLLECTOR CURRENT
I_B	i_B	BASE CURRENT
V	v	VOLTAGE
V_{CE}	v_{CE}	COLLECTOR EMITTER VOLTAGE
V_{EB}	v_{EB}	EMITTER BASE VOLTAGE
V_{CB}	v_{CB}	COLLECTOR BASE VOLTAGE

CCP105–5–146

Figure 9-8. Current and voltage conventions for PNP transistors.

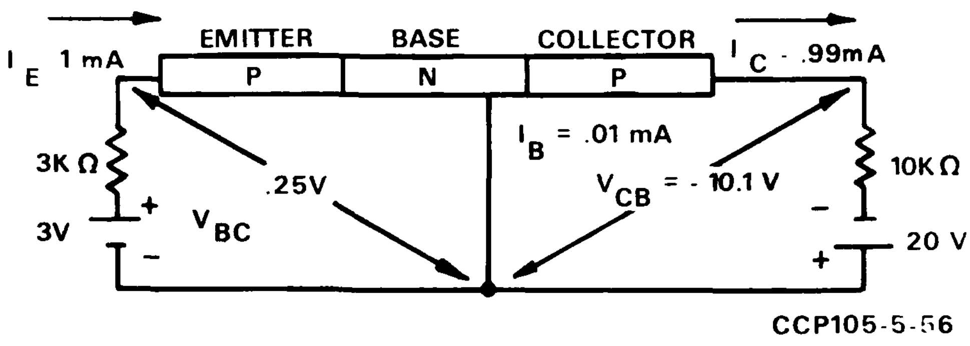


Figure 9-9. PNP transistor with collector and emitter biased functional diagram.

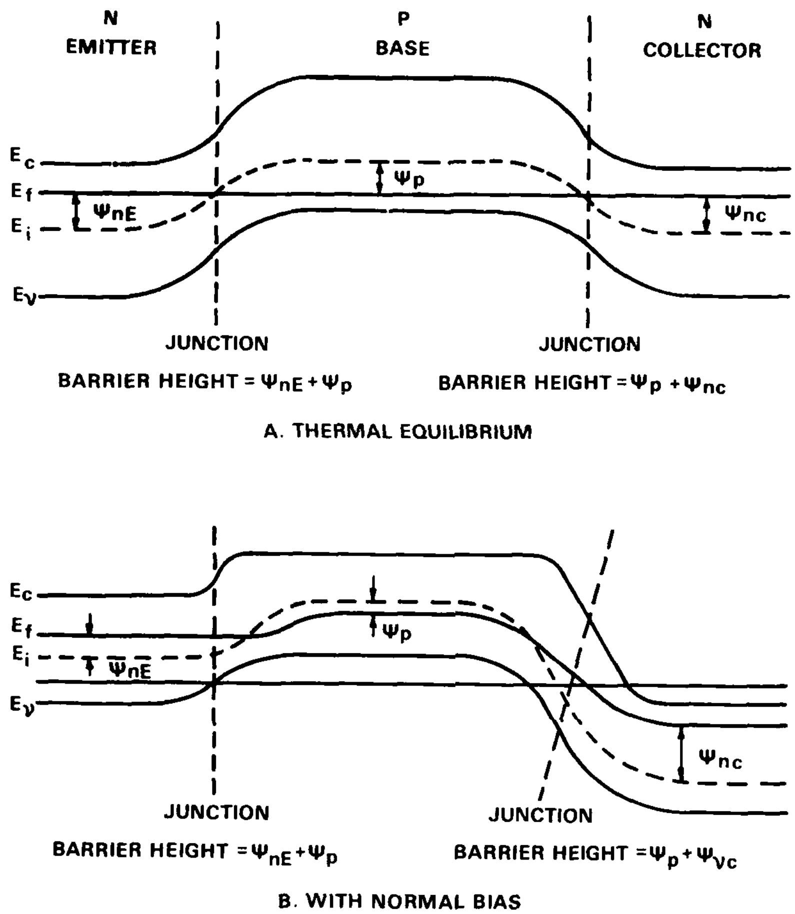
9-8. TRANSISTORS.

a. There are many models used today to explain the operation of semiconductors. The theory has been simplified by omitting the mathematical analyses and reducing the technical language to common words. However, in this section a semitechnical analysis is necessary to explain the workings of a transistor.

b. Once thermal equilibrium has been achieved in a junction no further carrier motion exists, unless energy from an external source is applied.

c. Figure 9-10 shows two energy level curves for the majority carriers (electrons) in an NPN transistor. Curve A shows the energy levels at thermal equilibrium. In order for electrons to exist in the P material, they must have a much higher energy level; whereas, electrons can exist in the N material at a lower energy level. Therefore, electrons must climb a potential "hill" in order to reach the P material and, as a result, the P material contains few electrons. In this state, no current flows across either the emitter or collector junctions, since the energy barrier effectively prevents the carriers in the N material from passing into the P material.

d. The B curve, in figure 9-10, shows the energy distribution across a normally biased transistor; that is, one with forward bias across the emitter base junction and reverse bias across the collector-base junction. Note that the difference in potential energy across the emitter-base junction is much less now than in the thermal equilibrium state. In this biased condition, a small increase in V_{EB} (emitter-to-base voltage) would cause the carriers to break across the barrier into the P material. With an abundance of electrons flowing into the base material let us look at the collector N material and see how how these electrons are transferred across the collector-base junction.



E_c	CONDUCTION BAND ENERGY LEVEL
E_v	VALENCE BAND ENERGY LEVEL
E_f	FERMI ENERGY LEVEL
E_i	INTRINSIC ENERGY LEVEL
Ψ_{nE}	CONTROL POTENTIAL EMITTER DEPLETION REGION
Ψ_p	CONTROL POTENTIAL BASE DEPLETION REGION
Ψ_{vc}	CONTROL POTENTIAL COLLECTOR DEPLETION REGION

CCP105-5-57

Figure 9-10. Energy level curves of NPN transistor.

e. The collector-base junction is reverse biased. This can be seen by the larger difference in energy across the junction. However, the application of reverse bias causes the N material to become highly positive, creating a strong attraction for negative carriers. As a result of this attraction nearly all the electrons from the emitter are pulled to the collector and are replaced by electrons from the external circuit. A very small number of holes are neutralized in the base material and cause a base current to flow, which is in the order of one percent of the total emitter current.

f. Therefore, it can be stated that $I_E = I_B + I_C$. That is, the total current from the emitter equals the base current plus the collector current. From this it is determined that the collector current will be slightly less than the emitter current. The transistor data sheets issued by the manufacturers for individual transistors will refer to emitter-to-collector current gain. This is signified by the Greek letter alpha (α) and is less than one. The value of alpha for junction transistors is in the order of 0.95 to 0.998. For example, if the alpha of a given transistor is 0.95, and the emitter current is 10 mA, the collector current is 10 X 0.95 or 9.5 mA. It follows that the remainder of the emitter current must appear in the base circuit; therefore, 10 −9.5 = 0.5 mA, the base current in this hypothetical case.

g. Since an ordinary transistor has three electrodes, there are three possible ways of connecting it into a circuit. In previous discussion the base has been common to the emitter and collector circuits. This is called, appropriately, the common base circuit and is shown in part A of figure 9-11. This configuration has a low input impedance and the phase of the signal remains unchanged. This circuit is very much like a grounded grid vacuum tube circuit.

h. Part B of figure 9-11 is a schematic of a common emitter transistor amplifier. Its vacuum tube counterpart is the grounded cathode amplifier. Like its counterpart, the common emitter amplifier, it is the most commonly used circuit in everyday practice. It requires only one battery supply if the proper emitter resistor is chosen and the input and output impedances are of medium value. In operation, the input signal is applied between the base and emitter and the output is taken from the collector-to-emitter circuit. The direction of current flow in the collector circuit is from the positive side of the battery through the transistor and through the load resistor, as shown. This results in a dc voltage across R_L, which is negative at the bottom. When a positive input signal is applied to the base, it will oppose the base-to-emitter bias. Because of this action, the hole current is reduced and the current in the output circuit is reduced. When the input signal goes negative, the opposite effect takes place and the output signal increases; the input signal is amplified at the output, but is of opposite polarity. This condition is called a 180-degree phase reversal and is also a characteristic of vacuum tube grounded cathode amplifiers.

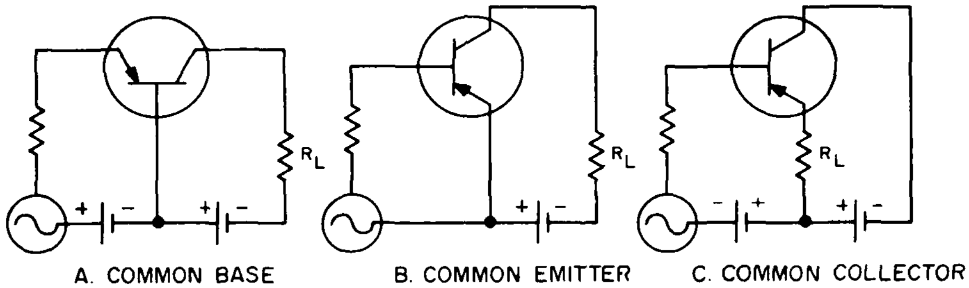


Figure 9-11. Three common transistor amplifier circuits schematic diagram.

i. Following the same line of reasoning for the common collector circuit (part C of fig. 9-11), we can see that, although the input and biasing are the same as the common emitter, the output is taken from the emitter circuit. The application of the input signal is in the same direction as the battery voltage and, therefore, no phase change takes place. When the input signal swings positive so does the output and vice versa. This transistor circuit is similar to a cathode follower vacuum tube circuit and many of their characteristics are the same. It has no phase reversal, a voltage gain of less than 1, and a high impedance input with low impedance output. Table 9-1 lists the characteristics of the three types of configurations.

9-9. NPN AND PNP TRANSISTORS.

a. Another arrangement of semiconductor material is the PNP transistor. The biasing of a PNP transistor is the same as that of the NPN. The emitter-base circuit is forward biased and the collector-base circuit is reverse biased. This requires that the voltages, used to operate a PNP amplifier, be the reverse of those used with an NPN amplifier. Characteristics and energy levels are identical for NPN and PNP transistors. The basic difference is that holes are the majority carriers in PNP devices and electrons are the majority carriers in NPN devices.

b. The schematic symbol representing an NPN transistor differs only in the direction of the arrowhead. The arrowhead points away from the base on the NPN transistor. The arrowhead points toward the base on the PNP transistor.

c. These two types of transistors are frequently used in circuits to complement one another, resulting sometimes in a reduction of the number of components required. The availability of complementary types of transistors allows

Table 9-1. Transistor amplifier configurations characteristics

Characteristic	Common base	Common emitter	Common collector
Voltage gain	Medium 100 - 200	High 300 - 600	Less than 1
Current gain	Less than 1	Medium 20 to 100	Medium 20 to 100
Power gain	Medium	High	Low
Input impedance	Extremely low	Low	Extremely high
Output impedance	Extremely high	Medium	Extremely low
Phase change - input to output	0 degrees	180 degrees	0 degrees

additional freedom in the design and simplification of circuitry. Figure 9-12 is an illustration of a complementary symmetrical circuit.

d. The arrows in figure 9-12 indicate the direction of current flow. Note that a single battery supplies the power for the three transistors in the normal bias arrangement and only three additional components are required.

9-10. TRANSISTOR APPLICATIONS.

a. This paragraph and paragraphs 9-11 and 9-12 concern some of the common applications of transistor theory.

b. Figure 9-12 illustrates a direct coupled amplifier using three transistors. The main advantage of a direct coupled amplifier is that it avoids the need for transformers and coupling capacitors and can amplify signals at frequencies all of the way down to direct current. Other commonly used coupling methods (fig. 9-13) are resistor-capacitor coupling, transformer coupling, and impedance coupling.

c. Another commonly encountered circuit is the oscillator. An oscillator can be described as an amplifier in which a portion of the output energy is returned to the input to produce a recurrent cyclic output. Compare an oscillator to a pendulum. When a pendulum is set in motion each successive swing will get smaller until the pendulum finally stops. However, a push each time the pendulum swings will cause it to continue swinging.

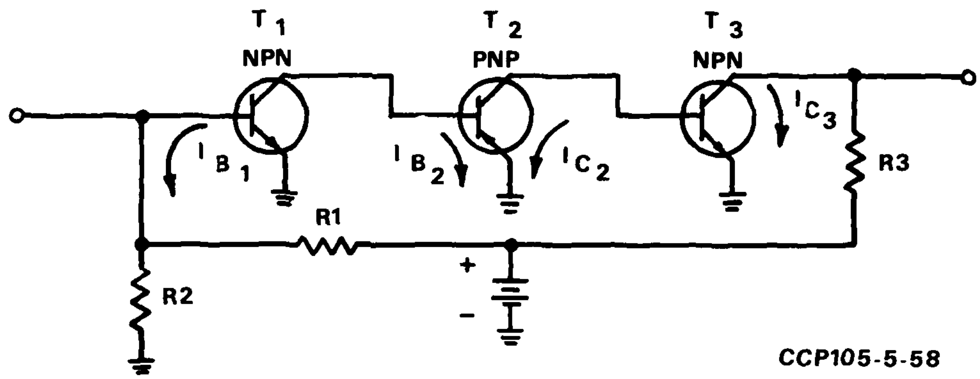


Figure 9-12. Simplified dc amplifier schematic diagram.

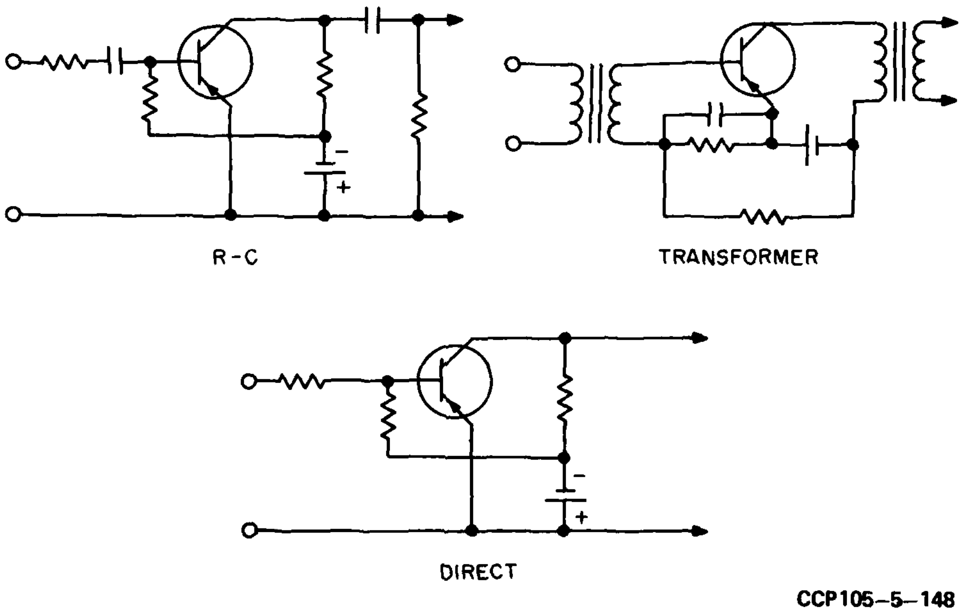


Figure 9-13. Coupling methods of af amplifier schematic diagram.

d. An electronic oscillator functions in much the same way. If a small portion of the output signal is fed back into the input at the right time an oscillation is set up that continues as long as the feedback exists. Figure 9-14 shows three common transistor oscillators.

e. Since an oscillator is basically an amplifier, each basic amplifier configuration can be made to oscillate with the addition of one of the accepted feedback configurations.

f. In computer circuitry one of the basic building blocks is the multivibrator. There are several types that have been developed from the original Eccles-Jordan multivibrator circuit. Two of these are the one-shot multivibrator (monostable) and the free running multivibrator (astable).

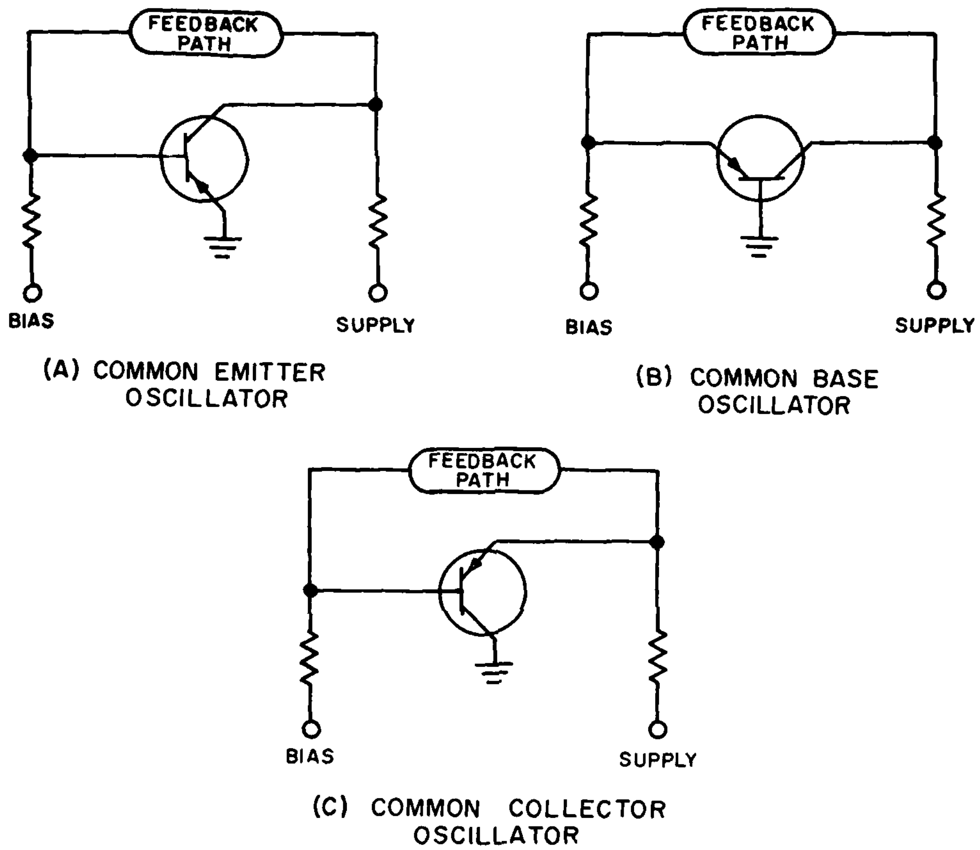
g. A multivibrator is a circuit that can be called an electronic switch. It has only two output levels, an on voltage and an off voltage.

h. A one-shot multivibrator is shown in figure 9-15. With no input trigger, transistor Q_2 conducts because of the positive voltage across the base resistor (R_b). Transistor Q_1 is cut off by its bias voltage across resistors R_1 and R_2. A negative trigger pulse causes Q_2 to turn off and Q_1 to turn on. Transistor Q_2 remains off until capacitor C_1 discharges bringing the base voltage of Q_2 to approximately that of the emitter voltage. At this point Q_2 starts to conduct and the circuit returns to its original state. The effect in the output is that of a switch being closed for a specific time in a dc loop.

i. Figure 9-16 illustrates the equivalent dc circuit and the resulting waveform. Consider the circuit again at time T_0, transistor Q_2 is on and Q_1 is off; at time T_1, transistor Q_2 is off and Q_1 is on. At time T_2, transistor Q_2 is again turned on and Q_1 is off. The circuit will remain in this state until another trigger pulse is injected.

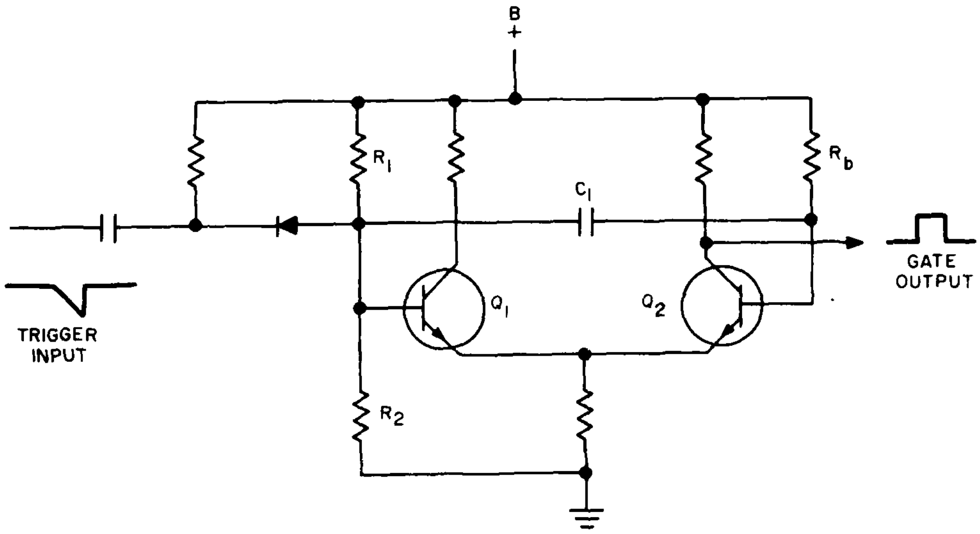
j. A bistable multivibrator, or flip-flop, is a similar configuration; however, the action is slightly different. A bistable multivibrator is shown in figure 9-17. When a trigger pulse is applied, transistor Q_2 will turn off and Q_1 will turn on. The flip-flop remains in this condition until another trigger pulse is applied, then Q_2 turns on again and Q_1 is turned off. Both types of multivibrators are widely used in computers as gate or switch circuits.

k. Just as the bistable multivibrator had two stable states and the monostable one stable state, the astable multivibrator has no stable states. It is described as an oscillator that produces a square or rectangular output. A schematic of the astable multivibrator is shown in figure 9-18.



