COMSAT TECHNICAL REVIEW
Volume 10 Number 2, Fall 1980

Nickel-Hydrogen Battery technology—Development and status*

J. D. DUNLOP AND J. F. STOCKEL

(Manuscript received July 23, 1980)

Abstract

A history of the development of the nickel-hydrogen (Ni-H$_2$) technology at COMSAT Laboratories is presented, starting with the initial exploratory effort in 1970 and extending to the present INTELSAT V Ni-H$_2$ batteries. The latter are based upon the proven reliability of the NTS-2 Ni-H$_2$ cell design and the manufacturing procedures developed to fabricate these cells.

This paper compares four different Ni-H$_2$ batteries: the NTS-2 battery, the INTELSAT V flight battery, a 50-Ah battery and a high-pressure 50-Ah battery. Although all these batteries use the same baseline cell design, the length of the pressure vessel and number of modules in the electrode stack are varied to achieve the desired ampere-hour capacity. Comparisons are made with respect to energy density, energy per unit volume, structural design, and heat transfer capability. The maximum energy density achieved was 60.1 Wh/kg for the high-pressure 50-Ah cell.

Introduction

Exploratory research and development of hermetically sealed Ni-H$_2$ cells was started at COMSAT Laboratories in 1970 based upon an

---

original invention [1]. Once initiated, this technology advanced very rapidly from the initial experimental proof-of-concept phase to the development phase. The basic concepts were verified using a variety of available electrode stack components. Since the sealed Ni-H$_2$ cell is a hybrid combining battery and fuel cell technologies, components were available from both technologies for use during the early experimental work. For example, nickel-oxide electrodes were taken from aerospace nickel-cadmium (Ni-Cd) cells, and platinum black hydrogen electrodes were taken from hydrogen-oxygen fuel cells; both battery and fuel cell types of separator materials were used. Performance of these experimental cells, combined with the very favorable electrochemistry of the system, indicated significant potential advantages of Ni-H$_2$ over Ni-Cd batteries for communications satellite applications.

After the validity of the principle was demonstrated, a research and development program was started with INTELSAT funding to develop a 50-Ah Ni-H$_2$ cell. The following major development areas were included:

a. design of a 50-Ah cell;
b. investigation of lifetime limiting mechanisms;
c. development of a hermetically sealed pressure shell;
d. development and optimization of the electrode stack components.

By 1975, the technology had advanced sufficiently to warrant a flight demonstration of an Ni-H$_2$ battery. INTELSAT and the Naval Research Laboratory (NRL) entered into an agreement whereby COMSAT would provide NRL with 35-Ah Ni-H$_2$ cells for use on the U.S. Navy Technology Satellite (NTS-2) spacecraft. This program united all the previous development efforts and culminated in a very reliable Ni-H$_2$ cell design.

Salient features of the NTS-2 cell design are the following:

a. Inconel 718 pressure vessel with plastic compression seals at each end for the negative and positive terminals;
b. electrochemically impregnated positive nickel-oxide electrodes (Bell Laboratories aqueous process) in a back-to-back design;
c. asbestos separator material;
d. platinum black negative electrodes with teflon backing.

The NTS-2 battery was the first Ni-H$_2$ battery in space and has performed exceptionally well to date [2]. Results from the NTS-2 flight

demonstration combined with laboratory test data showed that Ni-H$_2$ batteries offer significant advantages over Ni-Cd batteries for commercial geosynchronous satellites. As a result of these efforts, Ni-H$_2$ batteries are now scheduled for use on board the INTELSAT V F-4, F-5, and F-6 spacecraft and are proposed for the INTELSAT V-A program. INTELSAT V will provide the first application of Ni-H$_2$ batteries on board an operational communications satellite and will represent a significant accomplishment for COMSAT Laboratories.

Electrochemistry of Ni-H$_2$

Table 1 gives the Ni-H$_2$ cell electrochemical reactions for normal operation, overcharge, and reversal.

TABLE 1. ELECTROCHEMICAL REACTIONS

	Normal Operation	
Nickel Electrode	$NiOOH + H_2O + e^- \underset{charge}{\overset{discharge}{\rightleftharpoons}} Ni(OH)_2 + OH^-$	(1)
Hydrogen Electrode	$1/2\ H_2 + OH^- \underset{charge}{\overset{discharge}{\rightleftharpoons}} H_2O + e^-$	(2)
Net Reaction	$1/2\ H_2 + NiOOH \underset{charge}{\overset{discharge}{\rightleftharpoons}} Ni(OH)_2$	(3)
	Overcharge	
Nickel Electrode	$2\ OH^- \rightarrow 2e^- + 1/2\ O_2 + H_2O$	(4)
Hydrogen Electrode	$2\ H_2O + 2e^- \rightarrow 2\ OH^- + H_2$	(5)
Chemical Recombination of O$_2$	$1/2\ O_2 + H_2 \rightarrow H_2O$	(6)
	Reversal	
Nickel Electrode	$H_2O + e^- \rightarrow OH^- + 1/2\ H_2$	(7)
Hydrogen Electrode	$1/2\ H_2 + OH^- \rightarrow e^- + H_2O$	(8)

Normal operation

The net reaction shows hydrogen reduction of nickelic hydroxide to nickelous hydroxide on discharge with no net change in KOH concentration or in the amount of water within the cell.

Overcharge

Electrolysis occurs on overcharge with oxygen generated at the positive electrode and with hydrogen at the negative electrode. Oxygen is chemically recombined with hydrogen to form water at the catalytic platinum electrode. Again, there is no change in the KOH concentration

or the amount of water in the cell with continuous overcharge [combine reactions (4), (5), and (6)].

The oxygen recombination rate at the negative platinum electrode is very rapid so that even at high rates of overcharge there is no significant buildup of oxygen in the hydrogen (oxygen partial pressure of less than 1 percent is typical).

Reversal

During cell reversal, hydrogen is generated at the positive electrode and consumed at the negative electrode at the same rate. Therefore, the cell can be continuously operated in the cell reversal mode without a pressure change or a net change in electrolyte concentration. This is a unique feature of the system.

Self-discharge

The electrode stack is surrounded by hydrogen under pressure. A salient feature is that the hydrogen reacts electrochemically, but not chemically, to reduce the nickelic hydroxide. Actually, the nickelic hydroxide is reduced, but at such an extremely low rate that performance for any of the aerospace applications is not affected.

NTS-2 in-orbit data

The Ni-H_2 battery continues to supply all the energy storage requirements of the NTS-2 spacecraft. This spacecraft has been operating in orbit for three years, and the battery has powered the spacecraft during seven eclipse seasons and one short lunar eclipse. To date, the battery has accumulated approximately 420 cycles.

Figure 1 shows the longest eclipse end-of-discharge voltage for seven eclipse seasons. During the first two seasons, the spacecraft was operated at full load most of the time. However, subsequent to the second season, a failure of a major experiment was experienced, thereby reducing the energy demand from the battery. The depth-of-discharge (DOD) during the seventh eclipse season was increased to 40 percent.

Near the middle of the seventh eclipse season, the spacecraft failed to execute the yaw maneuver that orients the solar arrays toward the sun. As a result, the Ni-H_2 battery experienced a complete discharge. The yaw maneuver was properly executed during the next orbit and the spacecraft returned to normal operation under reduced power load.
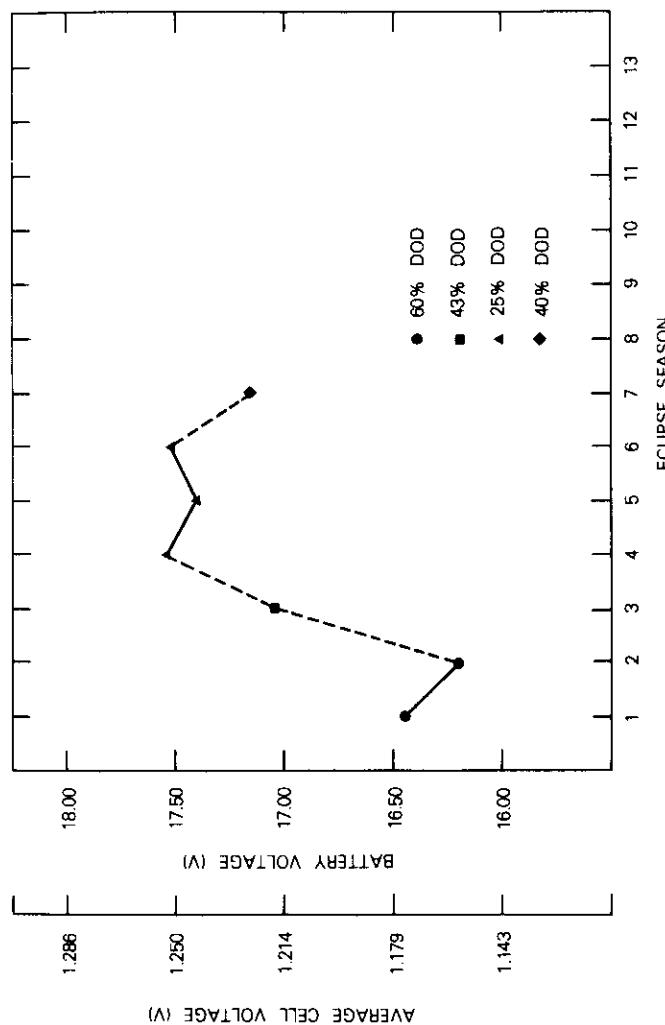
Figure 1. *NTS-2 Ni-H_2 Battery Longest Eclipse End-of-Discharge Voltage*

Figure 2 depicts this event by illustrating the battery voltage and current vs time. The eclipse began with the spacecraft entering the shadow and the battery was discharged for 1 hr. Upon reentering the sunlight, the solar array was disoriented, thereby causing its power output to vary. For approximately 1 hr, the solar array was charging the battery. During the next 2.25 hr, the battery was completely discharged to its cutoff of 14.0 V.
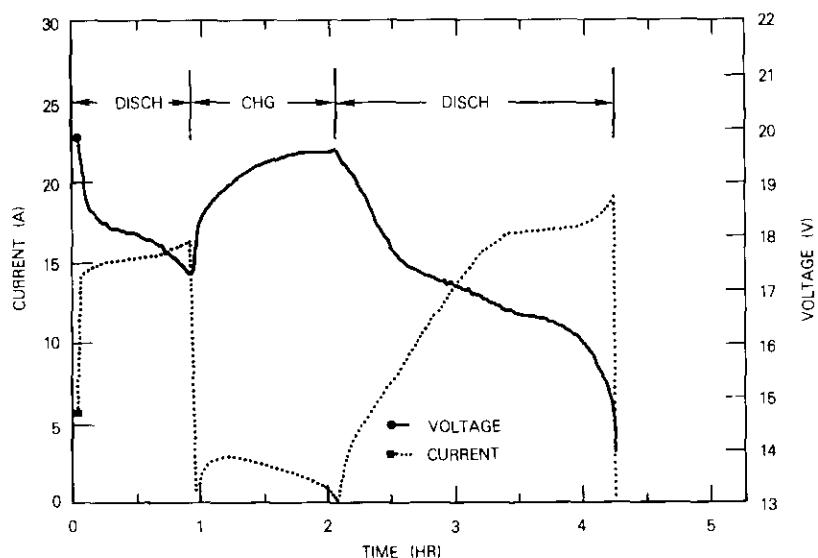

Figure 2. *NTS-2 Battery Current and Voltage Data*

The measured battery's capacity was 40.3 Ah for this complete discharge and the measured temperature was 17.8°C. Prelaunch measured capacity at 15°C temperature was 40.0 Ah. These data indicate no measurable loss in capacity of the Ni-H$_2$ battery after 3 years in orbit.

Comparison of different Ni-H$_2$ batteries

The NTS-2 battery, the INTELSAT V flight battery, a 50-Ah battery, and a high-pressure 50-Ah battery are compared in terms of energy density, volume density, structural design, and heat transfer. Each of these batteries uses the same baseline cell design.

Energy density of individual cells

Design of the Ni-H$_2$ cell was a compromise between maximum energy density and fabrication constraints [3], [4]. Early in the development program (1972), a cell design was selected which used a cylindrical pressure vessel 8.89 cm (3.5 in.) in diameter to store the hydrogen gas. The design goal was to develop a 50-Ah cell. Ampere-hour capacity can be changed by adjusting the length of the pressure vessel and the number of modules in the electrode stack while maintaining the diameter at 8.89 cm.

Table 2 presents Ni-H$_2$ cell weights for NTS-2, INTELSAT V, and the 50-Ah high-pressure cell, and a weight breakdown of the individual cell components. Table 3 shows the effects of ampere-hour capacity and temperature on energy density. As capacity increases, energy density improves. Electrode stack components become a larger percentage of the total weight, and the pressure vessel and miscellaneous hardware become a reduced percentage. Continuing to increase the cell energy above 62.5 Wh (50 Ah) does not improve energy density.

TABLE 2. CELL WEIGHTS

Parameter	INTELSAT V (30 Ah)		NTS-2 (35 Ah)		Laboratory Test Cell (50 Ah)	
	Weight (g)	Percentage of Total	Weight (g)	Percentage of Total	Weight (g)	Percentage of Total
Positive Electrode	282.4	31.7	348.6	33.9	480.0	40.0
Negative Electrode	32.2	3.6	72.0	7.0	54.7	4.6
Separators	23.0	2.6	35.1	3.4	39.1	3.3
Screens	3.4	0.38	8.0	0.8	5.6	0.5
Electrolyte	134.0	12.4	159.0	15.5	187.0	15.6
End Plates	35.1	3.9	34.0	3.3	35.1	2.9
Pressure Vessel, Internal Hardware, and Seal Assembly	379.9	42.7	371.7	36.1	397.3	33.0
Total	890.2	100.0	1028.4	100.0	1198.8	100.0

Figure 3 shows computed energy densities vs stored energy for Ni-H$_2$ cells as initially determined from a computer optimization study. For the Ni-H$_2$ cell, the energy density approaches 62 Wh/kg. In the computer model, the hardware (such as bus bars and terminals) was scaled directly to the ampere-hour capacity. Also shown in Figure 3 are the actual capacities achieved for an NTS-2 cell, INTELSAT V cell, and a 50-Ah cell. The hardware is identical within all of these cells and

was sized for the 50-Ah cell. This explains the tendency of the actual cells to diverge from the computed values as capacity is reduced.

TABLE 3. ENERGY DENSITY

Parameter	INTELSAT V	NTS-2	50-Ah Cell
Rated Capacity (Ah)	30	35	50
Cell Weight (kg)	0.890	1.028	1.198
Measured Capacity (Ah)			
20°C	31.9	38.5	52.8
10°C	34.8	43.0	57.6
0°C	35.3		
Weight Positive (g)	282.4	348.6	480.0
20°C (Ah/g)	0.113	0.110	0.110
10°C (Ah/g)	0.120	0.120	0.120
Average Discharge Voltage (V)	1.25	1.25	1.25
Energy at 10°C (Wh)	43.5	53.7	72.0
Energy Density (Wh/kg)			
20°C	44.8	46.8	55.1
10°C	48.9	52.2	60.1

Battery energy density

Table 4 contains battery weights and energy densities for INTELSAT V, NTS-2, and 50-Ah batteries. INTELSAT V represents the latest and most advanced battery design. The Ni-H$_2$ cells comprise 80 percent of the total battery weight. A reasonable goal for a 50-Ah battery design was determined by using the INTELSAT V weight ratio of 80 percent for cells to 20 percent for the battery packaging. The result shows a very significant improvement in energy density for a 50-Ah battery (48.1 Wh/kg compared to 39 Wh/kg), which reflects the increase in energy density attained by changing from the 30- to the 50-Ah size cell.

Depth of discharge

Usable energy density which is directly related to the DOD must be considered. To define DOD, it is first necessary to define capacity. It is recommended that capacity be defined as the capacity measured for cells during acceptance testing at an operating temperature of 10°C. For INTELSAT V, the DOD is 48 percent based on the measured capacity
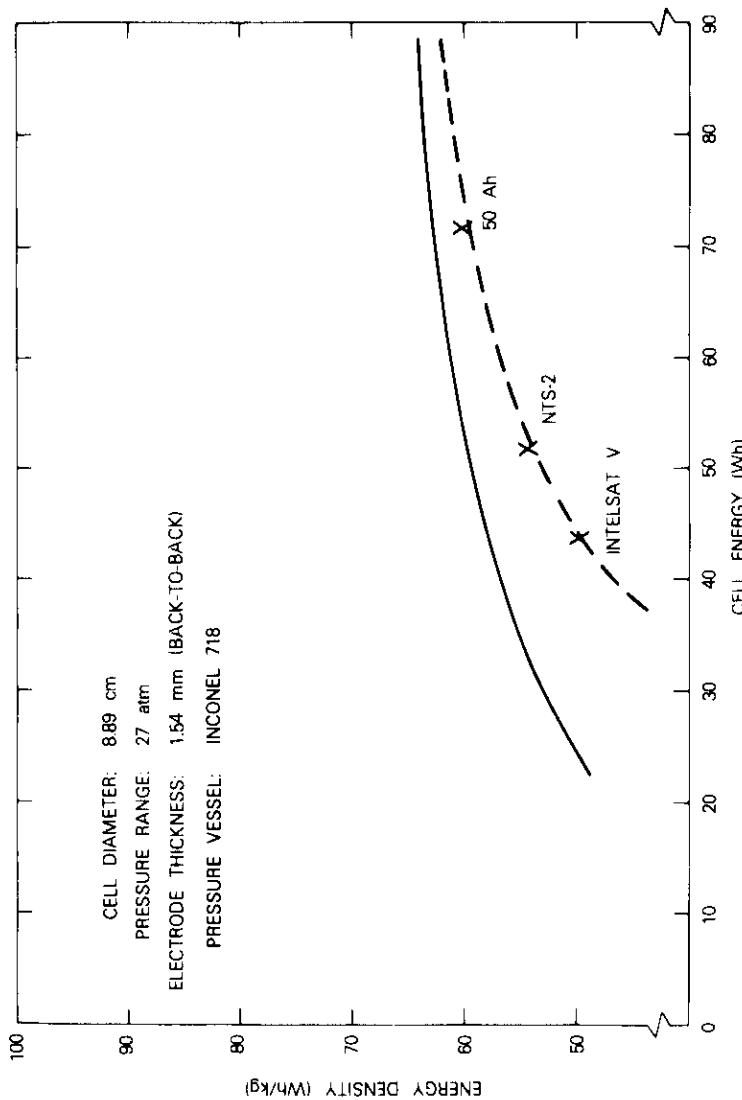
Figure 3. Energy Density of Ni-H$_2$ Cells as a Function of Cell Size

TABLE 4.   BATTERY WEIGHT and ENERGY DENSITY

	INTELSAT V		NTS-2		50-Ah Cell	
Parameter	Weight (kg)	Percent-age	Weight (kg)	Percent-age	Weight (kg)	Percent-age
Cell Weight	0.890		1.028		1.198	
Cell Number	27		14		27	
Total Cell Weight	24.03	80	14.3	70	32.35	80
Cell Mounting Shells	3.3	11	3.1	15.2	4.45	11
Base Plates	0.73	2.4	1.8	8.8	0.97	2.4
Diode Assembly	0.96	3.2	—		1.29	3.2
Mounting Hardware	0.10	0.3	0.2	1.0	0.12	0.3
Connectors etc.	0.23	0.8	1.0	4.9	0.32	0.8
Miscellaneous	0.77	2.6				
Total Weight	30.12	100.00	20.50	100.00	40.44	100.00
Energy (Wh at 10°C)	1174		750		1944	
Energy Density (Wh/kg)	39.0		36.5		48.1	

at 10°C (Table 5). Future Ni-H$_2$ battery systems are expected to use much deeper DOD in synchronous orbit (80 percent), which will significantly increase usable energy density.

It should be noted that the INTELSAT V batteries are discharged at constant power, not constant current. This means that the discharge current will increase as the battery voltage decreases. The battery voltage is expected to gradually decrease with life. Unfortunately, this means that the maximum DOD will occur at the end of life, which in

TABLE 5.   USABLE ENERGY DENSITY (INTELSAT V)

	Measured Capacity	
Parameter	at 20°C	at 10°C
Cell Capacity (Ah)	31.9	34.8
Energy on Discharge, Assuming 1.25 V/Cell Average (Wh)	1076.0	1174.0
Total Electrical Bus Load Estimated for F5, F6, and F7 for 1 Bus (W)	470.0	470.0
DOD Based on 1.25 V/Cell Average (%)	52.4	48.0
Total Energy Density (Wh/kg)	35.76	39.0
Usable Energy Density (INTELSAT V Mission) (Wh/kg)	18.7	18.7
Usable Energy Density at 80% DOD (Wh/kg)	28.6	31.2

turn will accelerate voltage degradation. Therefore, a rather conservative DOD value of 48 percent was chosen for INTELSAT V.

Safety

The safety factor is defined as the ratio of burst to working pressure. Presently, there is a 4/1 safety factor for NTS-2 and INTELSAT V cells. Based on NASA and Air Force specifications, this could be lowered to 2/1 [5]. The high-pressure cell design uses a 2.5/1 ratio, which reduces the volume and weight.

Stored energy per cell (dual stack cell)

It is necessary to store about 70 Wh of energy in an 8.89-cm-diameter pressure vessel to achieve an energy density of 60 Wh/kg (Figure 3). For a single cell stack with an average cell voltage of 1.25 V on discharge, a capacity of 56 Ah is required to achieve 70 Wh of stored energy.

Many spacecraft systems require batteries with capacities lower than 56 Ah; for example, INTELSAT V required a 30-Ah battery. For these applications, the single cell stack approach per pressure vessel sacrifices energy density. To improve energy density, a new cell design is proposed, which uses two 30-Ah cell stacks in series within one pressure vessel. With this design the stored energy is 70 Wh, thereby maintaining the higher energy density of 60 Wh/kg while providing a 30-Ah cell. In this case, the average cell voltage on discharge is 2.5 V.

Energy per unit volume

This section discusses the advantages of increasing both cell capacity and operating pressure (67 atm instead of 40 atm) in terms of volume density.

Cell energy per unit volume

With the present pressure vessel and electrode geometry, the outside cell diameter is fixed at 8.89 cm. Table 6 shows the volume density for an INTELSAT V, an NTS-2, and a 50-Ah cell. The cells with an upper operating pressure of 40 atm (600 psi) deliver approximately 0.064 Wh/cm^3. Changing the upper operating pressure to 67 atm (1000 psi) increases the energy per unit volume by 34 percent from 0.064 to 0.086 Wh/cm^3. By comparison, the energy per unit volume for the INTELSAT V Ni-Cd cell is 0.135 Wh/cm^3.

The high-pressure cell represents the maximum achievable energy

per unit volume in a cell design with a diameter of 8.89 cm. Table 6 indicates that almost all of the hydrogen is stored within the hemispherical ends, and that the electrode stack occupies the entire cylindrical section.

TABLE 6. CELL ENERGY PER UNIT VOLUME

Parameter	INTELSAT V	NTS-2	50-Ah Cell	50-Ah High Pressure
Cell				
Capacity at 20°C (Ah)	31.9	38.5	52.8	52.8
Maximum Pressure (atm)	40.8	40.8	40.8	68.0
Length Electrode Stack (cm)	4.2	5.08	7.0	7.0
Volume Electrode Stack (cm³)	260	315	434	434
Volume to Store Hydrogen (cm³)	426	514	686	409
Total Volume Required (cm³)	686	829	1120	843
Actual Cell Volume (cm³)	762	841	1120	841
Volume of Domes (cm³)	368	368	368	368
Diameter (cm)	8.89	8.89	8.89	8.89
Length of Cell (cm)	15.2	16.5	21.3	16.5
With Terminals (cm)	21	24	29	23
Energy at 10°C (Wh)	43.5	53.7	72	72
Energy per Unit Volume at 10°C (Wh/cm³)	0.057	0.064	0.064	0.086
Battery				
Number of Cells	27	7	27	27
Length (cm)	51.8	48.26	51.8	51.8
Width (cm)	52.1	24.13	52.1	52.1
Height (cm)	22.2	25.4	29	24.2
Volume (ℓ)	59.9	29.6	78.3	75.3
Energy at 10°C (Wh)	1174	376	1944	1944
Energy/Volume (Wh/cm³)	0.0196	0.0127	0.0248	0.0298

Figure 4. *NTS-2 Ni-H₂ Battery*

Battery energy per unit volume

The external dimensions for the NTS-2 battery are shown in Figure 4, and those for the INTELSAT V battery are shown in Figure 5. Energy-per-unit-volume data are presented in Table 6.

In the NTS-2 design, the cells are structurally contained so that they are separated. In the INTELSAT V design, the cells are closely packed, resulting in a 50-percent improvement over the NTS-2 design.

Increasing cell capacity improves the battery energy per unit volume significantly, as shown in Table 6. This is explained by noting that the compression seal terminals at both ends extend cell length by approximately 6 cm. In the NTS-2 and INTELSAT V battery designs, space for these terminals accounts for 20 to 25 percent of the total battery volume. Increasing cell capacity reduces the percentage of the volume occupied by the terminals.

The high-pressure cell also improves energy per unit volume. The energy per unit volume of the high-pressure 50-Ah cells is 0.0298 Wh/cm³ compared to 0.0196 for INTELSAT V cells, representing a 52-percent improvement.

Larger diameter positive electrode

The advantage of the larger diameter positive electrode is to reduce the temperature gradient across the hydrogen gap (ΔT gap). Table 7 indicates that a 2.3°C decrease in ΔT gap could be expected if a larger diameter positive electrode had been used in the NTS-2 cells.

Ni-H₂ cells have been fabricated with larger diameter positive electrodes. This development effort determined the extent to which
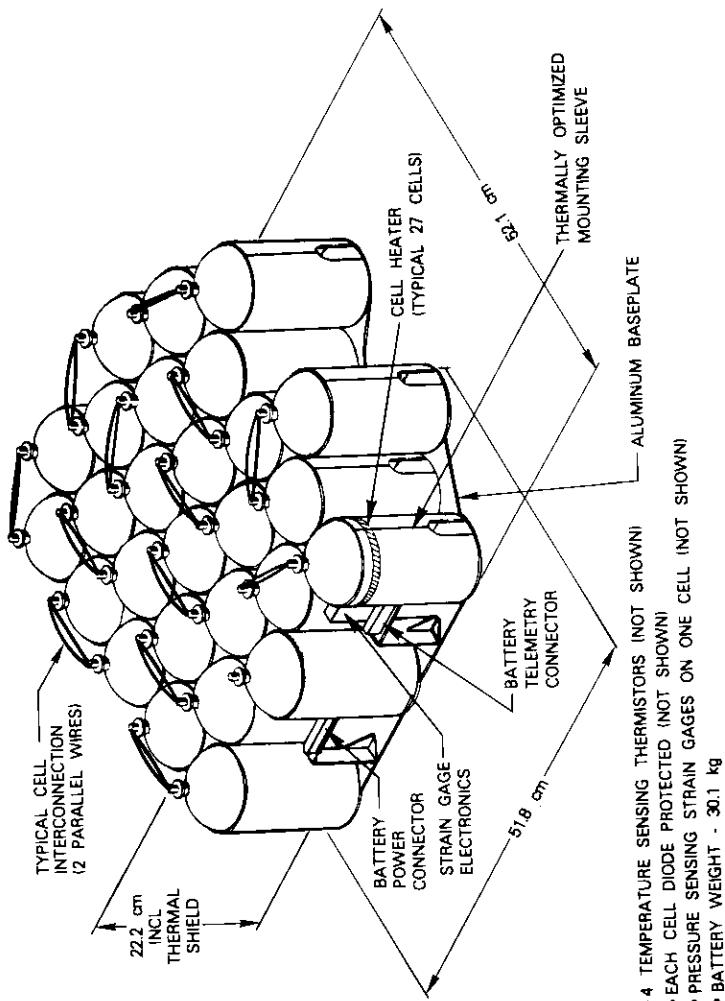
Figure 5. *INTELSAT V Ni-H₂ Battery*

TABLE 7.   DIMENSIONS FOR POSITIVE ELECTRODES

Cell Type	Diameter Positive Electrodes	Hydrogen Gap	ΔT Gap (From Fig. 5)
NTS-2 and INTELSAT V	8.380 cm	0.203 cm	6.8°C
New Design	8.53 cm	0.127 cm	4.5°C

the hydrogen gap could be reduced and still enable the construction of a cell without shorting or damaging the electrodes during assembly. The dimensions for a larger diameter positive electrode are given in Table 7.

Temperature gradients

Figure 6 shows the computed temperature gradient across the hydrogen gap and the temperature gradient from the center of the stack to the wall. These data are for the NTS-2 cell stack structure
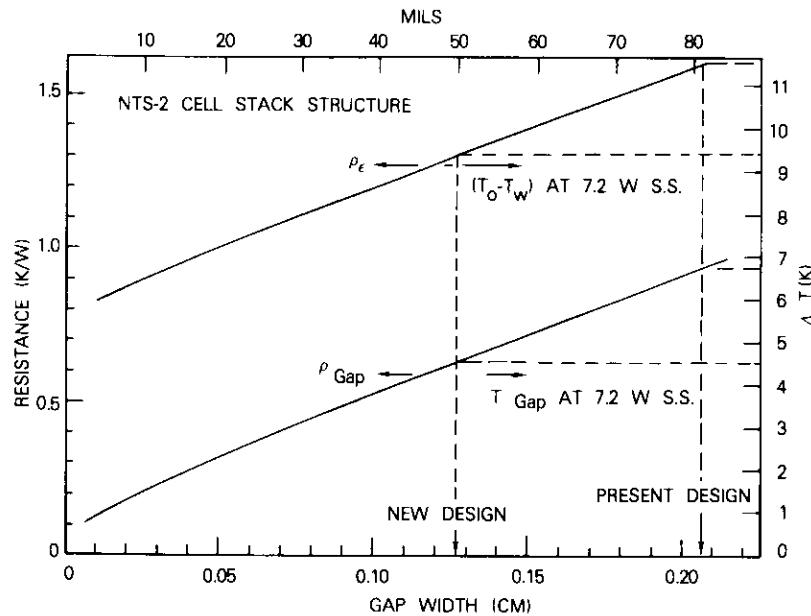

Figure 6. *Thermal Resistance* vs *Hydrogen Gap Width*

(35-Ah cell) with an anticipated maximum dissipation level of 7.2-W steady state [6], which is the estimated maximum heat dissipation on discharge. In actual transient situations, ΔT gap will never attain this value, but it is useful as a worst-case criterion to determine if water vapor transfer can occur from the stack to the wall. The temperature difference required for vapor transfer from a 30-percent KOH solution in the stack to pure water at the wall varies from 10°C with the wall at 0°C to 12°C with the wall at 20°C.

Conclusions

The current status of aerospace Ni-H$_2$ battery technology developed at COMSAT Laboratories has been presented. The NTS-2 spacecraft launched in June 1977 was the first satellite to utilize this new technology, and its performance to date has been excellent. The use of Ni-H$_2$ batteries on INTELSAT V will represent a major commitment to this new technology, which is expected to replace the Ni-Cd batteries used until now for geosynchronous communications satellites. The Ni-H$_2$ batteries planned for use in the next generation of INTELSAT communications satellites should achieve 48-Wh/kg usable energy density in comparison to the 39 Wh/kg of INTELSAT V batteries. This goal will be accomplished by storing at least 70 Wh of energy per pressure vessel and operating at higher depth of discharge.

References

[1] J. Dunlop et al., U.S. Patent 3,867,199, February 18, 1975.
[2] J. Dunlop and J. Stockel, "Orbital Performance of NTS-2 Nickel-Hydrogen Battery," *COMSAT Technical Review*, Vol. 7, No. 2, Fall 1977, pp. 639–647.
[3] J. Dunlop, J. Stockel, and G. van Ommering, "Sealed Metal Oxide-Hydrogen Secondary Cells," *Power Sources Five*, Proceedings of the 9th International Symposium, Brighton, England, 1974, pp. 315–329.
[4] J. F. Stockel et al., "A Nickel-Hydrogen Secondary Cell for Synchronous Orbit Application," 7th Intersociety Energy Conversion Engineering Conference, San Diego, California, 1972, *Proc.*, pp. 87–94.
[5] "Nickel-Hydrogen Battery Development Final Report," May 17, 1976, INTELSAT Contract IS-677, TRW Defense and Space Systems Group.
[6] G. van Ommering and J. F. Stockel, "Characteristics of Nickel-Hydrogen Flight Cells," 27th Annual Power Sources Conference, June 1976.
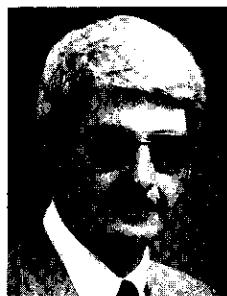
James D. Dunlop received a B.S. in electrical engineering from Rose Polytechnic Institute and an M.S. in engineering sciences from Purdue University. He is currently Manager of the Electrophysical Devices Department at COMSAT Laboratories, responsible for R&D programs on electrophysical devices with special emphasis on nickel-hydrogen battery technology.

Joseph F. Stockel received a B.S. in mechanical engineering from Drexel University and an M.S. in mechanical engineering from George Washington University. A member of the Technical Staff at COMSAT Laboratories in the Electrophysical Devices Department, he has been responsible for the design, fabrication, and testing of nickel-hydrogen cells.

.

Thermoelectrically cooled MESFET low-noise amplifier for earth stations

H.-L. HUNG AND N. L. HYMAN

(Manuscript received August 5, 1980)

Abstract

A new compact and practical low-noise amplifier (LNA) which uses thermoelectrically cooled GaAs MESFETs for the 11.2- to 12.2-GHz communications band is described. The LNA consists of a 3-stage cooled preamplifier followed by a 3-stage uncooled postamplifier, both of microwave integrated circuit design. A specially designed multistage solid-state thermoelectric cooler and a high-performance compact heat exchanger have been developed. The preamplifier, thermally isolated within an insulation-filled compact enclosure, achieves a physical temperature lower than 185 K in a room temperature environment.