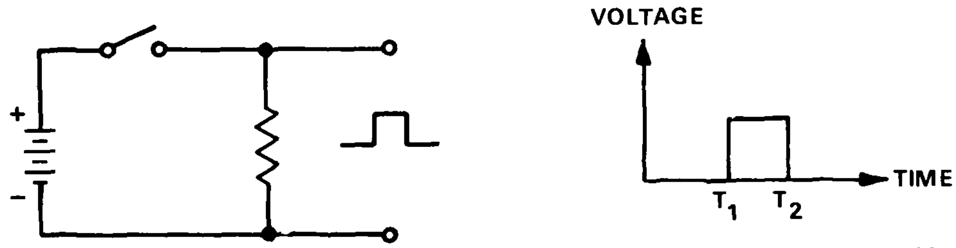
CCP105–5–149

Figure 9-14. Three basic oscillator configurations functional diagram.



CCP105-5-60

Figure 9-15. One-shot emitter coupled monostable multivibrator schematic diagram.



CCP105-5-61

Figure 9-16. Equivalent dc circuit and output waveform of one-shot multivibrator.

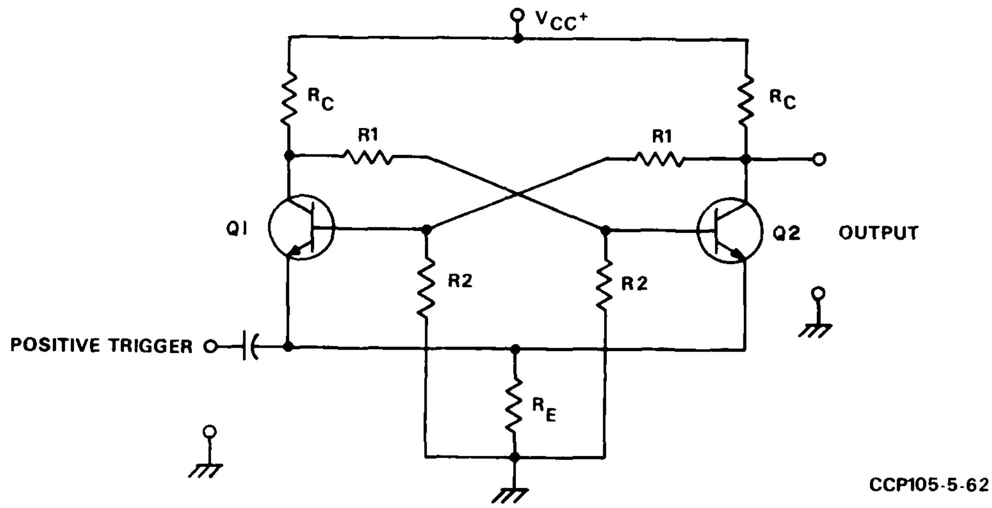


Figure 9-17. Emitter coupled bistable multivibrator schematic diagram.

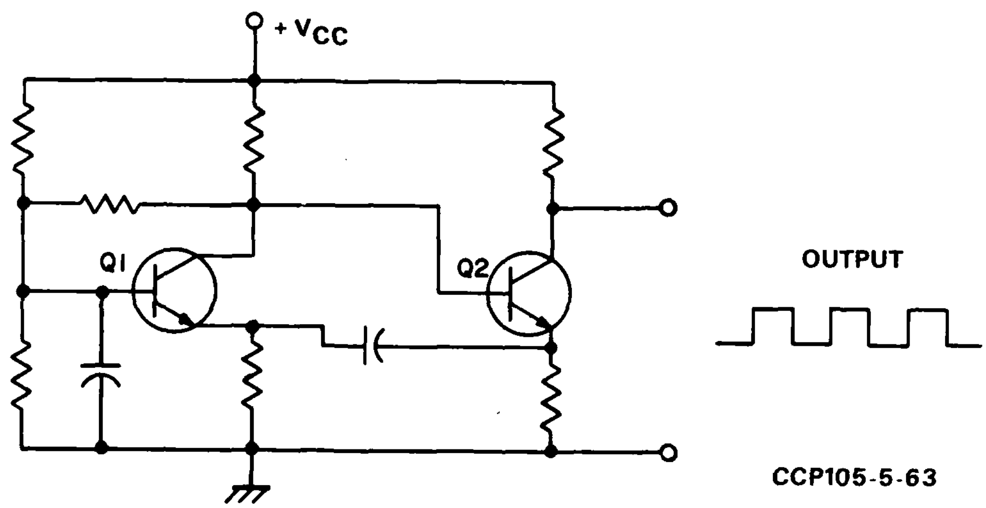


Figure 9-18. Emitter coupled astable multivibrator schematic diagram.

9-11. OTHER SOLID STATE DEVICES.

a. While transistor circuits are widely used, there are other semiconductor devices that are useful.

b. It is beyond the scope of this publication to discuss them in detail, but general information is presented.

(1) *Tetrode transistor.*

(a) Constructing an ordinary NPN junction transistor and attaching an extra connection to the P base section has created the tetrode transistor.

(b) By suitably biasing the second base connection, an improvement in frequency response is achieved. Tetrode transistors are used as mixers and oscillators.

(2) *Photodiode and phototransistor.*

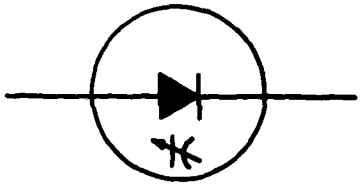
(a) Semiconductors have also been developed that are sensitive to light. These are called photodiodes and phototransistors. The energy of the light photons is transferred to the junction of the semiconductor materials.

(b) This gives the electrons sufficient energy to break away from the covalent bonds. Once free the electrons aid in charge transportation and increase the current flow through the semiconductor.

(3) *Varactor.*

(a) A varactor (figure 9-19) is a diode that is used as a capacitive element. The name varactor is a contraction of variable reactor. The effective capacitance of the varactor is varied by changing the bias voltage. Varactors are used in paramps, frequency multipliers, and electrically tuned circuits.

(b) When biased in the reverse direction, all semiconductor diodes exhibit a variation in capacitance as the reverse voltage is increased and decreased. This variation can be explained as follows. The junction of the diode has P semiconductor material on one side and N semiconductor material on the other side. The current carriers in the N material are electrons, while the current carriers in the P material are holes. The application of a reverse voltage attracts the electrons and holes away from the junction, leaving a depleted region in the neighborhood of the junction. The layer of electrons on one side of the junction and of holes on the other side can be considered the plates of a capacitor. Since variation of the reverse



CCP105-5-64

Figure 9-19. ***Varactor*** *symbol.*

voltage will cause the electrons and holes to move closer to and farther from the junction, the equivalent capacitance changes accordingly.

(4) *Tunnel diode.*

(a) The tunnel diode is ideally adapted for use in satellite communications since it can amplify at super high frequencies. It is resistant to nuclear radiation without additional hardening and can operate over a wide temperature range.

(b) The tunnel diode amplifies by using the negative resistance characteristic of the tunnel diode, as shown in figure 9-20.

(c) Note the linear increase in current with application of positive voltage until the negative resistance range is reached. At this point, increasing voltage causes a current decrease; likewise, a decrease in voltage causes an increase in current. Only the linear portion of the negative resistance range is used for amplification.

(5) *Silicon-controlled rectifier.*

(a) The silicon-controlled rectifier (SCR) is used extensively in satellite communications equipment for switching and controlling power. The SCR is physically constructed as a PNPN device with the configuration shown in figure 9-21.

(b) It is actually a semiconductor rectifier controlled by gating signals, its symbol is shown in figure 9-22. The SCR operates by virtue of its characteristic of not conducting much current in the forward direction until the anode voltage is more than a certain minimum value (referred to as the forward breakover voltage). The value of this breakover voltage is controlled by an external voltage applied to the gate. Thus, a small signal at the gate can trigger a relatively high current

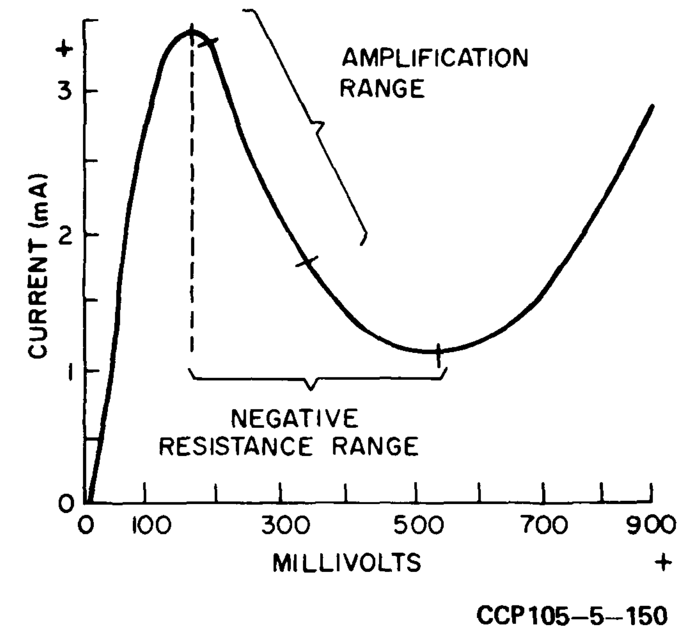


Figure 9-20. Tunnel diode amplification range curve.

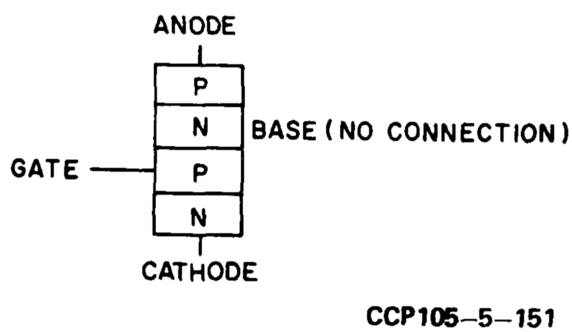


Figure 9-21. SCR construction diagram.

AD-A114 385 ARMY COMMUNICATIONS COMMAND FORT HUACHUCA AZ F/G 17/2

INTRODUCTION TO SATELLITE COMMUNICATIONS.(U)

APR 74

UNCLASSIFIED ACC-CCP-105-5 NL

4 OF 4

AD A114385

END

DATE FILMED

6-82

DTIC

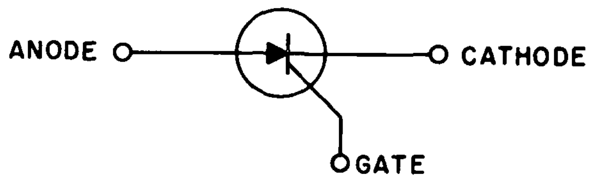


Figure 9-22. SCR symbol.

in the cathode-anode circuit. After the SCR is triggered by the gate signal, it continues to conduct regardless of the gate signal until the anode current is reduced below the holding current. Currents in excess of 100 amperes and voltages of 1,000 volts are possible with SCR's.

(6) *Thyristor.*

(a) The thyristor is a PNP device and differs from the SCR in that it may be turned off at any time by the gating signal.

(b) The name comes from thyratron which is a gas filled thermionic triode of somewhat similar characteristics.

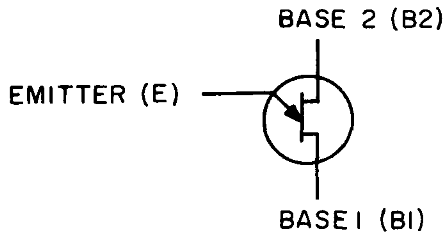
(7) *Unijunction transistor.*

(a) The unijunction transistor is usually used in SCR trigger circuits. It is also used in voltage and current sensing circuits, oscillators, timing circuits, and digital logic circuits. The unijunction transistor has an external emitter connection and two external base connections, its symbol is shown in figure 9-23.

(b) The operation of the unijunction transistor is similar to other negative resistance characteristic switching devices. The most useful features of the unijunction transistor are its stable firing point voltage, low firing current, and high pulse current capacity.

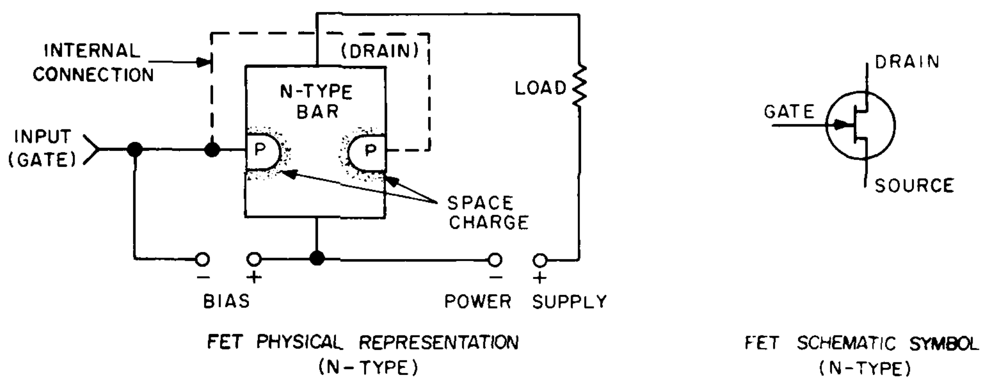
(8) *Field-effect transistor.*

(a) The field-effect transistor (FET) is a device in which an externally controlled field is used to control the resistance of a doped semiconductor bar, as shown in figure 9-24.



CCP105–5–153

Figure 9-23. Unijunction transistor symbol.



CCP105–5–154

Figure 9-24. FET construction diagram and symbol.

(b) The bar is constructed of N material with P gates. A bias applied to the P gates creates an electric field, which effectively increases the resistance of the bar and reduces current flow through the load circuit in proportion to the bias. A schematic comparison of the FET and vacuum tube amplifiers is shown in figure 9-25.

(c) Note the similarity between the N type FET and the vacuum tube biasing configuration. This similarity extends even further as the input resistance of an N type FET, like that of the vacuum tube, is also very high.

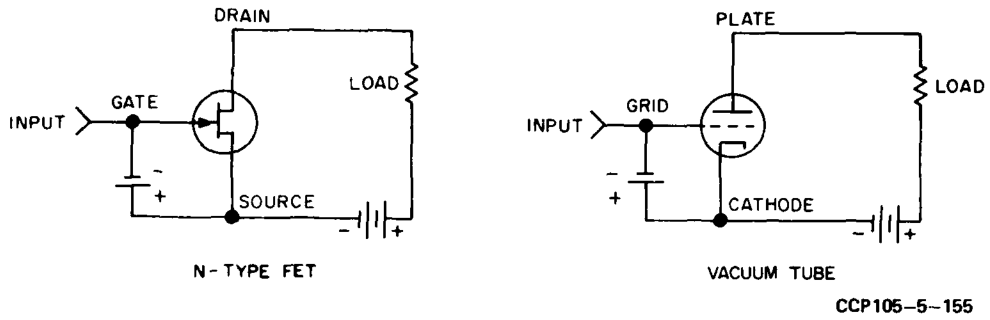


Figure 9-25. FET and vacuum tube amplifiers schematic diagram.

(d) The FET is also available with a P type bar and N type gates. In the P type bar FET, biases are the reverse of the N type FET. The type of FET can be identified by the schematic symbol, as shown in figure 9-26.

(e) Another type of FET is the metal oxide semiconductor (MOSFET). The advantage of using a MOS fabrication technique is that the gate is electrically isolated from the semiconductor material. This gives the MOSFET a much higher input impedance and makes it much more sensitive than a FET circuit. The structure of a MOSFET is shown in figure 9-27.

(f) The oxide is used as an insulator between the metal and the semiconductor. The gates create an electric field in the semiconductor, which allows the gate voltage to control the flow of current from source to drain.

(9) *Light-emitting diode.*

(a) A recent development has been the light-emitting diode (LED). The operation is very different than that of the photo diode. By applying the proper current and voltage to the LED, it will produce visible light. One of the first applications of the LED has been in display units.

(b) Arrays of LED's are arranged so that letters and numerals may be displayed by activating the appropriate diodes. Integrated circuit drivers are used with the arrays to convert input signals into the required diode activating signals.

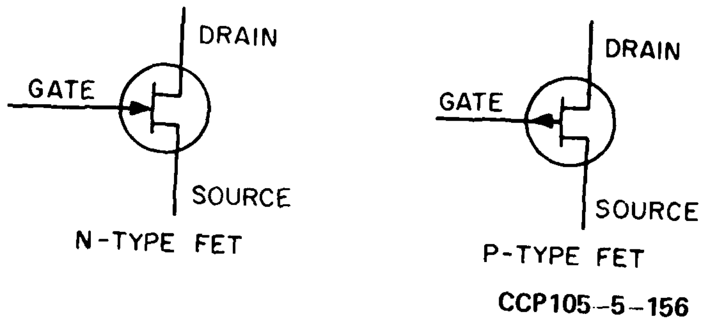


Figure 9-26. N type and P type FET symbols.

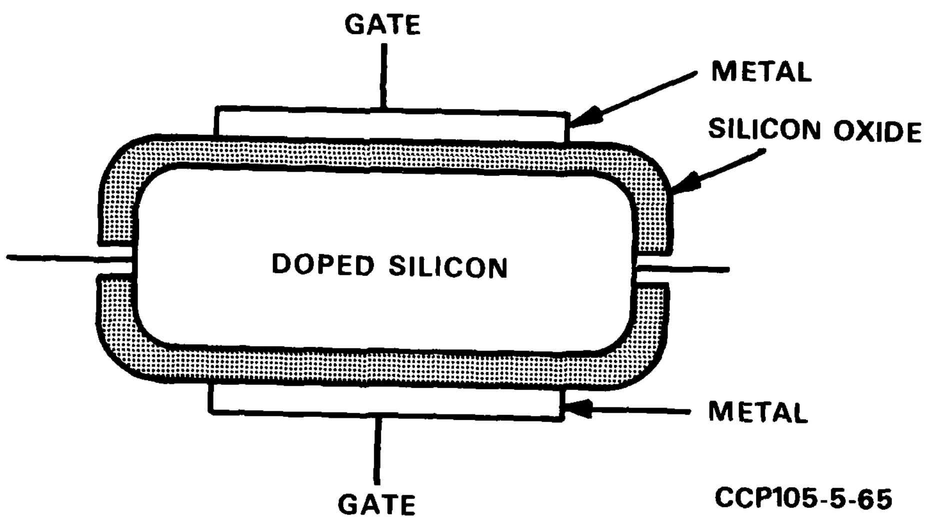


Figure 9-27. MOSFET construction diagram.

9-12. INTEGRATED CIRCUITS.

a. The development of integrated circuits (IC) has caused many changes in the field of electronics. Component densities have increased so greatly that a card of integrated circuits may replace a rack of conventional circuitry.

b. The IC has also permitted equipment to perform additional functions without markedly increasing equipment cost, size, or power consumption. The use of medium scale integration (MSI) and large scale integration (LSI) promise even greater reductions in equipment size.

(1) *Fabrication.*

(a) Integrated circuits are fabricated using many of the techniques originally developed for manufacturing ordinary single device semiconductors. Rather than place one device on a chip of semiconductor material, an integrated circuit will have many transistors and diodes on the same piece of material. The methods used in making the devices on the chip are also used to interconnect the devices into a circuit.

(b) Resistors, capacitors, and inductors can also be placed on the semiconductor chip along with the active devices. The types of circuits that can be built are limited only by the designer's imagination.

(2) *Linear integrated circuit.*

(a) The use of linear integrated circuits has become widespread in only a short time. Most new radios and televisions contain IC's to reduce size and cost. Many of the analog IC's require external components, but this is to allow the circuit designer to make the IC fit his particular needs.

(b) Some of the linear IC uses have been as audio amplifiers, if amplifiers, fm detectors, fm limiters, rf amplifiers, dc amplifiers, and phase detectors. IC's have also been used to a great extent as operational amplifiers. Operational amplifiers have been used as integrators, active filters, comparators, analog switches, as well as wideband dc amplifiers.

(3) *Digital integrated circuit.*

(a) The major impact of integrated circuits has been in the digital field. Digital IC's can replace circuits many times their size and cost, and can operate at speeds up to 50 MHz. IC's have been built to accomplish most standard logic operations. Besides AND, OR, NAND and NOR gates, inverters, flip-flops, buffers, registers, and counters are available. Many different types of circuitry are available to fill various speed, temperature, and cost requirements. Some of the more common are resistor-transistor logic, diode-transistor logic, transistor-transistor logic, and emitter-coupled transistor logic.

(b) Some examples of the reductions in size that digital IC logic can produce are seen in digital computers and counters. While computers have not decreased in size, their capabilities and reliability have increased in quantum leaps. Until recently, counters were a rarity because discrete component logic circuits were too bulky. Now with IC's, counters and digital display meters are readily available.

(4) *LSI and MSI.*

(a) LSI and MSI take the original idea of integrated circuits one step further. With a typical IC, only a few logic functions are contained on the chip. With LSI and MSI, large numbers of individual logic elements are fabricated on one chip and the elements interconnected as required.

(b) One example is in the building of semiconductor memories. MSI registers, capable of storing 1,024 bits, are now being used. Read-only memories and random access-memories also are being built using MSI techniques. Most circuits using LSI and MSI techniques are designed for specific applications. Plans are being made to place all the electronic logic circuitry needed for a desk calculator on one LSI chip. The biggest problem, at present, is producing the LSI and MSI chips. Their complexity makes production yield rates too low to permit low cost, high quantity manufacturing.

CHAPTER 10

DIGITAL TECHNIQUES

10-1. **GENERAL.** This chapter provides a familiarization with the basic fundamentals of digital logic including numbering systems, logic elements, and data storage.

10-2. NUMBERING SYSTEMS.

a. *Binary.*

(1) It is difficult to discuss or understand computers and logic circuits without some familiarity with the binary numbering system. The ten numbers used in the decimal system are too cumbersome to be handled by computer circuits; therefore, the binary system is used. The binary system uses only two numbers so that many simple devices, such as flip-flops or even a simple switch (on or off) can be used to represent the value of a binary digit. The decimal numbering system uses ten as a base. A number larger than nine (counting 0 as the first digit) uses some combination of the first ten digits. For instance, twelve is written as 1 followed by a 2. The binary system uses a base 2 rather than 10. It uses two digits, namely 1 and 0. Any number greater than 1 (counting 0 as the first digit) requires some combination of the first two digits. In the decimal system 10^2 is written as a 1 followed by 2 zeros (100), 10^3 is written as 1 followed by 3 zeros (1,000) or in general 10^n is written as 1 followed by n zeros. In the binary system 2^2 is written as a 1 followed by 2 zeros (100), 2^3 is written as a 1 followed by 3 zeros (1,000) and 2^n is written as 1 followed by n zeros.

(2) Use decimal number 117 to make a further comparison. This may be thought of as being comprised of 10^2 plus 10^1 plus 7. The binary equivalent, 1110101, may be thought of as being 2^6 plus 2^5 plus 2^4 plus 2^2 plus 2^0. Adding these binary equivalents (table 10-1) we can see that binary 1110101 equals 117 in base 10.

(3) To convert from binary to decimal, first find the highest power of 2 in the decimal number. In the example above, this is 2^6 or 64 (binary 1000000). Subtract 64 from 117 and find the highest power of 2 in the difference (53). This is 2^5 or 32 (binary 100000). Repeat the process until the remaining difference is 1 or 0. The sum of the binary equivalents of each power of 2 in the original decimal number is the binary equivalent (table 10-1).

(4) At first glance a number like 1110101 seems an awkward way of writing 117, and that writing larger decimal numbers in their binary equivalent form would

Table 10-1. Binary equivalents

Decimal		Binary equivalent
64	=	1000000
32	=	100000
16	=	10000
4	=	100
1	=	1
117		1110101

Table 10-2. Decimal, octal, and binary equivalents

Decimal	Binary	Octal
1	1	1
2	10	2
3	11	3
4	100	4
5	101	5
6	110	6
7	111	7
8	1000	10
9	1001	11
10	1010	12
16	10000	20
32	100000	40
64	1000000	100
128	10000000	200
256	100000000	400
512	1000000000	1000
1024	10000000000	2000
2048	100000000000	4000
4096	1000000000000	10000
8192	10000000000000	20000
16384	100000000000000	40000

be even more awkward. In reality the difference in the number of digits in a large decimal number and its binary equivalent is not as large as assumed. It requires, on the average, about three times as many digits to express a number in binary form as to express it in decimal (table 10-2).

(5) Despite its length, a binary number is well adapted for use in electronic and electrical circuits. The presence of voltage can be used to represent a 1, and an absence of voltage can represent a 0. This is called positive logic. Negative logic uses an absence of voltage to represent a 1, and the presence of voltage to represent a 0.

b. *Binary-Coded Decimal.*

(1) The binary-coded decimal (bcd) system is sometimes used in digital data systems. Binary numbers represent decimal digits in the bcd system. Four binary digits are needed to represent a decimal digit. Thus:

decimal	31	becomes
bcd	0011 0001	

(2) Since the fourth binary digit is used only for decimals 8 and 9, it becomes obvious that the four-digit code should be used with a base 16 numbering system for greater efficiency.

c. *Binary-Coded Octal.*

(1) Present-day digital data processing equipment often uses the binary-coded octal system (base 8 rather than the binary-coded decimal system (base 10)). The binary-coded octal numbering system (table 10-2) is used since a three-digit binary number efficiently utilizes the base 8, peripheral equipment is easily adapted to a three-digit binary code, and conversion from binary to octal is easily accomplished.

(2) Conversion from binary to octal is accomplished in groups of three, starting from the least significant digit and noting the decimal value of each group, for instance:

decimal	117	16394
binary	1 110 101	100 000 000 001 010
octal	165	40012

d. *Gray Code.*

(1) The gray code (also known as a cyclic or progressive numbering system) is a system where numbers differ in only one digit as shown in table 10-3.

(2) The gray code is often used to provide digital information concerning angular position by means of a coded commutator where the code is arranged in conducting and nonconducting segments as shown in figure 10-1.

(3) The gray code is superior to the binary-coded decimal system for angular position information as a mechanical misalignment, or phase lag, will not cause errors. With a slightly out-of-phase commutator based on the binary-coded decimal system, the transposition from decimal 3 (binary 011) to decimal 4 (binary 100) could possibly go through binary 111, binary 000, binary 001, binary 010, or binary 101.

Table 10-3. Gray code

Decimal	Binary	Gray (typical)
0	000	000
1	001	001
2	010	011
3	011	010
4	100	110
5	101	111
6	110	101
7	111	100

10-3. BOOLEAN ALGEBRA AND BASIC LOGIC.

a. Boolean algebra is a system of logic annotation where the logical functions are represented by symbols. In simplified form, some of these logical functions and their symbols are stated in table 10-4.

b. Using these symbols, it is possible to represent the logic employed in reaching certain conclusions. For example, in planning a ski trip there should be snow on the mountain, no snow on the roads, no rain, and food and lodging available. This can be annotated as follows: (snow on mtn) · ($\overline{\text{snow on rds}}$) · ($\overline{\text{rain}}$) · (fd & ldg) = ski trip.

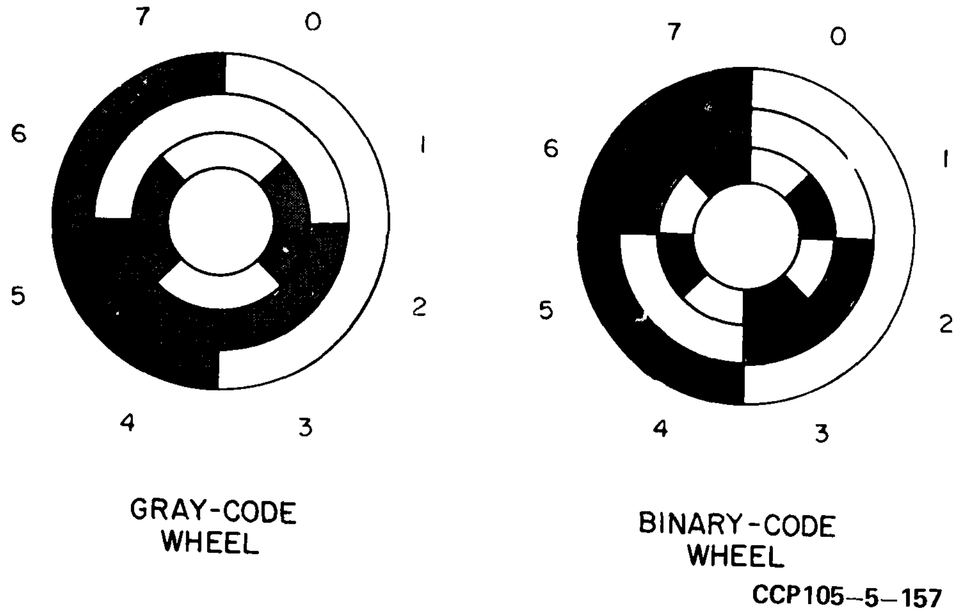


Figure 10-1. Gray and binary code wheels segment layout.

Table 10-4. Logic functions and symbols

Function	Symbol	Example
and	•	A • B
or	+	A + B
not (inversion)	–	$\overline{A}$

c. This can also be accomplished by inverting all the variables, and changing the functions. When a function is inverted AND becomes OR, OR becomes AND, NOT symbols are removed and placed on the variables that were previously not inverted. The equation now becomes: $\overline{\text{snow mtn}}$ + (snow on rds) + rain + $(\overline{\text{fd \& ldg}})$ = $\overline{\text{ski trip}}$.

d. A summary of basic Boolean identities is shown in table 10-5.

Table 10-5. Summary of basic Boolean identities

Fundamental laws

OR	AND	NOT
$A + 0 = A$	$A \cdot 0 = 0$	$A + \overline{A} = 1$
$A + 1 = 1$	$A \cdot 1 = A$	$A \cdot \overline{A} = 0$
$A + A = A$	$A \cdot A = A$	$\overline{\overline{A}} = A$
$A + \overline{A} = 1$	$A \cdot \overline{A} = 0$	

Associative laws

$$(A + B) + C = A + (B + C) \qquad (A \cdot B) \cdot C = A \cdot (B \cdot C)$$

Commutative laws

$$A + B = B + A \qquad A \cdot B = B \cdot A$$

Distributive laws

$$A \cdot (B + C) = A \cdot B + A \cdot C$$

De Morgans laws

$$\overline{A \cdot B \cdot C} = \overline{A} + \overline{B} + \overline{C} \qquad \overline{A + B + C} = \overline{A} \cdot \overline{B} \cdot \overline{C}$$

e. Schematic diagrams of logic circuitry use symbols for logic operations. Some of these schematic symbols are shown in figure 10-2.

f. With the use of these gates, a ski trip computer can be built to make decisions. Either of the equations can be used to design the computer. Figure 10-3 illustrates the computer circuit using an AND gate. Figure 10-4 contains the same computer circuit using an OR gate.

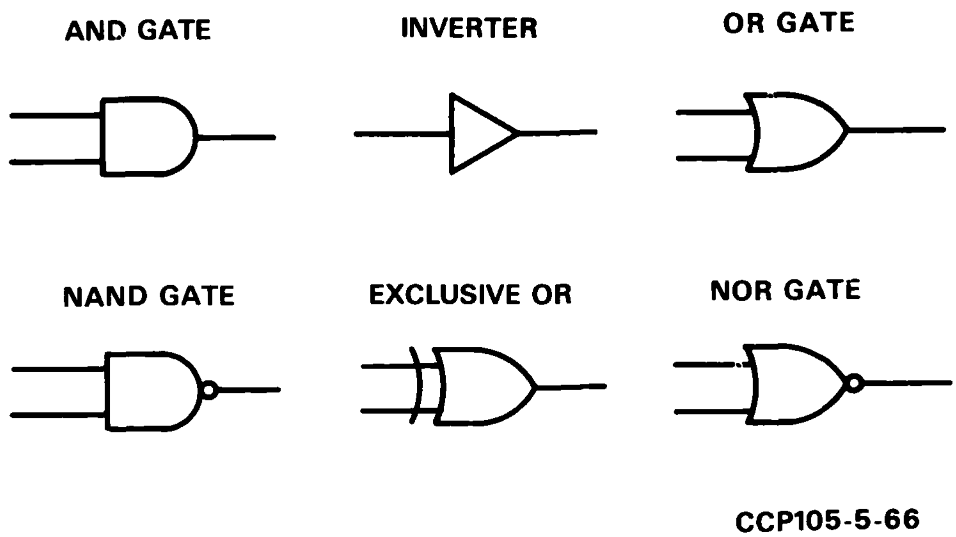


Figure 10-2. Logic operation symbols.

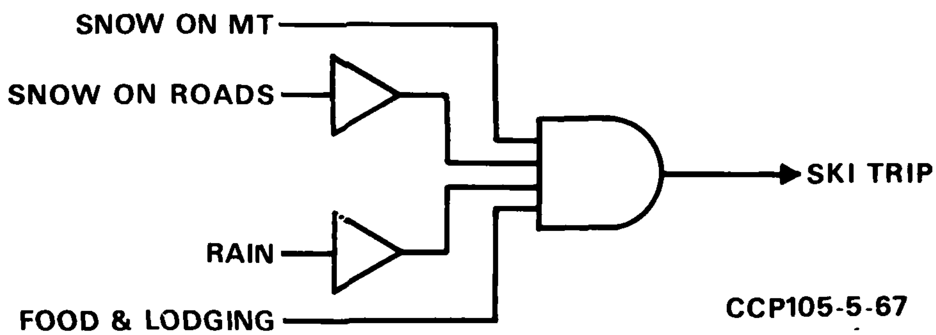


Figure 10-3. AND gate ski-trip computer logic diagram.

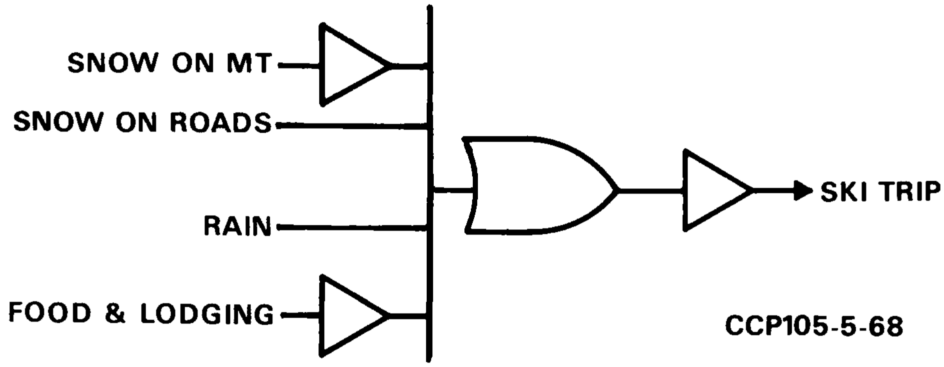


Figure 10-4. OR gate ski-trip computer logic diagram.

10-4. **LOGIC ELEMENTS.** In present-day digital logic systems there are a few basic circuits used, which are used repetitively in various combinations to perform required operations. These more or less standard units are generally referred to as building blocks. Digital logic systems operate on pulses and the circuits involved are generally used to control the movement of these pulses. The system will perform a variety of operations on these pulses such as storing, gating, delaying, inverting, and shifting.

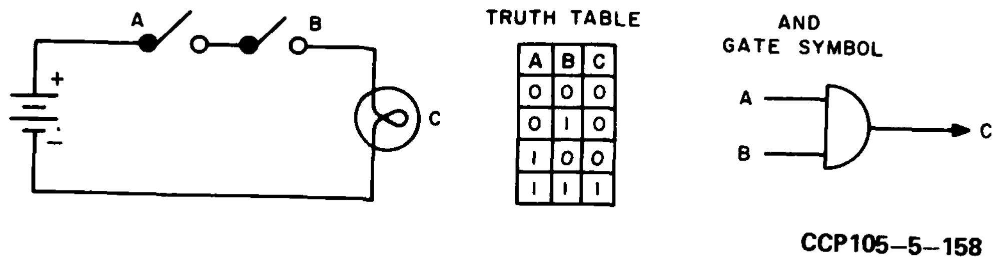
a. *AND Gate.*

(1) A simple form of AND gate is shown in figure 10-5. If lamp C is to light, switches A and B must both be closed at the same time. The Boolean expression for this function is $A \cdot B = C$ which is read as A and B equals C.

(2) To relate this example to the binary mechanics of digital systems, a closed switch is represented by 1 and an open switch by 0. A truth table is made to show the possible conditions of two variables. These binary conditions correspond to digital logic systems.

(3) In reality, as shown in figure 10-6, diodes are used for AND gates instead of relays.

(4) Operation of the AND gate depends on the inputs at A and B. If one input is at a positive voltage and the other is at ground potential (waveforms A and B), current will flow through R and the output will be at ground potential. If input A and input B are at a positive voltage (output waveform) no current will flow through R and the output will be at a positive voltage.



A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

Figure 10-5. Simple form of AND gate, truth table, and symbol.

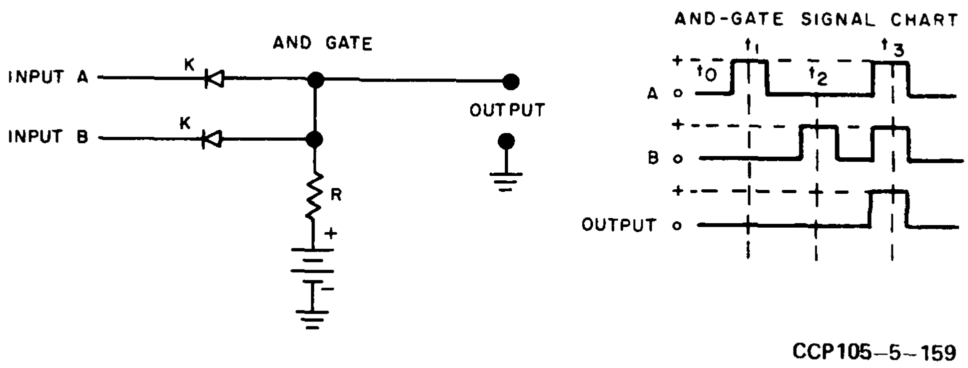
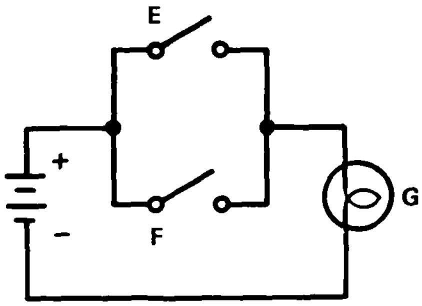


Figure 10-6. Diode AND gate schematic diagram and waveforms.

b. *OR Gate.*

(1) The OR function is another logical function which can be described with a switch configuration as shown in figure 10-7.

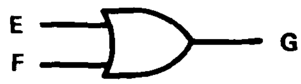
(2) If either switch E or switch F is closed, the lamp will light. The Boolean expression is E + F = G. A truth table for this function is included in figure 10-7. In reality, as shown in figure 10-8, diodes are used for OR gates.



TRUTH TABLE

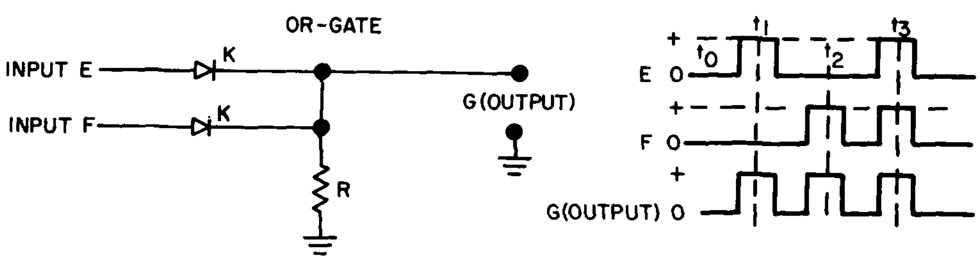
E	F	G
0	0	0
1	0	1
0	1	1
1	1	1

OR GATE SYMBOL



CCP105-5-69

Figure 10-7. Simple form of OR gate, truth table, and symbol.



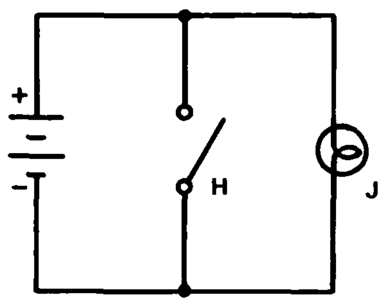
CCP105–5–160

Figure 10-8. Diode OR gate schematic diagram and waveforms.

(3) Operation of the diode OR gate illustrated depends on inputs E and F. If a positive voltage is applied to input E (waveform E), or input F (waveform F), or both (waveform G), current will flow through resistor R and the output will be at the positive voltage.

c. *Inverter.*

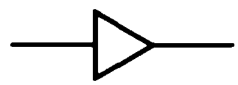
(1) Inverters are a logic function that changes only the state of whatever binary input is applied. This is accomplished by using the phase inverting characteristics of transistor amplifiers. The operation is shown in figure 10-9.



TRUTH TABLE

H	J
0	1
1	0

INVERTER SYMBOL



CCP105-5-70

Figure 10-9. Simple inverter diagram, truth table, and symbol.

(2) If the switch is closed, the battery is short circuited and the lamp will be off. When the switch is opened the lamp will light. This is equivalent to the Boolean expression H = J. A circuit that performs inversion uses a transistor, rather than diodes.

(3) The operation depends on the input signal H (fig. 10-10). When H is zero, V_B reverse biases the base of the transistor cutting the current flow. No collector current flows and J is equal to V_c. When H is positive the transistor is forward bias saturated, collector current flows heavily, and the output J becomes zero.

d. *NAND Gate.*

(1) The NAND comes from the combination of NOT and AND (fig. 10-11).

(2) The gate has the logic function of an AND gate with an inverter on its output.

e. *NOR Gate.*

(1) Just as NAND is a combination of two operations, NOR is a combination of the operations NOT and OR (fig.10-12).