A noise temperature of 120 K (noise figure of 1.5 dB) and a power gain of 54 dB at 12 GHz have been achieved with the preamplifier at 188 K. Noise temperature and power gain variations across the 11.7- to 12.2-GHz band are 5 K and 0.8 dB. The noise performance achieved represents the lowest temperature known to have been reported for a thermoelectrically cooled MESFET amplifier at 12 GHz, and is comparable to that of a conventional temperature-stabilized parametric amplifier. At 1-dB gain compression, 12-dBm output power has been obtained. The 6-stage uncooled amplifier, including an input waveguide isolator, exhibits a noise temperature of 190 K (noise figure of 2.19 dB) and a power gain of 50 dB.

Introduction

Recent significant improvements in GaAs MESFET technology and performance have extended the application of these devices in microwave communications systems [1]. The feasibility of their use in low-noise and high-power amplifiers, mixers, and oscillators has been demonstrated in practical components. In particular, low-noise MESFET amplifiers have been developed for use at frequencies ranging from UHF [2] to Ka-band [3].

Previously, large earth stations for satellite communications have extensively employed parametric amplifiers as the low-noise front end. There has been considerable demand to replace parametric amplifiers with GaAs MESFET amplifiers, which, although exhibiting a higher noise figure, offer lower cost, simpler and more compact configurations, and easier maintenance. Recent advances in MESFET technology have lowered the device noise figures to less than 2 dB at 12 GHz [4]. It will be shown that amplifiers using these new devices and operating at room temperature can provide performance suitable for some earth terminals. The noise temperatures achieved, however, are not comparable to those of the parametric amplifiers used in large earth stations.

At room temperature, the noise performance of GaAs MESFETs in microwave amplifiers is dominated by thermal noise. Cooling the MESFET to lower temperatures results in improved noise performance as well as power gain [5], [6]. The method of cooling for a specific system is determined by the noise temperature requirement. Very low temperatures can be achieved with cryogenic fluids such as liquid nitrogen [5], but supply and maintenance problems make it impractical for mobile or unattended earth stations. Passive cooling methods using radiation to space techniques [6] can achieve physical temperatures around 180 K but are limited to satellite applications.

Solid-state thermoelectric devices have previously been used for cooling small semiconductor detectors or for maintaining amplifiers slightly below ambient temperature. In earlier efforts at COMSAT Laboratories, a multistage thermoelectric device inside a vacuum enclosure was used to cool a 3-stage MESFET amplifier to physical temperatures as low as 190 K. However, the design, fabrication, preparation, and operational requirements for units that must maintain high vacuum integrity over several years do not render this approach cost effective. Thus, a maintenance-free, practical cooling package is

needed to provide a temperature sufficiently low to yield amplifier noise performance comparable to, if not better than, that of high-quality parametric amplifiers.

This paper describes the design of an LNA using GaAs MESFETS cooled by a multistage thermoelectric heat pump. The amplifier has been designed as the front end of an earth station receiver for the 11.7- to 12.2-GHz satellite communications band. The characterization of the MESFET devices and the design of the amplifier will be described. The thermal design of the hermetic package for maintaining the preamplifier at low temperatures will be discussed. The performance of the complete LNA will then be presented. Finally, design techniques and applications of this type of cooled MESFET LNA to other satellite communications bands will be examined.

GaAs MESFET devices

The LNA uses commercially available NE137 and NE388 MESFETS manufactured by the Nippon Electric Company. Simplified physical models of these MESFETs are shown in Figure 1a. Both devices have a nominal gate-length (l_g) of 0.5 μm. Their structures, however, are different. One has a recessed gate and the other a flat structure; consequently, they exhibit different noise performance.
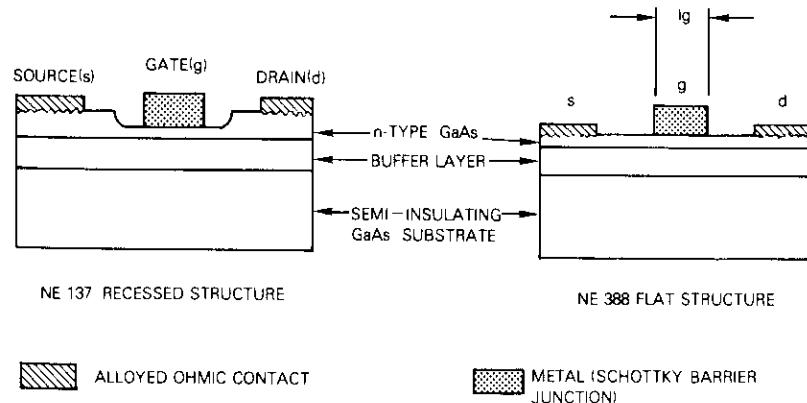
Figure 1a. *Simplified Cross-Section of GaAs MESFETs*

The noise properties of a GaAs MESFET can be described in terms of an equivalent circuit [7] as shown in Figure 1b. The circuit includes the effect of resistive parasitic elements on noise performance. Noise
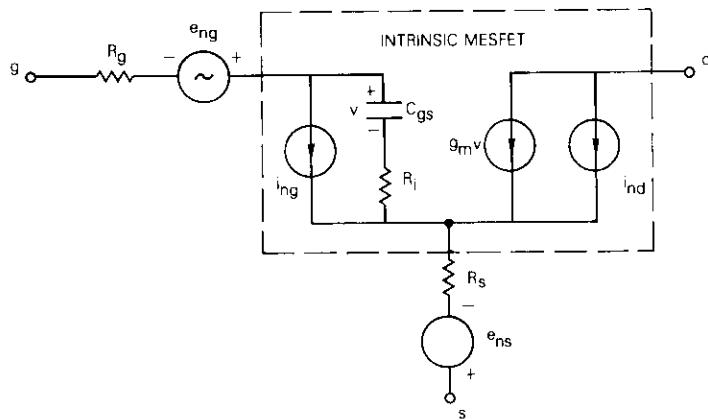
Figure 1b. *Equivalent Circuit of a GaAs MESFET Used in Noise Analysis*

sources e_{ng} and e_{ns} represent the thermal noise contributions of R_g, the AC gate metallization resistance, and R_s, the total source series resistance, respectively. The induced gate and drain circuit noise sources are i_{ng} and i_{nd}. The gate-to-source capacitance is C_{gs}, and R_i is the associated charging resistance. Transconductance, g_m, is assumed to be constant in the frequency band of interest.

For a MESFET operating at room temperature below the cutoff frequency, the expression for the minimum equivalent input noise temperature of the device, T_e, in terms of the equivalent circuit elements can be simplified [8] as follows:

$$T_e \simeq \alpha T_0 f C_{gs} \sqrt{\frac{R_s + R_g}{g_m}} \qquad (1)$$

where α is a coefficient which depends on the channel material, T_0 equals 290 K, and f is the frequency. The elements are evaluated at null gate bias and at a relatively low drain bias voltage in the current saturation region.

Equation (1) indicates the dependence of minimum T_e on the device parameters. The effect of gate length is implicit in C_{gs} and g_m. Without shortening the gate length, the minimum noise temperature of the device can be improved by reducing the parasitic resistances. The

deep recess structure with a cylindrical edge shape has enabled the reduction of the parasitic source resistance to half that of conventional flat-type MESFETs. A 0.5-dB improvement in device noise figure at 12 GHz has been achieved [4].

As previously mentioned, the dominant noise mechanism in a GaAs MESFET operating under optimum noise conditions is thermal. Measurements of parasitic resistances have indicated a linear dependence of R_g on decreasing temperature [4]. A 70-percent reduction in R_g has been observed for a temperature drop of 200 K; however, only an 18-percent reduction in R_s has been observed. This implies that the reduction of source resistance is important for MESFETs not only at room temperature but also at low temperatures. Also, transconductance increases as the device ambient temperature is lowered. The temperature dependence of these parameters on the overall device noise behavior is evident from equation (1).

The MESFETs have been extensively characterized at room and lower temperatures. Probe tests and measurements have ensured proper DC parameters and breakdown voltages for high-reliability performance. Small-signal S-parameters measured over a wide frequency band (2 to 18 GHz) have been used in computer-aided circuit analysis. Noise parameters have also been characterized at various temperatures using a noise measurement system with an in-band noise figure of less than 3 dB. Optimal source impedance for minimum noise operation has been used for the amplifier module design. Proper DC bias points have been established for minimum noise measure at low temperatures.

Electrical design

The complete LNA consists of six single-stage modules. Figure 2 shows a block diagram of the amplifier. The general design objectives have been to achieve the lowest noise measure of the input amplifier modules and to minimize RF losses of the input passive components.

Preamplifier

The preamplifier is a single-ended module design. This allows the selection of the best noise performance device for the first input stage, thus achieving the best overall performance. Initial input and output matching networks have been designed for a 10-percent bandwidth, using quarter-wavelength resonator broadbanding techniques. The overall performance across the band is then optimized for noise, power
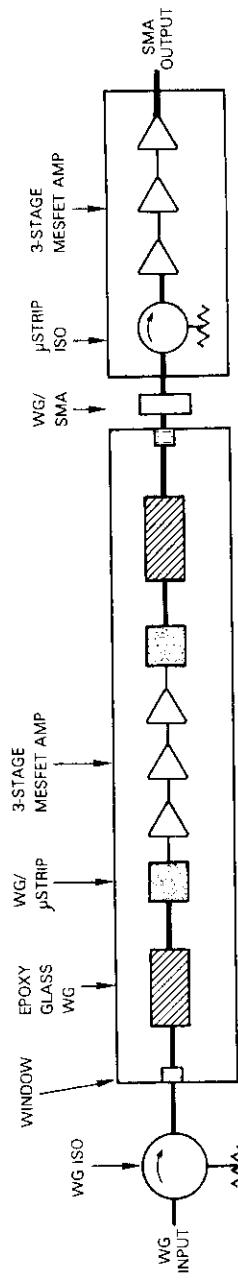
Figure 2. *Block Diagram of the LNA*

gain, and output return loss using computer-aided circuit analysis. The wideband (2- to 18-GHz) frequency response is then examined to ensure stable operation and flat gain response for both the single-stage and cascaded multi-stage amplifiers. The stability problem is especially critical at low temperatures, since power gain increases.

The GaAs MESFETs are mounted on low thermal-expansion Invar ground plates for direct heat transfer. The microstrip matching circuits are fabricated on 0.38-mm-thick fused silica substrates, which cause lower RF circuit losses than the commonly used alumina. Open shunt stubs provide circuit tunability to accomodate device parameter variations. Interstage mismatches are minimized by designing each stage with an output return loss better than 20 dB. Careful device selection has resulted in no significant variation of noise performance among modules. Single-stage modules with noise figures of 1.85 to 1.90 dB and a power gain of 8.5 dB have been achieved across the band.

The assembled 3-stage amplifier is shown in Figure 3. The preamplifier housing has been designed for minimum surface area, thus
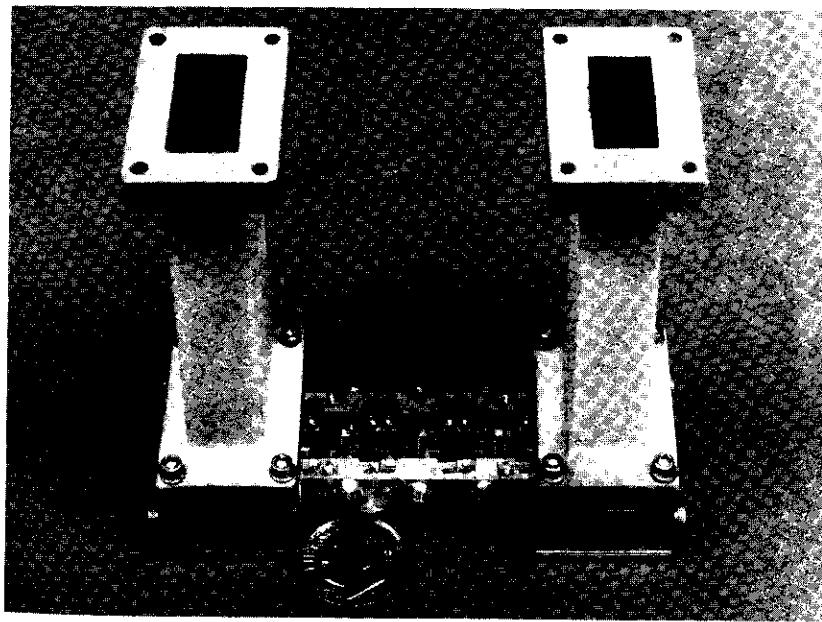

Figure 3. *3-Stage MESFET Preamplifier With Epoxy-Glass Waveguides*

reducing the heat flow from the surroundings. The housing includes two integrated waveguide-to-microstrip transitions with a 3-section coaxial transformer. This type of transition offers flexibility in the orientation of the waveguide with respect to the microstrip circuits. Figure 4 shows the RF performance of a pair of transitions through a 38-mm-long 50Ω reference microstrip line on a fused-silica substrate. Return loss close to 30 dB and insertion loss of less than 0.1 dB per transition have been measured in the 11.7- to 12.2-GHz band. At low temperatures, the frequency response of these waveguide-to-microstrip transitions has been observed to be stable.

A pair of specially designed epoxy-glass internally-metal-plated waveguides is used to thermally isolate the cooled preamplifier from the external warmer environment. These 76.2-mm WR-75 lightweight waveguides, also shown in Figure 3, exhibit an RF performance comparable to that of conventional metallic waveguides. The maximum loss of each waveguide has been measured to be ≈0.02 dB. Two hermetic quartz pressure windows are used at the interface between the epoxy-glass waveguides and the external RF components. Insertion loss of less than 0.04 dB and VSWR better than 1.08 have been measured.

At room temperature, the 3-stage amplifier exhibits a power gain of 24.6 ± 0.3 dB across the 11.7- to 12.2-GHz band. Output return loss is better than 20 dB. Optimal bias drain currents for the NE137 devices are 8 to 10 mA with the drain-to-source voltage optimized at 4 V. At low temperatures, the drain currents are set at higher values.

Postamplifier

The postamplifier shown in Figure 5 has a low-noise input stage module and two high-gain modules. Its noise figure is not as critical as that of the preamplifier, and therefore the relatively less expensive NEC NE388 devices are used. Since an optimal low-noise design does not result in a good input match because of the intrinsic properties of MESFETs, a microstrip isolator with small insertion loss (<0.4 dB) has been integrated into the input port. In-band ripples caused by RF mismatches at the interface between the preamplifier and the postamplifier have thus been minimized.

The high-gain modules are designed with good input and output matches. The MESFETs are biased at 50 percent of I_{DSS} to achieve high linear output power capability across the band. An interdigital coupler provides interstage DC isolation and the additional ability to
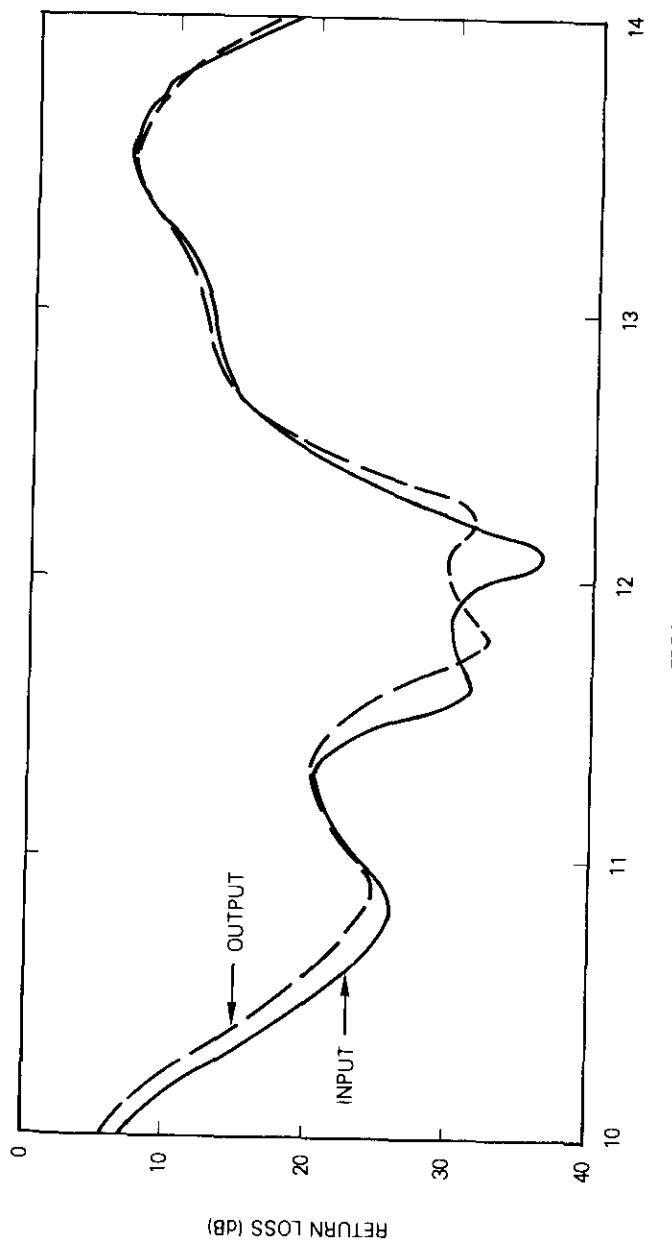

Figure 4. Return Loss of Two Waveguide-to-Microstrip Transitions Through a 38-mm 50Ω Microstrip Line
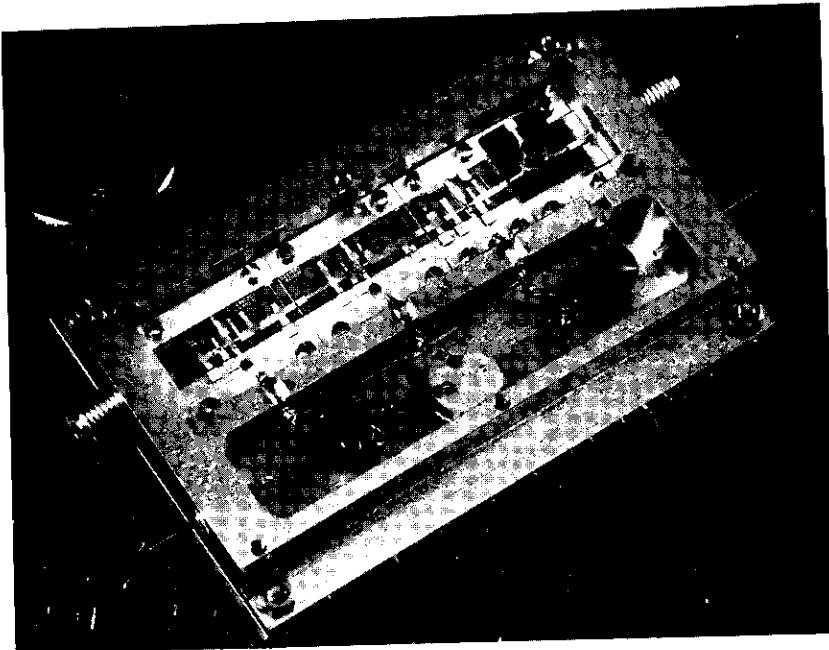
Figure 5. *The 3-Stage MESFET Postamplifier*

Figure 6. P_{out} vs P_{in} *of the 3-Stage MESFET Postamplifier*

shape the frequency response of the complete amplifier. Reliability has also been improved by eliminating the DC blocking chip capacitors. The postamplifier low-noise high-gain design contributes minimal noise to the overall amplifier performance and, at the same time, achieves high power gain and good linearity. Figure 6 shows the output power of the post-amplifier versus input power. At 1-dB compression point, 12 ± 0.3-dBm output power has been measured across the band with a small signal gain of 26.4 ± 0.2 dB.

A DC bias circuit, which requires a single 12-V DC input, provides the proper DC supply to the individual MESFETs. It offers variable biases to both the gate and drain terminals of each cooled device so that optimal values can be set at the operating temperature. The bias supply circuits maintain the drain current level through a feedback network, thus minimizing any gain drift effect due to device defects. The single 12-V DC supply also powers the regulator circuit of the postamplifier, which uses the simpler self-bias scheme.

Low-temperature evaluation

The noise performance and power gain versus temperature of the complete 6-stage amplifier at 12 GHz are shown in Figure 7. The measurements were performed by varying the preamplifier temperature while the postamplifier remained at room temperature. A cooling method similar to one previously reported [5] has been employed. The preamplifier is suspended in a dry atmosphere above liquid nitrogen, and the temperature is stabilized by a heating element. The low RF loss and VSWR together with the excellent thermal properties of the epoxy-glass waveguides contribute to the high accuracy of this low-temperature amplifier characterization method. Consistent and repeatable measurements of noise figure and power gain versus temperature have been achieved.
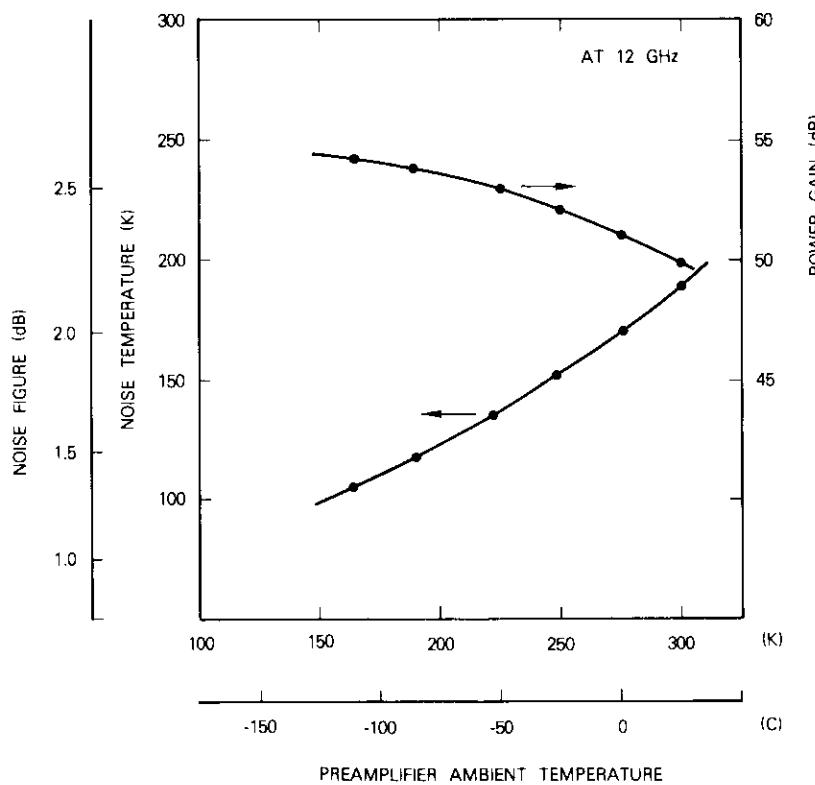
Figure 7. *Noise Figure, Noise Temperature, and Power Gain of the 6-Stage MESFET Amplifier With Input Isolator vs Ambient Temperature*

With a computer-aided circuit pretuning technique and an optimized device bias operating point, the best noise performance of the complete amplifier across the band of interest has been determined at low temperatures. Figure 7 shows the tradeoff between both noise temperature and power gain and the temperature needed to achieve a particular performance requirement. A noise temperature as low as 100 K and a power gain of 54.5 dB can be achieved if the preamplifier is cooled to 150 K. Noise performance improves at a higher rate near room temperature.

Thermal design

Thermal design objectives were to maintain the LNA preamplifier at a temperature of 188 K* in an inert atmosphere with minimum cost and size and maximum long-term reliability. The obvious choice of cooling device is the thermoelectric heat pump (TEHP), which is preferable to mechanical refrigeration and cryogenic liquids because of its solid-state reliability, smaller size and weight, and lower cost.

Maintaining the preamplifier at a low temperature requires good thermal isolation from the external environment, TEHP performance matching the thermal loading, and efficient rejection of the heat generated by the TEHP. The requirement of water vapor exclusion was satisfied with conventional hermetic sealing. Good thermal contact between the TEHP and the preamplifier under potentially high thermal strain conditions was achieved with a flexible high-conductance connector.

Insulation system

Figure 8 shows a schematic of the preamplifier thermoelectric package (PTEP).† The thermal paths between the PTEP external cover and the preamplifier are primarily through the intermediate space and the two waveguides.

Heat flow from PTEP walls to the preamplifier is minimized by a low-conductance material, which fills the intermediate space, and heat shields, which are thermally attached to the waveguides and to the appropriate TEHP stages. The insulation material is a low-density Perlite powder with interstitial krypton gas slightly above atmospheric pressure. Perlite is easy to keep clean and dry, handle, and pour. The low thermal conductivity of krypton gives the combination a conductance of 0.00017 W/cm K at 188 K. Other insulation systems offering this low conductance would be more costly or more difficult to fabricate and apply. Although the use of heat shields increases the

---

*This approximates the lowest temperature achievable with thermoelectric heat pumps of reasonable size and input power. Many applications may be satisfied with higher temperatures which, while not affecting the design practices described in this paper, would result in size, cost, and power reductions.

†The 6-stage thermoelectric heat pump was manufactured by Marlow Industries Inc.

heat loads in the lower TEHP stages, it lowers the top-stage temperature and reduces the required TEHP size.

The thin-walled, epoxy-glass waveguides exhibit satisfactory low conductance; two-thirds of the total conductance is due to the internal metallic plating. A thermal resistance as high as 440 K/W has been measured for the 76.2-mm-long waveguides.
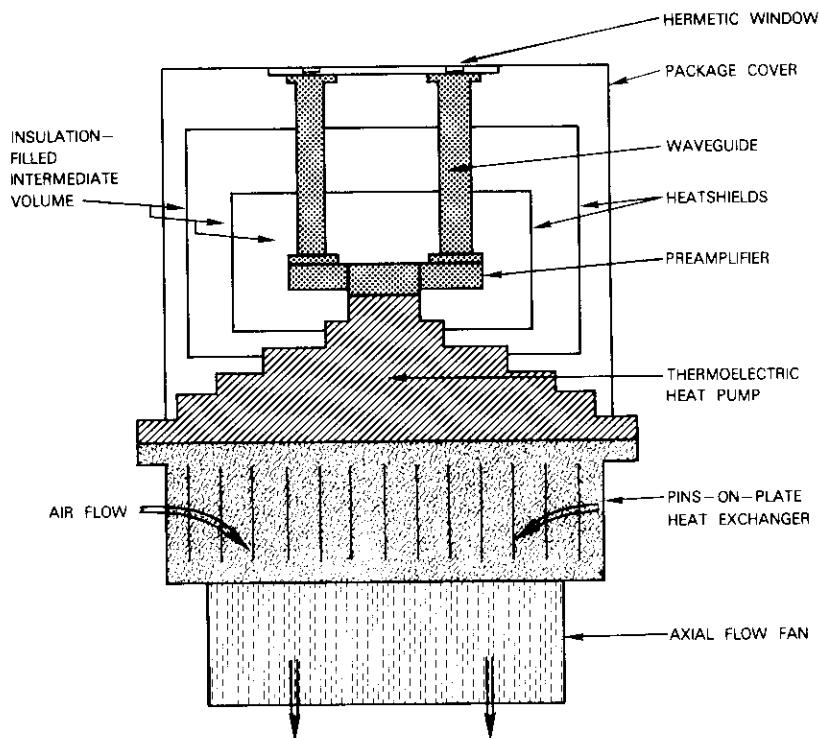

Figure 8. *Cross-Sectional Diagram of a PTEP*

PTEP thermal analysis

Predictions of heat flow and temperature distribution within the PTEP were based upon a familiar analytical aid, the thermal model. This model was used as a design tool by noting changes in the heat loads while varying parameters such as overall PTEP size, number of thermal shields, thermal conductance between shields and waveguides, and

insulation conductance. Heat flow into the externally loaded TEHP stages was computed and used to establish TEHP design specifications.

The important parameter of PTEP size was based on an optimum balance between insulation volume and TEHP capacity. Increased volume results in longer heat flow paths and higher thermal resistance; lower heat loads allow the use of a smaller TEHP. With an approximately cubical PTEP shape maintained for near optimum volume utilization, minimum size (19 × 19 × 15 cm) was defined as a square base which would contain the square bottom stage of a TEHP whose capacity exactly matched the PTEP heat loads.

Heat rejection to ambient air

The pumped heat loads and total power supplied to the TEHP are rejected as heat from the bottom-stage hot side to the ambient air. This heat transfer is through a heat exchanger designed for minimum temperature elevation above the ambient air within certain volume, weight, and cost limitations.

A new type of heat exchanger was developed based on efficient volume utilization. There are no space-consuming air ducts because of the approximately equal areas of the TEHP bottom stage and a high-performance axial flow fan. The TEHP, heat exchanger, and fan are concentric as shown in Figure 8. Relative to the conventional heat exchanger with rectangular fins, a higher surface area-to-volume ratio is realized with closely spaced, small diameter cylindrical pins. Optimum pin size and spacing pattern were analytically determined. The air flows inward at the four sides and exhausts through the fan. Measured effectiveness based on average heat exchanger temperature is 39 W/K with a small axial flow fan and 50 W/K with a larger centrifugal blower.

Thermal evaluations

The TEHP was evaluated separately with simulated heat loads and controlled bottom-stage hot-side temperature. Of particular interest were changes in top-stage cold temperature with variations in heat load into the different stages and in the temperature of the bottom-stage hot side. In addition to verifying satisfactory TEHP design and fabrication, these data were essential for an accurate parametric analysis of the insulation, thermal shields, and package size.

Figure 9 shows the relationships among bottom-stage hot-side and top-stage cold-side (preamplifier) temperatures and TEHP input power.
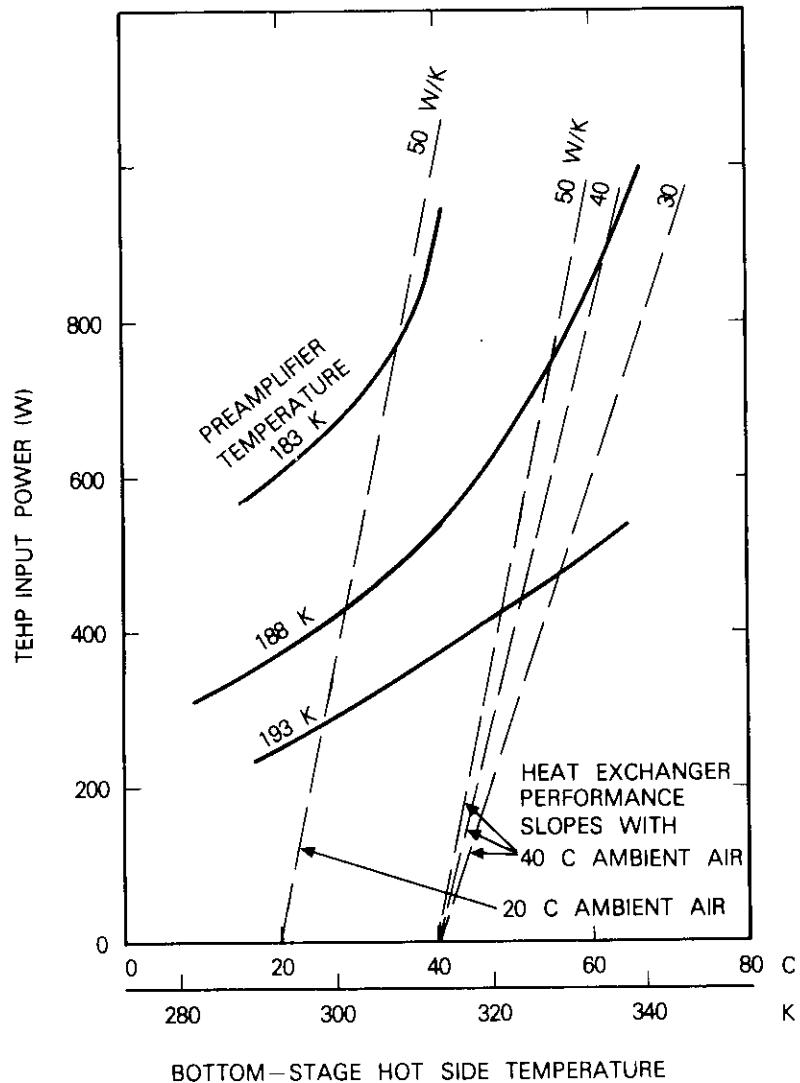
Figure 9. *Preamplifier Temperature as a Function of TEHP Input Power and Bottom-Stage Hot-Side Temperature With Superimposed Heat Exchanger Performance*

With this parameter format, achievable temperature regimes are conveniently defined by heat exchanger performance. Slopes corresponding to performances of 30, 40 and 50 W/K* with origins at ambient air temperatures of 20° C and 40° C, for example, intersect the curves to define the relationship between achievable preamplifier temperature and TEHP input power.

Overall performance

The completely assembled 6-stage LNA is shown in Figure 10, with a WR-75 waveguide input port and a coaxial SMA output. The postamplifier is located on top of the preamplifier thermoelectric package. Noise and power gain characteristics at a 300 K environmental temperature are given in Figure 11. With the preamplifier operating at 188 K, a noise temperature of 120 K† (1.5-dB noise figure) and a power gain of 54 dB have been achieved at 12 GHz. Variations in noise temperature across the 11.7- to 12.2-GHz band have been measured to be 5 K with a power-gain change of 0.8 dB. No significant shift in the overall gain-frequency response shape has been observed as the preamplifier temperature is reduced. The VSWR of the input is better than 1.12; the output port, without a matching isolator, achieves a maximum VSWR of 1.22 across the band.

With no special circuit design for temperature compensation, the noise temperature of the LNA varies by 6 K when the external ambient changes from 10°C to 40°C as indicated in Figure 12. The preamplifier can be cooled rapidly. In reference to final steady-state performance, the reduction in LNA noise temperature is 90 percent, and the decibel power gain is 99 percent within 15 minutes after activation at 300 K.

At room temperature, the 6-stage uncooled amplifer, including an input waveguide isolator, exhibits a noise temperature of 190 K (2.19 dB noise figure) and a power gain of 50 dB at 12 GHz.

---

*The 50 W/K represents the new design heat exchanger; the 30 W/K is typical of a plate-fin type of comparable size.

†The accuracy of the measurement was verified by comparing results (within 1 K) with two hot and cold noise generators. Measurements with a solid-state noise diode indicated even slightly lower noise temperature values.
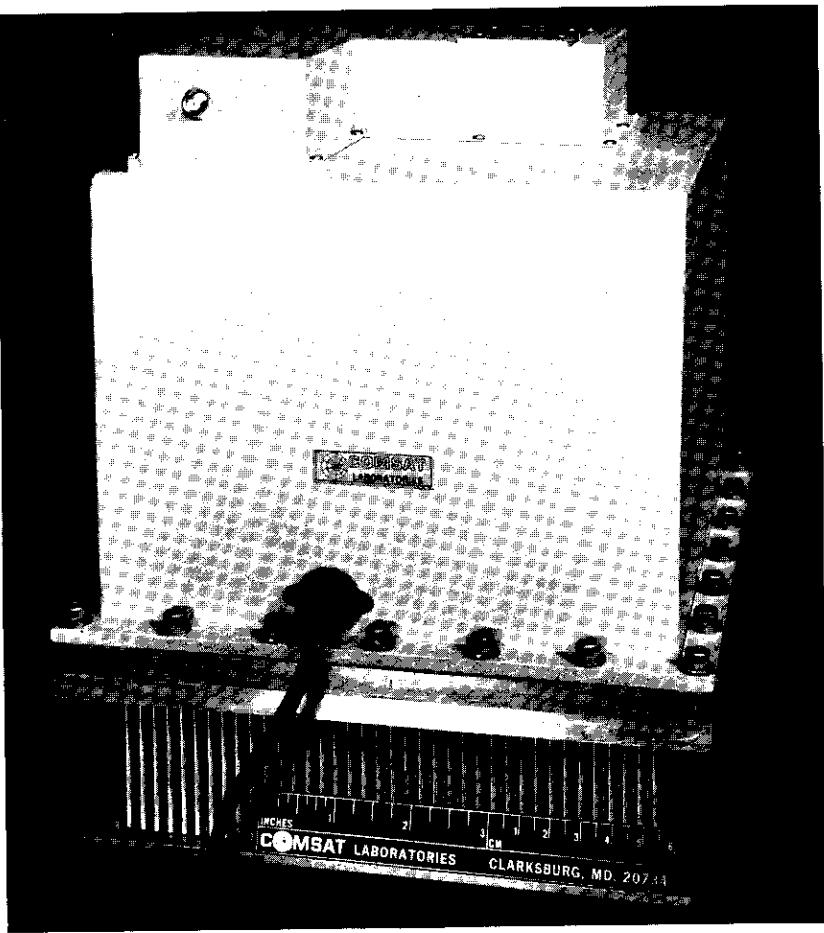
Figure 10. *Complete 6-Stage LNA*

Conclusions

A practical, fully engineered LNA, which is compact and low-cost, has been developed using GaAs MESFETS on microwave integrated circuits. The noise performance of the LNA is comparable to that of a temperature-stabilized parametric amplifier. The 120 K noise temperature and 54-dB gain represent the best values known to have been reported for a thermoelectrically cooled 12-GHz MESFET amplifier.
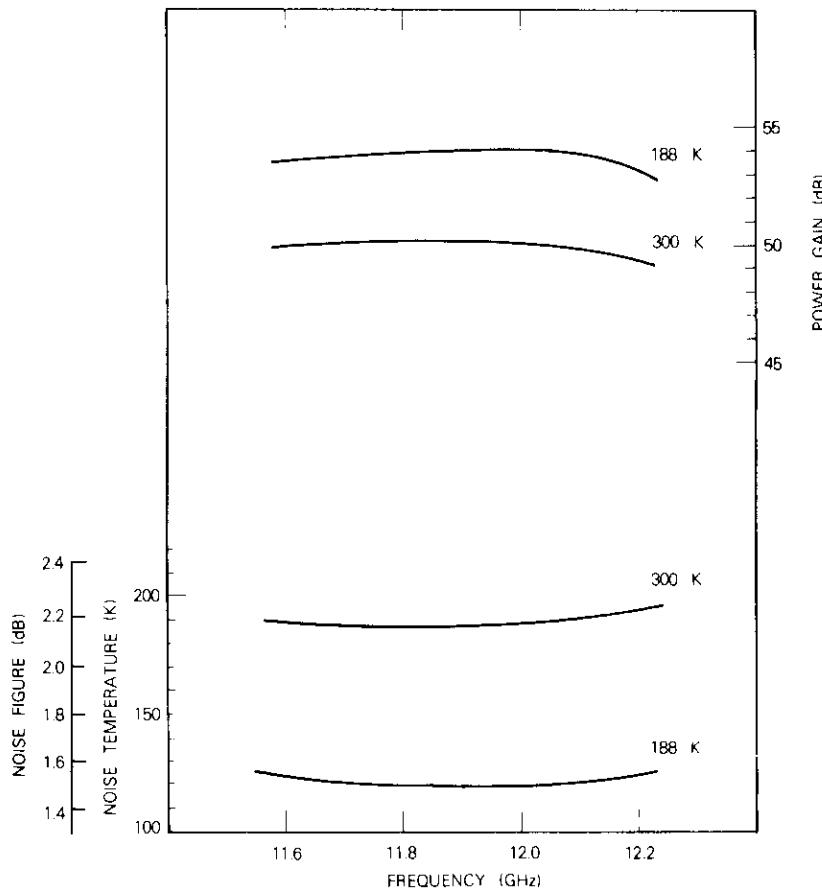
Figure 11. *Performance of the 6-Stage LNA (temperatures indicate preamplifier ambient values, postamplifier at room temperature)*

The MESFET's 12-dBm output power at the 1-dB compression point compares with −10 dBm for a typical parametric amplifier. The technology described can be readily extended to designs for other bands such as the widely used 4- and 11-GHz bands and the future 19-GHz band of satellite communications earth station receivers.

In comparison with conventional parametric amplifiers, MESFET LNAS offer improved gain stability versus temperature, simpler RF circuitry, easier maintenance, and lower cost. Cost-efficient cooled
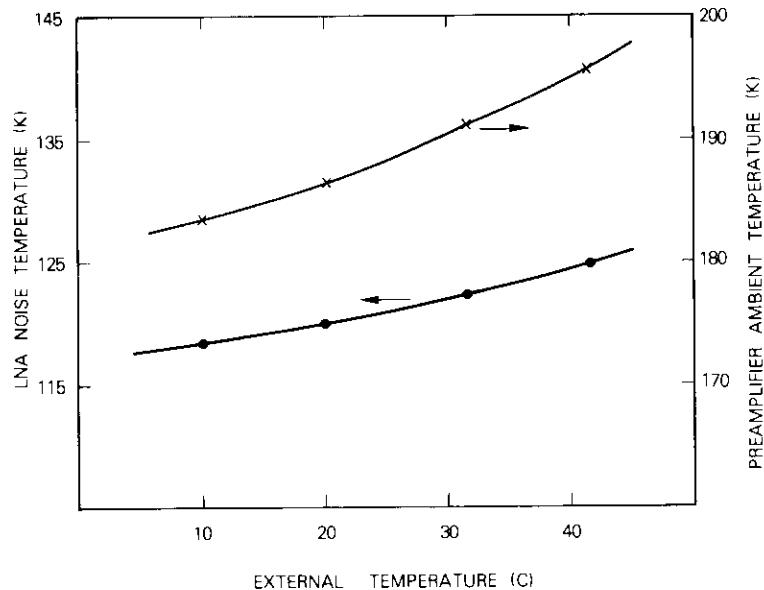
Figure 12. *Preamplifier Ambient and LNA Noise Temperature* vs *External Temperature at 12 GHz*

MESFET LNAS are expected to replace parametric amplifiers in earth terminals.

The achievement of noise temperatures below 200 K with the uncooled amplifier at 12 GHz demonstrates the feasibility of applying such amplifiers to receivers not requiring very low noise performance. With temperature stabilization or compensation of the complete amplifier, the uncooled LNAS are also applicable to smaller size satellite earth terminals as well as to radar systems.

Acknowledgment

References

[1] H. F. Cook, "Microwave FET's—A Status Report," 1978 IEEE International Solid-State Circuits Conference, *Digest of Technical Papers*, pp. 116–117.

[2] M. Mitama and H. Katoh, "U.H.F. Low-Noise GaAs MESFET Amplifier for Color T.V. Broadcasting Translator," *Electronics Letters*, Vol. 14, May 1978, pp. 319–321.

[3] J. M. Schellenberg et al. "RF Characteristics of Ka-Band GaAs FETs," 7th Biennial Cornell Conference, August 1979, *Proc.*, pp. 289–296.

[4] K. Ohata, H. Itoh, F. Hasegawa, and Y. Fujiki, "Super Low Noise GaAs MESFETs with a Deep Recess Structure," *IEEE Transactions on Electron Devices*, ED-27, June 1980, pp. 1029–1034.

[5] C. A. Liechti and R. B. Larrick, "Performance of GaAs MESFETs at Low Temperatures," *IEEE Transactions On Microwave Theory and Techniques*, MTT-24, June 1976, pp. 376–381.

[6] H.-L. Hung, R. E. Stegens, and M. Dressler, "Low-temperature Performance of GaAs MESFET Amplifiers at 14.25 GHz," *Proc.*, 6th Biennial Cornell Conference, August 1977, pp. 331–336.

[7] R. A. Pucel, H. A. Haus, and H. Statz, "Signal and Noise Properties of Gallium Arsenide Microwave Field-Effect Transistors," *Advances in Electrons and Electron Physics*, Vol. 38, New York: Academic Press, 1975, pp. 195–265.

[8] H. Fukui, "Optimal Noise Figure of Microwave GaAs MESFETs," *IEEE Transactions on Electron Devices*, ED-26, No. 7, July 1979, pp. 1032–1037.

Hing-Loi Hung received the S.B. degree in electrical engineering from M.I.T., and the M.S. and Ph.D. degrees from Cornell University. His research at Cornell was concerned with microwave avalanche devices including fabrication and modeling. During the spring term of 1974, he was an instructor with the University. Previously, he worked on microwave components at Bunker-Ramo Corporation, M.I.T. Research Laboratory of Electronics, and Chu Associates. Since 1978, he has been a professorial lecturer in engineering at the George Washington University. He joined COMSAT Laboratories in 1974 and is currently a Research Scientist with the Microwave Laboratory.

Nelson Hyman received a B.M.E. from Georgia Institute of Technology in 1958 and an M.S.M.E. from the University of Maryland in 1959. Prior to joining COMSAT Laboratories in 1969, he engaged in spacecraft thermal design with NASA/GODDARD Space Flight Center and Fairchild Industries. As a member of the technical staff in the Thermal Control and Environmental Testing Department of the Spacecraft Laboratory, he has participated in a variety of thermal analysis and evaluation tasks, from INTELSAT satellites and subsystems to earth terminal equipment.
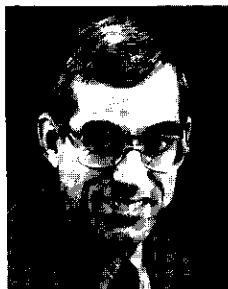
Optical transmission technology in satellite communications

R. W. KREUTEL, A. E. WILLIAMS, K. H. GREENE, Y. S. LEE, AND R. G. AUTRY

(Manuscript received June 17, 1980)

Abstract

This paper presents an overview of optical transmission technology for satellite communications engineering and describes its complementary and competitive impact on existing satellite technology. The state-of-the-art and projected development of optical components are described, as well as the performance of digital and analog optical systems. Complementary applications including interfacility, diversity, and intersatellite links and competitive applications such as long-haul fiber optic cables are addressed. Advanced applications such as optical-microwave interaction technology are briefly projected.

Introduction

Although optical transmission of information was one of the earliest forms of communication, line-of-sight requirements from transmitter to receiver and atmospheric effects made reliable communications difficult or impossible. Efforts to overcome these obstacles by the development of an optical guiding medium were accelerated by the invention of the laser [1] with its potential for enormous communication bandwidths.

Early optical waveguides utilized periodic arrangements of lenses or mirrors in cylindrical pipes. One fairly successful design utilized a gas-filled pipe in which the density of the gas was a function of the pipe radius [2]. The resulting refractive index profile provided continuous refocusing of the optical beams along the axis of the pipe. Although some of these schemes were moderately successful, they were awkward to implement and use.

In 1970, the situation was radically changed when the Corning Glass Corporation achieved a glass fiber optical waveguide attenuation of 20 dB/km through the development of ultra-pure glass manufacturing processes [3]. Extraordinary performance improvements have produced a fiber optic technology applicable in a wide variety of communications systems. Salient features of fiber optics include extremely wide bandwidth and low attenuation (0.2 db/km) [4]. In addition, optical fibers do not require grounding and are immune to crosstalk and electromagnetic interference (EMI). Furthermore, glass fibers are extremely small and lightweight and can be bent to a small radius.

Figure 1 shows the complementary and competitive influences of
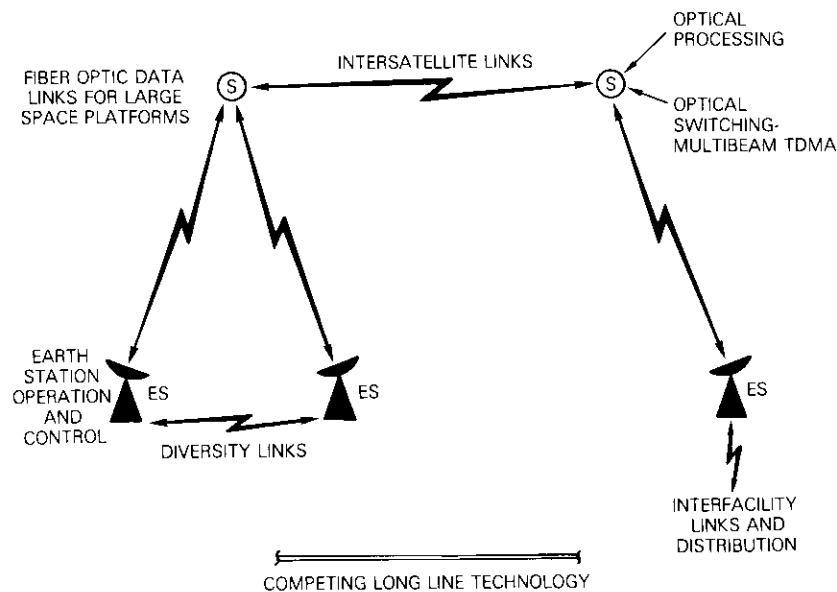
Figure 1. *Complementary and Competitive Optical Communications Applications*

optical technology on satellite communications system design. Potential complementary applications include earth station entrance links, interfacility links, customer-premises earth stations, and noise-resistant earth station circuits, in which fiber optics is advantageous because of its ease of installation and EMI immunity. For satellite systems operating at frequencies with a rain attenuation problem, diversity earth stations can be interconnected by optical fiber links. Free-space optical transmission is also attractive for intersatellite links. Onboard satellite applications may include distribution and switching (SS-TDMA) of signals and electronic control using lightweight and EMI-resistant optical fibers.

Competitive applications include high-density point-to-point trunking, which takes advantage of the long repeater spacings made possible by the low-attenuation and wide-bandwidth properties of optical fibers. In particular, long-haul fiber cables are now being placed in service, and transoceanic telecommunications cables appear feasible within the next decade.

This paper explores the impact of these applications on satellite communications systems. The first part of the paper reviews optical technology components and describes state-of-the-art specifications for fibers, sources, photodetectors, connectors, and splices. A discussion of the optical communications performance of analog and digital detectors, sensitivity calculations, system bandwidth, and attenuation limitations follows. The results are then used to determine the applicability of optical technology to satellite systems.

Optical technology

Sources

The semiconductor laser diode (LD) and high-radiance light-emitting diode (LED) are the principal sources employed in optical communications systems [5]. Typical optical output power characteristics as a function of temperature for LEDs and LDs are shown in Figures 2a and 2b. Improved laser diode fabrication technology has produced high optical power with excellent linearity. The output power of the LEDs varies about ± 1 dB as the operating temperature is changed by ± 30°C about room temperature, while LDs exhibit greater temperature sensitivity and normally require optical feedback control and/or direct temperature control circuits for temperature-stabilized operation.
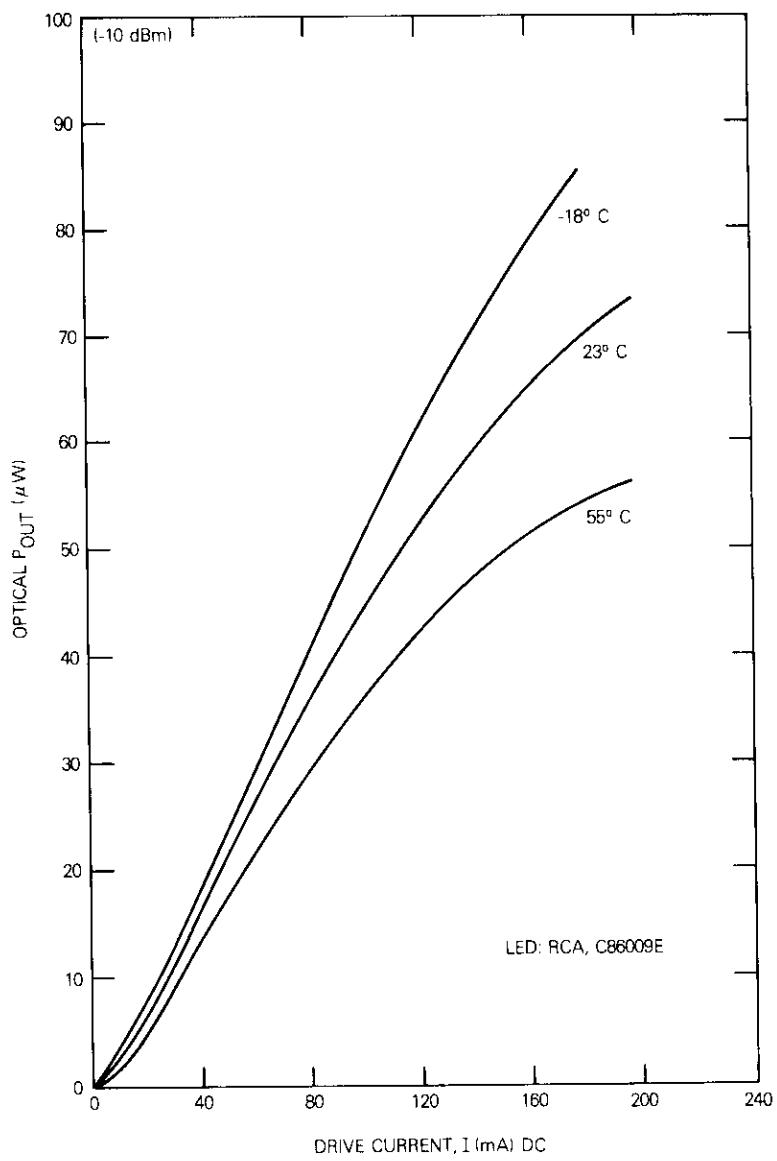
Figure 2a. *Typical Optical Output* vs *Current Characteristics of an LED*
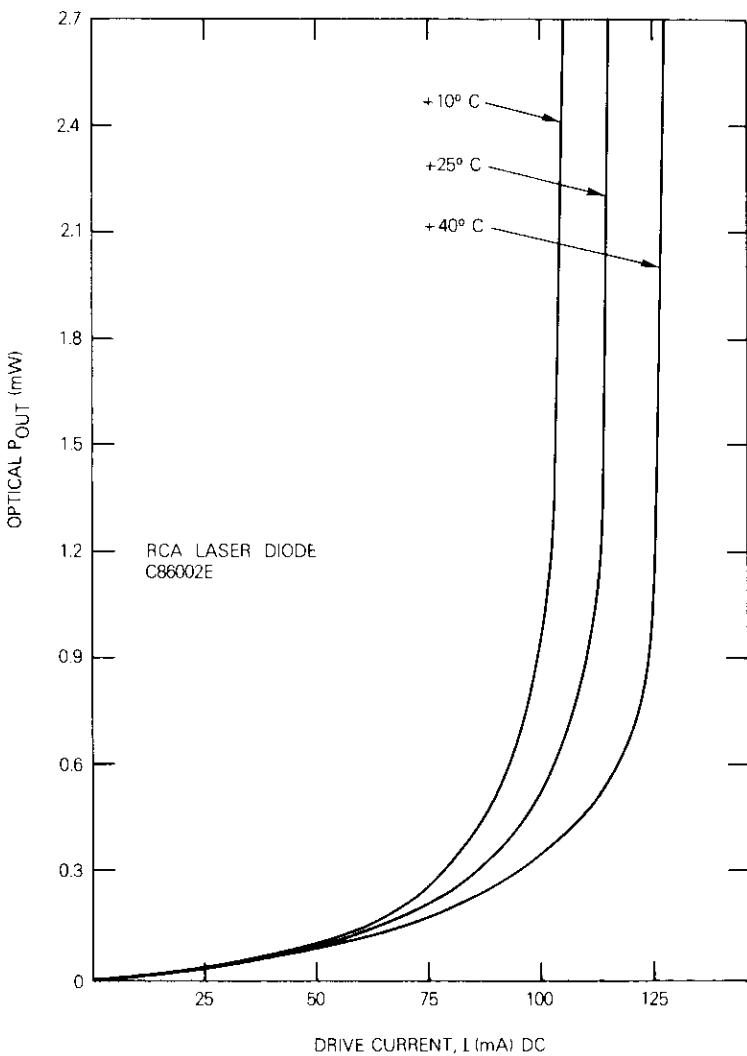
Figure 2b. *Typical Optical Output* vs *Drive Current Characteristics of a Laser Diode*

Figure 3 shows the bandgap energy and lattice constant of III–V compound semiconductor alloys [6]. Ternary compound GaAlAs
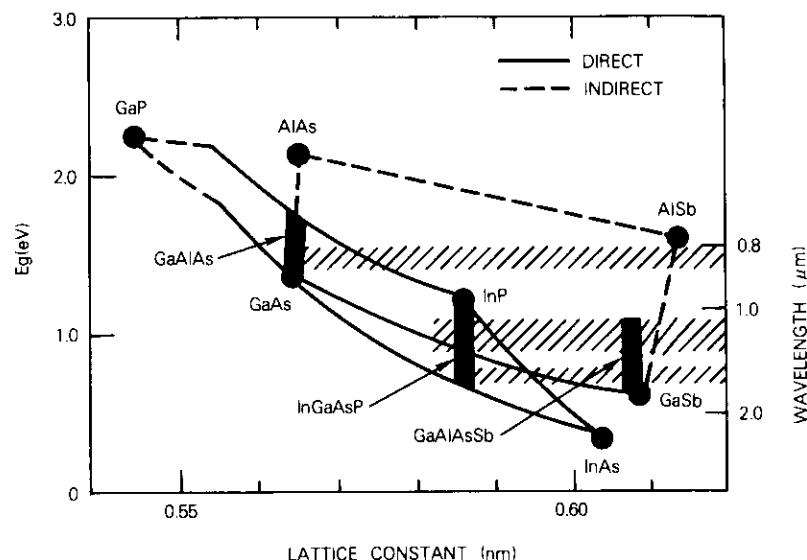
Figure 3. *Bandgap Energy and Lattice Constants of III-V Compound Semiconductor Alloys*

sources which operate in the 0.8- to 0.9-μm wavelength range are well developed and commercially available. Alternatively, new quaternary sources, such as InGaAsP/InP and GaAlAsSb/GaSb, are being developed for the longer 1.0- to 1.7-μm wavelength range. Table 1 compares the typical performance parameters of commercially available sources.

Extensive R&D efforts have improved the performance of GaAlAs single-mode lasers so that 40 mW of optical power at room temperature is achieved with the very narrow line width of <0.015 nm [7]. Commercial devices are also being developed for the longer wavelength region in which improved reliability is projected [8].

Sources suitable for space-based optical communications systems must meet minimum weight and volume and maximum power, efficiency, and reliability standards. Specifically, solid-state lasers employing Nd^{3+}-doped YAG and Nd^{3+}-doped YAlO have been considered suitable candidates for an optical intersatellite link design [9]. Recently, the GaAlAs pumped YAG laser has been shown to yield superior efficiencies. Table 2 compares three potential YAG laser pumping subsystems built for space applications.

TABLE 1. TYPICAL PROPERTIES OF OPTICAL SOURCES

Parameters	Laser Diode	LED
Optical Output Power (Fiber Pigtail Coupled)	3.0 mW	0.1 mW
Spectral Width	<5 nm	≈50 nm
Pulse Rise Time	<1 ns	3 ns
Direct Modulation Rate	>1,000 Mbit/s	<200 Mbit/s
L-1 Linearity	Excellent	Good
Temperature Stability	Compensation Required	Good
Predicted Lifetime	10^6 hr	$>10^8$ hr
Established Reliability for GaAlAs Diodes	≈50,000 hr (CW)	>50,000 hr

TABLE 2. YAG LASER PUMPING TECHNIQUES

Method	Weight (kg)	Prime Power (W)	Remarks
Lamp (K-Rb)	37	315	Limited Lifetime (About 5,000 hr)
Sun	32		Eclipse Consideration
GaAlAs Laser Diode	4.5	40	Compact, Potential Reliability

Detectors

A detector consists of a back-biased PIN junction which converts incident optical energy into hole and electron transitions between the valence and the conduction bands. An applied electric field then causes a current to flow. If the electric field is sufficiently high, secondary electron/hole pairs may be produced by ionization collisions to realize an internal gain. These diodes are called avalanche photodetectors (APDs) [10].

The principal determinant of photodiode operation and efficiency is the wavelength of the incident optical energy. For example, silicon is a particularly attractive detector at wavelengths below 1μm, where quantum efficiencies of up to 85 percent can be realized. Further, dark currents are low and the ratio k_1 of the hole and electron ionization

rates is small, resulting in a small excess noise factor and in very efficient avalanche photodiodes. At wavelengths above 1.1 μm, alternative semiconductors with smaller band gaps, such as germanium, must be employed. In addition, the same ternary or quaternary semiconductor alloys used for sources (Figure 3) can also be used for detectors. Considerable research is presently being devoted to the area of long wavelength, efficient avalanche photodiodes [11]. Table 3 summarizes the important performance parameters of silicon and the longer wavelength detectors.

TABLE 3. PHOTODIODE MATERIAL CHARACTERISTICS

Material	Maximum Operating Wavelength (μm)	Maximum Lattice Mismatch (%)	Quantum Efficiency (η)	Approximate Ionization Ratio (k_i)	Approximate Room Temperature Dark Current (nA)
Si	1.1	0	0.85	0.035	0.1
Ge	1.55	0	0.7	1.0	100
GaInAsP/InP	1.6	0	0.8	0.3	10
GaInAs/Inp	1.6	0	0.8	0.2	10
GaInSb/GaSb	1.6	0.3	0.8	2.0	10

Transmission medium

Electromagnetic energy transfer between a source and a detector can occur either in free space or along a guided transmission line. However, since atmospheric attenuation is normally very severe, unguided optical transmission is practical only in free space, and terrestrial optical communications must use guided transmission techniques. A description of fiber attenuation and bandwidth properties is essential to the communications performance characterization of these systems.

Optical fibers consist of a dielectric core surrounded by a cladding of lower refractive index [12]. The physical and refractive index cross section of the three main fiber types, step-index, graded-index, and single-mode, are shown in Figure 4. Optical ray transmission down the fiber and the process of continued total internal reflection at the core-cladding junction are also shown in Figure 4.
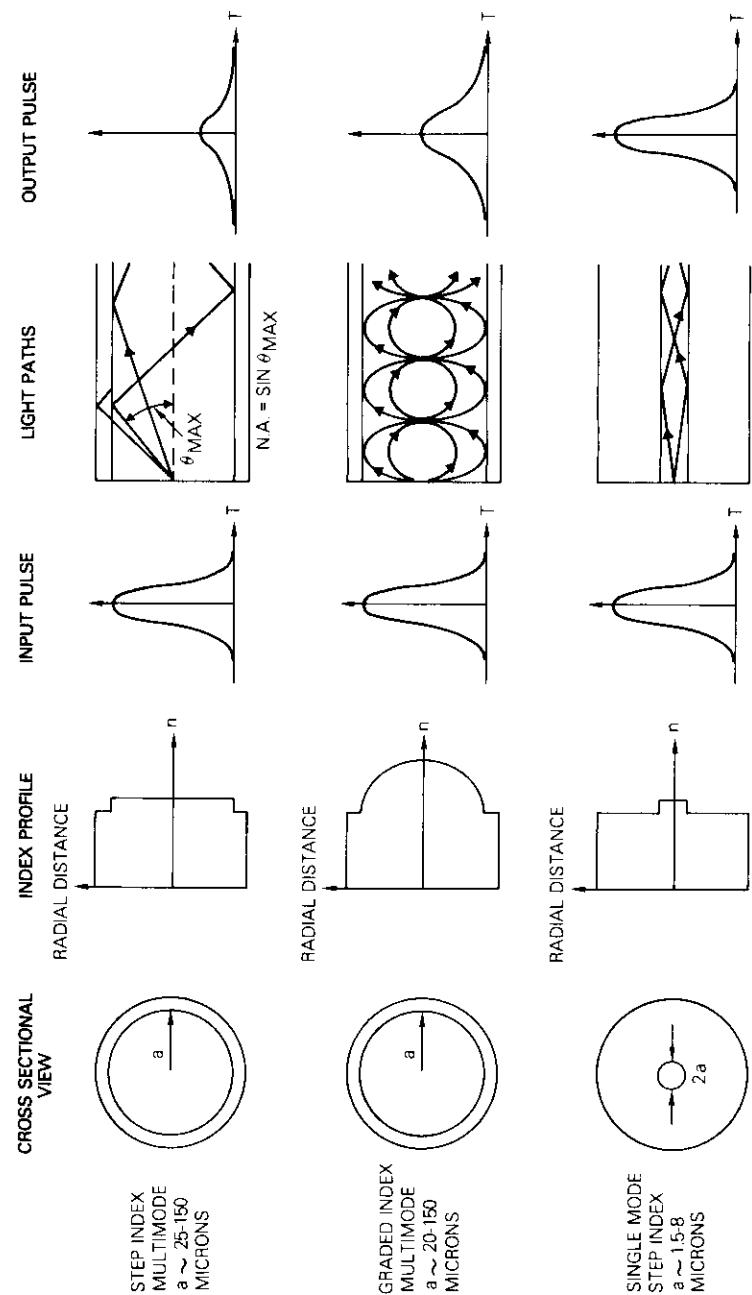
Figure 4. Type of Optical Fibers

The step-index fiber is constructed with a step change of about 1 to 3 percent in the refractive index between core and cladding. Typical fiber dimensions for the core radius range from 25 to 150 μm with the cladding radius about twice this amount. Since the core diameter is many times the optical wavelength (≈ 1 μm), the fiber behaves as a multimode transmission line. The different group velocity of each mode causes the dispersion of a pulse of energy as it travels down the fiber. This effect, known as intermodal dispersion (τ_m)*, is related to the optical bandwidth·distance product β by [13]

$$\tau_m \approx \frac{0.5}{\beta} \quad . \tag{1}$$

For example, a step-index fiber with a dispersion of 20 ns/km has an optical bandwidth of 25 MHz·km; alternatively a 5-km length of fiber will support a bandwidth distance product of about 5 MHz. The electrical or transmission bandwidth is generally less than the optical bandwidth, since optical bandwidth is proportional to the detector current, and electrical bandwidth is proportional to the square of the detector current. For a Gaussian fiber response, electrical bandwidth is approximately equal to $\beta/\sqrt{2}$.

Fiber bandwidth can be improved by at least two orders of magnitude using a parabolically varying core refractive index. This graded-index distribution continually refocuses the optical energy as it propagates down the fiber. Optical bandwidth·distance products in excess of 1000 MHz·km or intermodal dispersions of less than 0.5 ns/km can be achieved.

The bandwidth·distance product of fibers can be further improved by operating with a single propagating mode. For core radii ranging from 2.5 to 5 μm, only the fundamental electromagnetic mode can propagate, and intermodal dispersion is therefore eliminated. However, the bandwidth of the single-mode fiber is finite because additional dispersion mechanisms, which are known as intramodal (τ_s) and depend on finite source line widths, are generally present. The first intramodal effect occurs because of the wavelength dependence of the fiberglass refractive index and is known as material dispersion. The material

---

*Dispersion (τ) is defined in this paper to be the $1/e$ full width of a Gaussian pulse. An often used alternative definition is the rms pulse width σ, which is related to τ, by $\tau = 2\sqrt{2}\sigma$.

dispersion passes through zero and changes sign at 1.3 μm. The second intramodal effect, termed waveguide dispersion, occurs because each mode's group velocity varies as a function of frequency. It has been shown [14] that by combining the effects of material and waveguide dispersion the zero dispersive wavelength can be adjusted to realize multigigahertz single-mode fibers within the range from 1.3 to 1.6 μm.