(2) The NOR gate has the logic function of an OR gate with an inverter on its output.

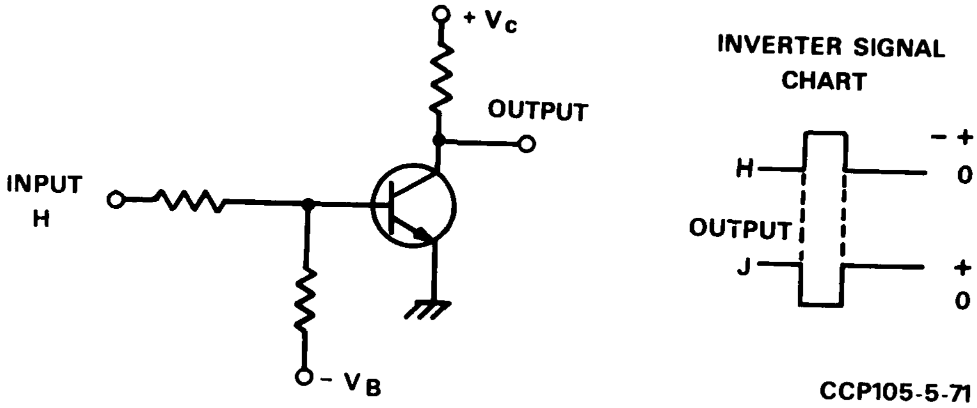


Figure 10-10. Transistor inverter diagram and signal chart.

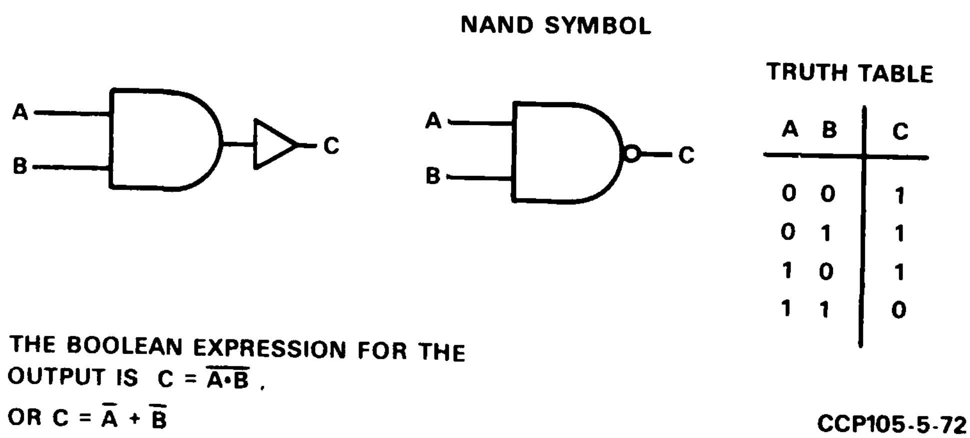
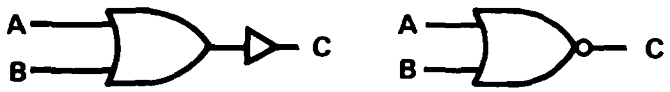


Figure 10-11. NAND gate symbols and truth table.



TRUTH TABLE

A	B	C
0	0	1
0	1	0
1	0	0
1	1	0

THE BOOLEAN EXPRESSION FOR THE OUTPUT IS $C = \overline{A + B}$, OR OR $C = \bar{A} \cdot \bar{B}$

CCP105-5-73

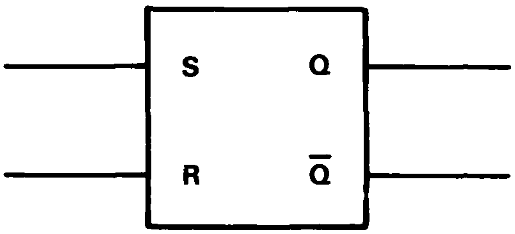
Figure 10-12. NOR gate symbols and truth table.

10-5. BISTABLE DEVICES.

a. Bistable devices, commonly called flip-flops, are used as counters, storage elements, or dividers. The circuits are called bistable because they have two stable states; a 1 state and a 0 state. The device can remain in either state indefinitely.

b. The flip-flops used in digital circuits are different from the gates just described in several important ways. One is that flip-flops are synchronous devices. No matter what inputs are present, the output will only change if a trigger pulse is present. The AND and OR gates will change output whenever the inputs change. The second is that the flip-flops have two outputs, generally called Q and $\bar{Q}$. These two outputs are always the opposite of each other. The last and most important difference is that flip-flops are sequential devices. That means that their output at a particular time will depend on the previous outputs, as well as the inputs. There are four basic types commonly used.

(1) The R-S flip-flop has two inputs R and S, from which it gets its name. The S input stands for set and causes the flip-flop to be in its 1 state (Q = 1). The R input stands for reset, and causes the flip-flop to be in its 0 state (Q = 0). This type of binary is sometimes called S-C for set and clear. The symbol and truth table for this type of flip-flop are shown in figure 10-13. The question mark in the table indicates that the state is indeterminate, and that the circuit is designed so that R and S can not be one simultaneously.



TRUTH TABLE

BIT TIME n		BIT TIME n+1
R^n	S^n	Q^{n+1}
0	0	Q^n
0	1	1
1	0	0
1	1	?

$$Q^{n+1} = (S + \bar{R}Q)^n$$

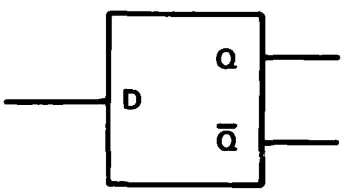
CCP105-5-74

Figure 10-13. R-S flip-flop diagram and truth table.

(2) The D flip-flop is a delay memory element. It has a single input, and an output equal to the previous input (fig. 10-14).

(3) The trigger (T) flip-flop only changes state when the input is 1. This permits the T flip-flop to be used as a divider (fig. 10-15).

(4) The J-K flip-flop is similar to the R-S flip-flop, except that both inputs may be 1 (fig. 10-16).



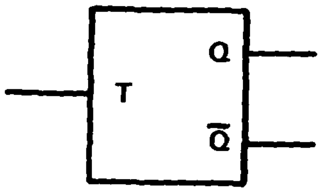
TRUTH TABLE

D^n	Q^{n+1}
0	0
1	1

$$Q^{n+1} = D^n$$

CCP105-5-75

Figure 10-14. Flip-flop diagram and truth table.



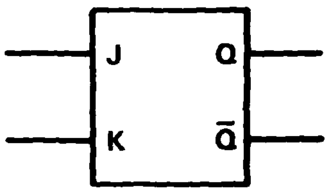
TRUTH TABLE

T^n	Q^{n+1}
0	Q^n
1	$\bar{Q}^n$

$$Q^{n+1} = (\bar{T}Q + T\bar{Q})^n$$

CCP105-5-76

Figure 10-15. T flip-flop diagram and truth table.



TRUTH TABLE

J^n	Q^n	Q^{n+1}
0	0	Q^n
0	1	0
1	0	1
1	1	$\bar{Q}^n$

$$Q^{n+1} = (\bar{K}Q + J\bar{Q})^n$$

CCP105-5-77

Figure 10-16. J-K flip-flop diagram and truth table.

10-6. ADDERS. Adders are necessary where arithmetic operations are being performed. Many types of adders, for different functions, are available. For the sake of brevity we will only look at the two most common adders.

a. *Half Adder.* The half adder is the basic building block for binary addition. It is called a half adder because it cannot be put in parallel to add numbers of more than one binary digit (fig. 10-17 and 10-18).

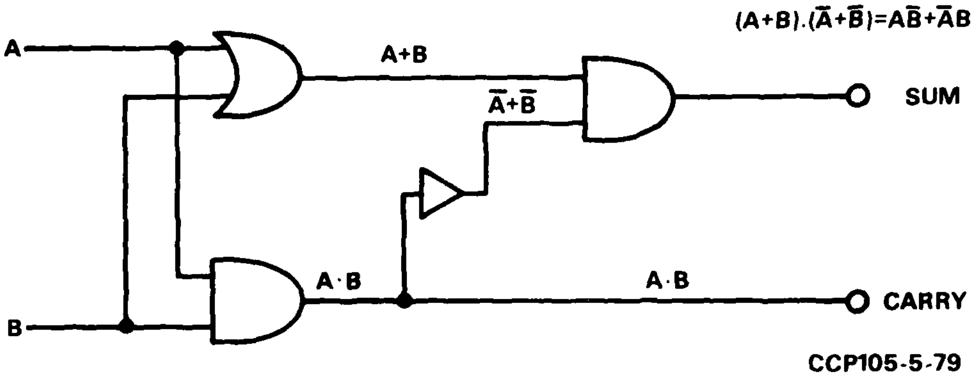
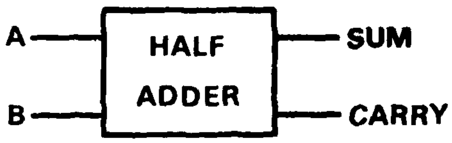


Figure 10-17. Half adder elements logic diagram.



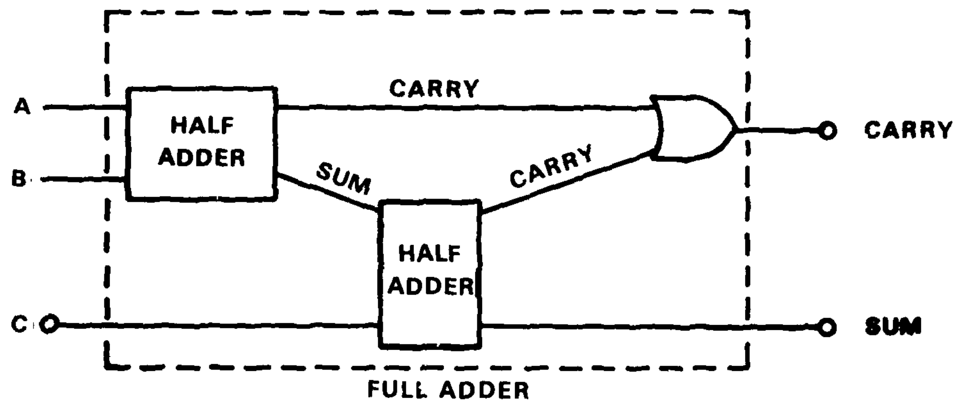
TRUTH TABLE

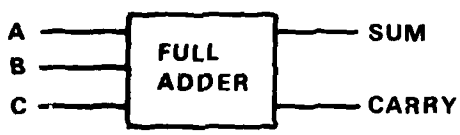
A	B	SUM	CARRY
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

CCP105-5-80

Figure 10-18. Half adder diagram and truth table.

b. *Full Adder.* Full adders have three inputs, which allows them to be placed in parallel to add two binary numbers of more than one digit. The basic building blocks are two half adders (fig. 10-19). A parallel adder, using four full adders, is shown. This adder can add two numbers consisting of four binary digits (fig. 10-20).





TRUTH TABLE

A	B	C	SUM	CARRY
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

$SUM = A(BC+\bar{B}\bar{C})+\bar{A}(\bar{B}C+B\bar{C})=\bar{A}\bar{B}C+\bar{A}B\bar{C}+A\bar{B}\bar{C}+ABC$

$CARRY=BC+AB+AC$

CCP105-5-81

Figure 10-19. Full adder logic diagram and truth table.

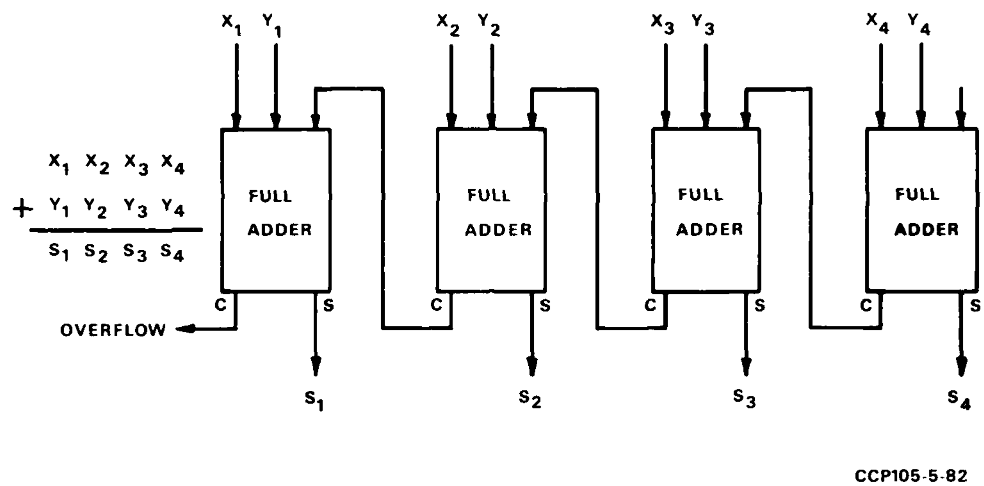


Figure 10-20. Parallel adder block diagram.

10-7. **DATA STORAGE.** In almost all types of digital systems there is a need for data storage. For a small computer, the storage may consist of flip-flops. For a large digital computer, the storage may include a magnetic drum, tape, or disc, ferrite core, or semiconductor memories.

a. *Magnetic Drums.*

(1) Magnetic drums are constructed with high precision and are coated or electroplated with a ferromagnetic material, such as cobalt-nickel alloy or ferric oxide. They normally rotate at a constant speed between 3,000 and 12,000 revolutions per minute. Read/write heads may be arranged either in parallel, or helically, along the length of the drum. Each head is associated with a track of stored data. The writing heads store the data as magnetized spots on the drum surface. The read function is similar to the readout, or playback, from a tape recorder and is nondestructive.

(2) One advantage of a drum system over tape storage is that a drum permits rapid random access. To find a piece of information on tape, the reel must be wound by the tape drive until the correct spot is reached. This may take several seconds and slows down the machine operation. For a drum, the maximum time to reach a piece of data would be the time required for the drum to make one revolution.

b. *Magnetic Tape.*

(1) Magnetic tapes are used extensively as storage devices in present day computers. In magnetic tape storage, data can be stored and read sequentially. The data is stored on a plastic tape coated with a ferromagnetic material. Read/write heads are generally ganged in parallel tracks across the width of the tape to permit parallel read and write of several rows of binary characters.

(2) The advantage of magnetic tape storage over other storage methods is the ease with which individual tapes can be replaced and stored. Storage capacity can be increased by employing a number of selectable tape reels per tape unit. At a satellite ground station the local computer will very likely receive its antenna pointing data from magnetic tape. Tape lengths are most commonly 2,400 feet on 10-inch reels.

(3) Information is magnetically recorded in blocks containing one or several computer words. Each block has an identifying address channel and a parity checking channel. A data recording device of this type is normally equipped with two read/write heads. The other head uses the spaces between the channels, resulting in an interlacing of the information that yields 12 recording tracks.

(4) The storage density of magnetic tape is approximately 200 characters per linear inch. The characters are distributed in several channels across the width of the tape, giving a theoretical storage capacity of 5,360,000 characters per standard reel. This does not consider the spaces between information bits or spaces lost in stops and starts.

c. *Magnetic Discs.*

(1) The disc operates on the same principles as the magnetic drums. The memory consists of several discs stacked one above an other with adequate spacing for read and write heads.

(2) The discs rotate at high speeds so that input-output cycle time is low. Discs offer more storage capacity than drums and are found in many computer applications.

d. *Ferrite Core Memories.*

(1) The standard means of providing computer storage has been the use of ferrite devices. Originally, the ferrite cores were large and bulky. Advances in ferromagnetic technology have been able to reduce the size, weight, and power requirements of the memories. Use of a magnetic core as a memory element is shown in figure 10-21.

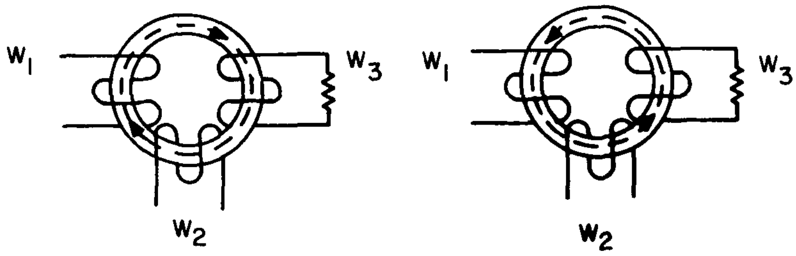


Figure 10-21. Magnetic ferrite core schematic diagram.

(2) The magnetic core can be considered as being related to a transformer core, but is much smaller in size. For example, these units are usually smaller than one-eighth of an inch. To understand the function of the magnetic core, first assume that a pulse of current of sufficient amplitude to saturate the core is applied to coil winding W_1. Magnetic lines of force are generated in the core in a direction dependent on the direction of the flow of current through W_1. The core remains magnetized in the same polarity, at a point close to saturation after a current pulse has passed through W_1. Now assume that a current pulse, flowing in the same direction as the current that passed through W_1, is applied to W_2. There will be only a small increase in the number of flux lines generated; consequently, only a small voltage is induced in W_3 and the indication in the detecting network will be very small. If the current pulse through W_2 is in the opposite direction to the current pulse flowing through W_1, the direction of the lines of force reverses and a large output appears in the detecting network.

(3) Binary bits of data are stored in a core element depending upon the direction of magnetization. For example, a core can be regarded as holding a 1 when magnetized in one direction and holding a 0 when magnetized in the other direction. W_1 may be referred to as the write winding since it writes a 1 into the core. W_2 simulates a read winding since the current through it changes the direction of the lines of force if a 1 is present in the core. This results in an output voltage, which indicates to the detecting windings that a 1 is present in the core. When the core is in the read function, data is shifted out of the core resetting the core to the 0 state. The core remains in this 0 state until the next write function, at which time data is again shifted into the core, setting the core to the 1 state.

e. *Semiconductor Memories.*

(1) The use of medium scale integration (MSI) and large scale integration (LSI) has permitted large semiconductor arrays to challenge ferrite cores as the main computer memory. By placing hundreds of flip-flops on a single chip, the cost per bit of a semiconductor memory is almost that of the ferrite memories.

(2) Some computer systems are using semiconductor memories at present. Both bipolar and metal oxide semiconductor (MOS) type devices are being used for the memories. Future developments in semiconductor design and fabrication will decide which type will prove the most useful.

f. *Future Memories.*

(1) While memories are adequately sized for the present, future needs indicate that larger memories will be required. Two types of advanced memories are being researched. The first method uses cryogenic temperatures to improve the characteristics of materials. At temperatures near 0 K, certain types of materials exhibit superconductivity. The superconductivity can be used to create a very compact and efficient memory. Other systems using cryogenic semiconductors and superconducting inductors are under study.

(2) The second method uses lasers to write and store data in an optical material. The laser would also be able to read the memory. Because of its high information rate, the laser offers a very fast read/write capability.

CHAPTER 11

TESTING PHILOSOPHY

11-1. GENERAL TESTING.

a. This chapter covers types of testing that will be required during, and subsequent to, the equipment installation phase of preparing a ground station for operation. Preoperational and operational testing are also covered. Guidelines will be presented for the development of necessary tests. A logical sequence is followed in developing all such specific tests, whether such tests are to be conducted as the equipment is installed or are to take place after installation, but prior to operational testing. This sequence can be divided into three elements; i.e., testing individual items of equipment, testing subsystems as they are developed, and testing the completed system.

b. Equipment testing can be defined as that testing which is required to determine whether or not a particular piece of equipment is functioning as a separate entity within the design characteristics of the equipment item, and to ensure that the applicable general and human engineering aspects are satisfactory. This phase of the testing can and should be initiated as the individual equipment items are installed. Improper plans for the placement of the equipment and the presence of shock hazards can be corrected more easily, if they are detected prior to installation of the equipment. Modification is difficult when more items have been combined into a developed system or subsystem.

c. As the satisfactory testing of individual items of equipment and their installation progresses, subsystems will gradually evolve. A subsystem is that grouping of equipment items designed to perform as a functional part of the whole system. Assuming individual items of equipment have been proved to be satisfactory, the objective of subsystem testing is to determine whether or not the items, when combined into a subsystem, will perform as designed. During this phase of testing more emphasis should be placed on functional tests than was the case in the equipment testing phase, although general and human engineering tests will still be required. If results indicate that changes to the individual items or revisions to the subsystem are necessary to ensure proper functioning, it is again more practical to make changes at this time rather than postpone the changes so that they must be made after the total system has been established.

d. System testing programs are those which are applied to the completed system. As discussed in this chapter, the system is considered to be complete when the ground station, with satellite link and ground intersite link(s) to the DCS switch, have been installed. All essential equipment and subsystem installations will have been made and the preliminary testing of these functions completed. Tests for the entire system should be designed to determine the ability of the system to meet its overall design objectives. Each specific test generated for this phase must consider the desired system end product, encompass the functions of the subsystems, and yield an evaluation of the composite function of the system. The design of these tests will be predicated, to some degree, on the equipment and facilities available with which they are to be performed and evaluated. The validity of the conclusions reached may be expected to be directly dependent upon the careful and accurate execution of the preceding testing phases.