Attenuation, which is another important property of optical fibers, is influenced by the degree of glass impurity. Recent manufacturing techniques [15] have reduced impurities to a few parts per billion to realize near-theoretical attenuation. This is determined on the short-wavelength side by the Rayleigh scattering limit and on the long-wavelength side by the phonon energy absorption band. For silica glass, a theoretical minimum of 0.18 dB/km occurs at 1.55 μm, and an experimental value of 0.20 dB/km has been reported [4]. Indeed, the state-of-the-art values of attenuation over the wavelength range of 0.8 to 1.55 μm are very close to the fundamental limit (Figure 5), and little improvement can be expected in the future.
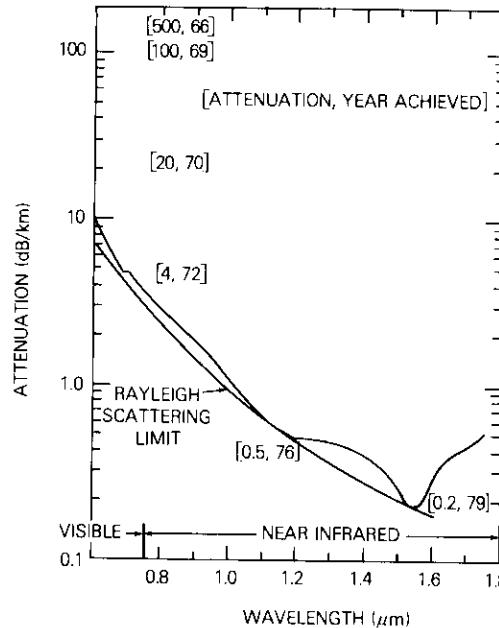

Figure 5. *Glass Fiber Attenuation* vs *Wavelength*

Connectors, splices, and couplers

Connectors, couplers, and splices have intrinsic loss mechanisms, some of which are inherent in the component design (e.g., coupler power splitting loss) and some of which can be overcome or reduced by careful component design and fabrication. For connectors and splices, the most serious loss mechanism is fiber misalignment. Figure 6 shows the three types of fiber misalignments that can occur. Lateral misalignment causes severe signal loss and is most difficult to correct, since the core radii range from 30 μm for a multimode fiber to 2.5 μm for a single-mode fiber. Longitudinal and angular misalignments result in much lower losses[16].

LONGITUDINAL DISPLACEMENT

CORE RADIUS
(2.5-30 μm, 0.0001-0.001 IN.)

LATERAL DISPLACEMENT

ANGULAR DISPLACEMENT
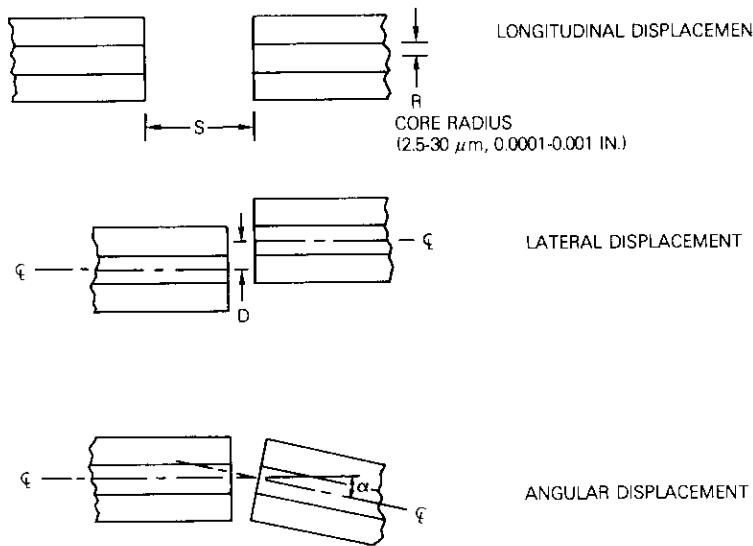
Figure 6. *Connector/Splice Fiber Misalignments*

CONNECTORS

Three types of connectors have been developed to date, each using a different technique to minimize fiber misalignment losses. Their characteristics are shown in Table 4.

For channel-centered connectors [17], the input and output fibers are inserted into ferrules or into an interstitial space among several alignment rods and epoxied into place; the fiber end faces are polished.

TABLE 4. CONNECTOR CHARACTERISTICS

Type	Material	Size[a] (in.)	Typical Loss (dB) With IMF[b]	Typical Loss (dB) Without IMF[b]
Channel-Centered	Metal	1	0.7	3
Cone-Centered	Plastic/Epoxy	1	0.4	2
Lensed	Metal With Lens	5	<0.5	—

[a] Ferrule-sleeve-ferrule overall length.
[b] Index matching fluid.

The ferrules or rods are then inserted into each end of a sleeve or of a second alignment rod cluster. Close machining tolerances of the assembly minimize fiber misalignment. Channel-centered connectors will usually withstand handling and environmental extremes because of their all-metal construction. However, the metal ferrules and sleeves or alignment rods will tend to wear as they rub against each other during disconnect/reconnect cycles. Because of the complicated construction, the close tolerances of each piece can compound to allow a total lateral fiber misalignment of several tenths of a core radius with the accompanying increase in loss.

Cone-centered connectors [18] consist of cone-shaped ferrules and connection sleeves with conical holes to guide the ferrules into alignment. A precisely drilled hole along the axis of each ferrule corrects lateral and angular misalignment, and grinding of the ferrule and fiber end faces minimizes longitudinal misalignment. The plastic or epoxy material cannot withstand the handling or environmental extremes tolerated by metal connectors; therefore, their molded construction and variations in material and fabrication conditions do not allow for close dimensional tolerances.

The best connector for reducing fiber misalignment losses is the lensed connector [19]. As the ferrules of this connector are inserted into the connecting sleeve, the fibers are guided into index matching fluid-filled cups in the sleeve. The bottoms of the cups form a lens which images the input fiber end onto the output fiber end. The bulkiness of this type of connector makes it inapplicable to restricted spaces, and the index matching fluid is subject to contamination and change of properties in environmental extremes. This is also true of index matching fluid used with channel- or cone-centered connectors;

however, they can be used without the fluid when conditions dictate (*e.g.*, in aircraft), whereas the lensed connector cannot.

SPLICES

Splices do not have the disconnect/reconnect requirement of connectors; therefore, they are made individually to minimize misalignment losses. The most reliable splices are of two general types: direct fiber end attachment via index matching adhesives and arc welded splices (giving typical splice losses of 0.6 dB and 0.1 dB, respectively).

COUPLERS

The purpose of couplers is different from that of connectors and splices, because couplers are designed to distribute incoming signals onto output fibers in a uniform manner, with loss minimization being a secondary consideration. Two types of couplers, mixing-block and blocked-fibers, are presently available. Mixing-block couplers [20] consist of a number of fiber pigtails attached to a glass block which distributes any incoming signal to all outgoing fibers. Blocked-fiber couplers [21], [22] are formed by gathering several multimode fibers into a bundle, heating a section of the bundle until the glass softens, and then pulling and twisting the softened section. Coupling is achieved because, as a signal propagates through the stretched biconical section, the higher order modes are coupled out of the core as it narrows and into the fused cladding. The signal is then returned to the cores of all the blocked fibers as they assume their normal cylindrical shapes. A similar "bottle coupler" has been developed for single-mode fibers [23].

Optic fiber link specifications

Before installation in the field, a fiber is encased in protective material, which may exert stress on the fiber and increase its attenuation. Practical experience has shown that the cabling procedure adds 0.25 to 0.5 dB/km to basic fiber attenuation. Further, installation may require a number of splices, connectors, and couplers, and these losses must be added to achieve a total cable loss. However, for long cables (>5 km), these component losses can be amortized as a function of system length, reducing the additional losses to less than 0.1 to 0.2 dB/km. Table 5 summarizes cable attenuation, optical bandwidth, and intermodal dispersion as a function of fiber type and wavelength. Estimates of intramodal (material and waveguide) dispersion values are also included for typical LED and laser sources.

TABLE 5. ATTENUATION AND BANDWIDTH CHARACTERISTICS

Wavelength (μm)	Fiber Type	Cable Attenuation (dB/km)	Bandwidth distance product (MHz·km)	Intermodal Dispersion (τ_m) (ns/km)	Intramodal Dispersion (τ_j)(ns/km)			
					LD (2-nm Line Width)		LED (35-nm Line Width)	
					Material	Waveguide	Material	Waveguide
0.8–0.9	Step-Index	2.50–4.0	25	20.0				
	Graded-Index		1000	0.5	0.2	<0.05	3	<0.7
	Single-Mode		>10⁵	~0.0				
1.3	Step-Index	0.75–1.0	25	20.0				
	Graded-Index		1000	0.5	0.0	<0.05	0.0	<0.7
	Single-Mode		>10⁵	~0.0				
1.55	Step-Index	0.45–0.7	25	20.0				
	Graded-Index		1000	0.5	0.0–0.05	<0.05	0.7	<0.7
	Single-Mode		>10⁵	~0.0				

Communications performance

The integration of various optical components in a communications system requires an evaluation of the capability of each component to satisfy bandwidth requirements and link power specifications. In a typical optical link consisting of an intensity modulated optical source, an optical transmission medium, and a photodetector, source output is directly modulated by a voltage or current, which can be either a baseband digital signal or a modulated carrier. Bandwidth is limited principally by source rise time and medium dispersion, and power is limited principally by noise in the photodetector and fiber attenuation.

In long-distance links, the attenuation or bandwidth reduction may be so large that the carrier-to-noise ratio (C/N) or the bandwidth at the receiver is insufficient. It is then necessary to employ repeaters consisting of a photodetector, which converts the optical to an electrical signal, and an optical source, which reconverts the signal for propagation along the next segment of the link. Repeaters restore the bandwidth and optical power of the signal at the cost of additive noise from the detection process. Unless regeneration of the electrical signal is employed at each repeater, the cumulative repeater effects may be severe. Regeneration of the electrical signal is readily accomplished when the signal format is baseband digital, but it can be more difficult for modulated carriers. The basic constraint determining the number of repeaters in a link is that repeaters must be sufficiently close to provide adequate bandwidth.

With respect to noise in the link, the contribution from each repeater increases as the spacing between repeaters increases because of the increased segment attenuation; however, as the spacing increases, the number of repeaters decreases. For a fixed total link length, there is a repeater spacing which maximizes the overall C/N. To determine this optimum spacing, it is first necessary to describe noise generation properties of a photodetector and to derive a general expression for the C/N at the receiver output. This expression will then be applied to the special cases of a baseband digital signal and a QPSK-modulated carrier.

C/N in photodiode detectors

The principal sources of noise in an optical link are the photodetector diode, which generates shot noise proportional to the signal current level, and the photodetector preamplifier, which generates thermal

noise proportional to the temperature of the diode load resistance. For baseband digital modulation, the average optical power required to achieve a given C/N [and thus a given bit-error rate (BER)] is shown in the Appendix to be given by*

$$P_O = \frac{h\nu}{2\eta} \sqrt{\frac{C}{N}} \left\{ F(\overline{M})R \sqrt{\frac{C}{N}} + \frac{2}{q} \sqrt{\left[qI_d F(\overline{M}) + \frac{2kT}{\overline{M}^2 R_e} \right] R} \right\} \qquad (2)$$

where
$h =$ Planck's constant (J·s)
$\nu =$ optical frequency (Hz)
$\eta =$ photodetector quantum efficiency
$C/N =$ carrier-to-noise ratio
$F(\overline{M}) =$ excess noise factor
$R =$ bit rate (bit/s)
$q =$ electron charge (C)
$I_d =$ photodetector thermally generated dark current (Amps)
$k =$ Boltzmann's constant (J/K)
$T =$ temperature (K)
$\overline{M} =$ average detector multiplication factor
$R_e =$ thermal noise equivalent resistance (Ω).

When the optical source is modulated by a QPSK carrier, the average optical power required to achieve a given C/N is

$$P_O = \frac{h\nu}{\eta} F(\overline{M})R \frac{C}{N} \left\{ 1 + \sqrt{1 + 2 \left[\frac{qI_d F(\overline{M}) + 2kT/\overline{M}^2 R_e}{q^2 F^2(\overline{M}) R\, C/N} \right]} \right\} \cdot \quad (3)$$

SYSTEM GAIN OPTIMIZATION

It is clear from equation (2) for the digital link and from equation (3) for the QPSK carrier link that some value of $\overline{M}$ maximizes the C/N at the detector output. Increasing $\overline{M}$ reduces the preamplifier thermal

---

*Equation (2), which has been derived from Gaussian statistics, predicts that the minimum number of photons received per bit is given by C/N. (Thermal noise and dark current are neglected.) That is, for a BER of 10^{-9}, 36 photons must be received per bit. While a more exact Poisson statistics analysis predicts 20 photons per bit, at higher noise levels the Poisson and Gaussian statistics become virtually equivalent.

noise, but increases the excess noise factor associated with multiplication of the photodetector shot noise. The excess noise factor is given by [24]

$$F(\overline{M}) = k_I \overline{M} + 2(1 - k_I) - \frac{(1 - k_I)}{\overline{M}} \qquad (4)$$

where k_I is the effective ratio of hole to electron ionization rates. When this is substituted into equation (2), the value of $\overline{M}$ which minimizes the optical power required to obtain a given C/N in a baseband digital link with negligible dark current is given by

$$\overline{M}_o = \sqrt{\frac{1}{k_I}\left[\frac{2}{q}\sqrt{\frac{2kT\pi C_j}{C/N}} + k_I - 1\right]} \qquad . \qquad (5)$$

Therefore, the minimum optical power is

$$P_{O_{min}} = \frac{qR(C/N)}{R_O}\left\{1 - k_I + \sqrt{k_I\left[\frac{2}{q}\sqrt{\frac{2\pi kTC_j}{C/N}} + k_I - 1\right]}\right\} \qquad (6)$$

where R_O, the detector responsivity is defined by $R_O = \eta q/h\nu$.

In equations (5) and (6), the parallel combination of junction capacitance, C_j, and effective resistance, R_e, must have a sufficiently small time constant to transmit the highest signal frequency (f_m), i.e.,

$$R_e C_j \geqq \frac{1}{2\pi f_m} \qquad . \qquad (7)$$

The minimum optical power required in a digital link is plotted against C/N in Figure 7 for typical values of k_I, T, R_O and C_j.

For QPSK modulation with negligible dark current, the maximum C/N for a given incident optical power is obtained when $\overline{M}$ is the solution of the following equation:

$$\overline{M}^3 + \frac{(1 - k_I)\overline{M}}{k_I} - \frac{8\pi kTf_mC_j}{k_IqR_OP_O} = 0 \qquad (8)$$

where f_m is the maximum signal frequency. Substituting a solution of this equation into equation (3) will yield the maximum C/N obtainable for a given P_O. Figure 8 shows the maximum C/N obtainable
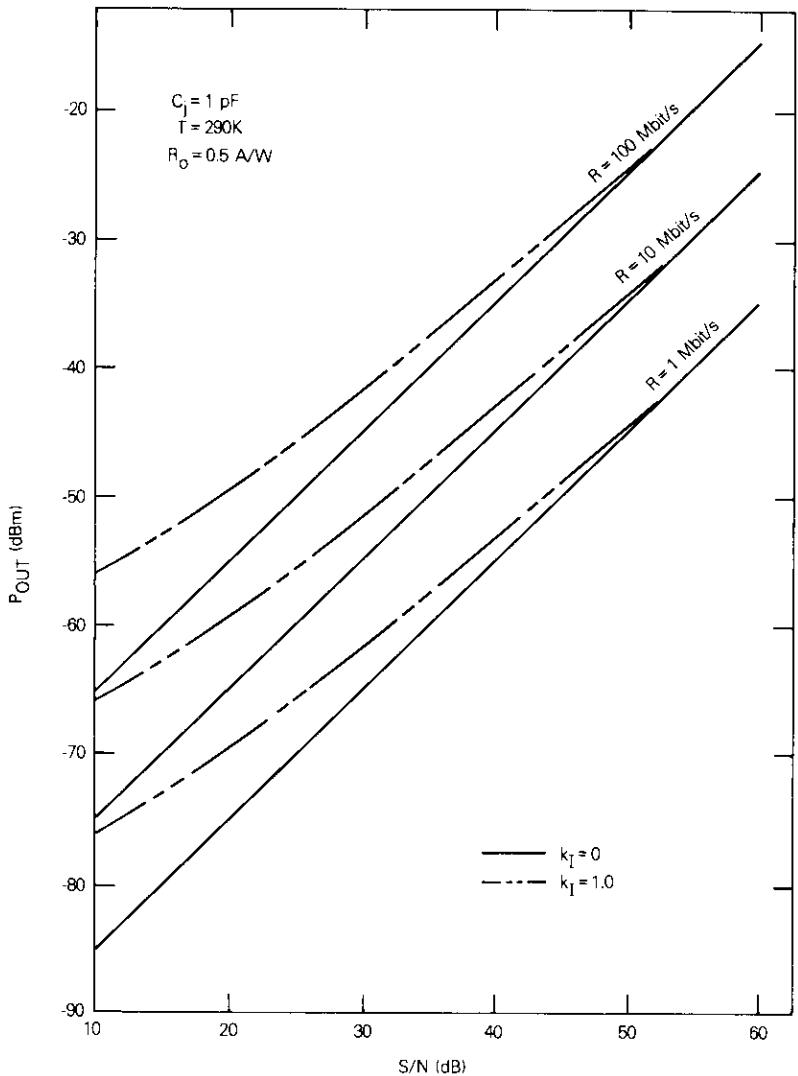
Figure 7. *Required Optical Power* vs *Signal-to-Noise Ratio*

in a QPSK modulation link for circuit characteristics the same as in Figure 7.
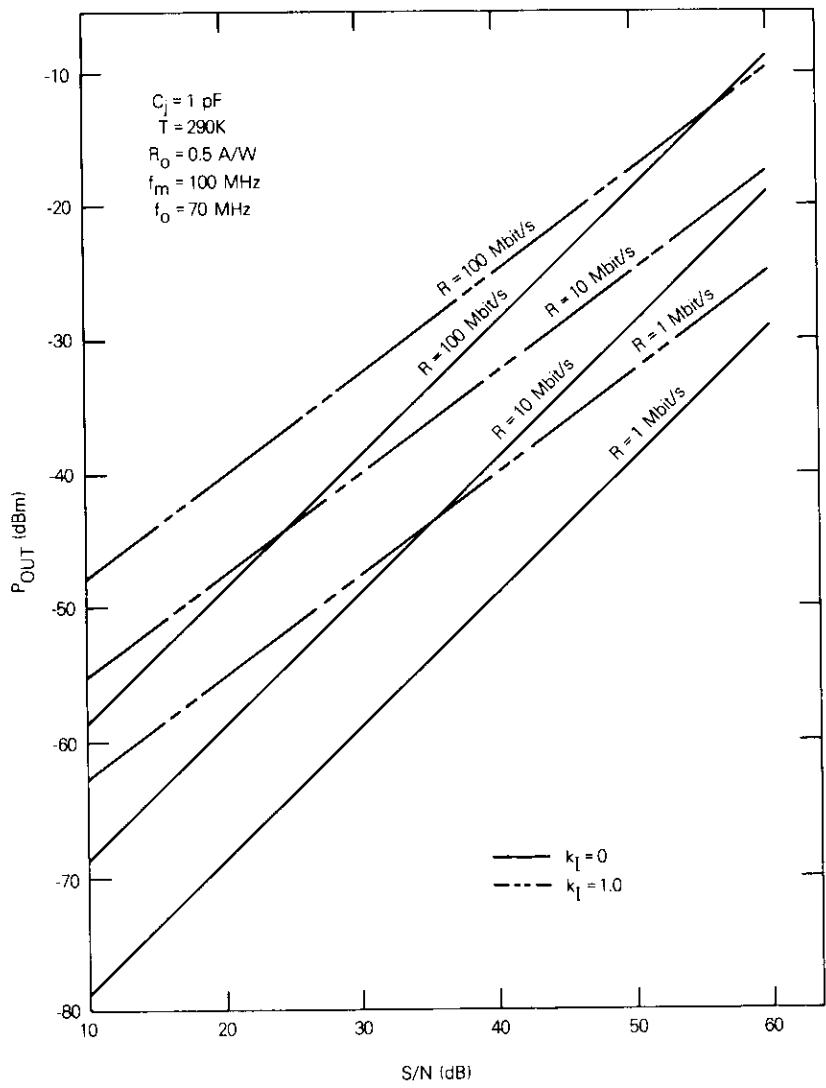
Figure 8. *Required Optical Power* vs *Signal-to-Noise Ratio QPSK Signal Modulation*

IF carrier transmission repeater spacing considerations

In applications such as earth station diversity links and long-distance

trunk lines, excessive attenuation and bandwidth reduction necessitates the use of multiple repeaters. Long-distance trunk lines employ pulse-code-modulated (PCM) encoded digital signals which are regenerated at each repeater. Regeneration of an IF band of modulated carriers, on a diversity link requires expensive demodulators and modulators at each repeater, unless a repeater spacing is selected to maximize the analog optical link C/N. This maximized ratio should be at least 15 dB higher than that of the received signal to incur no more than 0.1-dB degradation in overall C/N.

A multisection fiber optic link of length L may consist of m identical segments each of length ℓ transmitting a set of modulated carriers in a bandwidth B. The C/N at the output of the m segment link is

$$\left(\frac{C}{N}\right)_m = \frac{1}{m} \frac{R_O^2 P_O^2/qB}{4F(\bar{M})[R_O P_O + I_d] + 16\pi kTC_j f_m q\bar{M}^2} \qquad (9)$$

The signal attenuation over one segment is given by

$$\frac{P_O}{P_i} = 10^{-aL/10m} \quad . \qquad (10)$$

Consider two limiting cases of the photodetector: shot noise dominated (such as in a high-gain APD receiver) and preamplifier thermal noise dominated (such as in a PIN diode receiver where $\bar{M} = 1$). In the first case, equations (9) and (10) yield

$$\left(\frac{C}{N}\right)_m = \frac{R_O P_i}{4qF(\bar{M})B} \frac{10^{-aL/10m}}{m} \qquad (11)$$

whereas, in the second case

$$\left(\frac{C}{N}\right)_m = \frac{R_O^2 P_i^2}{16\pi kTC_j f_m B} \frac{10^{-2aL/10m}}{m} \qquad (12)$$

where P_i is the average optical power coupled into the fiber, and a is the fiber attenuation in dB/km.

Assume a diversity link transmitting a 45-MHz bandwidth signal over a 50-km distance with repeaters consisting of APD receivers and sources which couple 1 mW of optical power into the cable. For typical

receiver characteristics of R_O equals 0.5 A/W, $\bar{M}$ equals 100, and k_I equals 0.02. Equation (11) yields

$$\left(\frac{C}{N}\right)_{m_{dB}} = 66.4 - 10\log_{10} m - \frac{50a}{m} \quad . \tag{13}$$

From equation (9), $(C/N)_{m_{dB}}$ is plotted as a function of m in Figure 9 for a wide range of fiber attenuations. This figure shows that the C/N is near maximum over a broad range of repeater numbers (repeater spacings) for a given fiber attenuation and that a threshold exists. When the repeater number is above threshold, performance is near optimum. Maximum C/N occurs for a section loss of 4 dB. If threshold is defined as 3 dB below maximum, then the number of repeaters required to obtain the C/N at threshold is approximately one-third that required to obtain the maximum C/N. Moreover, if the required optical link C/N is 40 dB (or about 15 dB above the satellite to earth station clear sky maximum elevation angle C/N), only four repeaters will be required for a fiber with 2-dB/km attenuation.

If PIN diode receivers are used for the same link, then for a maximum signal frequency of 100 MHz, junction capacitance of 1 pF, and T of 290 K, equation (13) yields

$$\left(\frac{C}{N}\right)_{m_{dB}} = 84.9 - 10\log_{10} m - 100\frac{a}{m} \quad . \tag{14}$$

This is plotted in Figure 10. The optimum section attentuation is now 2 dB, but again C/N has a broad maximum. If the requirement is 40 dB, maximum C/N can be achieved with a 2-dB/km fiber and five repeaters.

Digital transmission communications performance

Link calculations for a digital optical communications system begin with the minimum optical power sensitivity P_o of an avalanche photodiode as given by equation (2). Minimum power and launch sensitivities for optimum avalanche silicon photodiodes at 0.85 μm and germanium photodiodes operating at 1.3 are shown in Figure 11 for a 10^{-9} BER.

For a power-limited optical link (no dispersion), the repeater spacing l(km) is given by the difference between optical power into the fiber P_i and P_o divided by the optical cable loss/km (a). With an appropriate
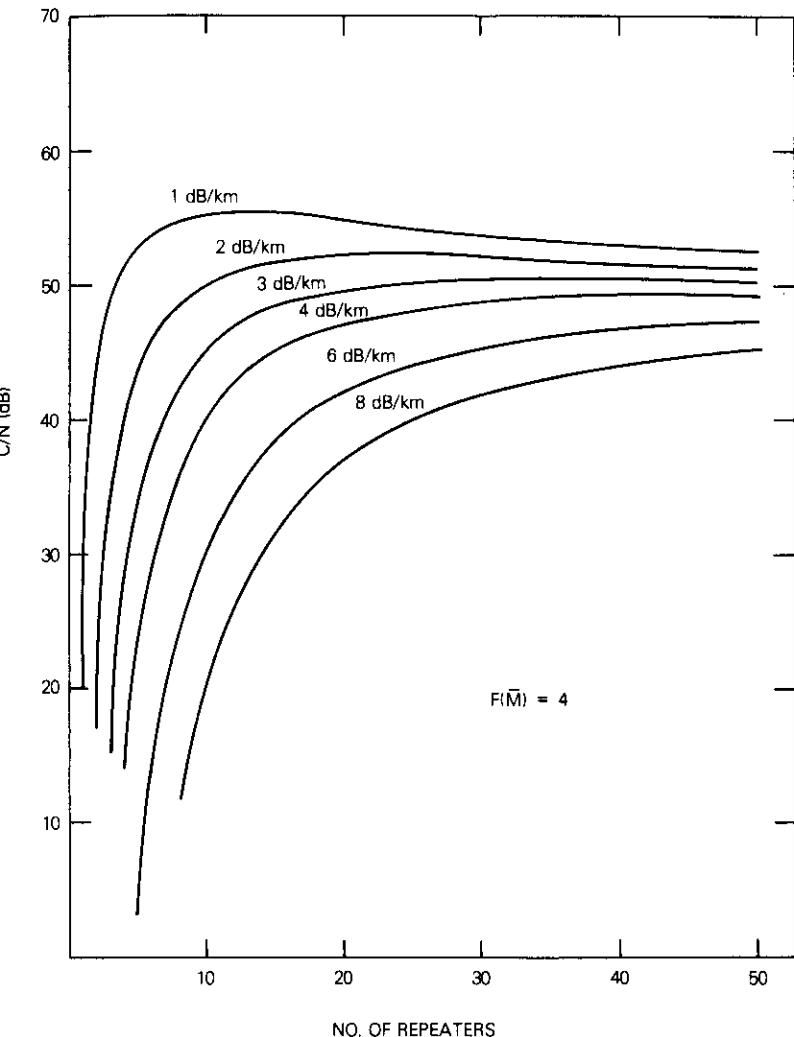
Figure 9. *Link Signal-to-Noise Ratio* vs *Number of Repeaters*

power margin P_m, the repeater spacing becomes

$$l = \frac{P_i - P_o - P_m}{a} \quad . \tag{15}$$
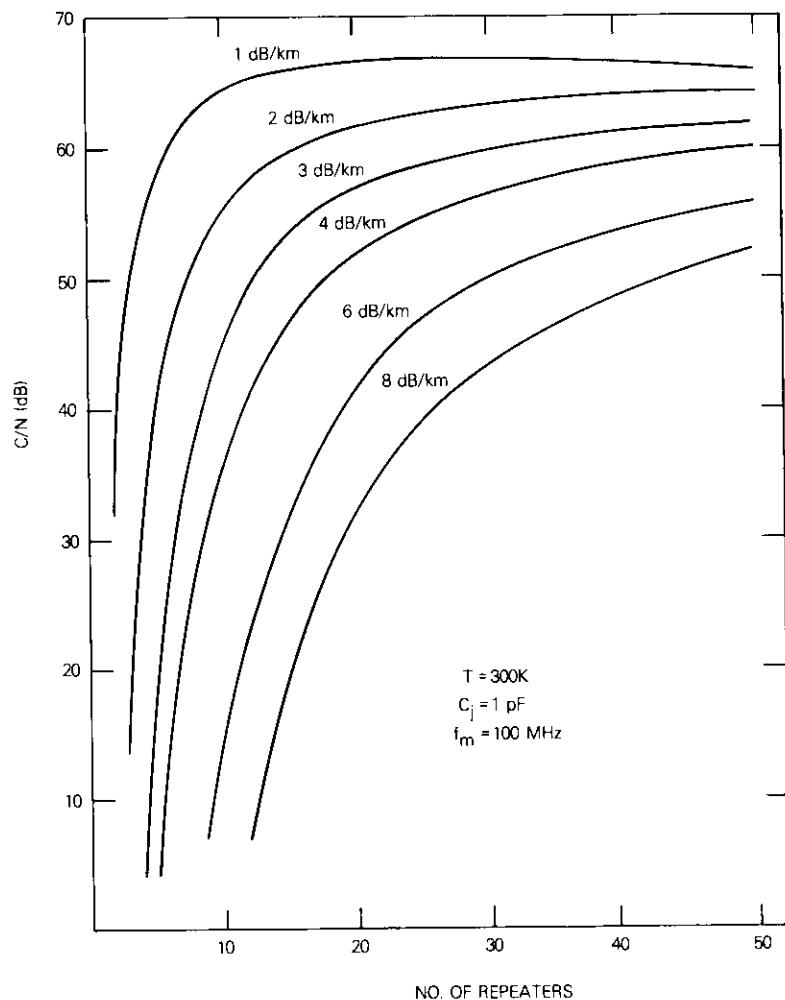
Figure 10. *Signal-to-Noise Ratio* vs *Number of Repeaters (50-km Link—PIN diode detectors 45-MHz signal bandwidth)*

For bandwidth-limited links, intersymbol interference will occur, and an additional equivalent power loss/km P_{isi} must be included in the numerator of equation (15). If Gaussian-shaped pulses are assumed, P_{isi} is given by [25]
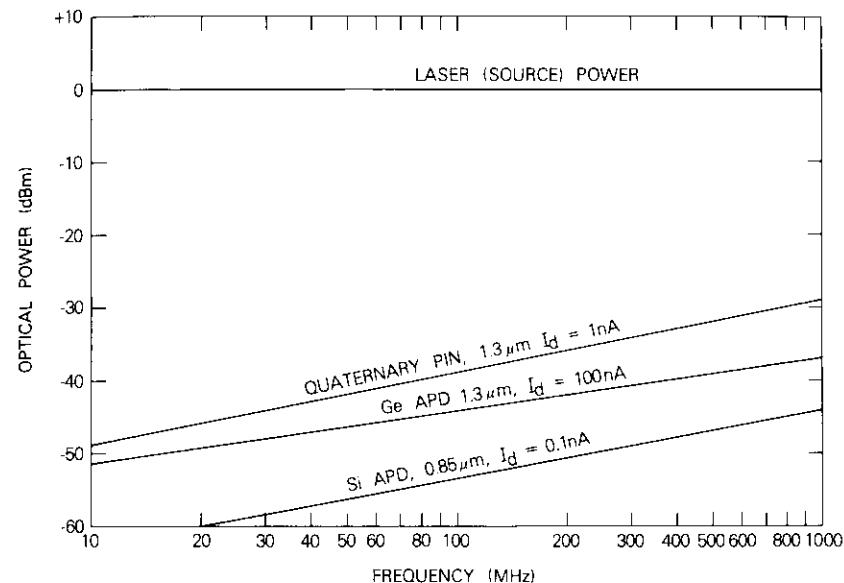
Figure 11. *Optical Detector Sensitivity* vs *Bit Rate*

$$P_{isi} = 2\left(\frac{\tau}{T}\right)^4 \qquad (16)$$

where $(\tau/T)^2 = (0.408)^2 + (\tau_m l/T)^2 + (\tau_s l/T)^2$

τ_m = intermodal dispersion, s/km

τ_s = intramodal dispersion, s/km

T = bit interval.

Intramodal dispersion is a combination of both material (τ_f) and waveguide dispersion (τ_w); the magnitudes depend on the source line width. The general expression for repeater spacing, therefore, becomes

$$l = \frac{P_i - P_o - P_m - P_{isi}}{a} \qquad . \qquad (17)$$

This equation has been used to derive the repeater spacing as a function of bit rate at three optical wavelengths—0.85, 1.3, and 1.55

μm. The results are plotted in Figure 12. Minimum optical power sensitivities (which correspond to a BER of 10^{-9}) shown in Figure 11, source power levels (P_t) of 0 dBm, and a power margin of 10 dB were used in all the link calculations. Values for the intermodal and
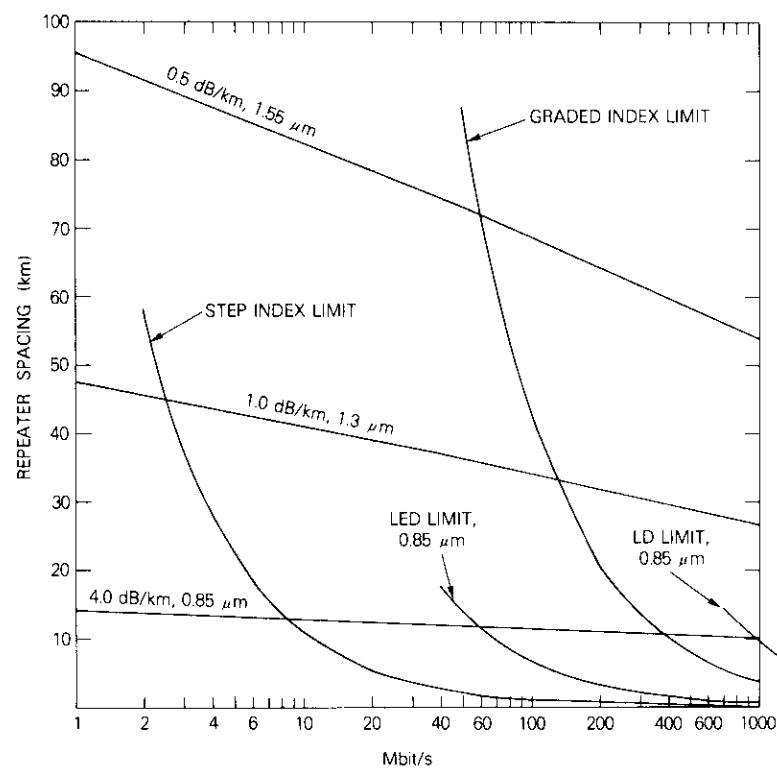
Figure 12. *Repeater Spacing* vs *Bit Rate*

intramodal dispersion parameters are summarized in Table 5. The results indicate the following:

a. The step-index multimode fiber is bandwidth limited to about 5 Mbit/s (2.5 MHz) for the 0.8- to 0.9-μm wavelength range technology. Although fiber attenuation is reduced at longer wavelengths, this advantage is offset by further bandwidth reduction at longer link lengths. Therefore, employing this fiber type at these

wavelengths would not be advantageous. LEDs offer more than adequate bandwidths and output powers for typical step-index multimode fiber links, and simple PIN receivers are normally employed.

b. The improved intermodal dispersion properties of graded-index fibers make them particularly useful in the 0.8- to 0.9-μm wavelength range for bandwidths up to about 100 MHz (200 Mbit/s). Operation at bit rates below 30 Mbit/s can employ LED sources, while applications near the upper bandwidth limit must employ LDS. PIN receivers can be used for these links, although APDS offer improved performance.

c. Graded-index multimode fibers operating at 1.3 μm can be employed in applications using moderate bandwidths of 50 Mbit/s (25 MHz) and repeater spacings of about 30 km. At this wavelength, material dispersion is negligible and only intermodal dispersion would be present.

d. Operation of single-mode fiber optic links in the 1.3- to 1.55-μm wavelength range eliminates material dispersion and allows for potentially large information bandwidths (>1 GHz) over repeater spacings of up to 100 km. This technology is clearly suited to long-distance trunk routes and future submarine cable applications.

e. The upper bandwidth limit of optical links is determined by the source or detector rise time. Modulation rates of 200 and 2,000 MHz can be achieved with LEDs and LDs, repectively, while PIN diode and APD receivers can operate at bandwidths greater than 4 and 1.5 GHz. The type of source and detector employed must be compatible with the desired operational bandwidth.

QPSK transmission communications performance

A link calculation for a QPSK-modulated signal is similar to the baseband digital calculation starting with the photodiode sensitivity given in equation (3). Consider, for example, a fiber optic link designed to transmit a 70-MHz QPSK carrier with a bit rate of 60 Mbit/s over 5 km at the required C/N of 40 dB. Further, assume a laser diode source, a graded-index multimode fiber with 4-dB/km attenuation, and a PIN diode receiver with a 290 K noise temperature, 0.5 A/W responsivity, and 1 pF junction capacitance.

For the above specifications with $\overline{M} = 1$ and $F(M) = 1$ and assuming negligible dark current, equation (3) gives a required incident optical

power of -21.3 dBm. The required C/N can be achieved by coupling 0 dBm from the source into the 4-dB/km fiber.

The link bandwidth will be limited principally by the intermodal fiber dispersion rather than by the laser source. For a laser spectral line width which results in 0.1 ns/km of material dispersion and 1 ns/km of intermodal dispersion, the total dispersion of 1.005 ns/km yields a bandwidth distance product of approximately 250 MHz/km. Thus, a link bandwidth of 50 MHz is realized over the 5-km run length. This link is, therefore, bandwidth rather than power limited, and three repeaters would be required to obtain the necessary bandwidth.

Applications

The following section will address the complementary and competitive applications of this technology in satellite communications. In addition, optical-microwave circuits, which should improve the flexibility and variety of onboard satellite optical signal processing, will be discussed.

Interfacility links

A particularly suitable application of fiber optics technology to satellite communications is interfacility links between earth terminals and signal processing sites as shown in Figure 13. Either analog or digital modulation may be employed, depending on the particular application. Two configurations are of major interest. The first is the intermediate frequency interfacility link (IF-IFL) where signal processing equipment is remotely located from the antenna site. The received down-link signal is down-converted to IF at the antenna site without further processing. The IF signal is then converted to optical power and transmitted on an optical fiber to the signal processing site where carrier recovery (for digital transmission), demodulation, and demultiplexing are performed.
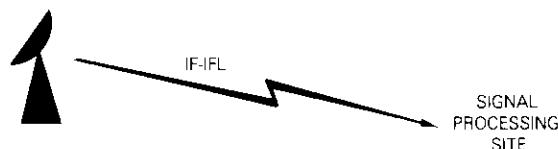

Figure 13. *Interfacility Link*

The second configuration is the local distribution interfacility link typically installed between a central earth station and several users, none of whom require a dedicated earth station. The multiplexed user data modulate an RF carrier at the central multiplexing unit.

Interfacility links may be either of these configurations or a combination of the two. Generally, they are less than 1 km long and do not require optical repeaters. Fiber optics is attractive in such installations because of its wide bandwidth and low loss characteristics, but more importantly because of its light weight, flexibility, and immunity from EMI.

A fundamental performance requirement of the interfacility link is that it must not measurably degrade the C/N (or, equivalently, the BER) of the satellite link. Thus, for example, in a satellite communications link which has a clear sky C/N of 25 dB in the IF bandwidth, the IF interfacility link with a noise-free input should have an output C/N of at least 40 dB. In the local distribution interfacility link, a minimum BER at the link input of 10^{-9} requires a fiber optic link BER of less than 10^{-10} for an error-free input.

Table 6 gives typical specifications for both IF and local distribution interfacility links. In the IF interfacility link, analog intensity modulation requires highly linear light output vs drive current optical sources. The excellent linearity of GaAlAs laser diodes makes them particularly suitable for high-capacity interfacility links. LEDs can also be used for frequencies up to 100 MHz, but require distortion compensation circuitry to correct L-I nonlinearity [26]. It was previously shown that a C/N of better than 50 dB can be achieved at a PIN receiver with a relatively high-loss (10 dB/km) graded-index silica fiber. An LED source will meet the specifications of the local distribution interfacility link,

TABLE 6. TYPICAL OPTICAL INTERFACILITY LINK SPECIFICATIONS

Parameter	IF-IFL	Local Distribution IFL
Carrier Frequency	70 MHz	Baseband
Signal Bandwidth	33 MHz	5 MHz (10 Mbit/s)
Carrier Modulation	4ϕ PSK-Burst Mode	
Optical Source Modulation	Analog Intensity Modulation	On-Off Intensity Modulation
Maximum Link Length	1 km	1 km
Input C/N or BER	25 dB	10^{-9}

and a high-loss cable of 10 dB/km will provide a link BER several orders of magnitude better than 10^{-9}. Further, complete local distribution links should be relatively inexpensive, since fiber optic bandwidth-distance of about 5 MHz·km can be achieved with step-index fibers.

The IF interfacility link applications have been tested in the laboratory. In this experiment, a 43-Mbit/s 4ϕ PSK-modulated burst carrier was transmitted over a 1-km graded-index fiber using a GaAlAs laser diode source and a PIN diode receiver. Negligible degradation in BER due to the fiber optic link was observed.

Earth station diversity links

Earth station spacial diversity has been proposed to overcome the effect of rain-induced propagation attenuation at higher frequencies. The distance between diversity stations can reach 50 km, and a basic scheme is shown in Figure 14. The total signal bandwidth is 500 MHz in the 11/14-GHz band, and exceeds 2 GHz in the 20/30-GHz band. The wideband, relatively long-range trunking requirements of diversity links are particularly suitable for digital or analog applications of optical fibers.

Nevertheless, state-of-the-art technology link designs are both bandwidth and power limited. Consider a typical 0.8- to 0.9-μm multimode graded-index fiber with 3-dB/km attenuation and a 1,000 MHz·km bandwidth-distance product. This allows repeater spans of only 10 to 12 km and bandwidths of only 100 MHz per fiber. Wavelength division multiplexing and demultiplexing are required to overcome these limitations. However, diversity links without repeaters or wavelength division multiplexing requirements can be designed with single-mode fibers in the 1.3- to 1.55-μm wavelength range. In this range, fiber cable loss is typically 0.5 to 1.0 dB/km, and the available bandwidth-distance exceeds 50 GHz·km (see Table 5 and Figure 12).

Digital fiber systems

Large numbers of digital optical systems employing the presently available 0.8- to 0.9-μm wavelength range technology and operating at T-2 and T-3 bit rates of 6.312 and 46.304 Mbit/s (96 and 672 telephone circuits), respectively, are being placed in service throughout the world. A recently announced ATT trunk system between Boston and Washington, D.C., will operate at twice the T-3 rate, i.e., 92.608 Mbit/s. This design will employ repeater spacings of about 7 km, laser diodes, graded-index fibers with 3-dB/km loss, and silicon APD detec-
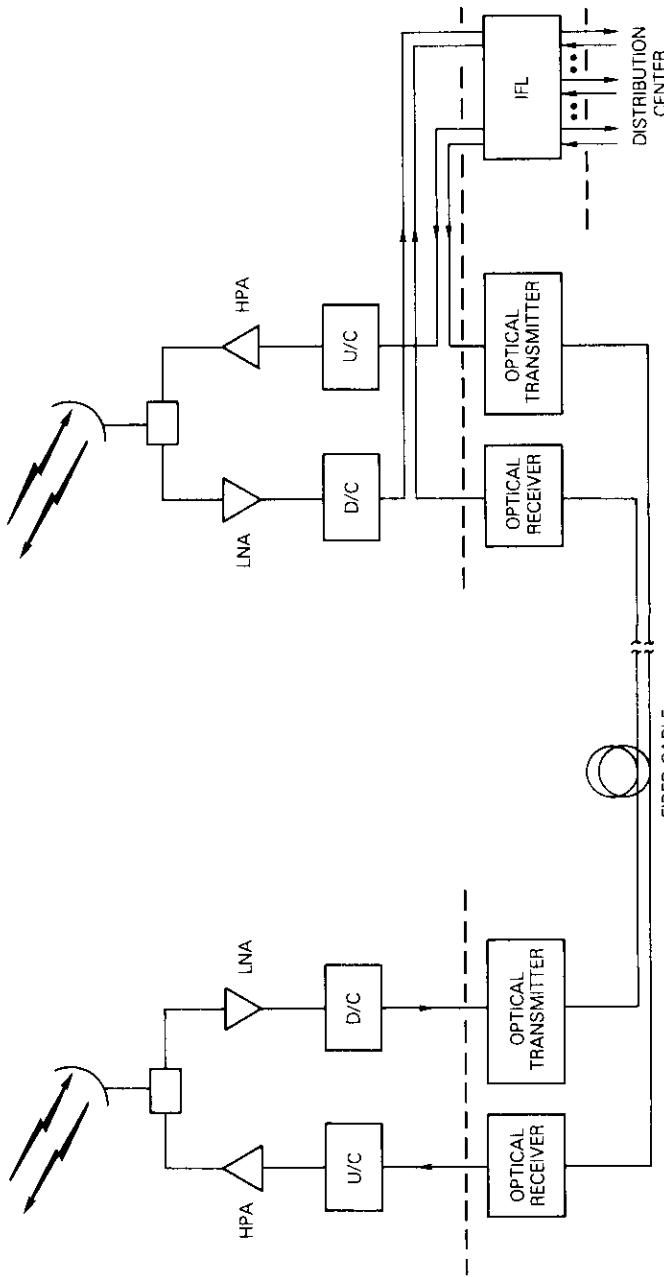
Figure 14. *Earth Station Diversity Link*

tors. While this design is considered conservative, it is projected to be much more economical than an equivalent coaxial cable system.

In the next few years, a more efficient longer wavelength single-mode technology should emerge. Optical submarine networks operating at the T-4 281-Mbit/s rate (or 4,000 circuits per fiber) are being planned for the 1988 to 1992 time frame. The initial capacities of 12,000 circuits can be expanded to 24,000 or 36,000 TASI circuits. Since this system will operate with only six fibers (three each way), the possibilities for increased capacity are obvious. By 1990, optical fiber cables will be used for long-distance communications between areas requiring high-capacity trunks. Since more advanced satellite systems will be available, continued use of alternate and diverse systems such as satellite and fiber optic cables, is expected to occur.

Optical intersatellite link

Future satellite communications systems, such as multi-mission geostationary platforms and satellite clusters or strings, will probably employ intersatellite links (ISL). Although microwave ISL systems [27] are currently being considered, optical technology presents a viable alternative.

Optical engineering feasibility models have already been successfully tested, both functionally and environmentally, using a solid-state Nd:YAG laser system [9] and a CO_2 gas laser [28]. Among the available laser sources for space application, the frequency doubled Nd:YAG laser (at 0.532 μm) system is attractive for several reasons. It is simple and efficient because it uses direct photodetection. Wideband traveling wave electro-optic modulators and high-gain, low-noise detectors are available; and GaAlAs laser diode-pumped lightweight YAG laser sources can be highly reliable. The CO_2 laser system at 10.6 μm requires heterodyne detection using cryogenically cooled detectors and added reference laser sources, and it is more complex than Nd:YAG laser systems. Further, the present 5000-hr lifetime of the gas laser is not sufficient for space qualification.

Figure 15 shows a basic block diagram of the duplex YAG laser optical ISL transreceiver using an internally frequency doubled Nd:YAG laser at 0.532 μm and a Nd:YAlO laser at 0.540 μm. A 4-Gbit/s ISL can be implemented using a 16-ary pulse-quadrature-modulated (PQM) optical transmitter. Traveling-wave type electro-optic LiTaO₃ modulators and optical delay circuits are used in the PQM transmitter. Direct photodetection or a 2-photodetector method for
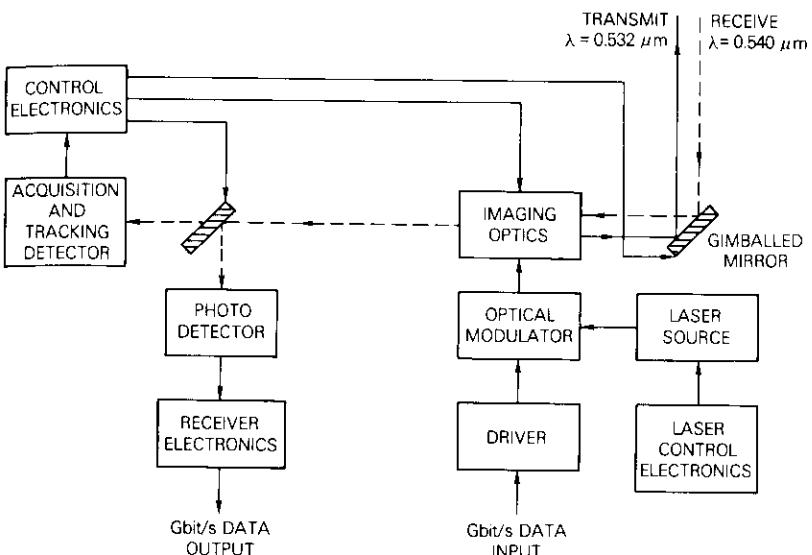
Figure 15. *Optical ISL System*

vertically and horizontally polarized components is possible. The optical demultiplexers are acousto-optic Bragg cells with PbMoO₄ crystals.

Figure 16 shows the optical ISL weight burden of the transceiver in Figure 15, based on the weight and power model of Reference 29. The weight of an individual subsystem can be expressed as follows:

$$W = A + BX^C$$

where X is either the optical antenna aperture size or laser power, and A, B, and C are constants determined from experimental design data [30]. The ISL link equation [9] was solved numerically for a GaAlAs-pumped YAG laser with an output of 300 mW and a pump efficiency of 1.2 percent. The required photodetector received optical power for a link BER of 10^{-6} was extrapolated from the M-ary optical channel analytical result [31] assuming an etalon filtered bandwidth of 0.1 nm, with solar conjunction. Improved gigabit logic components and photodetectors will reduce the overall system weight. The GaAlAs laser diode-pumped YAG laser has not yet been space qualified.
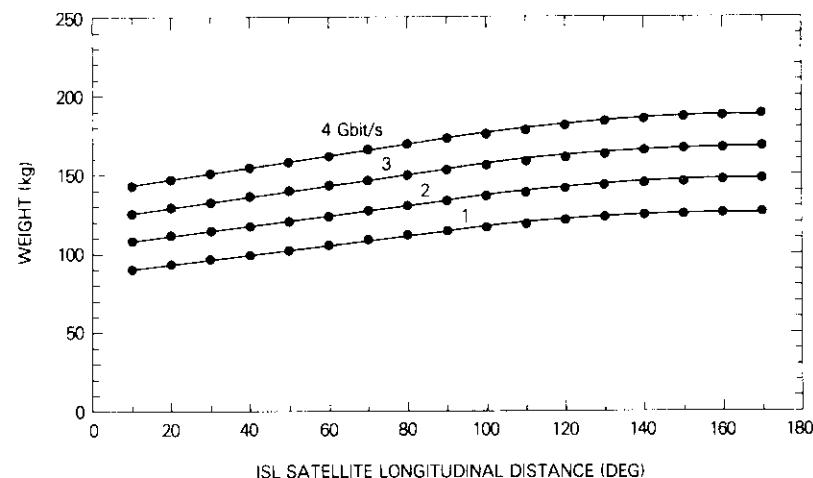
Figure 16. *Weight Burden of Optical ISL System*

Figure 16 shows that the weight of an optical ISL is only slightly dependent on ISL distance. This is in sharp contrast with the rapidly increasing microwave ISL weight burden and distance dependence, which can be attributed to the antenna size requirement [32]. The optical ISL is therefore attractive for high-capacity, long-range satellite configurations. The present technology of optical components indicates that a multi-gigabit rate optical ISL can be implemented by the 1990s.

Optical-microwave interaction technology

The rapidly developing optical-microwave interaction technology will be particularly advantageous for onboard satellite applications in which size, weight, power consumption, and reliability of the subsystems are important. Microwave semiconductor devices are being developed with built-in optical waveguides to couple optical signals directly into the device active regions. Typical applications include opto-GaAs FETs for optical mixing, optical injection locking of microwave signals, and optical detection. Opto-GaAs Gunn devices are also being developed for high-speed injection laser pulsed modulation. High-speed control of microwave signals at relatively high power levels with near-perfect isolation between the microwave circuit and input control signals can be achieved with this technology. In addition, small, flexible, lightweight optical fibers can be used to replace heavy,

bulky waveguides in many microwave communications and radar systems.

Opto-electronic switches with bandwidths from a few megahertz up to 4 GHz have been investigated, and the engineering feasibility of compact, high-performance, wideband switching has been demonstrated using hybrid electronic-optical circuit configurations [33]. Optical switching can be applied to advanced signal routing in high-capacity analog/digital channels and to onboard microwave switch matrices in future satellite-switched TDMA systems (Figure 17).

Many types of information processing functions can be performed using coherent optical signal processing technology. For instance, a very compact integrated optical RF spectrum analyzer consisting of injection laser diodes, a Ti:LiNbO₃ thin-film optical waveguide, geodesic lenses, surface acoustic wave transducers, and a detector array has been demonstrated [34]. Very high-speed electro-optic analog-to-digital converters in the 4- to 6-bit, 1- to 2-GHz range have been reported [35] using single-mode optical waveguides on electro-optic crystals ($LiTaO_3$). Holographic applications to image processing may also be possible.

Potential applications of optical-microwave interaction technology are also conceivable in space-based large antenna systems, in which lightweight, reliable, and movable feeds are important. Ultrahigh bit rate optical logic circuits that generate narrow clocking optical pulses by mode-locking of injection lasers is another possible application. Monolithic integration technology of optical, microwave, and electronic devices on semi-insulating semiconductor substrates has also been under investigation [36]. This technology may improve performance characteristics of existing communications subsystems and introduce new configurations in system design.

Conclusion

This paper has reviewed some aspects of current optical communications technology with emphasis on satellite communications applications. The wide bandwidth capability and EMI resistance of fiber optics make this technology particularly attractive for interfacility and diversity links and in long-haul ground communications. Analyses showing the applicability of fiber optics to these areas have been presented. Laser communications techniques have developed sufficiently to begin consideration of wide bandwidth, long-range optical
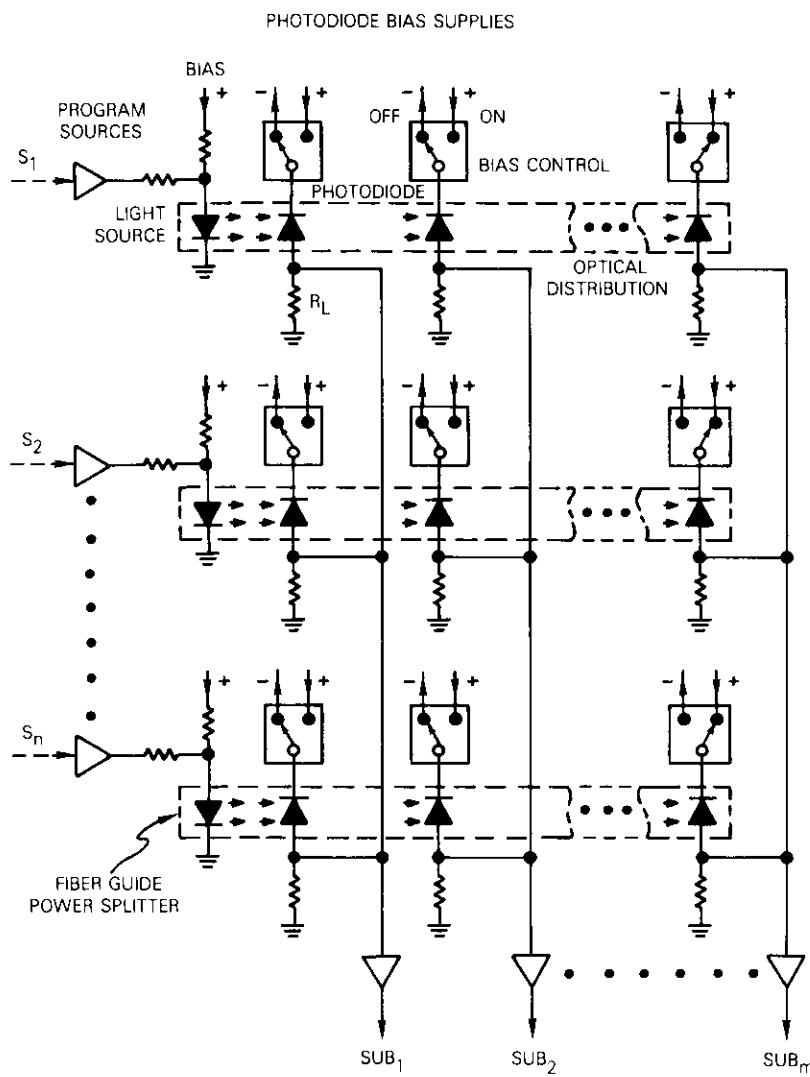
Figure 17. *Optically Switched TDMA System*

intersatellite links; developments in optical signal processing technology are introducing possible onboard satellite applications.

The future of optical communications technology is promising. Development of single-mode fibers and of reliable fiber optic components in the 1.3- to 1.55-μm wavelength range will increase the bandwidth and decrease the attenuation characteristics of optical fibers far beyond those of wire transmission lines. Future advances in laser technology may make optical intersatellite links more attractive than microwave links in the space environment, and optical signal processing has potential applications in switching and data handling.

References

[1] T. H. Maiman, "Stimulated Optical Radiation in Ruby," *Nature*, Vol. 6, 1960, p. 106.

[2] D. W. Berriman, "A Lens or Long Light Guide Using Convectively Distorted Thermal Gradients in Gases," *Bell Systems Technical Journal*, Vol. 43, 1964, pp. 1469–1475.

[3] F. P. Kapron, D. B. Keck, and R. D. Maurer, "Radiation Losses in Glass Optical Waveguides," *Applied Physics Letters*, Vol. 17, November 1970, pp. 423–425.

[4] T. Miya et al., "Ultra Low Loss Single Mode Fibers at 1.55 μm," *Review of the Electrical Communications Laboratories (NTT)*, Vol. 27, No. 7–8, July–August 1979, pp. 497–506.

[5] H. Kressel and J. K. Butler, *Semiconductor Lasers and Heterojunction LEDs*, New York: Academic Press, 1977.

[6] G. A. Antypas et al., "III-V Quaternary Alloys," 1972 Symposium Record on Gallium Arsenide and Related Compounds, pp. 48–54.

[7] D. Botez, "CW High-Power Single-Mode Lasers Using Constricted Double Heterostructures with a Large Optical Cavity (CDH-LOC)," International Conference on Integrated and Guided-Wave Optics, January 28–30, 1980, *Technical Digest*, pp. MC-2-1-2-4.

[8] M. Ettenberg and H. Kressel, "The Reliability of (AlGa)As CW Laser Diodes," *IEEE Journal of Quantum Electronics*, QE-16, No. 2, February 1980, pp. 186–196.

[9] M. Ross et al., "Space Optical Communications with the Nd:YAG Laser," *Proc. IEEE*, 66, No. 3, March 1978, pp. 319–344.

[10] H. Melchior, "Demodulation and Photodetection Techniques" *Laser Handbook*, F. T. Arecchi and E. O. Schulz-DeBois, Eds., Amsterdam, The Netherlands: North Holland Publishing Co., 1972.

[11] H. D. Law et al., "Sources and Photodetectors for Long Wavelength Optical Systems," International Conference on Communications, 1979, *Technical Digest*, pp. 53.6.1–53.6.6.

[12] W. G. French et al., "Optical Waveguides with Very Low Losses," *Bell Systems Technical Journal*, Vol. 53, 1974, pp. 951–954.

[13] J. Conradi, F. P. Kapron, and J. C. Dyment, "Fiber-Optical Transmission Between 0.8 and 1.4 μm," *IEEE Journal of Solid State Circuits*, SC-13, No. 1, February 1978, pp. 106–119.

[14] L. C. Cohen, C. Lin, and W. G. French, "Tailoring Zero Chromatic Dispersion into the 1.5–1.6 μm Low Loss Spectral Region of Single Mode Fibers," *Electronic Letters*, Vol. 15, No. 12, 1979, pp. 334–335.

[15] D. L. Myers and E. P. Partus, "Preform Fabrication and Fiber Drawing by Western Electric Product Engineering Control Center," *Bell Systems Technical Journal*, Vol. 57, No. 6, July–August 1978, pp. 1735–1744.

[16] T. C. Chu and A. R. McCormick, "Measurements of Loss Due to Offset, End Separation and Angular Misalignment in Graded Index Fibers Excited by an Incoherent Source," *Bell Systems Technical Journal*, Vol. 57, 1978, pp. 595–602.

[17] N. Suzuki et al., "A New Demountable Connector Developed for a Trial Optical Transmission System," International Conference on Integrated Optics and Optical Fiber Communications, 1977, *Technical Digest*, p. 351.

[18] T. Bowen and W. Schumacher, "Fiber Optic Cable Developments: Moving to Annul Coupler Mismatches," *Microwave Journal*, July 1979, pp. 55–59.

[19] M. A. Holzman, "A Detachable Connector for Multimode Graded Index Optical Waveguides," Deutsch Corporation Technical Reprint, No. 136, 1979.

[20] L. J. Coyne, "Coupling: Do it with a Star," *Laser Focus*, October 1979, pp. 60–66.

[21] B. S. Kawasaki and K. O. Hill, " Low Loss Access Coupler for Multimode Optical Fiber Distribution Networks," *Applied Optics*, Vol. 16, 1977, pp. 1794–1797.

[22] T. Ozeki and B. S. Kawasaki, "New Star Coupler Compatible with Single Multimode-Fiber Data Links," *Electronic Letters*, Vol. 12, 1976, pp. 151–152.

[23] S. K. Sheem and T. G. Giallorenzi, "Single Mode Fiber Multiterminal Star Directional Coupler," *Applied Physics Letters*, Vol. 35, 1979, pp. 131–133.

[24] R. J. McIntyre, "Multiplication Noise in Uniform Avalanche Diodes," *IEEE Transactions on Electron Devices*, ED-13, No. 1, January 1966, pp. 164–168.

[25] J. E. Midwinter, *Optical Fibers for Transmission*, New York: John Wiley and Sons, 1979.

[26] R. E. Patterson et al., "Linearization of Multichannel Analog Optical Transmitters by Quasi-Feed Forward Compensation Techniques," *IEEE Transactions on Communications*, COM-27, March 1979, pp. 582–588.

[27] G. R. Welti, "Intersatellite Link for Multiple-Access Telephony," EASCON 1978, *Conference Record*, pp. 432–440.

[28] J. H. McElroy et al., "CO$_2$ Laser Communication Systems for Near-Earth Space Applications," *Proc. IEEE*, 65, No. 2, February 1977, pp. 221–251.

[29] J. A. Maynard et al., "Multi-Gigabit Laser Communications for Satellite Cross-Links," 4th International Conference on Digital Satellite Communications, October 23–25, 1978, *Proc.*, pp. 155–159.

[30] L. S. Stokes, "Technology Forecasting for Space Communication," Final Report, NASA Contract NAS5-22057, June 1973.

[31] S. Karp et al., "Communication Theory for the Free-Space Optical Channel," *Proc. IEEE*, 58, No. 10, October 1970, pp. 1611–1616.

[32] G. Welti, COMSAT Laboratories, private communication.

[33] R. Ian MacDonald and E. H. Hara, "Switching with Photodiodes," *IEEE Journal of Quantum Electronics*, QE-16, No. 3, March 1980, pp. 289–295.

[34] M. K. Barnoski et al., "Integrated-Optic Spectrum Analyzer," *IEEE Transactions on Circuits and Systems*, CAS-26, No. 12, December 1979, pp. 1113–1124.

[35] F. J. Leonberger et al., "Design and Development of High Speed Electro-Optic A/D Converter," *IEEE Transactions on Circuits and Systems*, CAS-26, No. 12, December 1979, pp. 1125–1131.

[36] S. Margalit et al., "Monolithic Integration of Optical and Electronic Devices on Semi-insulating GaAs Substrates," 5th Topical Meeting on Integrated and Guided-Wave Optics, January 28–30, 1980, *Technical Digest*, pp. WC2.1–2.4.

Appendix

The complete optical receiver can be modeled as shown in Figure A-1. It consists of the primary electron generator, which produces a Poisson distribution of electrons at an average rate that is proportional to the incident optical power; an internal gain, M, that is a random variable with a mean value determined by the bias voltage on the detector; an ideal amplifier, A; and an output filter, $h_o(t)$. The diode has a junction capacitance, C_j, and a bias resistance, R_b. In parallel with the bias resistance is the amplifier input capacitance, C_a. Thermal noise, i_t, is associated with the equivalent resistance, $R_e = R_a \parallel R_b$.

The average rate of primary electron generation for the photodetector is proportional to the incident optical power:

$$\rho = \left(\frac{\eta}{h\nu} P_o + \frac{I_d}{q} \right) \qquad \text{(A-1)}$$
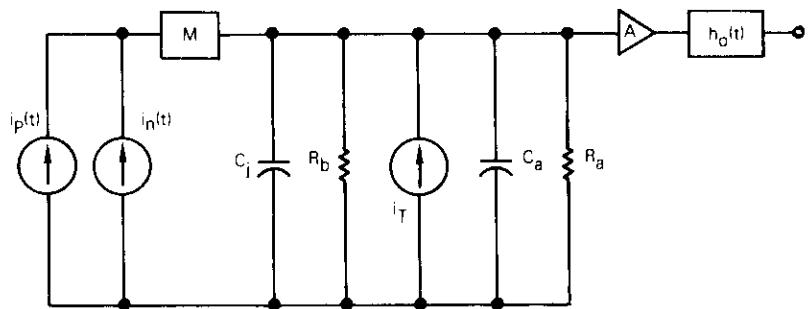
Figure A-1. *Optical Receiver Model*

where
$\eta =$ photodetector quantum efficiency
$h =$ Planck's constant
$v =$ optical frequency
$P_o =$ incident optical power
$I_d =$ dark current generated thermally in the photodetector (independent of signal level)
$q =$ charge on an electron.

For a Poisson distribution, the probability that K electrons are emitted in an interval $\Delta\tau$ is given by

$$P(K, \Delta\tau) = \frac{(\rho\Delta\tau)^K e^{-\rho\Delta\tau}}{K!} \quad . \tag{A-2}$$

If $h(t)$ is the impulse response of the combination of C_i, R_a, and $h_o(t)$, then the current produced at the output at time t because of K electrons produced in an interval $\Delta\tau$ centered on τ is

$$i(t, \tau) = M[qKh(t - \tau)] \tag{A-3}$$

where M is the instantaneous multiplication factor. The detected current, which is the average current value at t, is then

$$\overline{i(t, \tau)} = q\overline{M}\left[\sum_{K=0}^{\infty} KP(K, \Delta\tau)\right] h(t - \tau)$$

which, from equations (A-3), (A-2), and (A-1), is

$$\overline{i(t, \tau)} = \overline{M}\left[\frac{\eta q}{hv} P_o(\tau) + I_d\right] h(t - \tau) \Delta\tau \quad . \tag{A-4}$$

Therefore, the actual current resulting from the continuous optical power input is

$$i(t) = \overline{M}\int_{-\infty}^{\infty}\left[\frac{\eta q}{hv} P_o(\tau) + I_d\right] h(t - \tau)\, d\tau \quad . \tag{A-5}$$

Since only the signal portion of this expression is significant, terms involving I_d will be ignored in calculating signal power.