11-2. EQUIPMENT TESTING.

a. Before initiating the process of developing tests for equipment, subsystems, and systems, it is necessary to understand: (1) Why we make an issue of these tests when all the equipment is tested by the manufacturer, (2) how such tests can be conducted before installation is completed, (3) when preparation should be made to conduct the tests, (4) who prepares and conducts the tests, (5) what they consist of, and (6) where the information comes from that is used in preparing the tests.

b. One question often raised is why perform tests which partially duplicate those performed by the manufacturer? It is general knowledge that manufacturers are required to, and do, perform tests on equipment they produce and these tests ensure, to a considerable measure, the inherent operability of the equipment. However, several factors must also be considered. One of these is transport of the equipment. Items of equipment installed in a ground station will probably have been shipped a considerable distance, up to several thousand miles. They will have been handled many times both by men and machinery and may have been subjected to extremes of climatic conditions. Owing to the nature of the equipment, these factors alone make it necessary to recheck each individual item. Also, where humans are involved, mistakes can be made and an additional test will provide a margin of safety and check the accuracy of the first tests. Further, equipment can be technically perfect and, if not properly installed, still be unusable. Preventing this type of system failure is the main reason for testing the equipment for suitable installation. This type of test is particularly valuable during the initial phases of installation and, if accomplished in a suitable manner, will result in considerable savings in effort and will reduce lost time to a minimum at a later date.

c. A second aspect of the testing program that may appear strange is that much of the testing is done before the installation is complete. Testing can be conducted before installation is completed by proper planning; by reviewing the physical, human engineering aspects of the system, with or without the equipment; by conducting some basic electrical tests on the individual items as they are received and prepared for installation, and by conducting additional, general type tests during the process of installation.

d. Another point that should be noted is that test preparations should be initiated immediately upon receipt of information indicating that responsibility has been assigned for an operational station installation. At this time, sufficient information on the equipment should be available to permit planning and scheduling of many of the technical tests.

e. The activity, assigned the responsibility for installing and operating the station, will be responsible for determining the operational status of the station. In cases where this responsibility is divided; i.e., one activity makes the installation and another is charged with operations, it is the responsibility of the installing activity to turn over an operable station to the operating crew. This can only be assured if suitable tests have been performed prior to this turnover. In the final analysis, it is the individual, who has been assigned operational responsibility, who is responsible for determining that required tests have been properly prepared and successfully conducted prior to accepting the station as an operating unit.

f. These tests will be in two categories - general and specific. General tests are applicable to any type of equipment regardless of its function. They consist of such items as examination to ensure proper placement of the equipment, checks for shock hazard, evaluation of the ease of operation and maintenance, and evaluation of the degree of skill required to operate and maintain the equipment. Specific tests are primarily functional tests devised to determine whether or not the item tested will properly perform its intended function.

g. In viewing these requirements, one may ask where the information originates that is used in preparing the tests? No easy answer is available, in that the information required to prepare and conduct such tests must, of necessity, be drawn from many sources. Guidelines included in this and other documents will help orient the user. Equipment information may be obtained either directly or through engineering specifications from the manufacturer, from technical manuals, or from drawings. Functional aspects may be extracted from technical development plans, program definition phase documents, system descriptions, and operational plans. Previously conducted tests on other types of communication equipment will often serve as format and content guides and, of course, as with any aspect of communications, previous experience in this and associated fields is of prime value.

h. Keeping these general requirements in mind, let us consider a few of the assumptions and premises forming the basis for the testing material. It should be recognized that in a general document of this type, it is impossible to anticipate all of the conditions that might arise. It also should be recognized that in each testing area similar results may be obtained through different tests and that some areas may be consolidated. No attempt has been made to categorize the test items in order that they can be implemented most efficiently. It is assumed that, just as conditions and equipment vary, so will the tests vary to meet the objectives. It is also presumed that operating agencies will schedule and rearrange the installation and testing schedules in ways that will permit the most economical use of the equipment, time, and funds. Obviously, any testing plan is only as good as the results obtained.

(1) *Preparation of equipment tests.*

(a) The starting point for preparing a test plan is to conduct a research of pertinent documents to gather required information. These documents should include equipment lists, shipping and delivery schedules, installation schedules, equipment specifications and manuals, floor plans, equipment configurations, functional objectives (general and detailed), personnel lists and personnel assignment schedules, special equipment requirements (cranes, cherry pickers, etc.), construction schedules, and any additional information that will assist in preparing a test plan.

(b) In some cases, all of this material will not be readily available; however, sufficient information should be obtainable to prepare the basic portions of the plan. Most of the required information can be obtained prior to the physical arrival of the equipment at the ultimate site. By applying basic questions to this information, resolving the answers, and documenting the results it will be found that, in most cases, the equipment tests can be prepared. These types of preparatory questions, answer sources, and actions are shown in table 11-1.

(2) *General tests (equipment items).*

(a) The terms "general" and "specific", as used in application to the tests described in this chapter, are more or less arbitrary designations and are used to denote separate testing areas. It will be found that in actual testing some overlap will occur. Also, other designations may be used depending on the activity or individual concerned. For the purpose of this chapter; however, general tests will be construed to mean those tests dealing with general characteristics of the equipment involved that can be expected to be applicable to any item of equipment, whether it be a teletypewriter or a receiver using phase-locked loop principles. This type of test includes such things as proper placement of the equipment for ease of operation and maintenance, adequate lighting, and average time required to place the equipment in operation. In preparing general tests, basic rules can be followed and as the tests are conducted those items applicable to the equipment under test should be selected and used. As an example, adequate lighting for a teletypewriter would be an important test inasmuch as it can be assumed that

Table 11-1. Sample planning guide

Question	Answer source	Action
What equipment items must be tested?	Shipping list, bill of materials, equipment lists.	Compile lists of all equipment by type, number, nomenclature.
What general tests can be accomplished prior to installation?	Station layouts, floor plan, equipment configuration, technical manuals.	Check planned layouts, look for possible discrepancies in power hookup, operating and maintenance capabilities.
What specific tests can be accomplished prior to installation?	Technical manuals.	Categorize preinstallation checks and instructions.
What general tests can be accomplished during installation?	Station layouts, floor plans, equipment configurations, technical manuals.	Check planned power, associated or nearby equipment for interference, available instructional material.
What specific tests can be accomplished during installation?	Technical manuals, equipment specification.	Categorize installation checks and instructions, compile technical characteristics and parameters.
What general tests can be accomplished after installation?	Technical manuals.	Check operating instruction, compile lists of human engineering items.
What specific tests can be accomplished after installation?	Technical manuals, equipment specifications.	Categorize operation and maintenance checks and instructions and apply technical characteristics.

Table 11-1. Sample planning guide--continued

Question	Answer source	Action
When must preparations be completed and when can tests start?	Shipping and delivery schedules, installation schedules, construction schedules.	Determine when equipment items will be on site, plan installation, schedule accordingly.
Who will conduct tests?	Personnel lists, personnel assignment schedules.	Determine planned arrival of personnel by grade, MOS, and number, and schedule tests accordingly.
What results should be obtained by *tests?*	Equipment specification, technical *manuals, functional objectives.*	Categorize technical and operational *parameters of equipments as designed.*
What reporting procedures should be used?	Local policy, local directives.	Determine if procedures are established -- if not, institute.
What format should be used?	Local directives and policy guidance manuals, previous tests on associated equipment.	Check for local formal instructions--if none, prepare.

an operator will be constantly reading printed material at this machine. In the case of the receiver mentioned previously, lighting of any operation and maintenance required can readily be obtained by means of trouble lamps or work bench lighting, if there is not already sufficient lighting at the receiver location. As with so many other areas, testing is a mixture of proper planning, sufficient knowledge, and good common sense.

(b) Details of the general test should be ample to permit those conducting the test to know what result is desired by the originator. While format is primarily a matter of individual preference or local policy, certain items should be reflected in the test itself. These basic items are:

1. Objective – What is the objective of this particular test and what is to be determined?

2. Equipment – What equipment is being tested; where is it; what part of its functions are of interest?

3. Scope – What does the test consist of; what particular portions are being tested; what areas are being tested?

4. Methods – What means and methods will be used to arrive at the results and conclusions?

5. Results – What are the results obtained by conducting this test? Keep in mind that this portion must respond to each test statement, whether or not the results are satisfactory.

(c) Contents of the general test can be illustrated by a sample general type test that is particularly suitable during the early installation phase. This sample test format is shown in table 11-2 and illustrates the details concerned in this type of test. Areas of the general test are many and varied, but basically reflect these ideas. Some areas to be considered are listed with the idea in mind that this is not intended to be a complete list, but rather, is shown to generate thinking about these and other areas.

1. Human engineering – shown in table 11-2.

2. Operational suitability – covering time required to place in operation, type of personnel required, difficulties encountered during placement and during operation, procedures required, and complexity of operation.

3. Interference – covering interaction of test item with environmental and external equipment.

Table 11-2. Sample of general test

	General Test No. -- Human Engineering
Objective –	To determine the adequacy of the human engineering aspects.
Equipment –	The particular piece of equipment under test should be listed with complete nomenclature, including serial number, if applicable, and associated equipment used to conduct the test.
Scope –	Location of associated controls and their accessibility for operation and maintenance of the equipment -- comments on the type, location, and accessibility of controls, indicators, and connectors.
	Visibility and legibility of indicators and controls -- comments on ease of reading indicators, control indices, and termination or connector markings, including adequacy of alarms.
	Logic of operational progressions -- comments on the logical action required of the user in operating the test item.
	Complexity and extent of the operator's physical motions -- comments on the degree and complexity of the physical motion performed by the user in normal operation of the test item.
	Psychological user reactions -- general comments of the user as to color, packaging, and style, noise of operation, misleading indications, heat and light radiations, hazardous elements.
	Physiological effects -- comments on operator fatigue and/or hazards induced by acoustics, heat and light radiation or conduction.
	Nonessential features -- comments by test personnel as to nonessential features, or those features that could be regrouped for greater operating efficiency.
Methods –	Visual observation of the equipment, placement, lighting, heat, controls, indicators, connectors, packaging, color, noise.
Results –	All operator comments should be complied and together with visual observations, applied to the appropriate areas of the scope.

4. Extended operations – covering time and/or number of continuous operations before malfunction, frequency of malfunction, and repetitive malfunctions.

5. Instructional material – covering adequacy for installation, operation, and maintenance.

6. Shock hazards – covering grounding, interlocks, leakage, and exposed high potential voltages.

7. Personnel training – covering basic types of personnel required, numbers, MOS, and additional training required.

8. Logistic support – covering adequacy of supplied material (tools, test equipment, supplies) required to permit continual operations.

(d) As a basic rule, it can be assumed that general tests will be employed predominantly during the equipment testing phase, with less emphasis placed on specific tests until such time as the required associated equipment is installed. This, of course, does not eliminate specific tests from this particular phase, as many functional tests can be conducted with a single item of equipment. Some types of specific test will be covered in the following paragraphs. At this time it will be seen that, in some cases, general tests can be performed concurrently with the conduct of specific tests. This, of course, is the prerogative of the responsible individual conducting the tests. It is recommended that this consolidation be made whenever possible.

(3) *Specific test (equipment items).*

(a) As mentioned previously, this portion of testing can be and often is called by various names. Some people prefer to call these tests functional tests, technical tests, or operating tests. Essentially their objective is to establish the technical and design capability of the equipment to meet its functional criteria. As with the general tests, format is a matter of individual expedience or local policy. For continuity; however, specific tests should consist of the same general format as the general tests. The details making up a specific test will not be listed here; however, a sample test format is shown in table 11-3.

(b) Areas covered by specific tests will be many, and proper selection must be made for the item of equipment to be tested. Some of these areas are as follows:

1. Associated equipment – power, compatibility with associated equipment, and compatibility with communications section equipment.

Table 11-3. Sample of specific test

Specific Test No. - Power

Objective — To determine suitability of power source, distribution, control, and alarm features.

Equipment — The particular piece of equipment under test should be listed, with complete nomenclature, including serial number, if applicable, and associated equipment used to conduct test.

Scope — Determine power requirements under all conditions (standby, with light, medium, and heavy loads).

Determine suitability of power distribution to and through equipment.

Determine effect of power failure and start-up.

Determine effect of various power sources on equipment (prime, backup, regulated, nonregulated, and commercial generator).

Determine effectiveness of power controls and alarm features.

Methods — Operate equipment under all types of load conditions; observe effect on power source and equipment.

List all internal power components within or associated with the equipment; compare rated capability with requirements.

Operate equipment; stop equipment and shut down power (both basic power and individual component supplies at different times). Allow power to remain off for varying periods of time; reapply power and observe effect on equipment.

Operate equipment on basic source; switch to backup source, nonregulated source, regulated source, commercial source, generator source, and observe effect on equipment and power source.

Observe function of power controls and alarm features during conduct of test.

Results — All comments, observations, measurements, difficulties, or inadequacies should be compiled and applied to appropriate area.

2. Internal characteristics – signal characteristics, error rate, component reliability, message accountability, adjustment limits, operational limits, message handling capacity, mechanical and electrical limitations, and alarm features.

3. Associated characteristics – maintenance requirements, personnel requirements, installation requirements, and logistic requirements.

11-3. SUBSYSTEM TESTING.

a. The second step in preöperational testing is that of subsystem testing. Theoretically, this phase of testing will begin as the equipment testing phase ends and as the various individual items of equipment are assembled and tied together to form a functional grouping. Practically, it will be found that certain subsystem tests will develop and be conducted as installation progresses and before or during equipment testing.

b. A fine line cannot be drawn between one form of testing and another. There will be times when logical functioning of subsystems will make desirable integration of subsystem testing with individual items of equipment and other subsystems. Under this premise, there may be grouping of certain subsystems for testing in which these subsystems are closely related to each other and the function of one is dependent on the other. Normally, general tests can be applied to subsystems as they begin to reach their final configuration and functional tests will be prepared and conducted as the functions of the subsystem dictate.

(1) *General tests (subsystems).*

(a) The basic rules of applying general tests to subsystems are the same as those discussed previously for general testing of equipment items. In many instances, general tests performed on individual items will be sufficient and further testing of this type will not be required. In other cases, modified general testing covering some parts of tests previously applied will be indicated.

(b) As an example, general testing of a teletypewriter circuit would be a waste of time after the same tests have been conducted on the individual machines, on the other hand, a complete antenna subsystem could require some of the same general tests that were applied to individual parts of the same subsystem. Regardless of the decisions made for general type testing of subsystems and the extent to which these tests are carried, the sample tests (tables 11-2 and 11-3) are applicable and the same general rules are valid.

(2) *Specific tests (subsystems).*

(a) In preparing specific tests for subsystems, keep in mind that the objectives are to test the functional characteristics of the subsystems. In most cases, the function of any one subsystem is directly related to one or more other subsystems. Consequently, the results of tests of one subsystem may apply to other subsystems.

(b) As can be seen from the inputs and outputs, shown in table 11-3, some questions about associated systems can be answered while conducting the specific tests on the antenna subsystem. For instance, the transmitter, servo, and equipment control subsystems must put out the proper signals to test the antenna. Conversely, tests to determine if the antenna is functioning properly in supplying information and signals to the receiver, computer, and display subsystems could be conducted without the benefit of an actual hookup to these systems.

(c) Consider the example of the antenna subsystem, shown in figure 11-1, and evolve some possible areas for specific tests. Even though the diagram shows a very simplified version of a typical antenna system, some areas immediately stand out. From these it is possible to establish certain functions that this subsystem must perform and assign specific tests that must be conducted to determine that the system is performing its functions properly.

(d) As with any antenna system, a basic function of the system is to receive rf from the transmitter subsystem and radiate this rf to the distant receiving station. The reverse of course is also a function; i.e., the antenna receives rf from the distant transmitting station and distributes it to receiver components of the local station. Thus, we have the basic ingredients for two specific tests.

(e) Carrying this reasoning one step further, it is obvious that the antenna is directional and steerable and must be pointed in the proper direction. This opens up an area for the testing of proper control functions, either from programmed directions inserted through the servosystem or manual directions from the equipment control console.

(f) The third basic area is positioning readout or notification of where the antenna is pointing at any time. Several specific tests can be prepared in this area. The intent is not to write definitive tests, but rather to establish basic rules and guidelines for formulating these tests. The examples of specific tests derived from the simplified block diagram are to be construed only as examples. It is assumed that this general type of reasoning will be followed for all subsystems. As definite information becomes available for specified items of equipment, more functional characteristics will be apparent and more complete specific tests will be possible. It is the responsibility of the individual concerned with testing to gather all the necessary data available for the particular items to be tested. At this point, the

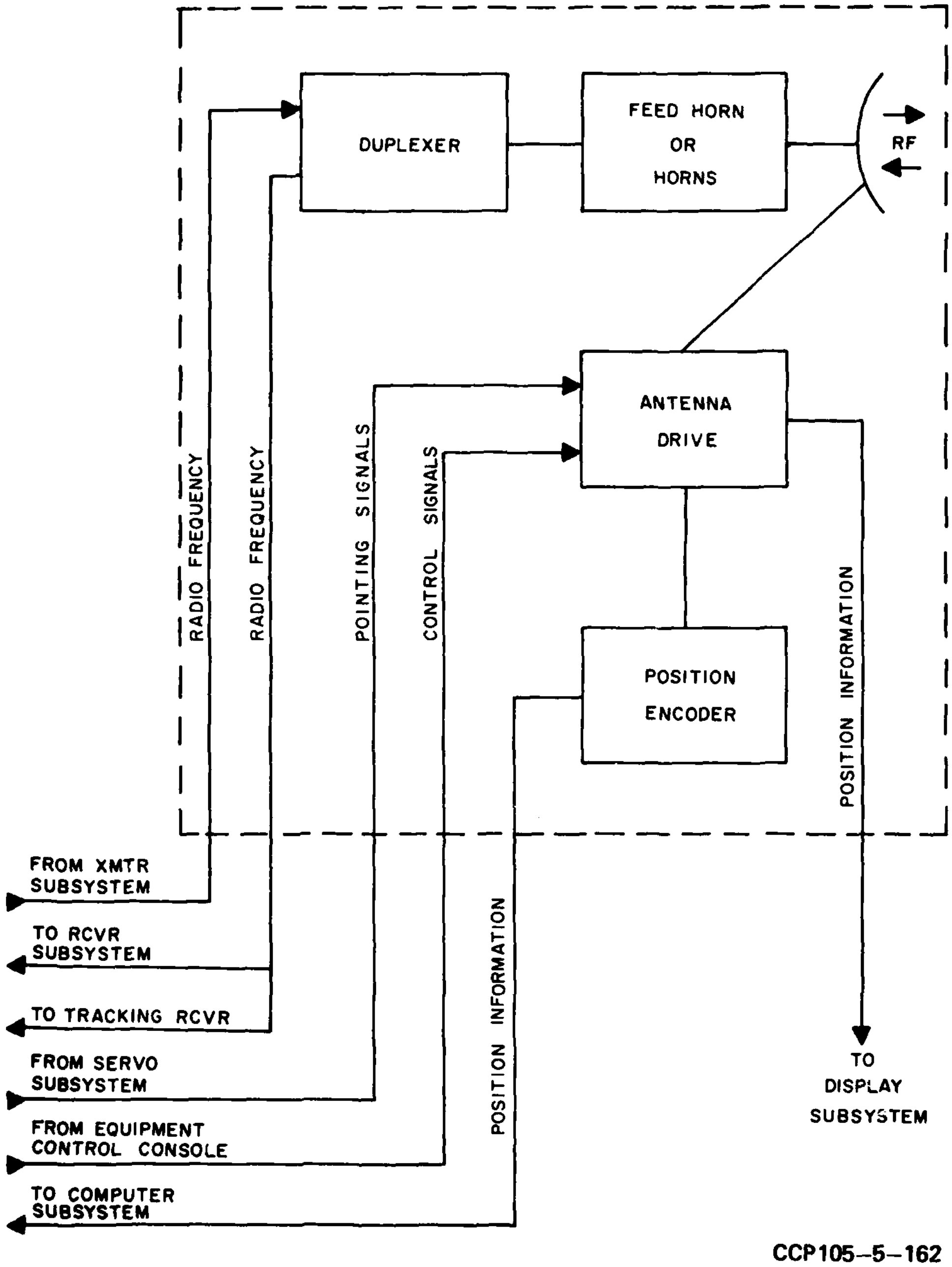


Figure 11-1. Typical antenna subsystem block diagram.

application of good basic common sense will result in sound and practical specific tests that will determine the suitability of the subsystems to perform their intended functions.

11-4. SYSTEM TESTING.

a. In paragraph 11-2, various steps were presented for planning and implementing testing. If the different phases of testing have been conducted as installation progresses and if results are conclusive for each step, the system testing should fall into place with little trouble. This, of course, is not meant to imply that system testing is of minor importance or that it requires less effort. On the contrary, this portion of the testing process is the ultimate goal and on these tests is based the final determination of whether the system is GO or NO-GO.

b. In this phase of the testing, the objectives will be primarily functional in nature; that is to say, general tests are to all practical purposes completed. Throughout the equipment and subsystem tests, it was obvious that a shift was occurring from emphasis on general testing in the equipment stage to functional testing in this, the system stage. As a result, this section will only discuss functional tests.

c. Figure 11-2 is a symbolic representation of a typical ground station layout showing links. This illustration represents a ground station with four satellite link terminals and two links connected to separate DCS switches. These switches in turn are shown tied to various subscribers.

d. Another representation of the same basic ground station is shown in figure 11-3. This is a simplified block diagram showing substantially the same breakdown. In this figure, the communications channels are depicted by the large solid arrows and the importance of proper subsystem testing becomes apparent. If all subsystems are functioning properly, the system should function as designed. It is, of course, very necessary to plan and perform the proper tests to assure that this is the case.

e. Needless to say, there is no set method of starting or conducting tests that will determine the proper functioning of such a complex system. Responsible individuals must apply known facts and requirements in the process of evolving suitable tests. For the purpose of illustration, system testing has been divided into three general test areas: testing the interconnect link, testing the satellite link, and testing the complete linkage from the DCS subscriber. Figure 11-4 shows two simplified earth terminals, each with one intersite link to the DCS switch and; hence, to a DCS subscriber. The purpose is to test the compatibility and capability of the interconnect link to meet the message handling requirement of the subscriber to and from the ground station.

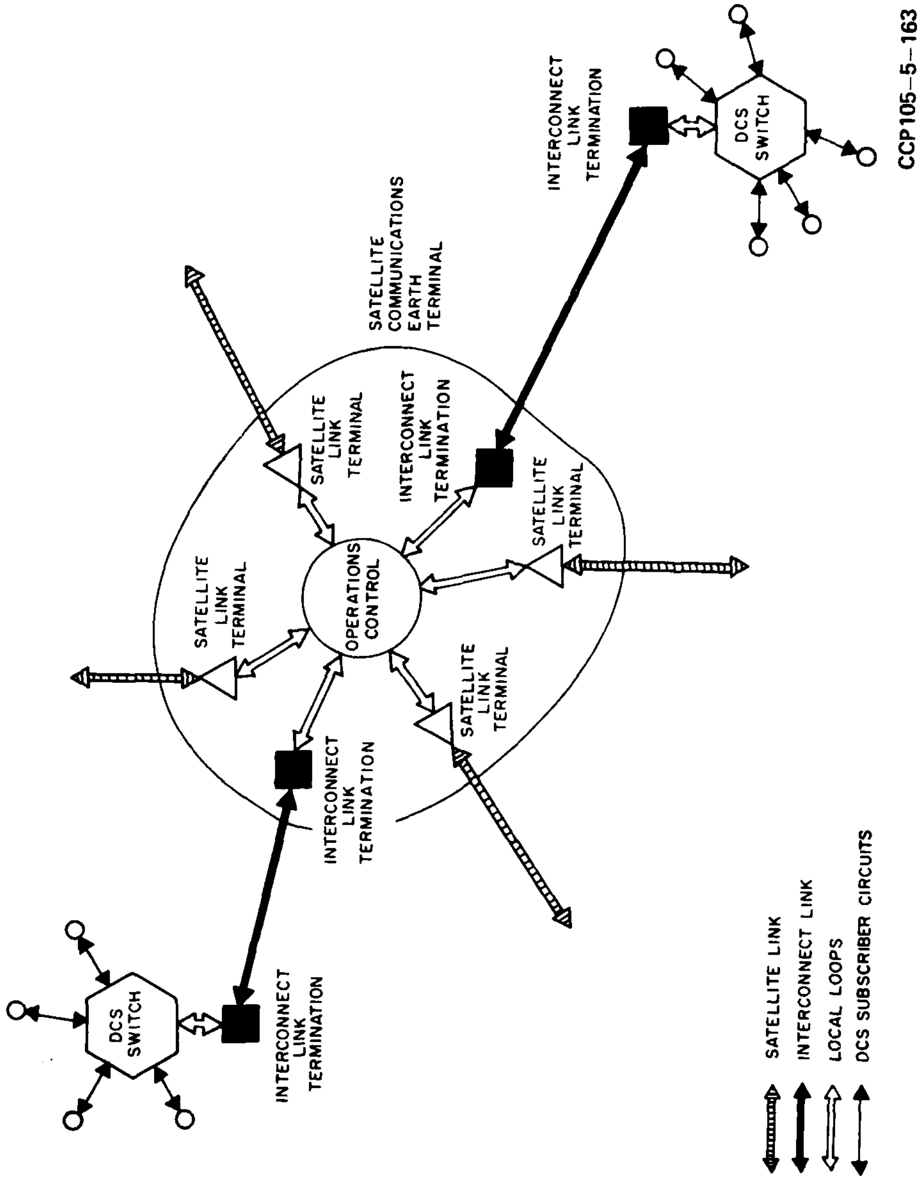


Figure 11-2. *Typical ground station layout diagram.*

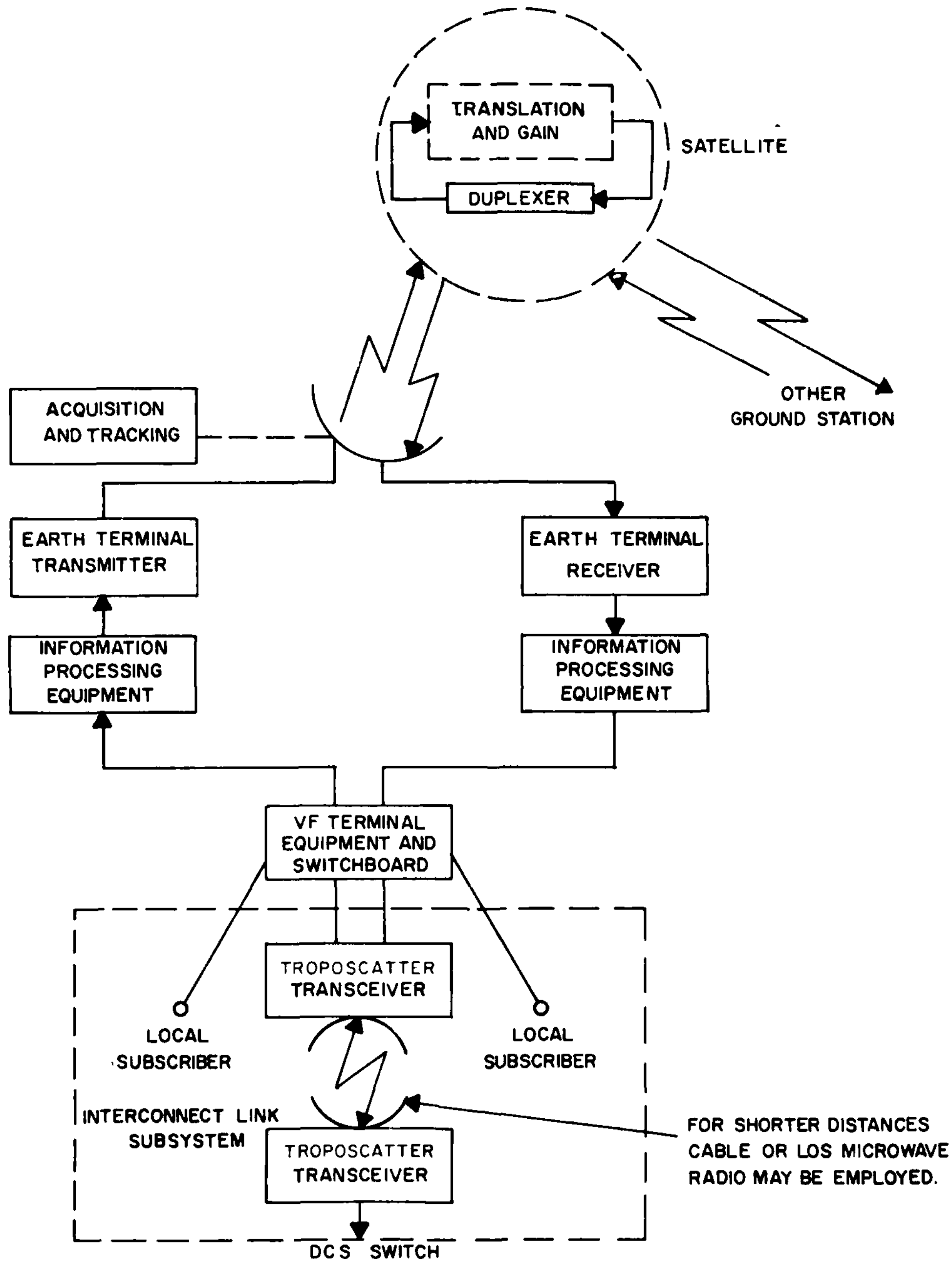


Figure 11-3. Typical earth terminal layout diagram.

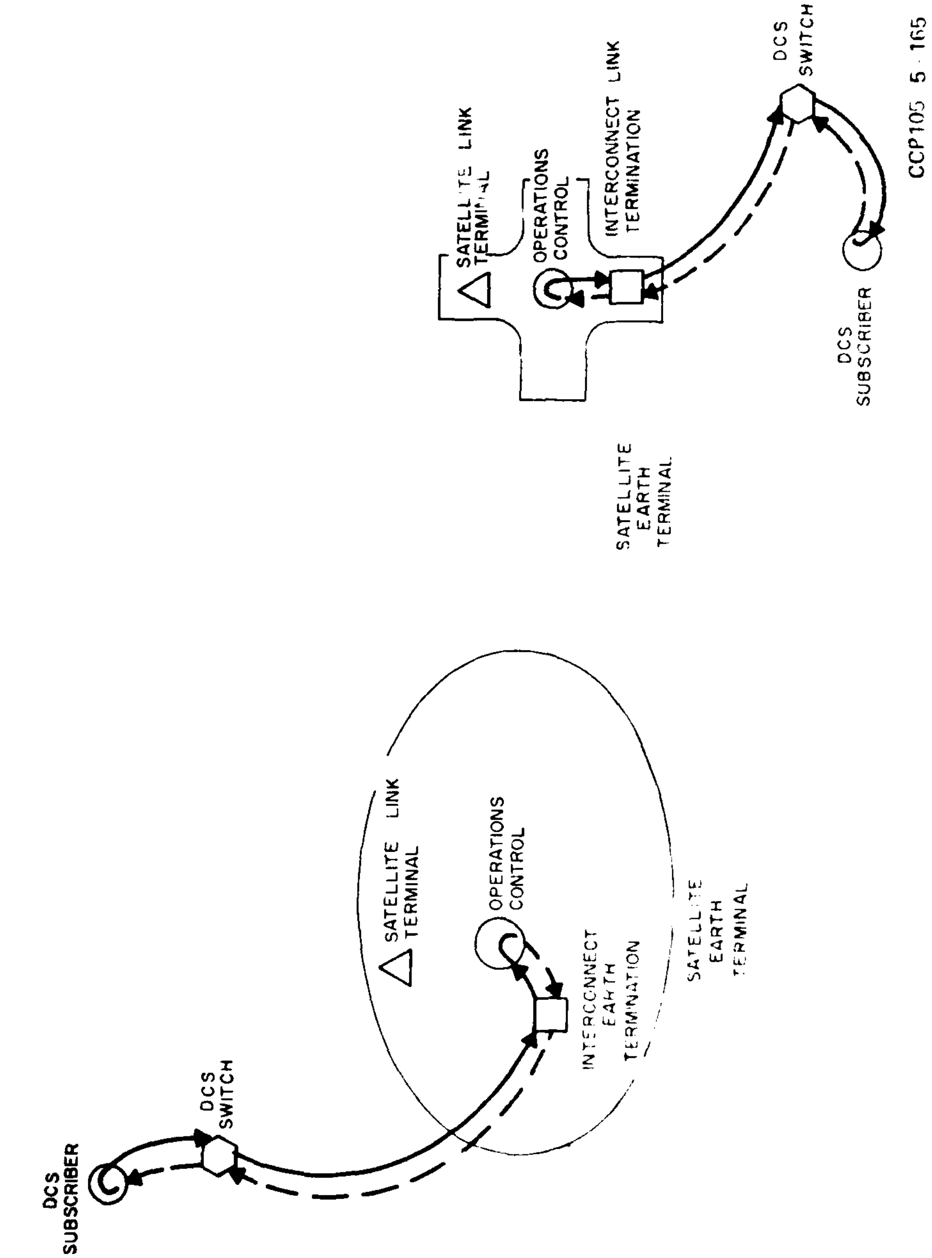


Figure 11-4. Earth terminal ... DCS ... interconnect link layout diagram

f. In figure 11-5 the same two earth terminals are shown; the purpose is to test the capability of the satellite link between the earth terminals to meet the imposed requirements. Figure 11-6 shows a complete satellite communications system. All links are depicted. Starting with the subscriber, the flow is through the DCS switch, to the earth terminal, to the satellite, to the appropriate distant earth terminal, through this terminal to the designated DCS switch and; hence, to the ultimate designee. Any message traversing this path would travel the entire length of the communication system from subscriber to subscriber.

g. In preparing system tests, as with previous tests, the guiding principle must be: What is the system supposed to do? All communications and design specifications for the system must be gathered and correlated. From them the tests, aimed at proving the capability of the system to meet these specifications, can be documented. This documentation should include the objective of the particular test, methods used to prove the objective, results obtained, and necessary changes, if required.

11-5. OPERATIONAL TESTING.

a. As installation tests near completion, operational tests will be phased into the program, often in conjunction with the more advanced installation tests. Operational tests are those tests devised and conducted to determine the operational capabilities of the system. As such, they may be divided into two testing areas consisting of preoperational tests and traffic tests.

b. In both areas, several methods of performing the tests may be used, depending on the availability and configuration of equipment. Paragraph 11-6 deals with these testing areas and the methods most commonly found, and will establish guidelines for specific tests.

11-6. TESTING AREAS.

a. *Preoperational Tests.*

(1) Just as it is impossible in most cases, to turn on a television set or radio and have the desired reception immediately available without some adjustment, so it is impossible to walk into a satellite communication station, turn on the power and start communicating through the satellite. Certain procedures must be followed; certain measurements and checks must be made, certain levels must be established, and all systems must be checked out before attempting to communicate.

(2) To the communications man, this is normal practice, regardless of whether the type of communications is hf radio, microwave, wire, or cable. The most noticeable difference to the satellite communications man, under the present

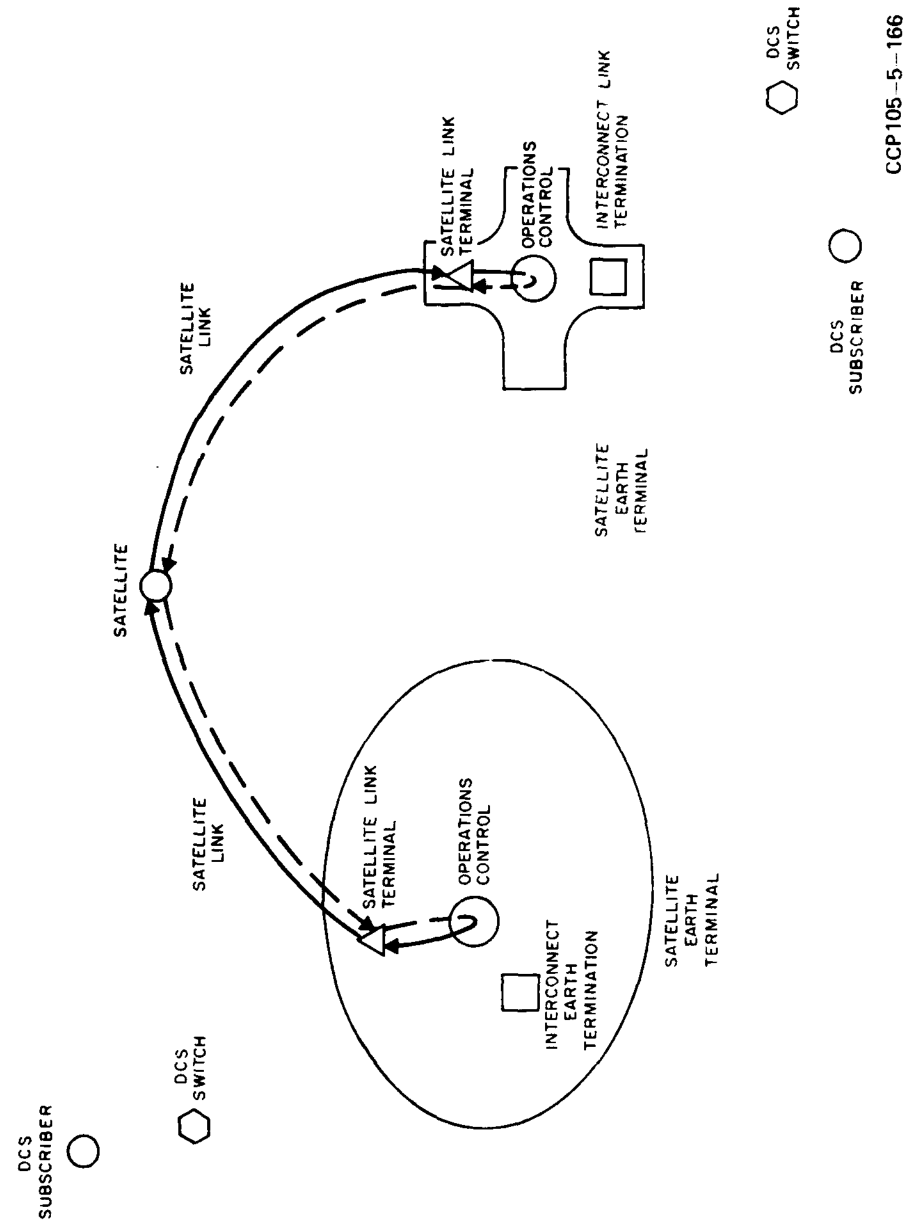


Figure 11 5 Satellite interconnect link layout diagram.

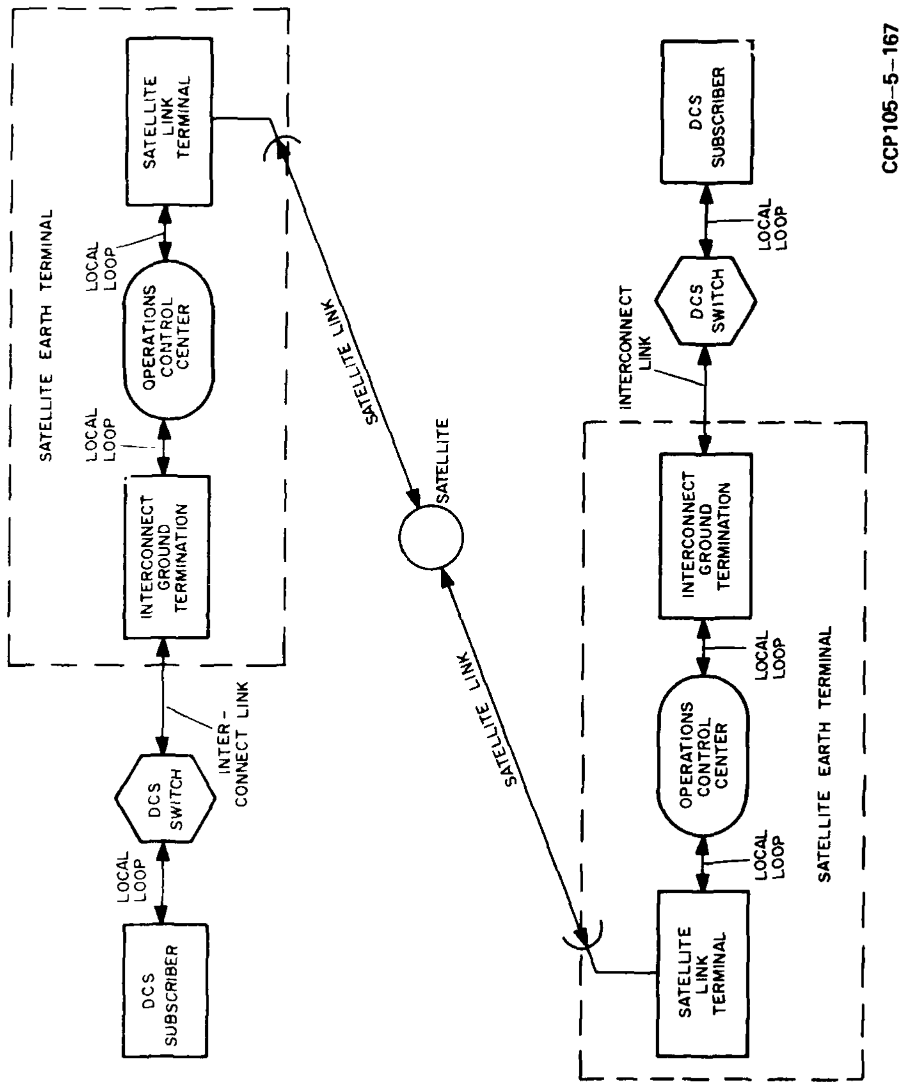


Figure 11-6. Satellite communications system block diagram.

state-of-the-art, will be the extensiveness of the tests required and the amount of time necessary to conduct these tests. General types of preoperational tests are identified in paragraph 11-8.

b. *Traffic Tests.*

(1) After ensuring that the station is ready for operation, as determined by the preoperational tests, the next step is to pass traffic through the system. Many types of testing may be considered at this point.

(2) The guiding principles will be the requirements for specific types of traffic and the capabilities of the equipment to meet these requirements. As with preoperational tests, general types of traffic testing are identified in paragraph 11-8.