In general, $P_o(t)$ assumes the following form

$$P_o(t) = P_o[m_T + m(\tau)] \tag{A-6}$$

where P_o is the incident optical power without modulation, $m(t)$ represents the information signal, and $m_T + m(\tau)$ is in the range 0 to 2. If the optical source is modulated by a PSK carrier, then $m(t)$ equals $M_o \cos[\omega t + \theta(t)]$ and m_T equals 1. If the optical source is digitally modulated, then $m(t)$ assumes values of $2 - m_T$ and $-m_T$, where m_T is the detector decision threshold level which yields equal probability of error for both the logical one and zero. The use of equation (A-6) in equation (A-5) yields the mean square current value:

$$\overline{i^2} = \left(\frac{\eta q\overline{M}P_o}{hv}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t - \tau)h(t - \mu)$$

$$\cdot \left\{m_T^2 + m_T[\overline{m(\tau) + m(\mu)}] + \overline{m(\tau)\, m(\mu)}\right\} d\tau\, d\mu$$

$$= \left(\frac{\eta q\overline{M}P_o}{hv}\right)^2 \left[m_T H^2(0) + \int_{-\infty}^{\infty} S_m(t)\, |H(f)|^2\, df\right] \quad . \tag{A-7}$$

Only the second term carries information energy. The signal power in a 1Ω resistor is therefore

$$\overline{i_s^2} = \left(\frac{\eta q\overline{M}P_o}{hv}\right)^2 \int_{-\infty}^{\infty} S_m(f)\, |H(f)|^2\, df \quad . \tag{A-8}$$

The method used to evaluate the mean square noise current is similar to the one used to derive the signal power. Equation (A-3) yields

$$\overline{i^2(t, \tau)} = \overline{q^2M^2K^2H^2(t - \tau)}$$

$$= q^2\overline{M^2}h^2(t - \tau)\cdot[(\rho\Delta\tau)^2 + \rho\Delta\tau] \quad . \tag{A-9}$$

The increment of noise power is then obtained by substituting equation (A-4) into equation (A-9):

$$\overline{i_n^2(t, \tau)} = \overline{i^2(t, \tau)} - \overline{i(t, \tau)^2} \tag{A-10}$$

$$= q^2[(\overline{M^2} - \overline{M^2})\rho^2\Delta\tau^2 + \overline{M^2}\rho\Delta\tau]h^2(t - \tau)$$

The first term is proportional to $\Delta\tau^2$ so that as $\Delta\tau$ approaches $d\tau$, it becomes negligible. Integrating equation (A-10) then yields the average noise power at t:

$$i_n^2(t) = q^2\overline{M^2} \int_{-\infty}^{\infty} \left[\frac{\eta P_o(\tau)}{h\nu} + \frac{I_d}{q}\right] h^2(t - \tau)\, d\tau \quad . \tag{A-11}$$

The thermal noise power, due to the noise current i_T in Figure A-1, is assumed to be white and Gaussian with a spectral density given by

$$N_o = \frac{2kT}{R_e} \qquad -\infty \le f \le \infty \tag{A-12}$$

where $R_e = R_a R_b / R_a + R_b$

k = Boltzmann's constant

T = resistor temperature.

The thermal noise power at the output is then

$$i_T^2 = \int_{-\infty}^{\infty} N_o |H(f)| \, df \quad . \tag{A-13}$$

Combining equations (A-8), (A-11), and (A-13) then yields the average C/N:

$$\frac{C}{N} = \frac{i_s^2}{i_n^2 + i_T^2} \tag{A-14}$$

$$= \frac{\left(\dfrac{\eta q P_o^2}{h\nu}\right) \int_{-\infty}^{\infty} S_m(f) |H(f)|^2 \, df}{\left\{q\left[\dfrac{\eta q P_o}{h\nu}(m_T + m) + I_d\right]F(\overline{M}) + \dfrac{2kT}{\overline{M^2}R_e}\right\} \int_{-\infty}^{\infty} |H(f)|^2 \, df}$$

where $F(\overline{M})$ equals the excess noise factor $= \overline{M^2}/\overline{M}^2$, and $H(f)$ must be sufficiently large to pass the modulating signal $m(t)$. Therefore, it is approximated that

$$\frac{\int_{-\infty}^{\infty} S_m(f) |H(f)|^2 \, df}{\int_{-\infty}^{\infty} |H(f)|^2 \, df} \cong \frac{\int_{-\infty}^{\infty} S_m(f) \, df}{2B_n} \tag{A-15}$$

where B_n is the double-sided noise bandwidth of $H(f)$. With $m_T + \overline{m} = 1$, the general form of average C/N is

$$\frac{C}{N} = \frac{\left(\dfrac{\eta q P_o^2}{n\nu}\right) \int_{-\infty}^{\infty} S_m(f) \, df}{\left\{2q\left[\dfrac{\eta q P_o}{h\nu} + I_d\right]F(\overline{M}) + \dfrac{4kT}{\overline{M^2}R_e}\right\} B_n} \quad . \tag{A-16}$$

Equation (A-16) yields an average C/N in terms of the modulation spectrum, $S_m(t)$. For PSK modulation of a carrier with a larger bit period than the carrier cycle, this equation can be used to accurately determine the theoretical BER. However, for digital modulation, because of the dependence of noise on signal levels, the signal-to-noise ratio (S/N) for a logical one will, in general, differ from that for a logical zero unless m_T is chosen to equalize the two. This can be illustrated through the following analysis.

Figure A-2 shows the current corresponding to a logical one followed by a logical zero with a decision threshold at $R_o P_o m_T$. The noise current during a logical one is $n_1(t)$ with an rms value of σ_1. The noise current during a logical
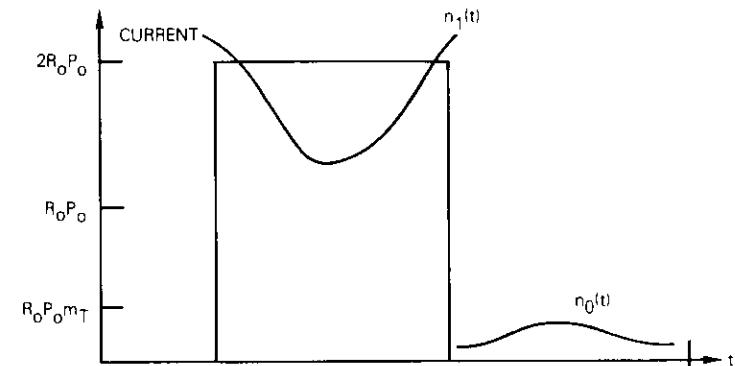
Figure A-2. *Receiver Current Response For a Logical One Followed by a Logical Zero*

zero is $n_0(t)$ with an rms value of σ_o. The probability of error if a one has been transmitted is the probability that $n_1(t)$ descends below $R_o P_o m_T$; the probability of error if a zero has been transmitted is the probability that $n_0(t)$ exceeds $R_o P_o m_T$. For equally probable signal messages assuming the shot noise is Gaussian,

$$P(\epsilon) = \frac{1}{2}[P(n_1 < -2R_oP_o + R_oP_om_T \mid 1)$$

$$+ P(n_o > R_oP_om_T \mid 0)] \tag{A-17}$$

$$= \frac{1}{2}\left\{\frac{1}{\sqrt{2\pi}} \int_{R_oP_o(2-m_T)}^{\infty} e^{-x^2/2}\, dx\right.$$

$$+ \left.\frac{1}{\sqrt{2\pi}} \int_{R_oP_om_T/\sigma_o}^{\infty} e^{-x^2/2}\, dx\right\} \quad.$$

For equal probabilities of error,

$$\frac{R_oP_o(2 - m_T)}{\sigma_1} = \frac{R_oP_om_T}{\sigma_o} \tag{A-18}$$

or

$$m_T = \frac{2\sigma_o}{\sigma_o + \sigma_1} \quad. \tag{A-19}$$

Choosing this value for the decision threshold equalizes the S/Ns for logical one and zero. In this case, the S/N, given by equation (A-16) will equal the individual bit S/N. The optical power required to achieve a given S/N (and thus a given BER) can be obtained from equation (A-16) by noting that

$$\int_{-\infty}^{\infty} S_m(f)\, df = \overline{m^2(t)} \tag{A-20}$$

$m_1 = 2 - m_T$ and $m_o = -m_T$. Therefore,

$$\overline{m^2(t)} = \frac{(2 - m_T)^2 + (m_T)^2}{2} = \frac{2(\sigma_o^2 + \sigma_1^2)}{(\sigma_o + \sigma_1)^2} \quad. \tag{A-21}$$

The rms noise power, σ_o^2, occurs when a logical zero is transmitted. This is evidenced by equations (A-11) and (A-13) in which $P_o(\tau) = 0$ is given by

$$\sigma_o^2 = \left[2qI_dF(\overline{M}) + \frac{4kT}{\overline{M}^2R_e}\right]B_n \quad. \tag{A-22}$$

When a logical one is transmitted, $P_o(\tau) = 2P_o$ and the rms noise power is

$$\sigma_1^2 = \left[2q(2R_oP_o + I_d)F(\overline{M}) + \frac{4kT}{\overline{M}^2R_e}\right]B_n \tag{A-23}$$

$$\sigma_1^2 = 4qR_oP_oF(\overline{M})B_n + \sigma_o^2 \quad. \tag{A-24}$$

Then equation (A-16) yields

$$\frac{S}{N} = \frac{4(R_oP_o)^2}{(\sigma_o + \sigma_1)^2} \quad. \tag{A-25}$$

After algebraic manipulation, equations (A-22) to (A-25) yield

$$P_o = \frac{h\nu}{2\eta}\sqrt{\frac{S}{N}}\left[F(\overline{M})R\sqrt{\frac{S}{N}} + \frac{2\sigma_o}{q}\right] \tag{A-26}$$

where the bit rate, R, is approximately $2B_n$.

For the case of analog modulation, e.g., when the optical source is modulated by a QPSK carrier,

$$P_o(t) = P_o[1 + \cos(\omega_c t + \theta)] \tag{A-27}$$

where θ assumes the values of $n\pi/2$, $n = 0, 1, 2$, and 3. In this case, $m(t)$ equals $\cos(\omega_c t + \theta)$, and

$$\int_{-\infty}^{\infty} S_m(f)\, df = \frac{1}{2}$$

so that the C/N becomes

$$\frac{C}{N} = \frac{R_o^2P_o^2/2}{2qR_oP_oF(\overline{M})B_n + \sigma_o^2} \quad. \tag{A-28}$$

When equation (A-22) is substituted into (A-28) and the resulting equation is solved for P_o, the optical power required to achieve a given C/N is

$$P_o = \frac{h\nu}{\eta}F(\overline{M})R\frac{C}{N}\left\{1 + \sqrt{1 + 2\left[\frac{qI_dF(\overline{M}) + (2kT/\overline{M}^2R_e)}{q^2F^2(\overline{M})R(C/N)}\right]}\right\}$$

assuming $B_n \approx R/2$.

Randall W. Kreutel received B.S.E.E. and M.S.E.E. degrees from Northeastern University and the D.Sc. degree from the George Washington University. Prior to joining COMSAT in 1966, he was a Research Engineer with the Antenna and Microwave Lab of Sylvania Electronic Systems. At COMSAT Laboratories, he was Manager of the Antenna Department from 1968 to 1977, and a Senior Staff Scientist from 1977 to 1979; he is presently Director of the Optical Communications Laboratory. He is a member of Commissions A and B of URSI, OSA, Eta Kappa Nu, and Sigma Xi, a Senior Member of IEEE, and an Associate Fellow of AIAA.

Albert E. Williams received a B.E. in electrical engineering from the University of Western Australia, Perth, in 1962, and a Ph.D. from University College, London, England, in 1966. He is currently a Senior Staff Member of the Optical Communications Laboratory at COMSAT Laboratories, developing advanced optical technology applicable to satellite communications systems. He was a joint recipient of the Institute (London) Sylvanus P. Thompson Premium award in 1966, and is currently a member of the IEEE (London) and a Senior Member of the IEE.
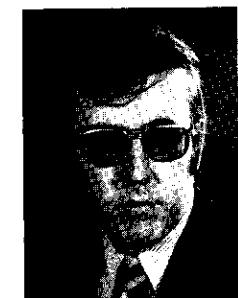
Kenneth H. Greene received a B.S. from Polytechnic Institute of Brooklyn and an M.S. and Ph.D. from the University of Connecticut. He joined COMSAT in 1969 as a member of the Systems Simulation Laboratory participating in the simulation of the INTELSAT IV satellite system and of the domestic satellite system. As a member of the Modulation Techniques Department, he was responsible for the field trials and operation of the shipboard terminal for the MARAD Experiment. His principal activities have involved modem development for data collection systems. He is presently Manager of the Optical Communications Department.

Young Soo Lee received a B.S. (1963) and an M.S. (1967) in electronic engineering from Seoul National University, and M.S.E.E. and Electrical Engineer degrees from M.I.T. (1969). Since September 1969, he has been a member of the technical staff at COMSAT Laboratories specializing in the development of microwave circuits and subsystems for communications satellites. He has contributed to the development of an onboard regenerative repeater for future digital satellite communications systems, and is presently concerned with fiber and integrated optical communications technology. He is a Senior Member of IEEE and a member of the Optical Society of America and the National Space Institute.

Roy Gene Autry received B.S. degrees in physics and electrical engineering from M.I.T. and an M.S. in physics from the University of New Mexico. Prior to joining COMSAT's Optical Communications Laboratory in 1979, he worked for six years at the Air Force Weapons Laboratory and for three years with McDonnell Aircraft Company. He is a member of APS.

Optical intersatellite links

A. K. SINHA

(Manuscript received May 27, 1980)

Abstract

In this paper, major system components and performance criteria for laser intersatellite links are identified, and the transmission characteristics are modeled analytically. Five types of laser transmitters are considered: HeNe, GaAs, CO_2, Nd:YAG, and frequency doubled Nd:YAG. Under the assumptions of a Poisson distribution for the photon arrival rate and quantum statistics of black body radiation for the background solar noise input, a parametric study of the link is conducted using a computer program developed for this investigation. Antenna size requirements and achievable bit rates are determined as functions of the mutual separation between the satellites, and the bit error rate for a binary pulse gate transmission system is calculated under various operating conditions. Finally, inherent system tradeoffs and the comparative viability of laser communications for high-capacity intersatellite links in a global digital satellite communication network are assessed.

Introduction

Interest in optical communication has grown considerably in recent years. Studies of various optical modulation/demodulation techniques and their impact on communications channel performance have been reported [1]–[3]. Specific applications have also been analyzed, *e.g.*, NASA's proposed TDRSS system [4] involving a CO_2 laser and the

U.S. Air Force's Project 405B involving an Nd:YAG laser [5].

This paper presents preliminary results of a study of laser intersatellite links (ISL) for geostationary communications satellites and related parametric tradeoff analysis. Although some discussion of other laser systems, such as CO_2, HeNe, and GaAs, is included for comparison, this paper primarily considers digital data transmission using an Nd:YAG laser. Both the fundamental and the frequency doubled (FD) modes are presented for the Nd:YAG laser. Since all of the above-mentioned laser systems are relatively well known in terms of characteristics such as available power, size, beamwidth, and reliability, attention is focused upon communications performance and systems tradeoffs for ISL applications.

The basic system parameters of a laser ISL system are identified and their theoretical interrelations explained to introduce the notations and the system model involved. Numerical results obtained by using a computer program developed on the basis of this model are presented. Systems comparisons are given and potential applications indicated.

Optical ISL system model

General

Figure 1 is a block diagram of an optical ISL system. The optical wavelength carrier from the laser source is modulated by the input signal (e.g., digitized voice or data) in the optical modulator and transmitted to the receiving satellite, where the modulated carrier is detected. Sources of noise include solar radiation, thermal noise, quantum noise, and dark current noise. Since the noise due to background solar radiation predominates, only this noise will be considered [6]-[8].

In its simplest form, an optical detector consists of a photoemissive surface (cathode) of a suitable material generating (primary) electrons whose number and velocity depend on the frequency and intensity of the incident radiation as well as on the cathode type. The incoming radiation is focused upon the cathode by an optical device (mirror or lens) which will henceforth be termed an "antenna" with a characteristic aperture diameter. Direct photoemissive detection can be conveniently used for digital signals in combination with photomultiplication and gating. In particular, this type of detection provides a convenient interface with the microwave components when a slow wave structure tube (TWTA) using a photoemissive cathode is employed [7].
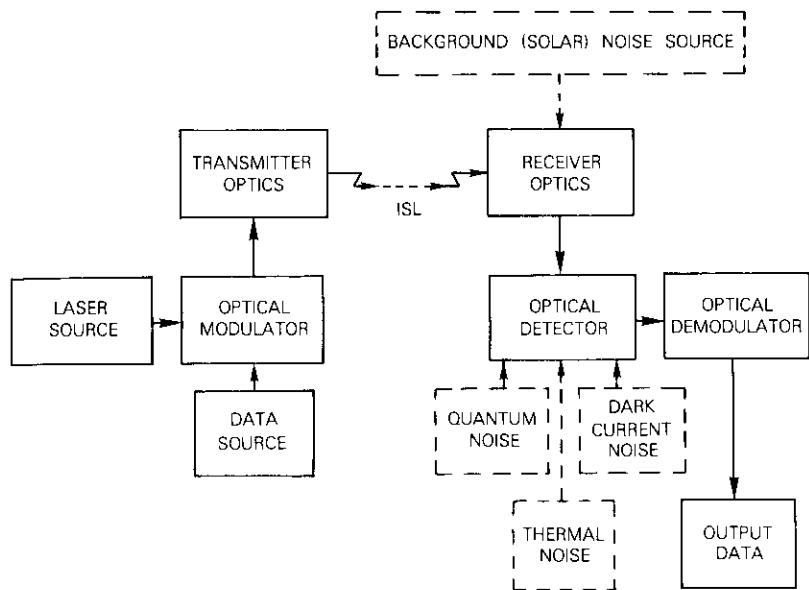
Figure 1. *General ISL System Model*

In a photoemissive detector, no electrons are emitted from the cathode unless the frequency of the incident radiation exceeds a threshold value characteristic of the cathode. Above threshold, the number of emitted electrons is determined by the quantum efficiency (number of electrons emitted per photon or quantum of radiation) and is also proportional to the intensity (power) of the radiation (number of photons incident per second). An optical filter applied over the aperture of the antenna accepts only radiation of the appropriate wavelength range and limits the received noise to the solar radiation within the filter passband.

At optical frequencies, the black body radiation of the earth is negligible in comparison to that of the sun (6,000 K). The peak of the spectral density curve of the solar radiation lies around 0.5-μm wavelength, which is close to the wavelength emitted by the FD Nd:YAG laser. In other words, the background solar radiation noise is unavoidable, whereas thermal noise is negligibly small.

Dark current noise represents electrons emitted from the cathode surface operating at a temperature T in the absence of radiation (signal

or solar noise) impinging on it. The magnitude of the dark current is given in terms of Richardson's law, and can easily be evaluated for typical detector systems. For instance, for an Sb-K-Na-Cs cathode optical detector combining a high quantum efficiency with a low dark current [7], the dark current density is typically about 10^{-15} A cm^{-2} at T = 300 K, which corresponds to about $10^{-15}/(1.6 \times 10^{-19}) \approx 6.25 \times 10^3$ electrons cm^{-2} s^{-1}. On the other hand, the number of photo-electrons due to the incident solar radiation at a wavelength of 0.53 μm is about 10^8 electrons s^{-1}. Thus, only for cathode areas larger than about 10^4 cm^2 can the dark current become increasingly significant. Since the value of the cathode area is several orders of magnitude smaller, the dark current contribution can be neglected in comparison to the noise due to the background solar radiation.

It has been shown [8] that for typical detectors the dark current noise contribution is smaller than the thermal noise contribution for (dark) current densities less than 0.13×10^{-12} A cm^{-2} per unit of noise bandwidth. This corresponds to an electron emission rate of about 10^{15} cm^{-2} s^{-1} due to dark current for a noise bandwidth of 1 GHz, which is many orders of magnitude larger than the dark current value for typical detectors. Since the thermal noise is much less than the solar radiation noise, the predominance of the latter is reconfirmed.

While the preceding discussion has focused upon direct photo-emissive detectors, most of the arguments also apply to other detector types such as photoconductive or photovoltaic detectors. The predom-inance of the solar radiation noise diminishes as the signal wavelength increases from the optical to the infrared region, with thermal noise ultimately becoming dominant in the microwave range. In certain cases, depending on the operating wavelength, temperature, and cathode type, etc., the dark current noise may become appreciable and should be appropriately considered, unless the cathode tempera-ture is maintained at a sufficiently low level by means of cryogenic or radiation cooling. However, detectors requiring such measures are not suited to ISL applications.

The quantitative results presented in this paper apply to FD Nd:YAG lasers; extrapolations for longer wavelength lasers such as HeNe, GaAs, Nd:YAG, and CO_2 are less accurate. In particular, for the CO_2 laser (10.6 μm) the characteristic noise bandwidth and path loss are much smaller, largely compensating for the reduced antenna gain. However, since the dark current of the photoconductive detectors used in this case is large and detector cooling is required, the CO_2 laser is not well suited to ISL applications despite the large power

available (~500 mW). HeNe lasers, while exhibiting high reliability, have limited output power (~2 mW) and hence can be considered only for very close range ISLs. GaAs lasers suffer from low output power (~5 mW) as well as from beam collimation and divergence difficulties. Of the Nd:YAG and FD Nd:YAG laser systems, the latter is obviously advantageous since higher antenna gain values lead to superior per-formance.

The signal-to-noise ratio estimates in the following analysis pertain to the input of the optical detector; hence, further degradations due to other significant noise sources must be included to obtain the corresponding estimates for the detector output or for the end-to-end link performance. In particular, the noise contribution of the microwave interface and components must be properly considered to determine the overall link budget of the ISL system. In view of the related degradations, the bit error rate (BER) and energy per bit to noise power density (E_b/N_o) values presented here may be regarded as rather optimistic upper bounds. However, these results constitute a useful starting point for further analysis of specific system configurations including different types of up-link and down-link transmissions and modulation and multiple-access schemes.

Finally, pulse gate binary modulation (PGBM) was chosen because of its inherent simplicity for operational implementation as well as for analytical treatment. Alternative modulation methods could also be considered. For instance, pulse quarternary modulation (PQM) could provide higher system efficiency. Such modulation methods deserve separate consideration and have not been included in this paper. It is sufficient to note that the advantage of binary transmission is the absence of degradations arising from the nonlinearity of the system components.

Expressions for evaluating system performance in terms of the basic system parameters will be derived in the following.

Signal power

The carrier-to-noise power density ratio (C/N_o) for an ISL link can be written as the ratio of the received signal power to the noise power density due to solar radiation within the waveband of interest. The received signal power can be conveniently written as

$$P_s = P_t G_t \left(\frac{4\pi R}{\lambda}\right)^{-2} G_r L^{-1} \tag{1}$$

where P_t = transmit laser power
         G_t = transmit antenna gain
         G_r = receive antenna gain
         λ = wavelength
         R = distance
         L = insertion loss of the optical passband filter.

From the geometry of an ISL between two geostationary satellites

$$R = 2(R_o + H_o) \sin\left(\frac{\phi}{2}\right) \tag{2}$$

where ϕ = angular separation between the two satellites
         R_o = radius of the earth (6,370 km)
         H_o = geosynchronous altitude (35,800 km).

With the well-known relationship

$$G_x = \eta_x \frac{4\pi}{\lambda^2} A_x = \eta_x \left(\frac{\pi D_x}{\lambda}\right)^2 \tag{3}$$

where x = t or r for transmit or receive aperture, respectively
         η = efficiency
         A = aperture area
         D = aperture diameter

the received signal power can be written as

$$P_s = P_t \eta_t \eta_r \left(\frac{\pi D_t D_r}{4R\lambda}\right)^2 \quad . \tag{4a}$$

Assuming equal diameter and efficiency for the transmit and receive apertures leads to

$$P_s = P_t \left(\frac{\pi \eta D^2}{4R\lambda}\right)^2 \quad . \tag{4b}$$

Channel noise

Solar noise power density can be approximated by a black body

radiating at temperature T = 6,000 K. In terms of Planck's quantum statistical distribution for black body radiation [6], the noise power per unit frequency bandwidth in the vicinity of frequency ν is [6]

$$P_n = \frac{h\nu}{e^{h\nu/kT} - 1} \tag{5}$$

where h = Planck's constant (6.63×10^{-34} J-s)
         k = Boltzmann's constant (1.38×10^{-23} J/K)
         ν = frequency (Hz).

Figure 2 gives a plot of P_n as a function of ν in the range of interest for the present paper.
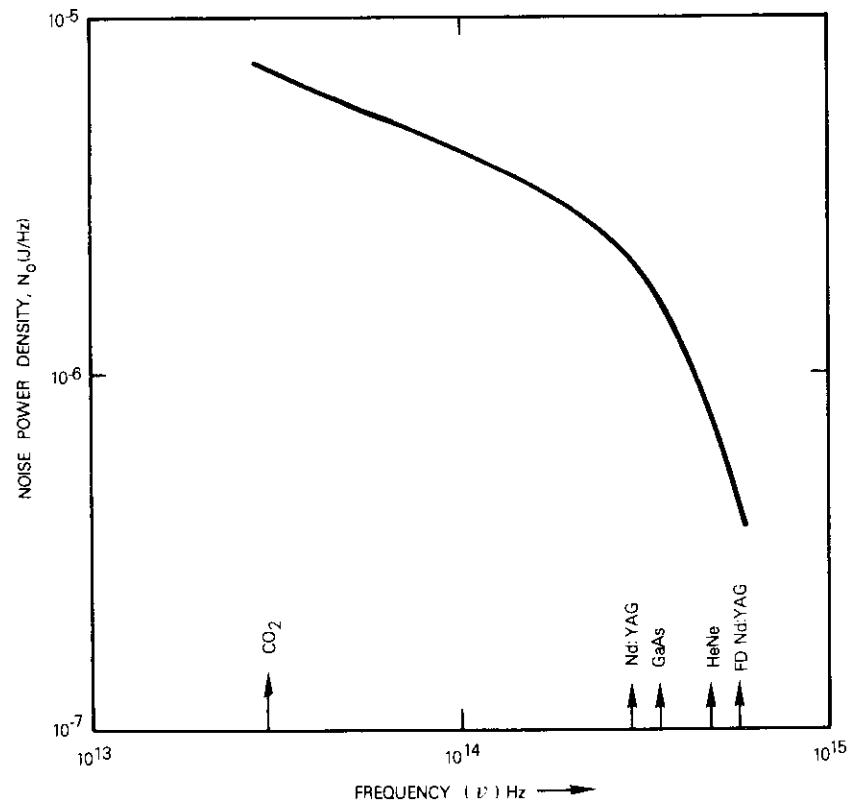

Figure 2. *Spectral Density of the Solar Radiation (Noise)*

If it is assumed that other noise contributions (internal or dark current noise, thermal noise, quantum noise, etc.) are negligible, C/N_o in dB at the input of the optical detector is

$$\frac{C}{N_o} = 10 \log_{10} \left(\frac{P_s}{P_n} \right) \text{dB} \quad . \tag{6}$$

Consequently, E_b/N_o in a PGBM system is

$$\frac{E_b}{N_o} = 10 \log_{10} \left(\frac{P_s \tau}{P_n} \right) \tag{7}$$

where

$$\tau = \frac{1}{R_b} \tag{8}$$

is the bit duration for a bit rate of R_b.

Bit error rate

The BER is given by the sum

$$P_e = p\bar{P}^{(1)} + (1 - p)\bar{P}^{(2)} \tag{9}$$

where $p = a\ priori$ probability for the signal to be present
$\bar{P}^{(1)}$ = probability that the receive current is below the threshold value for bit detection when a signal is present within the bit duration
$\bar{P}^{(2)}$ = probability that the receive current is at or above the threshold value when no signal is present during the bit duration.

Clearly, p can be taken as the fraction of time at which the ISL transmission takes place; that is, $100p$ represents the percent duty cycle of the ISL system. Further, assuming that the current is related to the mean number of photoelectrons detected within the bit duration, and that the photon arrival rate obeys a Poisson distribution [6], [7] yields

$$\bar{P}^{(i)} = \sum_{l=m_i}^{m_i'} P^{(i)}(l) \quad , \qquad i = 1, 2 \tag{10a}$$

$$P^{(i)}(l) = (n_i)^l \frac{e^{-n_i}}{l!} \quad , \qquad i = 1, 2 \tag{10b}$$

where $P^{(i)}(l)$ = probability of l photoelectron counts
n_1, n_2 = mean numbers of photoelectrons involved in the two cases distinguished above [i.e., corresponding to $\bar{P}^{(1)}$ and $\bar{P}^{(2)}$, respectively]
m_i, m_i' = lower and upper limits, respectively, of the actual count of individual photoelectrons detected in each of the two cases above ($i = 1, 2$).

Assuming a threshold n_t for bit detection results in

$$m_1 = 0, m_1' = n_t - 1, m_2 = n_t, m_2' = \infty \tag{11}$$

The mean numbers n_1 and n_2 can easily be determined by noting that the photoelectrons per second corresponding to the input power due to the laser signal and the solar noise radiation are, respectively,

$$n_s = \eta_q \frac{P_s}{h\nu} \tag{12a}$$

$$n_n = \eta_q \frac{P_n}{h\nu} (\Delta B)_n \tag{12b}$$

where $(\Delta B)_n$ is the detector bandwidth (assumed to be wider than the signal bandwidth), and η_q is the quantum efficiency of the detector.

To further characterize the impact of the detector performance on the error statistics, it is convenient to introduce the probability of pulse detection due to the signal, p':

$$p' = \frac{n_s'}{n_s' + n_n'} \tag{13}$$

where n_s' denotes the maximum number of signal photoelectrons

received when a pulse is desired and n_n' the minimum number of photoelectrons when no pulse is desired. Then,

$$n_1 = (n_s p' + n_n)\tau \qquad (14a)$$

$$n_2 = [n_s(1 - p') + n_n]\tau \quad . \qquad (14b)$$

In particular, for $p' = 1$ (unit probability of pulse detection), $n_1 = (n_s + n_n)\tau$ and $n_2 = n_n\tau$, as expected. The quantity

$$r_e = \frac{n_s'}{n_n'} = \frac{p'}{1 - p'} \qquad (15)$$

denotes the maximum achievable ratio of the number of signal photoelectrons received when a pulse is desired to that received when no pulse is desired. This ratio is called the modulation extinction ratio [2].

The value of $(\Delta B)_n$ depends on the efficacy of the filter employed to reduce or eliminate noise radiation at wavelengths adequately separated from the characteristic wavelength of the laser. For instance, for the FD Nd:YAG laser, filters permitting radiation only within a wavelength range of 1Å are available. This corresponds to $(\Delta B)_n = 10^{11}$ Hz. For convenience, this value will be assumed for the other types of lasers as well, except that a value $(\Delta B)_n = 10^9$ Hz will be assumed for the CO_2 laser, since the detector in this case is a photoconductive device at IF or DC with a much narrower bandwidth.

The optimum threshold

The threshold value, n_t, of the number of photoelectrons for optimal bit detection is easily obtained by requiring that the probability for receiving n_t or more photoelectrons when the signal is present must exceed the probability of receiving an equal number of photoelectrons when the signal is absent [9]:

$$pP^{(1)}(j) \geq (1 - p) P^{(2)}(j); j \geq n_t \quad . \qquad (16)$$

This condition yields the following expression for the optimum threshold corresponding to the equality in equation (16):

$$n_t = \frac{n_s + \ln[(1 - p)/p]}{\ln[1 + (n_s/n_n)]} \quad . \qquad (17)$$

For a given value of p, since

$$\ln \frac{1 - p}{p} = c_0 \text{ (constant)} \qquad (18)$$

the minimum number of signal photoelectrons for achieving the optimum threshold, $\hat{n}_s$, is given by the condition

$$\left.\frac{\partial n_t}{\partial n_s}\right|_{n_s = \hat{n}_s} = 0 \qquad (19)$$

leading to the relation

$$(\hat{n}_s + n_n) \ln \left(1 + \frac{\hat{n}_s}{n_n}\right) = \hat{n}_s + c_0 \quad . \qquad (20)$$

This transcendental equation can be solved graphically as illustrated in Figure 3. Plots of the expressions on the left- and right-hand sides of equation (20) as functions of $\hat{n}_n$ are illustrated in Figure 3a, with n_n and p (and hence c_0) as parameters. A plot of the right-hand expression of equation (20) is obviously a straight line with unit gradient and vertical intercept equal to c_0. A plot of the left-hand expression is a higher order curve passing through the origin. The point of intersection of the two plots corresponds to the minimum value, $\hat{n}_s$, to achieve optimum threshold detection.

Figure 3b presents linear plots for $p = 0.05$, 0.10, 0.25, and 0.50 corresponding to the laser transmission duty cycles of 5, 10, 25, and 50 percent, respectively. Nonlinear plots of the left-hand expression corresponding to $n_n = 0.5$, 1.0, 2.0, etc., are also shown. For a given combination of n_n and p, the abscissa of the point of intersection between the corresponding curves clearly yields the minimum value, $\hat{n}_s$. A plot of this minimum value as a function of n_n with p as a parameter (and vice-versa) is provided as the inset in Figure 3b. From this result and equation (12), the minimum power, $\hat{P}_s$, to achieve optimum threshold detection and the corresponding e.i.r.p. can easily be determined.

$$\text{e.i.r.p.} = 10 \log_{10} (\hat{P}_s G_t) \text{ dBW} \quad . \qquad (21)$$

In the inset of Figure 3b, the solid lines represent the variation of

$$y_1 = (n_s + n_n) \ln \left(1 + \frac{n_s}{n_n}\right)$$

$$y_2 = n_s + c_0$$
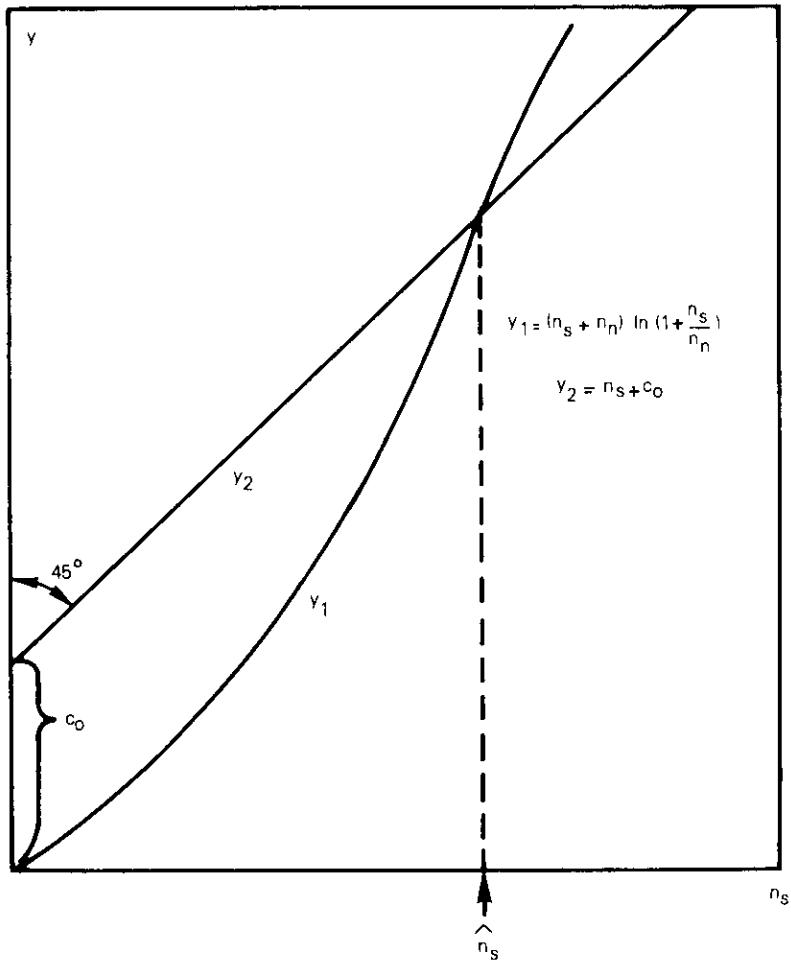
Figure 3a. *Graphic Solution of Transcendental Equation (20)*

$\hat{n}_s$ with n_n (bottom scale) for specified values of p, while the dashed lines represent the variation of $\hat{n}_s$ with p (top scale) for specified values of n_n. At a given frequency, the number of photoelectrons is directly proportional to the corresponding incident power [equation (12)]. Therefore, it is clear from the figure that the minimum signal power to achieve the optimum threshold detection increases with the background noise level, but decreases with increasing duty cycle, as
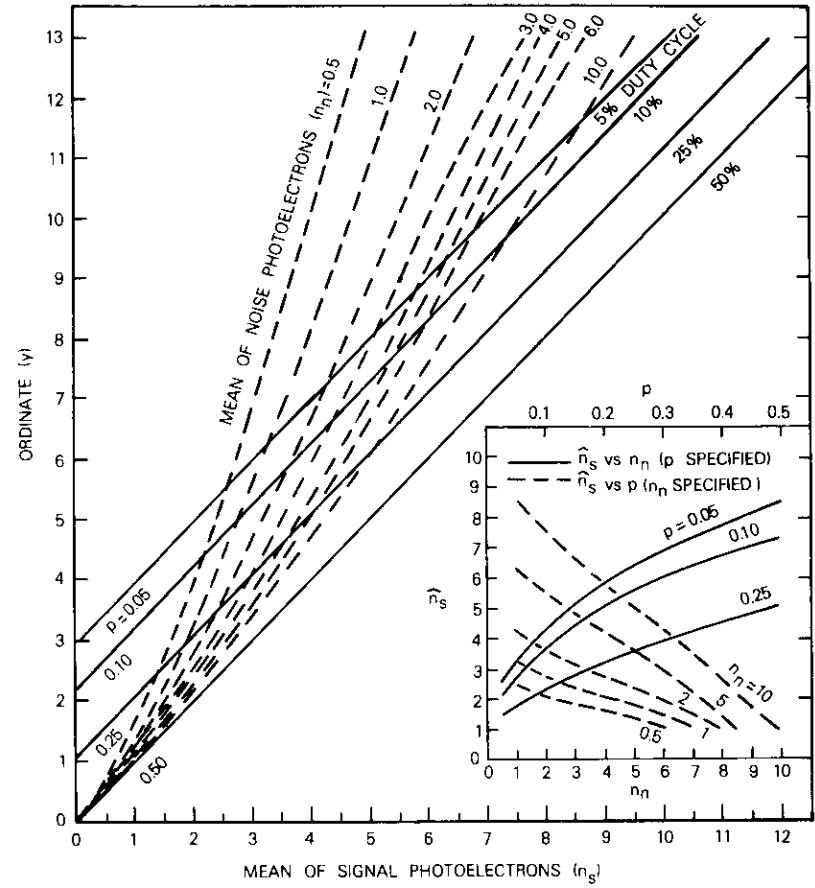
Figure 3b. *Plot of the Transcendental Equation for Determining the Optimum Signal Level, $\hat{n}_s$*

expected. Furthermore, as the noise level decreases, the variation of the optimum threshold minimum signal power level as a function of duty cycle becomes relatively less pronounced. Similarly, as the duty cycle increases, the variation of the optimum signal power level as a function of the noise power level becomes less pronounced. In other words, the optimal signal power is more sensitive to duty cycle at high noise levels and to noise level for low values of duty cycle. Therefore,

in the PGBM mode, channel performance as measured by the bit error rate can generally be categorized as either duty cycle limited (or, more generally, power limited) or noise limited in the extreme case of a very low duty cycle or a very high noise level, respectively.

Numerical results and discussion

Numerical results obtained by using a simple computer program developed to implement the model described in the preceding section will now be presented. The results for the FD Nd:YAG laser ISL system will be presented in detail; results for other laser systems will be indicated mainly for comparison. Unless otherwise specified, a 1-Gbit/s bit rate will be assumed with unity probability of pulse detection. The assumed parameter values include those given in Table 1.

TABLE 1. ASSUMED PARAMETER VALUES

	CO$_2$	GaAs	HeNe	Nd:YAG	FD Nd:YAG
Wavelength, λ (μm)	10.6	0.9	0.63	1.06	0.53
Average Transmit Power, P_t (mW)	500	5	2	100	100
Quantum Efficiency, η_q	4×10^{-1}	4×10^{-3}	5×10^{-2}	10^{-3}	2×10^{-2}
Noise Bandwidth, $(\Delta B)_n$ (GHz)	1	100	100	100	100

Power levels correspond to conservative estimates of available power output for the laser systems under consideration. The quantum efficiency factors are based on the data in Reference [7]. Finally, it is assumed that the efficiencies of the transmitting and receiving optics are equal ($\eta_t = \eta_r = 0.7$), and that there is a 3-dB insertion loss for the passband filter.

Bit error rate as a function of E_b/N_o for Nd:YAG and FD Nd:YAG is given in Figure 4. It is clear that the lower frequency for Nd:YAG implies an increase of over 10 dB in E_b/N_o to achieve the same BER value.

Figures 5 and 6 present plots of BER as a function of aperture diameter and intersatellite angular separation, respectively, for FD Nd:YAG operating at a 10-percent duty cycle; other parameters are as specified in the respective figures. It can be seen that there is a sharp drop in the BER as the aperture diameter is increased above a certain minimum value for a specified angular separation, or as the
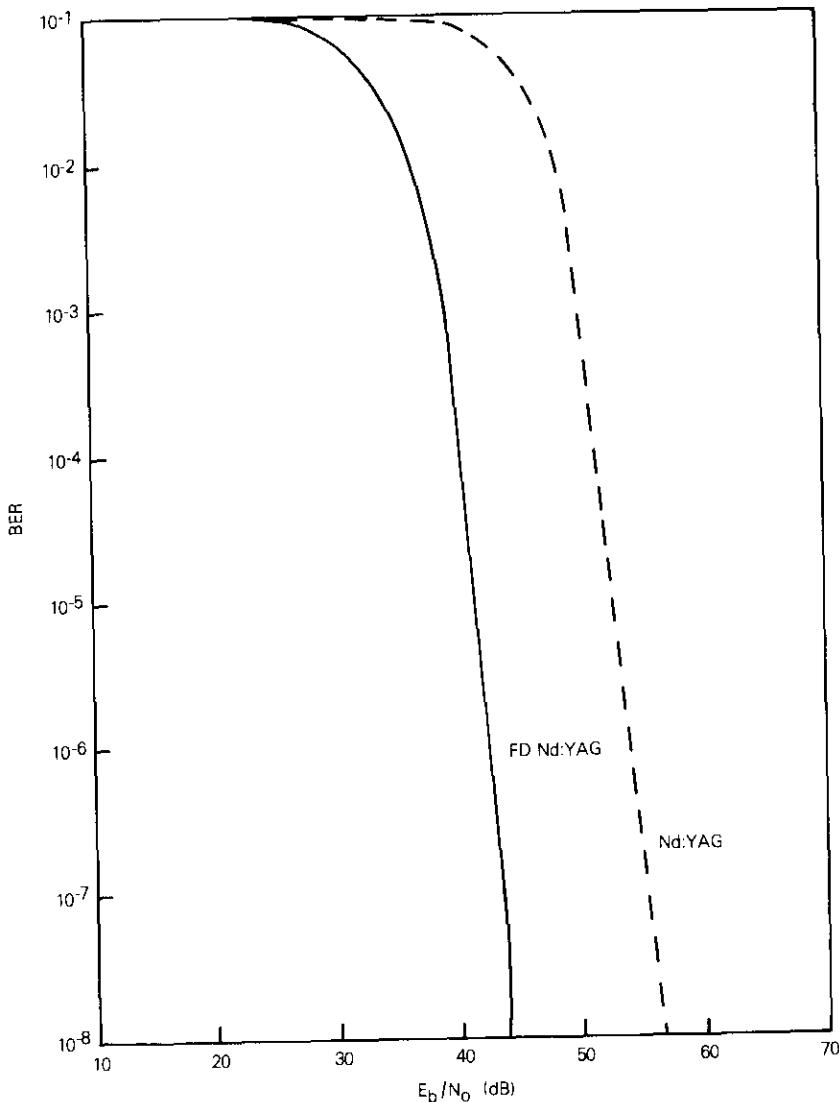
Figure 4. *BER versus E_b/N_o ($p = 0.1$)*

angular separation is reduced below a certain maximum value for a specified aperture diameter. For instance, a BER of 10^{-6} or less is
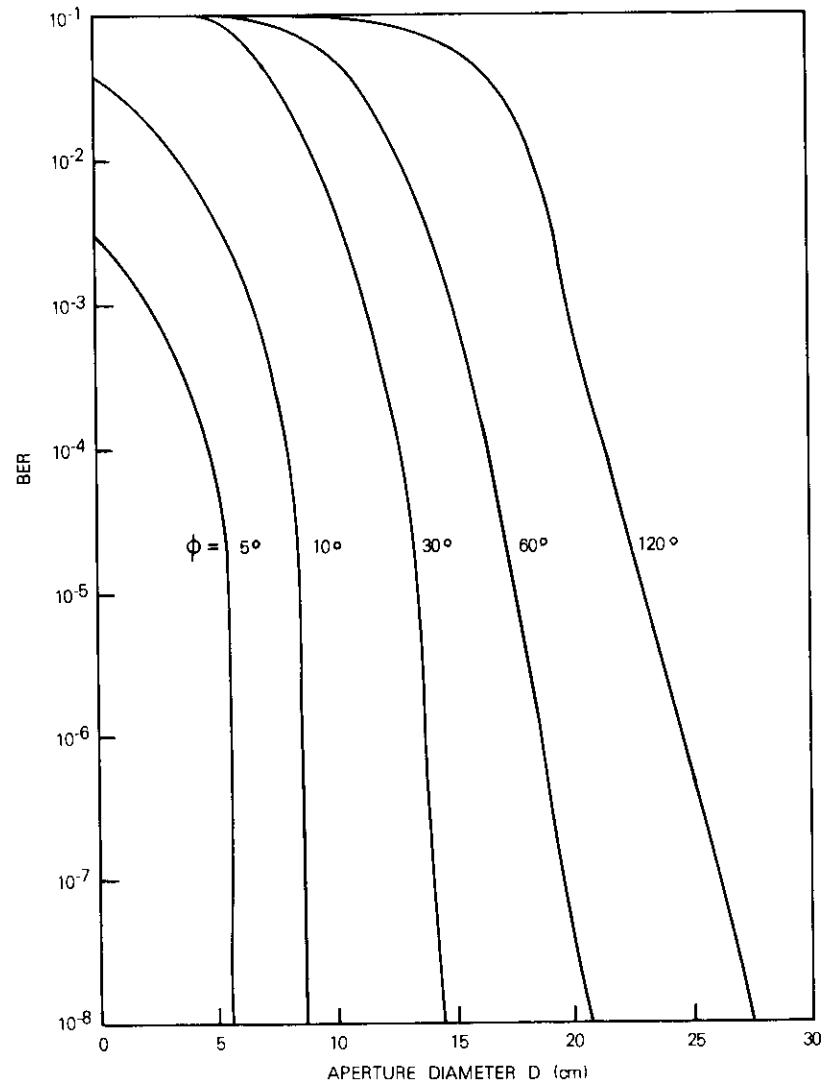
Figure 5. *BER* versus *Aperture Diameter (FD Nd:YAG,*
R_b = 1 Gbit/s, p = 0.1)

Figure 6. *BER* versus *Angular Separation (FD Nd:YAG,*
D = 10 cm, p = 0.1)

achievable at 5° intersatellite separation with an aperture diameter of
about 5.5 cm, and at 120° separation with an aperture diameter of
about 25 cm. Similarly, from Figure 6, for an aperture diameter of 10

cm, BER decreases only slowly with decreasing angular separation up
to about 35°, and then drops rapidly below 10^{-6} for angular separations
of 16° or less.

Link performance as a function of duty cycle is indicated in Figure 7, which gives plots of BER vs p for 60-Mbit/s data rates transmitted
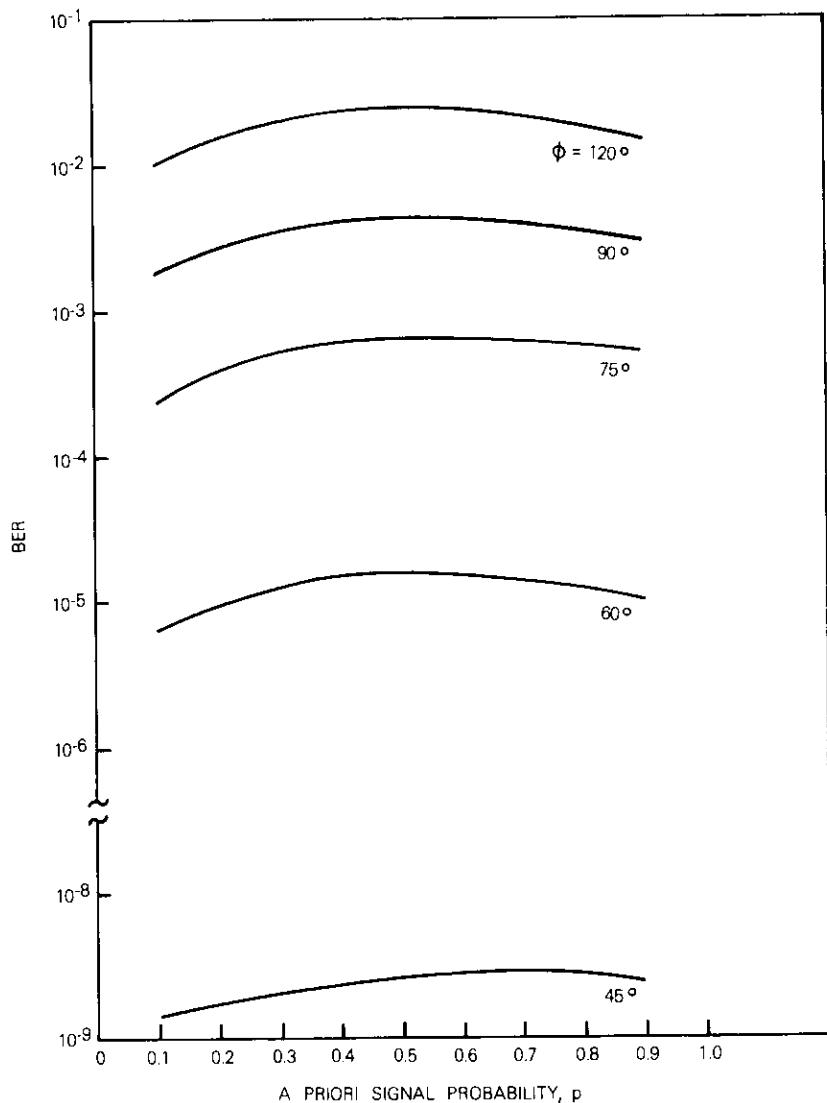
Figure 7. *BER versus a Priori Signal Probability (FD Nd:YAG, D = 10 cm, R_b = 60 Mbit/s)*

with an FD Nd:YAG laser using an aperture diameter of 10 cm. Results for several values of angular separation (ϕ) are shown. It is apparent that channel performance is relatively insensitive to changes of duty cycle. Nevertheless, an improvement in performance occurs at low or high values of the duty cycle, while a slight degradation in performance (increase in BER) is encountered in the intermediate range. This behavior can be understood by examining the variation of the threshold number of photoelectrons (n_t) as a function of the duty cycle (p) for specified values of the other parameters, as shown by the solid curves in Figure 8.

Figure 8 also includes the values of the signal photoelectron number (n_s) and the noise photoelectrons (n_n) within the bit period for comparison (dashed and dashed-and-dotted curves, respectively). Note that n_s is independent of p but depends on ϕ, whereas n_n is independent of both ϕ and p. The value of n_t decreases with increasing p, as expected [equation (17)]. Obviously, for specified values of n_s and n_n, the bit error contribution due to the error probability $\bar{P}^{(1)}$ related to failure in signal detection [see equations (10) and (11)] decreases with decreasing n_t. On the other hand, the contribution due to the error probability $\bar{P}^{(2)}$ resulting from false detection increases with decreasing n_t. It can thus be concluded from equation (9), that in the low value range of p, the term $\bar{P}^{(2)}$ dominates the remaining terms $p[\bar{P}^{(1)} - \bar{P}^{(2)}]$, causing the BER value to increase with increasing p. However, in the high value range of p, an inversion in the dominance of these terms causes the BER value to decrease with increasing p. This leads to a peak in the BER value in the intermediate value range of p.

In other words, for better channel performance, it is preferable to operate with a sufficiently low duty cycle, corresponding to a high value of the optimum detection threshold, because the background noise is less likely to be falsely detected as a signal. Alternatively, a sufficiently high duty cycle yields better performance by virtue of a larger *a priori* probability for the signal to be present, thereby reducing the chance of failure in signal detection when the channel is active, or of false detection when the channel is inactive. An intermediate value of the duty cycle yields the worst performance and is therefore not an optimal choice.

For a comparative evaluation of the laser systems mentioned earlier, Figures 9 and 10 provide values of E_b/N_o as a function of angular separation (Figure 9) and aperture diameter (Figure 10) (assuming 10-percent duty cycle and 1-Gbit/s data rate) with the value $E_b/N_o = 40$
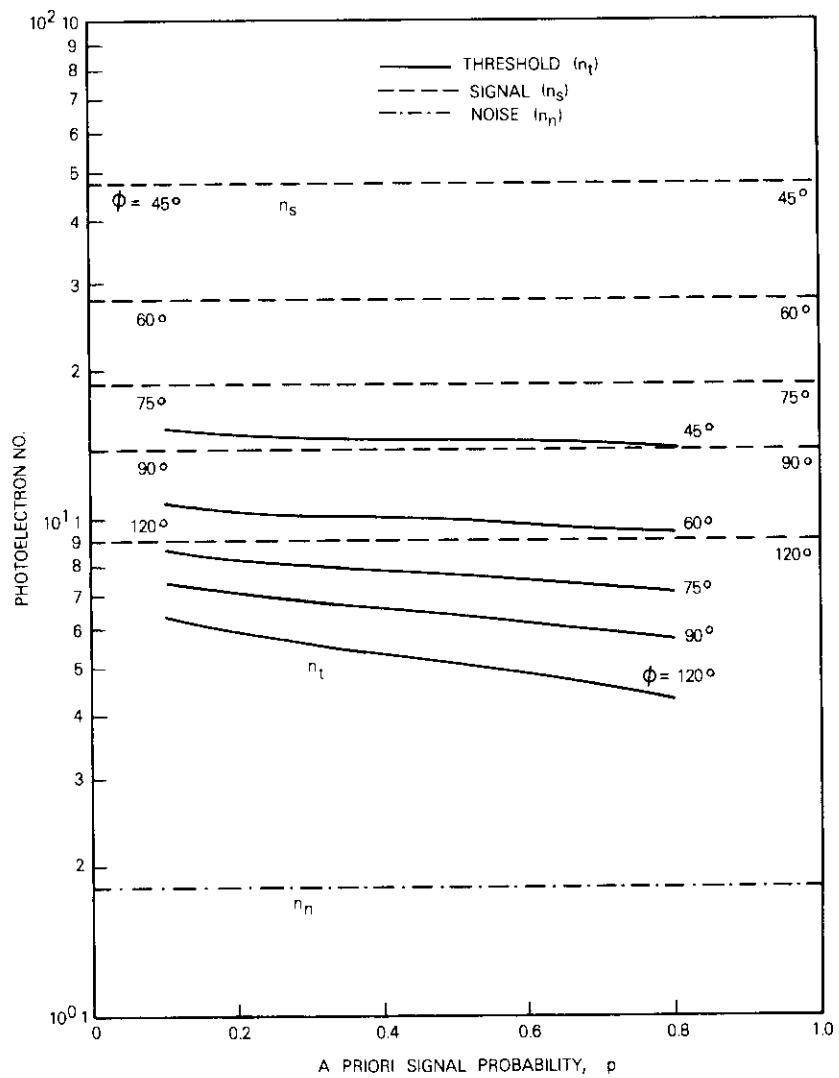
Figure 8. *Photoelectron Numbers per Bit Period* versus *a Priori Signal Probability (FD Nd:YAG, D = 10 cm, R_b = 60 Mbit/s)*

dB marked as a nominal threshold value corresponding to BER 10^{-6} (see Figure 4).

It may generally be observed from the results obtained that the FD
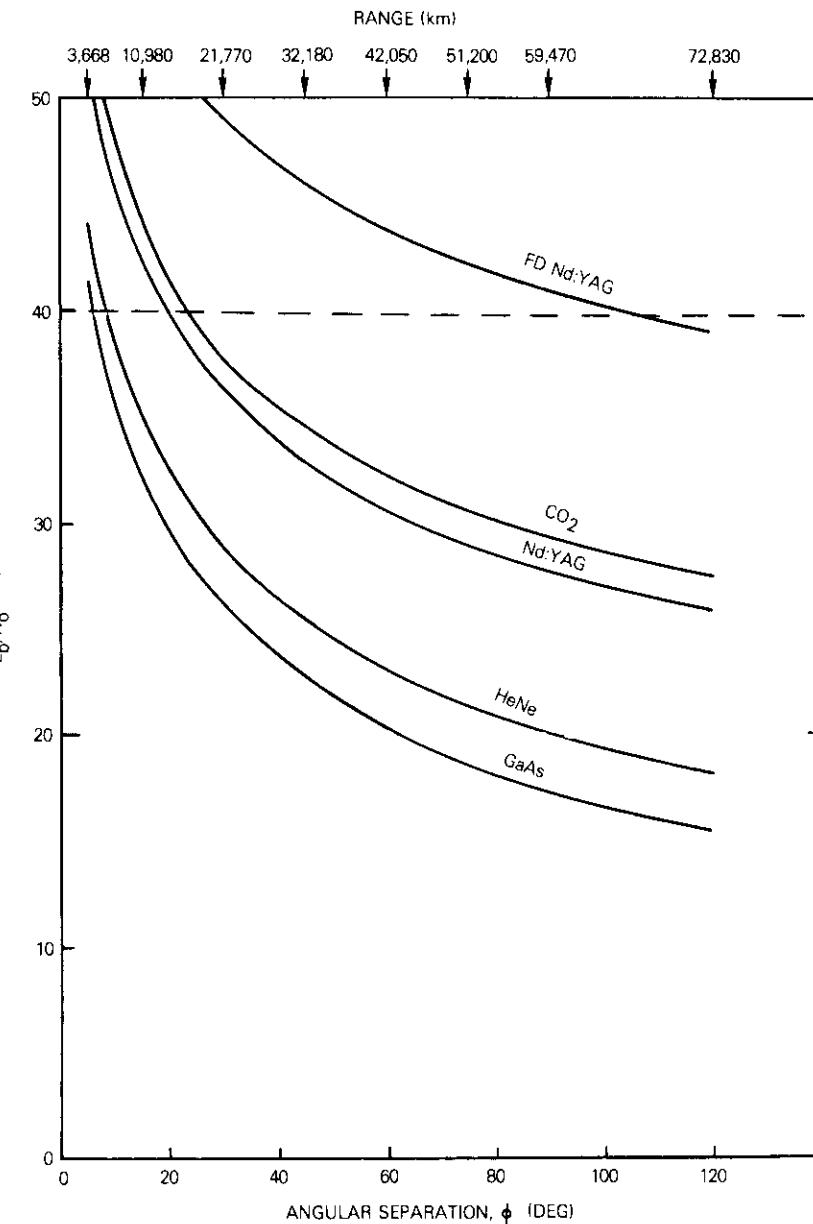
Figure 9. E_b/N_0 versus *Angular Separation (D = 20 cm, R_b = 1 Gbit/s, p = 0.1)*
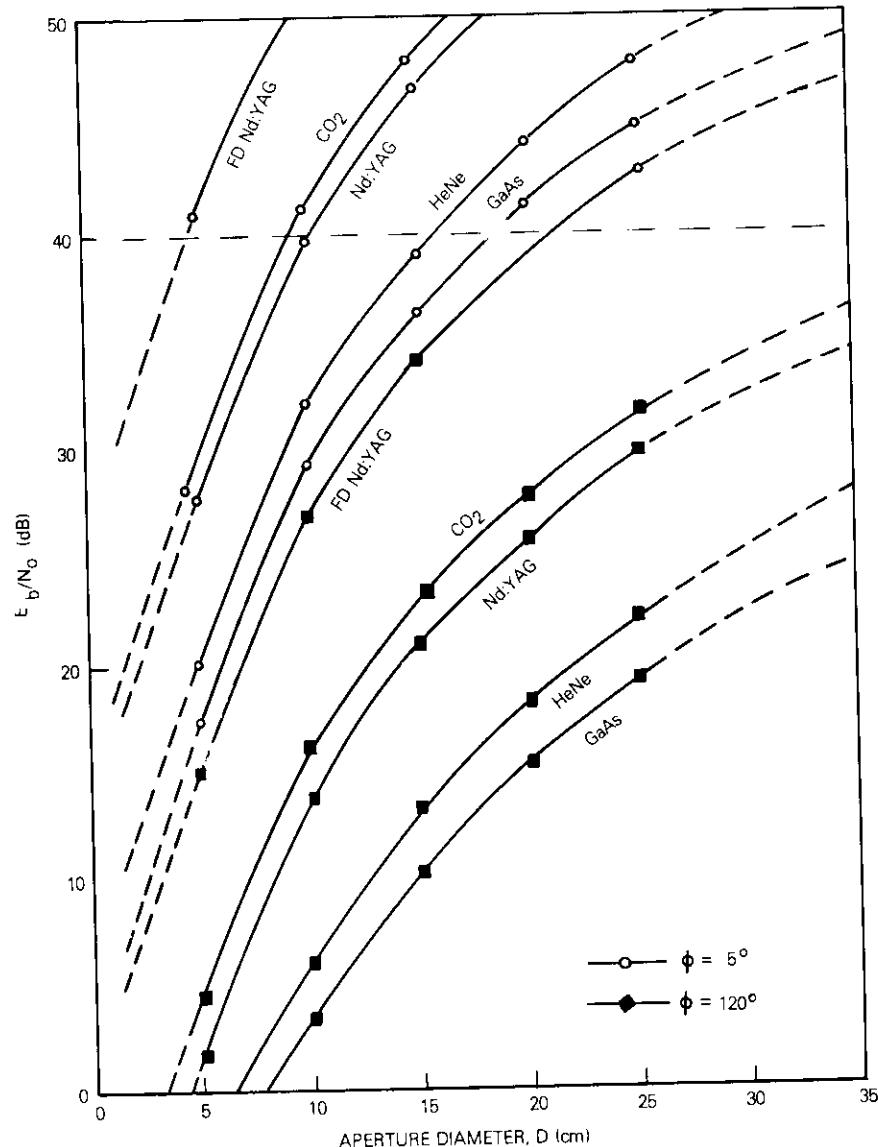
Figure 10. E_b/N_0 versus *Aperture Diameter* $(R_b = 1 \ Gbit/s, \ p = 0.1)$

Nd:YAG laser system employing moderate size ($\gtrsim$20-cm) antennas
would exhibit excellent performance characteristics for optical wave-

length ISLs over all conceivable values of angular separations and over
a wide range of operating conditions. The Nd:YAG, CO_2, HeNe, and
GaAs laser systems, which are applicable only to small angular
separations, provide marginal performance unless the transmit power
or the aperture size is drastically increased, with repercussions on
satellite weight, complexity, and cost.

The superior performance of the FD Nd:YAG laser system for ISL
application arises from the high antenna gain and reduced solar
background radiation noise, which more than compensate for the path
loss. The solar radiation in the wavelength range of 5,300 Å might be
turned into an advantage by using a sun-pumped FD Nd:YAG laser
transmitter. Alternatively, a semiconductor laser diode or LED could
be used advantageously instead of, or in addition to (as a backup during
a solar eclipse), the sun-pumped FD Nd:YAG laser [10]–[13]. Optical
fiber could perhaps be used for channeling the filtered solar energy for
optical pumping of the laser.

A factor in favor of the CO_2 laser is its comparatively high power
conversion efficiency ($\approx$10 percent), which provides higher power
output from the laser transmitter for a given amount of primary (DC)
power. In other words, for a given mass allocation for the power of the
ISL subsystem, higher e.i.r.p. can be realized by the CO_2 laser compared
to the Nd:YAG laser (efficiency $\approx$ 0.5 percent). Since satellite mass
is ultimately one of the major determinants of space segment cost and
system economy, the CO_2 laser system may prove to be more cost
effective for certain ISL applications. In general, apart from the
communications performance, factors such as the overall spacecraft
mass-power tradeoff, lifetime and reliability, and system economy
must be considered in greater detail to assess the relative viability of
different lasers for specific ISL systems.

Applications

Intersatellite links have been called the missing link of international
satellite communications. The placement of satellites over the Atlantic,
Indian, and Pacific Ocean regions, primarily to provide transoceanic
communications, has intimately impacted the overall system design,
performance, and cost of the space and earth segments. While this
inherent feature of the system is unlikely to change in the near future,
introduction of ISL capability may fundamentally alter the system
configuration and its design philosophy. Most probably, ISL imple-

mentation will contribute to the evolution of future generations of satellites by improving system performance and especially by reducing earth segment costs through elimination of double hop connections as well as multiple earth station antennas and related equipment.

ISL technology, once sufficiently advanced and perfected, will play increasingly significant roles in international communications as well as in a variety of other innovative applications. For instance, a novel approach to the international satellite system is placement of the satellites in the geosynchronous arc mainly over continental land masses instead of over the ocean regions. In this case, transoceanic international communications would be primarily provided through the ISL, while the satellites would directly provide domestic, regional, and international communications within continental regions. Placing the satellites over continental regions could simplify antenna beam coverage patterns and significantly reduce the number of onboard antennas required. Furthermore, the e.i.r.p. of the satellite could be enhanced for a given primary power, while retaining high values of earth station antenna elevation angle. Both effects (fewer onboard antennas and effective enhancement of the e.i.r.p.) have favorable implications for the mass requirements of the satellite's communications and power subsystems, and hence its fabrication and launch costs (as well as its lifetime) on one hand, and for earth station G/T requirements (and hence cost) on the other. Multiple satellites (a "satellite cluster") could be deployed over a given continental region, with ISLs among them as necessary.

Increasing demand for domestic and regional communications satellites by many countries in the world poses challenging problems related to the growth of the international system as well as the optimal utilization of the geostationary orbit. ISLs will alleviate these problems by

    a. permitting domestic and regional communications to be provided within the overall framework of the global system without proliferation of satellites in geostationary orbit;

    b. freeing the geostationary arcs over the ocean regions for other uses (navigational, maritime, and aeronautical satellites);

    c. providing a basic functional role for international communications in the form of high-capacity transoceanic trunks;

    d. avoiding double hops and/or a multiplicity of earth station equipment.

In particular, item *c* above may be viewed as a basis for justifying

the amount of investment required to fully develop ISL technology. Furthermore, the above system concepts can be advantageously integrated with other advanced techniques, such as all-digital transmission and SS/TDMA. More specifically, in the SS/TDMA mode, depending on the requirements for domestic/regional and intracontinental/transoceanic international traffic, transponders could be switched to provide corresponding (domestic, regional, continental) coverage or to carry the ISL traffic to the satellite(s) of adjacent continental regions. Alternatively, the transponder(s) could be used in a dedicated mode for the ISL traffic. Obviously, a large degree of dynamic flexibility is offered by this approach to accommodate all categories, levels, and fluctuations of traffic.

Other potential ISL applications include the tracking and data relay statellite system (TDRSS) and centralized world weather watch, *i.e.*, real-time meteorological observations on a global basis. Further investigation of the system concept, tradeoffs, and performance characteristics is required, however, to adjudge the practical viability of these and other practical applications.