11-7. **TESTING METHODS.** At the present time, three basic methods of conducting operational tests are in use. In many stations, use is made of all three of these methods, depending on the equipment available. These methods are discussed below.

a. *Collimation Tower.*

(1) In this method, a tower is errected at a known distance and direction from the center of the antenna pedestal. The baseline from the antenna to the tower is very accurately surveyed. Equipment that will simulate the performance of the satellite is placed in this tower. Although most types of system tests can be performed using the collimation tower and the satellite simulator, one of the principal tests commonly associated with the collimation tower is that of boresighting the antenna. During this procedure, the electrical axis of the antenna is aligned with the azimuth and elevation indicators. This is done by transmitting a beacon signal from the tower, locking onto this signal, and, using the manufacturer's alignment procedures, adjusting the azimuth and elevation indicators to the precise azimuth and elevation determined by the survey.

(2) In addition to alignment of the electrical axis, alignment of an optical axis is often provided. The object is to bring the optical axis into line with the electrical axis. This is usually accomplished by mounting a telescope on the antenna, adjusting it when the antenna is locked onto the transmitted signal, and recording the exact position. This alignment is useful in checking the pointing accuracy of the antenna at positions other than the collimation tower. Inasmuch as many factors may cause a shift in the axis of the antenna, alternate known points around the compass must be used as reference. These reference points generally consist of star shots. The optical axis of the antenna is pointed at a known star; readings are taken from the azimuth and elevation indicators and compared to the previously established reference information.

b. *Flyby Tests.*

(1) This method of testing is aptly named inasmuch as it consists of placing a satellite simulator in an aircraft and flying by the station.

(2) As with the collimation tower, practically all of the system tests can be accomplished through the use of the flyby procedure. This method is particularly good for checking the tracking capability of the system.

c. *Satellite Tests.*

(1) This is, of course, the ultimate method of testing the system; however, its obvious drawback is that a satellite must be available in order to perform such tests.

(2) When this situation exists, whether the satellite in use is the planned subject or another with similar characteristics, the final checkout of the station can be accomplished.

11-8. **TEST GUIDELINES.**

a. *Preoperational Tests.*

(1) Owing to the variety of equipment in use in satellite communications earth terminals, specific tests cannot be devised; however, general tests will be shown for the preoperational area.

(2) In preoperational testing, five general areas are normally considered. These areas are set forth in b through f below, together with an explanation of some types of common tests for each area.

b. *Servo and Tracking Area.*

(1) Control and calibration checks are used to visually check all control functions to assure smooth operation. When this is accomplished, the antenna and servosystem is calibrated against a known reference, usually the collimation tower. Such items as error rate of the synchros, elevation and azimuth tracking error, elevation and azimuth tachometer signals, and elevation and azimuth torque bias are checked against established known references.

(2) Tracking receivers are checked by testing such items as tracking sensitivity and lock-on sensitivity in both manual and automatic modes. Additional measurements are made of such things as am detector output (no signal input), discriminator output (locked condition), and balance error.

c. *Communication Receiving Equipment.*

(1) Paramp gain, calibration, and functional checks and recordings will be made of the receiver front end and will be evaluated against known references to determine a GO or NO-GO condition. Recordings will be made of the following: paramp chamber temperature, pump level, diode bias current, repeller voltage of the pump klystron, diode bias voltage, beam voltage, beam current, pump power and pump frequency, and paramp gain.

(2) Receiver calibration tests will be performed to determine input to the paramp, S/*N* ratio, receivc, threshold (receiver breaks out of lock), adequacy of the afc capture range, and functioning of the search feature. System distortion, response, and modulator frequency deviation tests will be performed. All modes will be checked during this process.

d. *Communication Transmitting Equipment.*

(1) Transmitter runup procedure, as is standard with all high power transmitters, will be accomplished in steps. As these steps are accomplished, visual note will be taken of all warning devices.

(2) As the runup proceeds, readings and recordings will be made of the various functions. These include: low power heat exchangers, exciter checkout and calibration, high power amplifier checkout and calibration, and low power amplifier checkout and calibration.

e. *Associated and Test Equipment.*

(1) Associated equipment is used throughout all tests and checks. This equipment consists of various types of recorders (graph and magnetic), timing, and frequency standards. All of this associated equipment must be calibrated and reference levels set before attempting to perform tests with the basic equipment.

(2) Test equipment, used to obtain system readings and references, must be constantly checked against known calibrated sources to ensure accuracy.

f. *Traffic Tests.*

(1) Although the preoperational tests are representative and are not intended to be complete, they will upon completion, indicate the beginning of the last type of testing to be conducted prior to instituting the operational communications system. This is the actual running of traffic through the system on a test basis prior to turning the system over to the final users. During this testing all types of traffic, for which the system was planned, will be used.

(2) This traffic will include voice, teletypewriter, data, facsimile, and television (if the system has the capability). Also the system's secure features and antijamming provisions, if applicable, will be tested. Definite deviations of test runs or specific tests cannot be predetermined and will be, in each case, an individual matter which will depend on the requirements and the capability of the system, as well as, on the actual operational conditions prevalant. Upon completion of the traffic tests, the system can be handed over to operational personnel.

CHAPTER 12

COMMUNICATION SYSTEM MEASUREMENTS

12-1. GENERAL.

a. Correct measurement techniques are usually acquired through experience. Many documents presenting various methods of measurement are available, but field or laboratory experience provides the best background. However, extensive experience is not necessary for the avoidance of many of the mistakes commonly made.

b. Improper terminations, lack of a termination, or mismatched impedences are common errors which result in erroneous data and repeated tests. Unless the person performing the test is absolutely sure of the test circuit, he should not proceed.

c. Lack of adequate grounding, no ground at all, or ground loops will cause the test to become invalid.

d. Overdriving the input of an amplifier will always result in high distortion, excessive noise and, probably, hum. The inputs and outputs of amplifiers should be monitored regularly to avoid this common mistake.

e. Impedance mismatching will always result in a decrease of the power being transferred and should always be avoided, unless the specific test intentionally calls for it.

f. Equipment terminals may be balanced or unbalanced. Transformers and other devices are available for interconnecting the two types. Care should always be taken to ensure proper conversion.

g. Tip and tip jacks on a double plug have grooves along one side to denote polarity. It is good practice to have the grooves positioned consistently on one side to avoid the possibility of crossing wires when interconnecting equipment.

h. Voice frequency circuits are parallel circuits, in contrast with teletype circuits which are series circuits. This difference greatly affects test circuit wiring, and care is necessary to achieve the correct hookup.

i. One guideline should definitely be followed. If the test or data does not make sense or give reasonable numbers when performed, either run it again or, if time is short, note all factors and suspicions and submit them with the test data. This cannot be overemphasized.

12-2. AUDIO FREQUENCY MEASUREMENTS.

a. *Amplifier Frequency Response.*

(1) Amplifier frequency response measurements ensure that the amplifier will pass the required band of frequencies.

(2) Frequency response measurements are accomplished by injecting a fixed-level signal at the input and measuring the output, while varying the frequency across the desired band. A typical amplifier frequency response test configuration is presented in figure 12-1.

(3) The signal generator provides the range of frequencies desired. The attenuator regulates the signal generator output to the required amplifier input, as measured by voltmeter No. 1. Voltmeter No. 2 measures the outputs of the amplifier as the signal generator is varied through the required amplifier passband.

b. *Audio Distortion.*

(1) Audio distortion measurements are required to ensure that audio distortion will not exceed levels at which deterioration of the signal will occur.

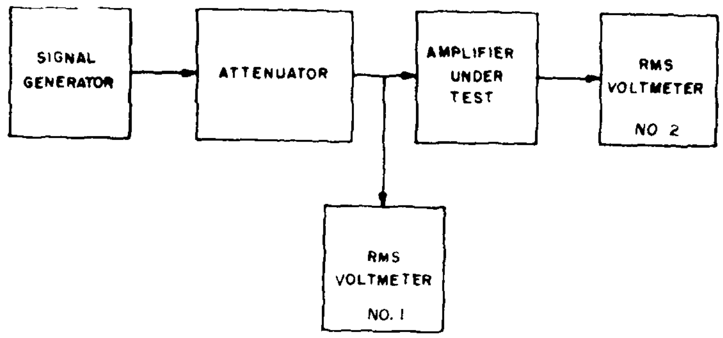
(2) Audio distortion measurements are accomplished by injecting various frequencies at a specified level into the amplifier under test. The amplifier output is mixed with the output of another signal generator at the same signal level and frequency and 180 degrees out of phase, as shown in the figure 12-2 test configuration.

(3) The audio signal generator at the amplifier output is tuned for a minimum reading (null) on the meter or oscilloscope. The difference between the indicated null and a zero indication on the meter or oscilloscope is caused by the distortion of the amplifier. One type of distortion analyzer is an instrument having a mixer, meter, and audio signal generator. Several types of distortion analyzers exist.

12-3. RADIO FREQUENCY MEASUREMENTS.

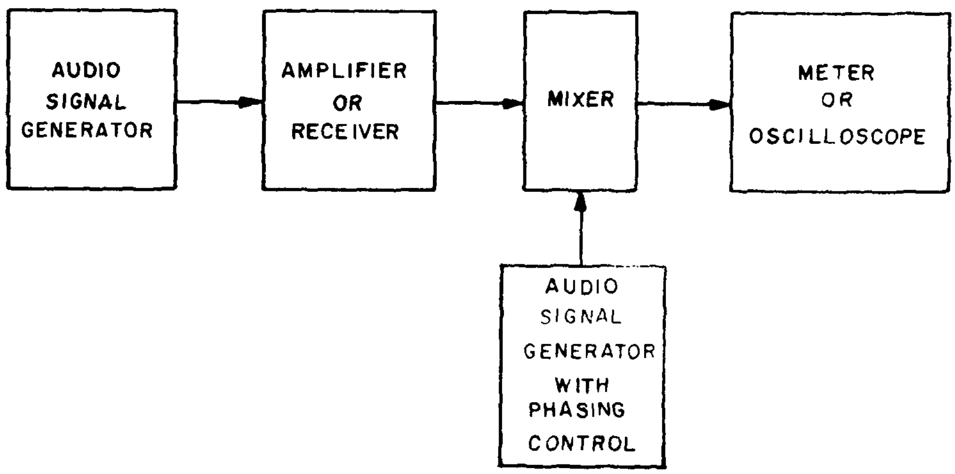
a. *Deviation at Frequency Modulator Output.*

(1) Measurements of deviation at the fm modulator output are required to ensure that the proper deviation for the specified modulation index exists. While direct reading deviation meters do exist, the Bessel null method is accurate over a wider range of frequencies.



CCP105–5–168

Figure 12-1. Amplifier frequency response test configuration block diagram.



CCP105–5–169

12-2. Audio distortion test configuration block diagram.

(2) The Bessel null method of deviation measurement makes use of the characteristic that when peak deviation is 2.4 times the modulating frequency, the carrier will null and maximum deviation of the carrier, for that specified input frequency and level, will occur. A simplified measurement configuration is shown in figure 12-3.

(3) The audio signal generator supplies audio frequencies at variable power levels to the modulator. The frequencies and power levels at which the null occurs are plotted and a deviation curve is established for the modulator.

b. *Carrier-to-Noise Measurements.*

(1) During the Syncom program the communications and beacon carrier levels were measured by precalibrating the receiver agc voltage in terms of carrier power at the input of the paramp. During subsequent satellite tests, the agc voltage was recorded and converted into dBm using the calibration data.

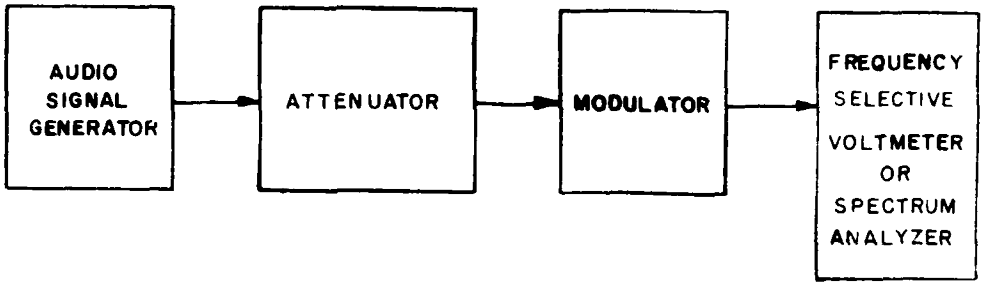
(2) This method has two serious shortcomings; the front end receiver gain changes and the agc response varies due to carrier modulation, which affects the validity of the calibration data.

(3) A new method has been devised in which the IF C/N ratio is measured and absolute carrier level obtained by computation, using the system noise temperature. Since gain variations in the paramp and IF strips preceding the detector do not alter the ratio, the first problem is eliminated. However, modulation of the carrier will probably affect the result and attention to this fact will be necessary during the test program. The theory and procedure for this new method are given below.

(a) *Theory of carrier level measurement using IF carrier-to-noise.*

1. Assume that some measurement technique has determined that the IF C/N ratio is 7 dB, using an amplitude detector with a 10-kHz filter. Suppose also that the system noise temperature of the receiving system has been measured and is 200 K referenced to the input of the paramp.

2. In chapter 4, it was determined that the average noise level for a 10-kHz band at 200 K is:



CCP105-5--170

Figure 12-3. Test configuration for fm output measurement block diagram.

$$N = kTB \text{ (mW)}$$

$$= 10 \log_{10} TB - 198.6 \text{ (dBm)}$$

$$= 10 \log_{10} (200 \times 10^4) - 198.6$$

$$= -135.6 \text{ dBm} \qquad (12.1)$$

3. Since the noise level referenced to the paramp input across 10 kHz is -135.6 dBm and the C/*N* ratio (which will be the same at the paramp as at the detector input) is 7 dB; the communications carrier level at the input to the paramp has to be -128.6 dBm. It must be remembered that the ratio will hold constant only when referenced to the same bandwidth. This is a basic principle of the measurement technique and must be clearly understood before proceeding further.

(b) *Measurement procedure.*

1. The diagram (fig. 12-4) illustrates the basic circuit used in the measurement of the C/*N* ratio using a fictitious receiver with a predetection IF of 10 MHz. The IF is tapped off, buffered through an amplifier and sent to the test van. Once in the van, the signal is padded and sent into a mixer. The mixing frequency is selected by switch S-1 from one of two local oscillators. The output of the mixer is applied through a 10-kHz wide filter, centered at 500 kHz, and finally transmitted to a linear detector. The detector provides a dc output voltage linearly related to its input power in dBm over a 40-dB range.

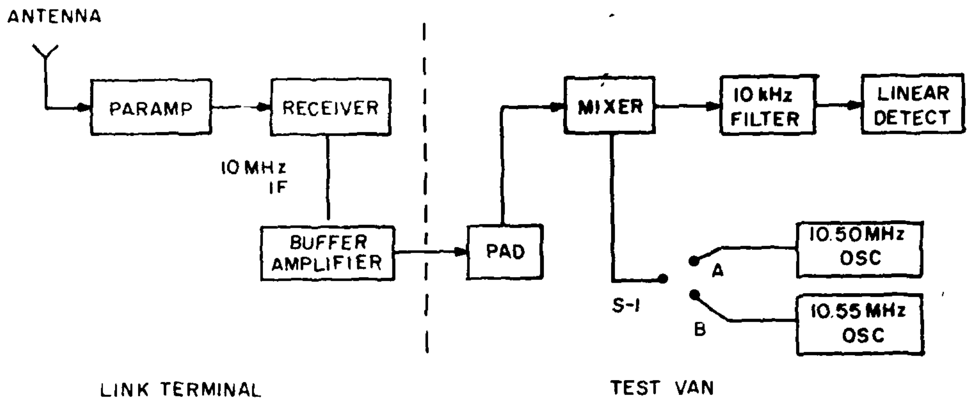


Figure 12-4. C/N ratio measurement circuit block diagram.

2. The measurement procedure begins by placing switch S-1 in position A. This causes the carrier (centered at 10 MHz) to be translated down to 500 kHz, passed through the filter, and detected. The dc voltage is measured and recorded.

3. S-1 is placed in position B, which causes a 10-kHz portion of the IF signal, centered at 10.05 MHz, to be translated down and appear at the output of the filter. Since the carrier, unmodulated in this case, was centered at 10.00 MHz, only noise is detected. The dc voltage is measured and recorded.

4. The voltage difference between the two readings, when converted to dB using the calibration data of the detector, is the C/N ratio of the IF for a 10-kHz noise bandwidth. The diagrams (fig. 12-5) illustrate the frequency translation scheme.

5. The system noise temperature will be measured using equipment available within the link terminal. Using the *k*TB formula for the 10-kHz bandwidth, carrier level in dBm is easily obtained.

6. The entire process can and will be automated using a computer for several bandwidths, and for several IF frequencies of interest. The several bandwidths will be switch-selected, with contact closure indications to inform the computer which B to use in the computation.

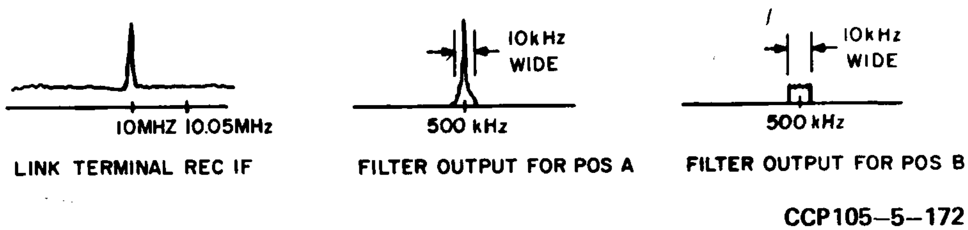


Figure 12-5. Frequency translation diagram.

c. *Test Tone-to-Noise (Audio Signal-to-Noise) Ratio.*

(1) Test tone-to-noise measurements provide an indication of the quality of the audio signal. This is accomplished by comparing the signal and noise on a modulated carrier with the noise on an unmodulated carrier. The higher the ratio of signal plus noise-to-noise, the higher the quality of the signal. The measurement is accomplished using the test configuration in figure 12-6.

(2) A reading is taken on the true rms voltmeter of the demodulator output with no modulation on the carrier (carrier plus noise). This reading is compared with a reading taken at the demodulator output of a carrier modulated at a specified level. The ratio of signal plus noise-to-noise (carrier only) is indicative of the quality of the received signal.

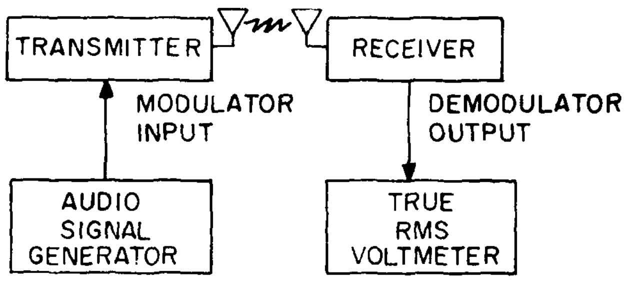
d. *Amplifier Gain, Attenuator, or Cable Loss.*

(1) A gain or loss measurement will indicate the specific quantity of gain or loss, as applicable, of an amplifier, attenuator, cable, other component, or an entire system. This measurement is accomplished using a test setup similar to the configuration shown in figure 12-7.

(2) With impedance properly matched, a comparison is made (on specified frequencies) between input and output; this indicates gain or loss.

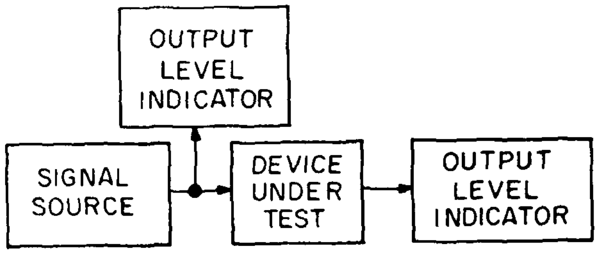
e. *Envelope Delay Distortion.*

(1) Envelope delay distortion measurements provide an indication on nonlinear phase shift characteristics in the amplifier passband. Nonlinear phase shift has a negligible effect on am; however, it highly degrades fm signals since fm is actually a form of phase modulation and the nonlinear phase shift causes unwanted modulation which is, in itself, distortion.



CCP105–5–173

Figure 12-6. Test tone-to-noise measurement configuration block diagram.



CCP105–5–174

Figure 12-7. Gain/loss measurement configuration block diagram.

(2) The test configuration for measuring envelope delay distortion is shown in figure 12-8.

(3) The signal generator is adjusted to the center frequency of the passband and the variable delay is adjusted so that both indications are coincident. The signal generator carrier frequency is adjusted (in equal increments) to various points throughout the passband. The delay between the two oscilloscope traces (differential delay) is plotted resulting in an envelope delay distortion curve.

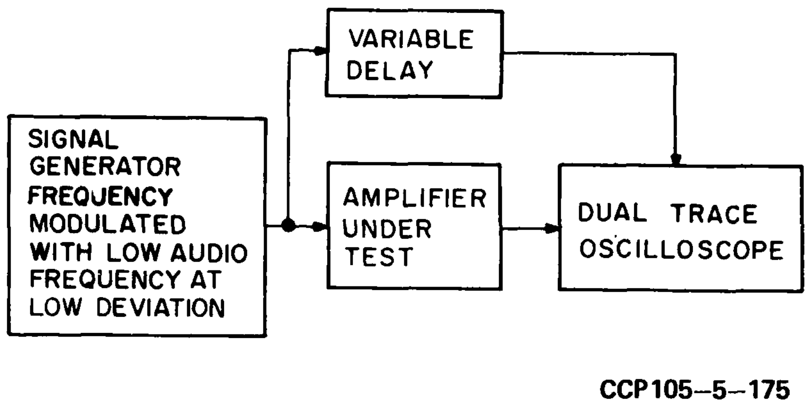


Figure 12-8. Envelope delay distortion measurement configuration block diagram.

Commercial envelope delay distortion analyzers use a sweep generator which permits the entire passband to be analyzed simultaneously.

12-4. MICROWAVE MEASUREMENTS.

a. *Voltage Standing Wave Ratio.*

(1) The voltage standing wave ratio (vswr) measurements give a qualitative indication of the condition of transmission lines, waveguides, and their terminations. On transmitter-to-antenna waveguides or coaxial lines, a high vswr indicates high power reflected back from the antenna and can damage the output amplifier. On receiving lines, a high vswr results in a poor system noise temperature; in most cases these transmission lines precede a paramp or other low noise amplifier.

(2) The vswr is usually measured using the configuration shown in figure 12-9. The signal generator provides an am audio signal modulating the signal generator output at the required frequency. The swr probe is moved along the slot and picks up relative amplitudes corresponding to the standing wave pattern. The ratio of maximum to minimum pickup (swr) is indicated directly on the swr indicator. Most transmitters have directional couplers and swr indicators built into the waveguide at the transmitter output.

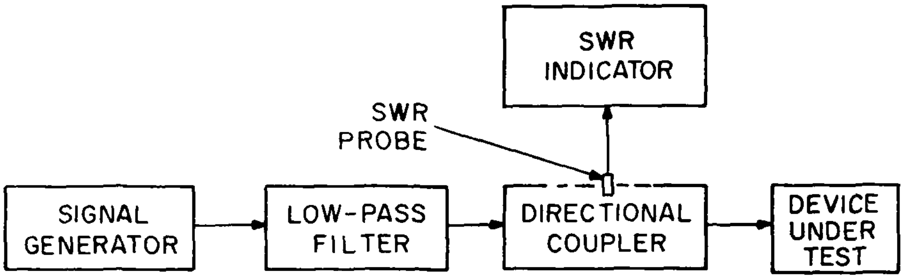


Figure 12-9. Test configuration for vswr measurement block diagram.

b. *Noise Temperature.*

(1) The noise temperature measurement indicates the amount of noise added to the signal by the receiving system. Normally, the noise temperature of the system is predetermined through design and selection of components. However, it must be measured from time to time to provide an indication of whether the noise temperature has changed or deteriorated over a period of time.

(2) Aside from direct-reading automatic equipment, which is sometimes provided as an integral part of a system, the Y-factor method is the preferred method of measurement.

(3) The test configuration for the Y-factor measurement system is as shown in figure 12-10. The Y-factor of noise temperature measurement is the most accurate method known. Using correct equipment and procedures, accuracy is within 0.1 dB of the reading obtained.

(4) With the noise generator off, the attenuator is adjusted to set a reference level on the meter. The attenuator reading is recorded. The noise generator is turned on and the attenuator readjusted to return the meter to its original reference. The difference between the NOISE SOURCE OFF and the NOISE SOURCE ON attenuator reading is the Y-factor. This Y-factor can be converted to noise temperature in kelvin by calculation or by reference to charts and graphs.

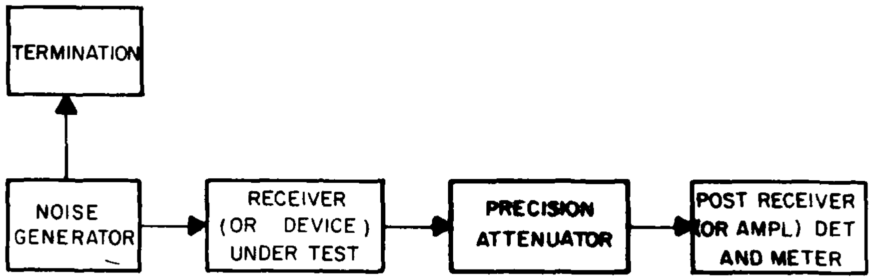


Figure 12-10. Y-factor noise measurement configuration block diagram.

12-5. HIGH SPEED DATA ERROR RATE.

a. *General.*

(1) High speed data error rate or bit error rate (ber) is an important measure of the quality of a digital circuit.

(2) Most digital circuits have a requirement to pass data with x number of errors in y bits measured over m time period at a data rate of r bits per second; i.e. one error in 1,000 bits per one minute time interval at a data rate of 1,000 bits per second, meaning a total of 60 errors would be allowed in a one minute time interval.

b. *Test Method.*

(1) The ber of a digital (data) circuit can be evaluated using the test configuration in figure 12-11.

(2) A pattern is produced by a random pattern generator at station A and transmitted to station B where an identical pattern is produced. The transmission test set compares the two patterns and registers any differences as errors. These errors are counted and/or recorded over a specified time interval. This reading, or recording, would then be compared to the specified error rate.

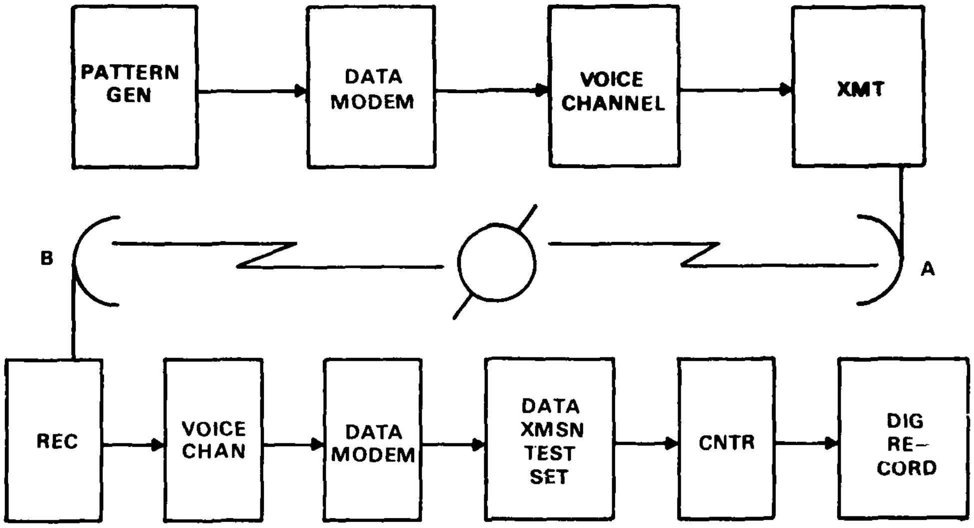


Figure 12-11. Bit error rate digital circuit configuration block diagram.

12-6. PHASE JITTER.

a. *General.*

(1) One factor that can influence the ber is phase jitter. The degree of influence is determined by the type of system (synchronous or asynchronous).

(2) Phase jitter is characterized by rapid incremental changes in phase of a given frequency transmitted through a voice frequency channel.

b. *Test Method.*

(1) The test configuration in figure 12-12 may be used to evaluate the presence and the amount of phase jitter.

(2) Phase jitter $[|\Delta\phi|^\circ]$ is obtained by displaying the receive signal $[f_t + |\Delta\phi|^\circ]$ on channel A of the dual trace oscilloscope and comparing it to the reference signal $[f_t]$ that is injected into channel B by the oscillator. Phase jitter is the deviation $[|\Delta\phi|]$ of the received signal $[f_t + |\Delta\phi|]$ about the reference signal $[f_t]$. $|\Delta\phi|^\circ$ is calculated as:

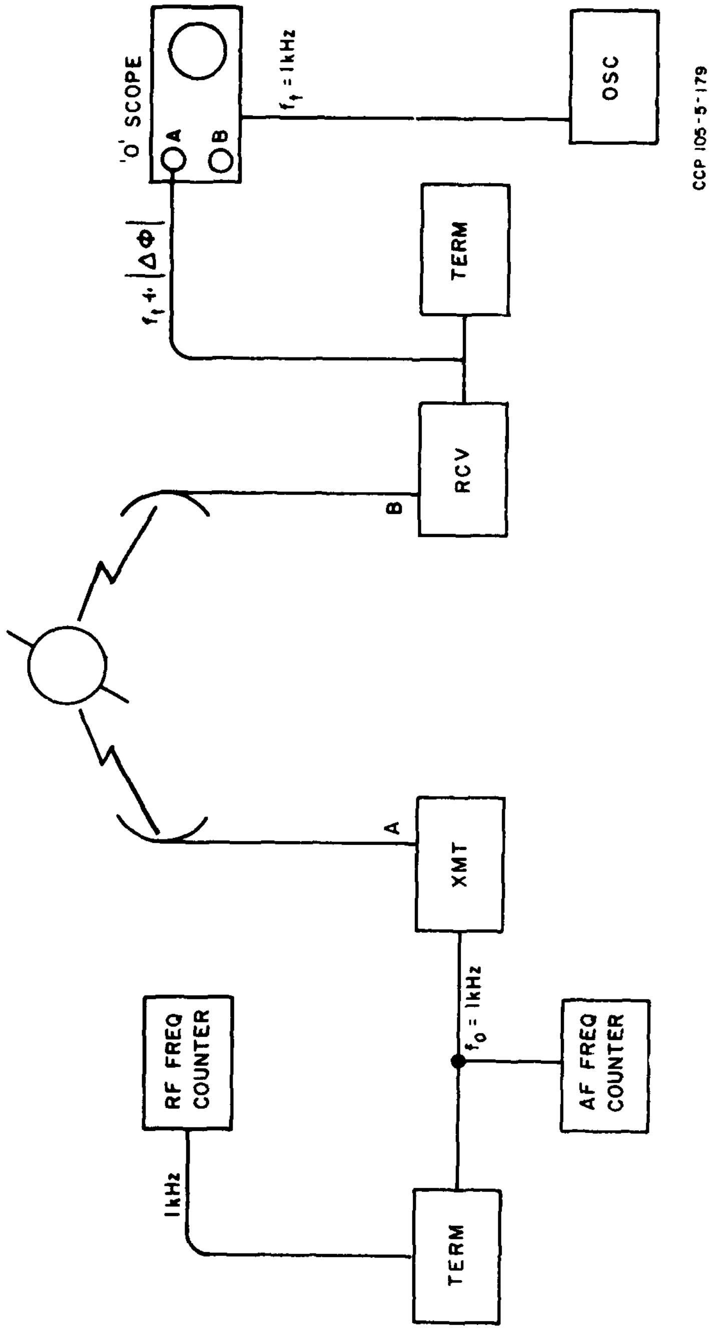


Figure 12-12. Phase jitter evaluation configuration block diagram.

$$|\Delta\phi|^{\circ} = (360^{\circ}/T)\,(|\Delta\phi|\ HD)\,(HSR) \qquad (12.2)$$

where

T is $\frac{1}{f_t}$ for f_t in MHz

HD is oscilloscope horizontal divisions

HSR is oscilloscope horizontal sweep rate

12-7. IMPULSE NOISE MEASUREMENT.

a. *General.*

(1) Impulse noise is another factor that will influence the ber of the data system. The noise spike may alter one or more bits in a message and cause an error.

(2) The impulse noise of a system can be evaluated using the test configuration in figure 12-13.

b. *Measurement Technique.*

(1) The data channel is terminated in its characteristic impedance at the transmit station. The corresponding channel at the receive end of the system is terminated with the impulse noise counter.

(2) The channel is monitored for a specified time interval. At the end of the time interval, the number of hits can be read directly off the impulse noise counter.

12-8. DELAY DISTORTION MEASUREMENT (ENVELOPE DELAY).

a. *General.* A third factor that influences the ber of a data system is delay distortion. The amount of delay distortion will be specified for a given data rate.

b. *Test Method.*

(1) The test configuration (fig. 12-14) may be used to evaluate the amount of delay distortion in a given channel.

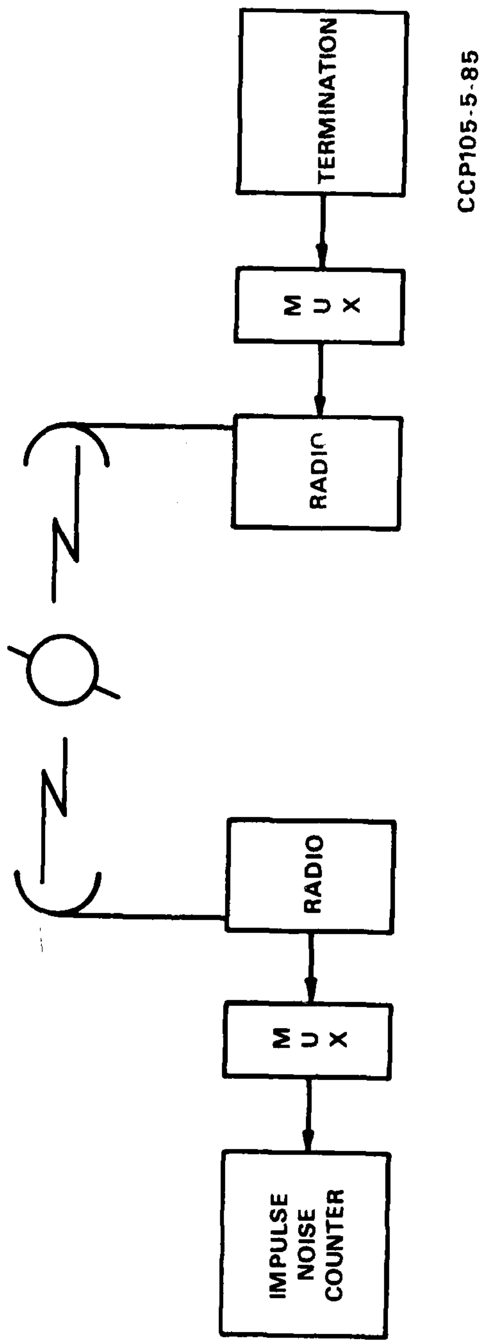


Figure 12-13. Impulse noise evaluation configuration block diagram.

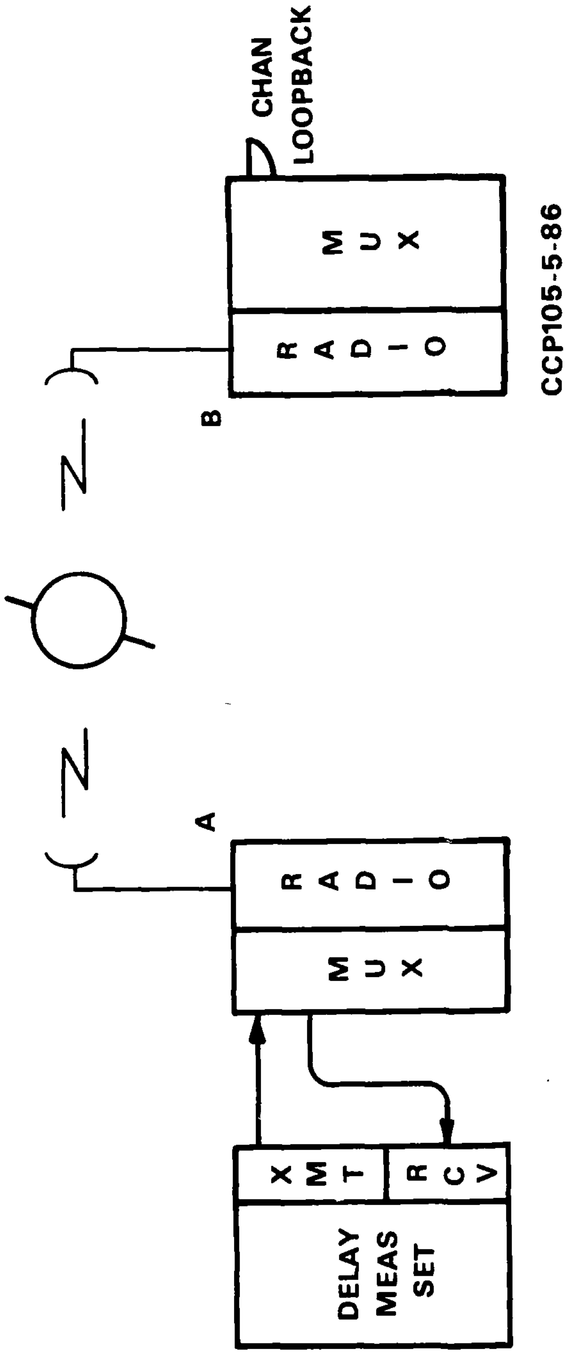


Figure 12-14. Delay distortion test configuration block diagram.

(2) The data channel under test is looped back at the B terminal. A delay measuring set is used at the A terminal. The measuring set is used as directed in the operating instructions for the set. The amount of delay distortion (envelope delay) may be read directly from the measuring set. The amount of delay read is divided by two (2) to give the delay for a one-way path.

CCP 105-5 15 April 1974

(ACC-FD)

FOR THE COMMANDER:

OFFICIAL:

THOMAS C. MUSGRAVE
Colonel, GS
Chief of Staff

DANIEL L. LEDDA
Colonel, AGC
Adjutant General

DISTRIBUTION

25 - ACC-POA
10 - ACC-FD
5 - ACC-LOG
1 - ACC-PER
1 - ACC-COMPT
1 - ACC-ENGR
1 - ACC-IG
5 - ACC-AG-ASD
2 - ACC-ACO
300 - ACCH-PCA-ASP
1 - USACC Liaison Office, Rm 2D-576, Pentagon Bldg, Washington, DC 20310
1 - Signal School LNO USASCS Ft. Monmouth, NJ 07703
1 - USACC LNO Ft. Monmouth, NJ 07703
1 - USACC Safeguard Comm Agency
1 - 11th Signal Group (USACC)
5 - 521st Signal Co, Ft. Huachuca, AZ 85613
100 - USACEEIA, Ft. Huachuca, AZ 85613
5 - USACEEIA WH, Ft. Huachuca, AZ 85613
25 - USACEEIA-CONUS, Ft. Ritchie, MD 21719
2 - USACEEIA-CONUS Field Office EAST, Ft. Meade, MD 20755
2 - USACEEIA-CONUS Field Office CENTRAL, Ft. Houston, TX 78234
2 - USACEEIA-CONUS Field Office WEST, Presidio SF CA 94129
10 - USACEEIA-EUR, APO New York 09056
10 - USACEEIA-PAC, APO San Francisco 96557
5 - USACC-ALASKA

15 - USACC-CONUS
5 - USACC-FORCES, Ft. McPherson, GA 30330
10 - USACC-TRADOC, Ft. Monroe, VA 23651
5 - USACC-AMC, Alexandria, VA 22333
10 - USACC Satellite Communications Station, Camp Roberts, CA 93451
10 - USACC East Coast Telecommunications Center, Ft. Detrick, MD 21701
10 - USACC Satellite Station, Lakehurst, P.O. Box 85, Lakehurst, NJ 08733
10 - USACC-EUR
2 - USACC-EUR Area Maint and Sup Fac, APO NY 09086
2 - USACC-EUR SC Spt Agency Mediterranean, APO NY 09019
2 - USACC-EUR Signal Facility Tehran, APO NY 09205
2 - HHD Signal Group 106 USACC-EUR, APO NY 09131
10 - USACC-PAC
2 - USACC Signal Support Agency, Hawaii, APO San Francisco 96557
10 - USACC-JAPAN, APO San Francisco 96343
5 - 1st Signal Brigade (USACC), APO San Francisco 96301
10 - USACC-THAILAND, APO San Francisco 96346
2 - USACC-SOUTH
2 - HQDA (DAMO-TCS) Washington, DC 20310
2 - HQDA (DAMA) Washington, DC 20310
400 - US Army Southeastern Signal School, Ft. Gordon, GA 30905
25 - US Army Signal Center & School, Ft. Monmouth, NJ 07703
5 - Lexington-Bluegrass Army Depot, Lexington, KY 40507
2 - US Military Academy, West Point, NY 10996
3 - US Army Satellite Communications Agency, Ft. Monmouth, NJ 07703
2 - US Army Infantry School, Ft. Benning, GA 31905
2 - US Army Artillery School, Ft. Sill, OK 73503
2 - US Army Air Defense Artillery School, Ft. Bliss, TX 79906
2 - SATCOM Detachment, Landstuhl, Germany, APO NY 09180
2 - SATCOM Detachment, Korea, APO SF 96301
2 - USACC Long Lines, Okinawa, APO SF 96331
5 - USACSA, Ft. Monmouth, NJ 07703
2 - USACSA(H), Ft. Huachuca, AZ 85613
5 - DCA, Washington, DC 20305
5 - Defense Communications Engineering Center, Reston, VA 22070
5 - Air Force Communications Service, Richards-Begaur AFB, MO 64030
2 - Rome Air Development Center, Rome, NY 13440
5 - S Army Electronics Command, Ft. Monmouth, NJ 07703
5 - US Naval Electronics Systems Command, Washington, DC 20360
5 - US Naval Telecommunications Command, 4401 Mass Ave, Washington, DC 20390
2 - Joint Tactical Communications Office (TRI-TAC), Ft. Monmouth, NJ 07703
2 - US Naval Academy, Annapolis, MD 21402
2 - US Air Force Academy, Colorado Springs, CO 80900

ILMED

8