Conclusions

A system model for an optical ISL has been presented for communication between geosynchronous satellites. Transmission channel performance measures have been formulated for PGBM digital transmission on the basis of solar background radiation as the predominant noise, and Poisson statistics for photon counting. In addition, an expression has been derived for the minimum transmitted power level for optimum threshold detection of the PGBM signal, and a graphical method to solve the implicit (transcendental) equation has been illustrated. The concept of duty-cycle-limited and noise-limited regions of system performance has been introduced, and preliminary numerical results have been provided.

Representative types of performance tradeoffs in terms of the basic physical parameters of the system, such as angular separation, aperture diameter, duty cycle, and bit rate, have been presented for the FD Nd:YAG laser ISL system. Some results for other typical laser systems, such as CO_2, GaAs, HeNe and Nd:YAG, have also been included for comparison.

It is generally concluded that an FD Nd:YAG laser system with transmitter power of about 100 mW and aperture diameter of about 25

cm would provide adequate ISL channel performance for all practical values of angular separation between the two satellites. The advantage of using the FD Nd:YAG laser system stems not only from the inherent superiority of its ISL channel performance, but also from the economy in the satellite power budget which would be implied by the use of a laser diode or solar radiation for optical pumping of the laser transmitter. The compact, lightweight, and efficient FD Nd:YAG system appears attractive for the practical implementation of an optical ISL providing high-data-rate, wide-bandwidth intersatellite communications with diverse applications.

Acknowledgment

References

[1] H. Hance et al., "Modulation/Demodulation Techniques for an Optical One Gigabit/Second Intersatellite Transmission System," International Telemetry Conference, Washington, D.C., 1971.

[2] G. M. Lee et al., "Bit Error Probability for Baseband and Subcarrier Optical Communications Systems," *IEEE Transactions on Communications,* COM-22, No. 1, January 1974, pp. 55–61.

[3] H. Yanai, M. Yano, and T. Kamiya, "Direct Modulation of a Double Heterostructure Laser Using a Schottky-Barrier-Gate Gunn-Effect Digital Device," *IEEE Journal of Quantum Electronics,* QE-11, No. 7, July 1975, pp. 519–524.

[4] J. H. McElroy et al., "Spaceborne CO_2 Laser Communications System," EASCON 1975, p. 176-A–176-P.

[5] J. D. Barry et al., "1000 Mbits/s Intersatellite Laser Communication System Technology," *IEEE Transactions on Communications,* COM-24, April 1976, pp. 470–478.

[6] R. M. Gagliardi and S. Karp, *Optical Communications,* New York: John Wiley & Sons, 1976.

[7] W. K. Pratt, *Laser Communication Systems,* New York: John Wiley & Sons, 1969.

[8] M. Ross, *Laser Receivers,* New York: John Wiley & Sons, 1966.

[9] T. F. Curran and M. Ross, "Optimum Detection Thresholds in Optical Communications," *Proc. IEEE,* Vol. 53, November 1965, pp. 1770–1771.

[10] N. P. Barnes, "Diode-Pumped Solid-State Lasers," *Journal of Applied Physics,* Vol. 44, No. 1, January 1973, pp. 230–237.

[11] R. B. Chesler and S. Singh, "Performance Model for End-Pumped Miniature Nd:YAG Lasers," *Journal of Applied Physics,* Vol. 44, No. 12, December 1973, pp. 5441–5443.

[12] G. I. Farmer and Y. C. Kiang, "Low-Current-Density LED-Pumped Nd:YAG Laser Using a Solid Cylindrical Reflector," *Journal of Applied Physics,* Vol. 45, No. 3, March 1974, pp. 1356–1371.

[13] F. W. Ostermayer, Jr., "LED End Pumped Nd:YAG Lasers," *IEEE Journal of Quantum Electronics,* QE-13, No. 1, January 1977, pp. 1–6.

Ashok K. Sinha received a B.Sc. and an M.Sc. in physics from Patna University, India, and a Ph.D. in physics from the University of Maryland. In 1974 he joined COMSAT Laboratories and became involved with analytical and computer-based modeling, optimization, performance analysis, and long-range planning methodology of communications satellite systems. Recently, he has been concerned with technical analysis and planning of overseas (ARABSAT, SATCOL, and CHINASAT) and domestic (SSTV) satellite systems with COM-SAT's Systems Technology Division. He is a member of the International Union of Radio Science (URSI), the International Association of Geomagnetism and Aeronomy (IAGA), the American Physical Society (APS), and IEEE.

2-bit soft-decision weighted erasure decoding for binary block codes

L. N. LEE AND M. I. WENG

(Manuscript received May 30, 1980)

Abstract

The design and implementation of weighted erasure codecs are described for binary block codes with 2-bit soft-decision demodulated signals. In terms of implementation, weighted erasure soft-decision decoding is a simple and cost-effective approach to achieve a given bit-error-rate performance for high-speed digital transmission. A design case for a (24, 12) extended Golay code is presented, and the measurement results are reported. It is demonstrated that a simple rate ½ codec with about 90 chips of standard MSI transistor-to-transistor logic integrated circuits can achieve a coding gain of 3 dB at a bit-error rate of 10^{-6} over a simulated thermal noise channel. In its current form, this codec can be operated at a clock rate up to 8 Mbit/s. With certain modifications, the total chip count can be further reduced and the clock rate can be increased to about 13 Mbit/s.

Introduction

It is well known that the performance of digital communications systems can be substantially improved by forward error correction (FEC) coding techniques. Since the discovery of the single error-correcting Hamming code in 1949, coding theory has evolved in two directions: more powerful codes which can be decoded without brute-

397

force computation, and improved decoding algorithms for simpler implementation or better performance. While two different approaches, block and convolutional encoding, have evolved, the development of practical decoding algorithms has centered around algebraic decoding algorithms for cyclic and majority decodable codes, and probabilistic decoding procedures for convolutional codes.

Algebraic decoding algorithms use the mathematical structure embedded in a code and operated directly on the received sequence. Probabilistic decoding procedures usually calculate the likelihood of message transmission based on reliability information extracted from the receiver. Generally, algebraic decoding algorithms are simpler and therefore faster, but less powerful; whereas probabilistic decoding algorithms are more complicated and slower, but more effective. However, since decoding complexity generally increases rapidly with the error-correcting capability, it is often easier to use a probabilistic decoder for a simpler code, and an algebraic decoder for a much more complex code to achieve a given performance objective. This is particularly true when the coding improvement requirements are high, and is probably the main reason that sequential decoding [1] and Viterbi decoding [2] with convolutional codes have received much attention during the past decade. Since superior performance of the probabilistic decoders results from using reliability information extracted from the receiver, the performance of algebraic decoders may also be improved by using such information.

Forney's minimum distance decoding [3] and Chase's algorithm [4] are two techniques which exemplify the above-mentioned trend. However, both algorithms require many iterations and thus lose the principal attractions of an algebraic decoder, namely, simplicity and speed. A different algorithm, weighted erasure encoding, which is established on the same principle as Forney's minimum decoding, was proposed by Weldon in 1971 [5]. It offers the possibility of implementing very simple decoders that operate at high speed and uses extracted reliability information to further improve performance.

In 1973, Goathal [6] proposed a set of weights which are linearly related to the likelihood ratio and satisfy all of the constraints required by the decoding procedure. He then demonstrated that the method is equivalent to the maximum likelihood decision rule. However, it was still believed that weighted erasure decoding is considerably more complex than hard-decision decoding because it may require the addition and multiplication of real numbers. This paper shows that

such a requirement is unnecessary in the special case of 2-bit soft-decision decoding, and that weighted erasure decoding can be implemented directly with standard logic circuits of reasonable complexity.

Weighted erasure decoding

A typical digital communications system consists of a modulator which accepts binary digits and outputs corresponding waveforms, a waveform channel, and a demodulator that estimates the binary digits transmitted based on the received waveforms. In some cases, the demodulator also indicates reliability and is called soft-decision to distinguish it from the hard-decision demodulator which does not provide reliability information.

It is often assumed that the digital communications system is linear and has sufficient bandwidth; therefore, it can be modeled as a discrete memoryless channel with two inputs and q outputs as shown in Figure 1, where $q = 2$ corresponds to the case of hard-decision, and $q > 2$ soft-decision. Without loss of generality, the q outputs can be ordered according to their likelihood ratios, so that the most reliable outputs are located at the two extremes. That is,

$$\frac{\Pr(0/i)}{\Pr(1/i)} > \frac{\Pr(0/i + 1)}{\Pr(1/i + 1)} \quad , \qquad 0 \leq i \leq q - 1 \quad . \tag{1}$$

In Reference 1, Weldon assigns each of the q outputs with a weight, ω_i, with the constraints

$$0 = \omega_0 \leq \omega_1 \leq \omega_2 \leq \cdots \leq \omega_{q-1} = 1 \tag{2a}$$

and

$$\omega_i + \omega_{q-i-1} = 1 \quad . \tag{2b}$$

Given the set of weights (ω_i), $0 \leq i \leq q - 1$, it is always possible to find a set of r positive real numbers (V_j), $1 \leq j \leq r$, such that

$$\omega_i = \sum_j a_{ij} V_j \quad , \qquad 1 \leq j \leq r \tag{3a}$$

and

$$\omega_{q-1} = \sum_j V_j \qquad 1 \leq j \leq r \tag{3b}$$
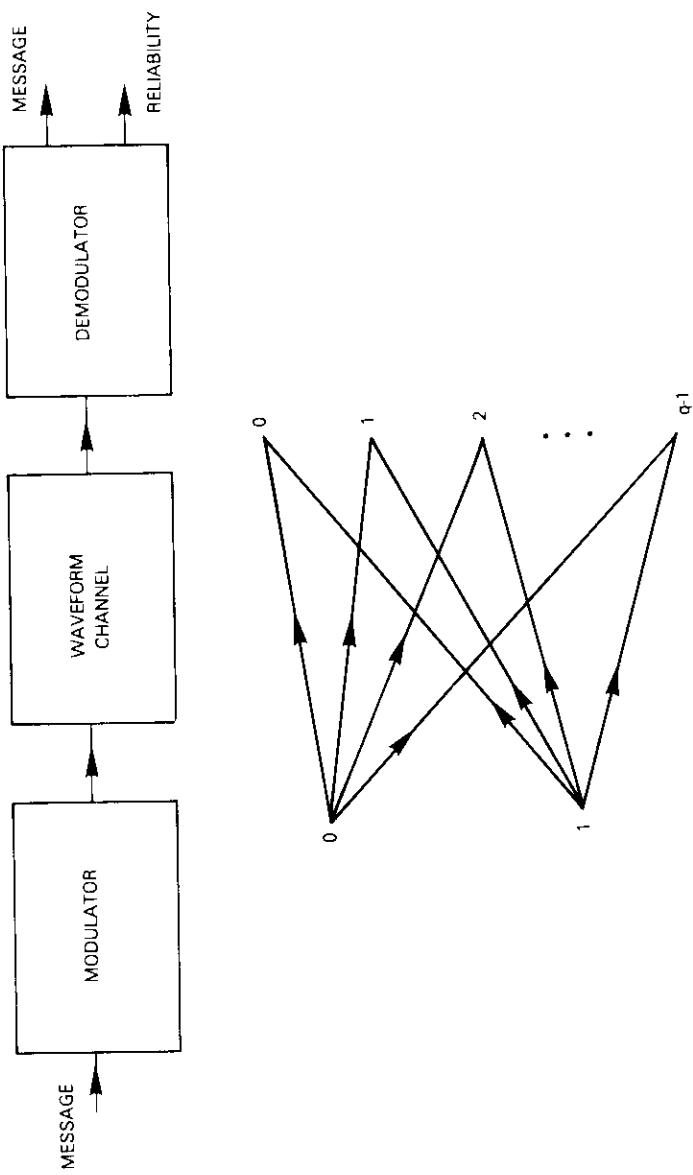
Figure 1. *Digital Communications System with Soft-Decision Demodulator and its 2-Input q-Output Discrete Memoryless Channel Model*

where a_{ij} are binary numbers of value 0 or 1. Obviously,

$$r \geq \log_2 q$$

and should be kept as small as possible because the implementation complexity of the decoder is directly proportional to r.

For each demodulator output i, a corresponding binary r-tuple $a_i = \{a_{i1}, a_{i2}, \ldots, a_{ir}\}$ can be determined. The weighted erasure decoder sends each of the r binary digits to one of r independent hard-decision subdecoders. Upon receiving n binary digits where n is the block length of the code, each hard-decision subdecoder decodes independently and counts the number of errors, e_j, that it detects or attempts to correct. A reliability indicator for the decoded block

$$R_j = \max [0, d - 2e_j] \qquad (4)$$

where d is the minimum distance of the code, is obtained for each independent subdecoder. The final decision is made according to the outputs and reliabiltiy indicators of all r hard-decision subdecoders on a bit-by-bit basis. The decision logic first calculates the product of the reliability indicator, R_j, and its corresponding weight, V_j, for the set of hard-decision subdecoders, S_o, which output 0; it performs the same function for the set of hard-decision subdecoders which output 1. The decision rule is to choose 0 if

$$\sum_{S_1} R_j V_j < \sum_{S_0} R_j V_j$$

and 1 otherwise. Goathal [6] observed that by choosing the weights

$$\omega_i = \frac{1}{2} \frac{\log [P_0 P_{q-i-1}/P_i P_{q-1}]}{\log (P_0/P_{q-1})} \qquad (5)$$

the decoding procedure is maximum likelihood, where $P_i = \text{Pr}\ (i/0) = \text{Pr}\ (q - 1 - i/1)$.

Weighted erasure decoding for 2-bit demodulation

A number of simplifications can be made in 2-bit soft-decision decoding. First, since $\omega_1 + \omega_2 = 1$,

$$V_1 = \omega_1 \qquad (6a)$$

$$V_2 = \omega_2 \qquad (6b)$$

and the r-tuples a_i are the binary representations of i. Thus, $r = \log_2 q = 2$ and only two hard-decision subdecoders are necessary (8-level or 3-bit soft-decision demodulation usually requires three). If the subdecoders produce inconsistent decisions, the decision rule involves only a comparision of the reliability indicators; that is, decoder 1 is chosen if $R_1 V_1 > R_2 V_2$, and decoder 2 otherwise.

Since the reliability indicator holds constant for the entire block, the decision rule can be made on a block-by-block basis instead of bit by bit. Furthermore, the reliability indicator is only a function of the number of errors detected or corrections attempted by the decoder. If the weights ω_1 and ω_2 or V_1 and V_2 are determined, the decision rule is only a function of the number of detected errors (e_1 and e_2) or attempted corrections. Thus, it can be implemented with digital logic without comparing the value of reliability indicators. Usually, if the code can correct or detect up to t errors, a $t + 1$ by $t + 1$ matrix can completely describe the decision rule. Sometimes simple combinatorial logic or a comparator may suffice.

Depending on the decoding method and the speed requirement of the system, decoder complexity can be further simplified by allowing the two subdecoders to share portions of the hardware. Generally, clock generation or conversion circuits, synchronization circuits, and output buffers can always be shared by both decoders. For simple block codes, these circuits often dominate decoder complexity. In some cases, almost all except the input buffer and syndrome generators can be shared; consequently, the overall soft-decision weighted erasure decoder is only slightly more complex than a hard-decision decoder.

Design case for the (24, 12) extended Golay code

The (24, 12) extended Golay code has been selected to demonstrate the feasibility of simplified weighted erasure decoding partially because it is a simple block code with good error-correcting capability. This code is based on the (23, 12) three error-correcting Golay code with generator polynominal $g(X) = 1 + X + X^5 + X^6 + X^7 + X^9 + X^{11}$. An overall parity check bit is added to the last position of the 24-bit block so that the code can detect four errors. The encoder implemen-

tation is fairly straightforward and will not be discussed. The decoder implementation consists of two hard-decision decoders and weighted erasure decision logic, which will be treated separately.

Figure 2 is a general block diagram of the hard-decision decoder. List decoding is used instead of the standard decoding technique [7], [8] to decode the (23, 12) Golay code; it is more cost-effective because of rapid advances in solid-state technology during the past decade. The syndrome corresponding to each of the 2,048 correctable error patterns is computed in advance, and the most likely error pattern is


Figure 2. *General Block Diagram of a Hard-Decision Decoder for the (24, 12) Extended Golay Code*
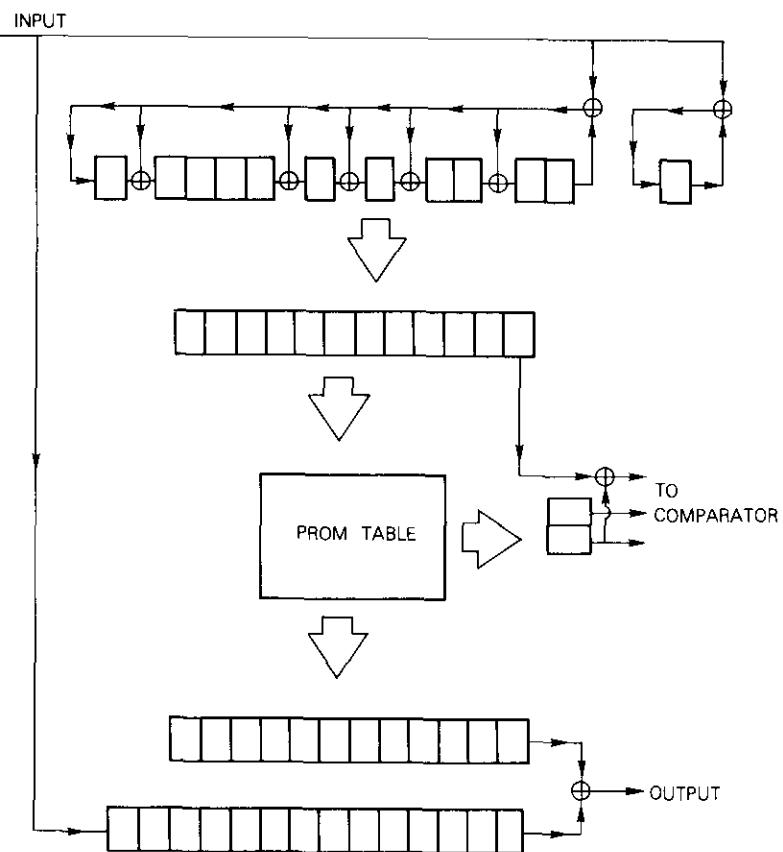
stored in a programmable read-only memory (PROM) at the appropriate address. Only the information portion is stored because it contributes to decoder output. To detect four errors, the parity of the entire 24-bit block is computed and compared with that of the error pattern. Four errors are detected if the two are inconsistent. The parity of each error pattern must be stored because only the portion of error patterns corresponding to the information bits is stored.

Unlike ordinary hard-decision decoders, the individual subdecoders must determine the total number of errors corrected or detected and inform the weighted erasure decision logic. This total must also be stored for each syndrome. Fortunately, the parity of the error pattern is the least significant bit of the total, and two bits are sufficient for the triple-error-correcting, four-error-detecting Golay code for both purposes. Because PROMS are organized in 4- or 8-bit wide forms, 32 K-bit storage may be needed to accommodate the necessary information. The syndrome generator and the input buffer are standard [9] and will not be discussed. When four errors are detected, the information must also be delivered to the weighted erasure decision logic. This is achieved by combinatorial logic circuits between the parity of the entire block and the total from the PROM.

The weighted erasure decision rule is based on weighting the decision of the two decoders; however, weighting is again a function of the transition probabilities of the 2-input, 4-output discrete memoryless channel, which are determined by the noise characteristics, the signal-to-noise ratio, and the quantization thresholds. Based on the assumption that the noise is additive white Gaussian with one-sided spectral density (N_o) and standard deviation $(\sigma = \sqrt{N_o/2})$, Table 1 shows the value of optimum weighting as a function of the signal energy-to-noise ratio, and the thresholds which are set at $t\sigma$, 0, and $-t\sigma$. It is clear from the table that weighting as given by equation (5) depends basically on the thresholds, and only slightly on the signal energy-to-noise density ratio for the entire range of interest. The value of ω_1 is greater than $\frac{1}{3}$, which implies that the second subdecoder should be chosen if it detects only one-half as many errors as the first subdecoder. If both subdecoders detect or correct the same number of errors, the first decoder must be chosen because the value of ω_i is less than $\frac{1}{2}$.

For a double error-correcting code, these observations are sufficient to determine the decision rule. When e_1 and e_2 are the number of errors detected or corrections attempted by the primary and secondary subdecoders, respectively, the secondary subdecoder is chosen if

$$\omega_2(d - 2e_1) < \omega_1(d - 2e_2) \quad ,$$

or equivalently,

$$\omega_1 > \frac{d - 2e_1}{2(d - e_1 - e_2)}$$

because $\omega_2 = 1 - \omega_1$. For the three-error-correcting, four-error detecting (24, 12) extended Golay code, $d = 8$. The weighted decision rule depends on the region to which the value of ω_1 belongs: between 1/3 and 2/5, between 2/5 and 3/7, or between 3/7 and 1/2. Table 2 illustrates the decision rule for each region. Reexamining the values of ω_1 in Table 1 reveals that these three regions can be roughly mapped to $t \geq 0.5$, $0.3 \leq t \leq 0.5$, and $t \leq 0.3$ for a practical range of E/N_o.

TABLE 1. THE VALUE OF ω_1 AS A FUNCTION OF SIGNAL ENERGY-TO-NOISE RATIO AND THRESHOLD GAIN (t) IN AN ADDITIVE WHITE GAUSSIAN NOISE CHANNEL
($\omega_2 = 1 - \omega_1$)

t	E_b/N_o					
	0 dB	2 dB	4 dB	6 dB	8 dB	10 dB
0.05	0.485	0.485	0.486	0.486	0.487	0.488
0.10	0.472	0.472	0.472	0.473	0.474	0.476
0.15	0.459	0.460	0.460	0.461	0.463	0.465
0.20	0.447	0.448	0.449	0.450	0.452	0.454
0.25	0.437	0.437	0.438	0.440	0.442	0.445
0.30	0.427	0.427	0.428	0.430	0.432	0.435
0.35	0.417	0.418	0.419	0.421	0.423	0.427
0.40	0.409	0.410	0.411	0.412	0.415	0.419
0.45	0.401	0.402	0.403	0.405	0.407	0.411
0.50	0.394	0.395	0.396	0.398	0.400	0.404
0.55	0.387	0.389	0.399	0.391	0.394	0.398
0.60	0.381	0.382	0.383	0.385	0.388	0.392

A major difference between 2-bit soft-decision weighted erasure decoding and 3-bit Viterbi or sequential decoding is that performance is sensitive to the thresholds for the former. Note that the soft-decision improves performance only when the primary subdecoder performs incorrectly, the secondary decoder decodes correctly, and the weighted decision logic chooses the second decoder. This occurs only if the

probability of the second decoder being correct is reasonably large. Consequently, t must be reasonably small. Computer simulation results seem to indicate that the optimal value for t, which is generally a function of the minimum distance of the code and the signal energy-to-noise ratio, is above 0.3 for all practical ranges. The decision rule should therefore be that of (a) or (b) in Table 2. However, the simulation results seem to indicate that the performance degradation resulting from using a simple comparator rule of (c) instead of (a) or (b) is so insignificant that the comparator serves as a good approximation to the optimal rule in all cases. Figure 3 shows a general functional block diagram for the entire weighted erasure decoder.

TABLE 2. WEIGHTED DECISION RULES FOR THE (24, 12) EXTENDED GOLAY CODE AS A FUNCTION OF ω_1*

(a) $1/3 < \omega_1 < 2/5$					(b) $2/5 < \omega_1 < 3/7$						(c) $3/7 < \omega_1 < 1/2$				
		e_1					e_1						e_1		
e_2	0	1	2	3	4	e_2	0	1	2	3	4	e_2	0	1	2	3	4
0	I	I	II	II	II	0	I	I	II	II	II	0	I	II	II	II	II
1	I	I	I	II	II	1	I	I	II	II	II	1	I	I	II	II	II
2	I	I	I	II	II	2	I	I	I	II	II	2	I	I	I	II	II
3	I	I	I	I	II	3	I	I	I	I	II	3	I	I	I	I	II
4	I	I	I	I	I	4	I	I	I	I	I	4	I	I	I	I	I

* I = Choose the primary subdecoder, II = Choose the secondary subdecoder.
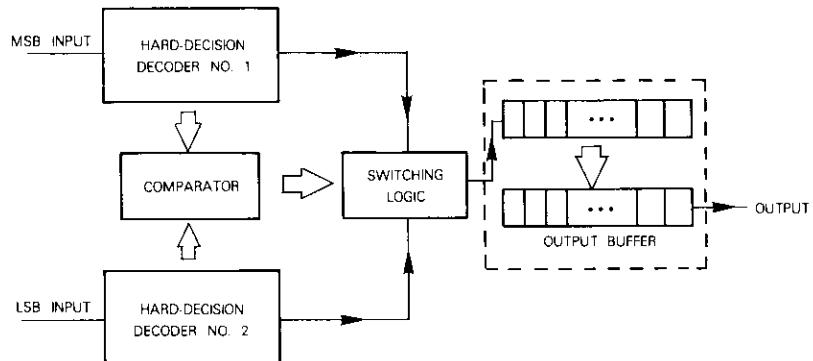

Figure 3. *General Functional Block Diagram for a Complete Weighted Erasure Decoder*

After the logic design was completed, a prototype of the codec was built on two 7.5 × 5.5 in. (18.3 × 13.8 cm) wire-wrapped boards. The weighted erasure codec is implemented with approximately 90 chips of TTL integrated circuits and four PROM chips. A photograph of the prototype is shown in Figure 4.
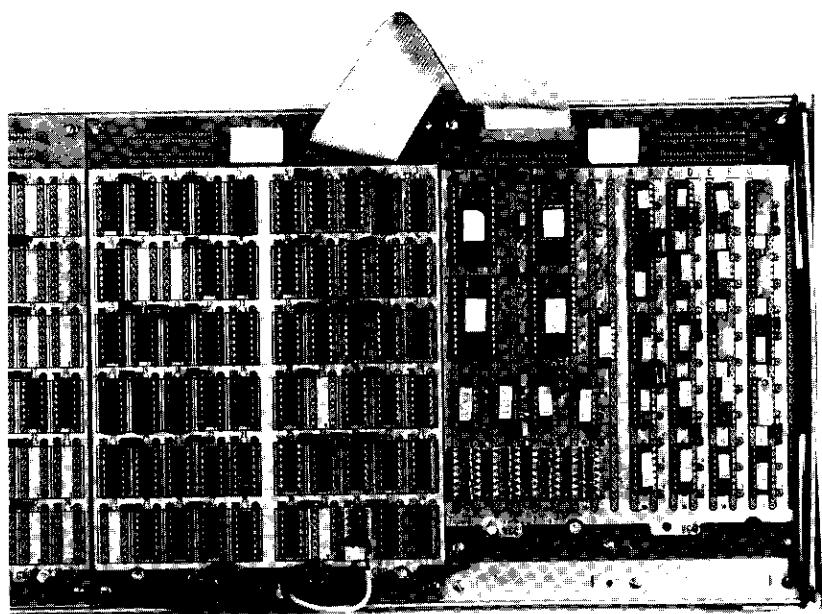
Figure 4. *Prototype WED Codec*

Performance measurement

The prototype was connected to a soft-decision modem and test equipment to measure performance. Figure 5 is a general block diagram for the test setup. A pseudorandom bit stream was generated by the test set and sent to the modem. After being differentially encoded, the bit stream is externally routed to the FEC encoder and returned to the modem. The BPSK modulator then modulated the bit stream into 70-MHz IF. In the meantime, thermal noise was passed through a 19-MHz low-pass filter so that the filter output was essentially a white Gaussian noise throughout the 19-MHz band. The IF signal and the
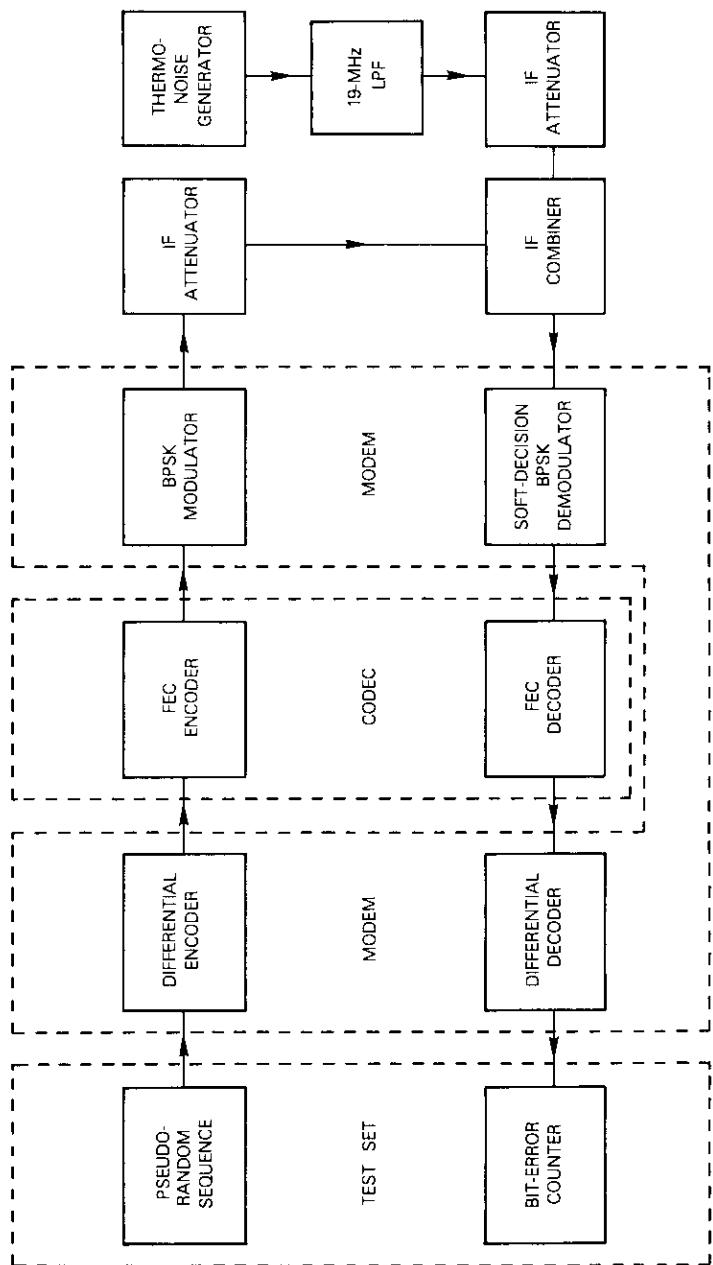
Figure 5. *Test Configuration for the Performance Measurement*

noise were combined by an IF combiner after appropriate attenuation in order to simulate the actual received signal before soft-decision demodulation.

The output of the integrate-and-dump circuit is quantized into eight levels with uniformly spaced thresholds as shown in Figure 6a, where E is the average signal energy for each channel bit. A few logic gates are necessary to convert the modem output to a form acceptable to the weighted erasure codec. Because the simulation results indicate that the optimal set of thresholds is about $\{3/10, 0, -3/10\}$ of the signal amplitude, three thresholds between 001 and 000, 000 and 100, and 100 and 101 of the original modem output are selected. Figure 6b shows the input to the weighted erasure decoder with respect to the amplitude of the integrate-and-dump circuit. After the received signal is decoded, the bit stream is routed back to the modem. The differentially decoded sequence was returned to the test set for bit-error-rate calculations.

a. 3-BIT MODEM OUTPUT

b. 2-BIT WEIGHTED ERASURE DECODER INPUT
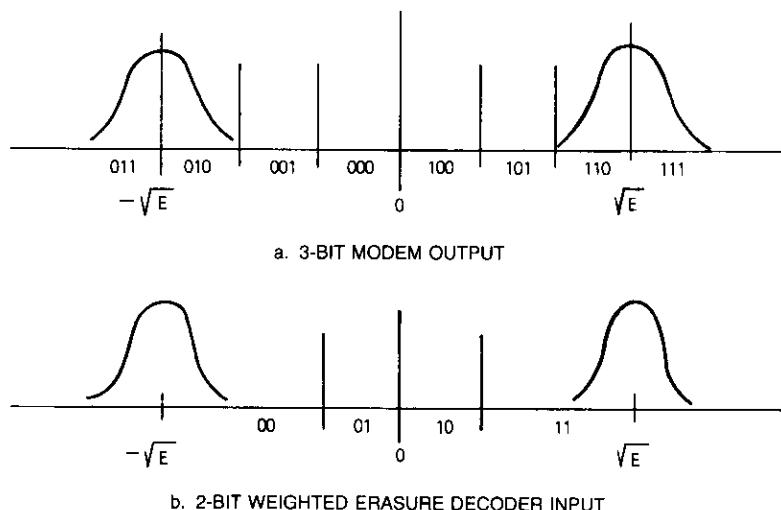
Figure 6. *Soft-Decision Thresholds*

By measuring the output power of the attenuated signal and the 19-MHz low-pass filter, the signal energy per information bit to one-sided power spectral density ratio, E_b/N_o, can be calibrated by $E_b/N_o = (C/N)$ (B/R), where C/N is the carrier-to-noise ratio, R is the information

rate, and B is the filter bandwidth. The E_b/N_o is then adjusted by changing the amount of attenuation to the desired level for error rate measurements. The decoder was successfully tested at slightly higher than 4-Mbit/s information rates or 8-Mbit/s channel signaling rates, but the actual measurements were performed at 1-Mbit/s information rates or 2-Mbit/s channel rates.

A computer simulation program was prepared to examine the performance of the weighted erasure decoder. A pseudorandom number generator whose output is equivalent to a white Gaussian noise passing through the soft-decision quantizer is used to create additive noise at the receiver input. For comparison, differential decoding is added at the output of the weighted erasure decoder (to be consistent with modem used in measurements). Figure 7 plots the computer simulation results, with quantization thresholds set at $(-0.3\sqrt{E}, 0, 0.3\sqrt{E})$. It also includes the performance upper bound of the uncoded Gaussian channel, which is twice the error probability obtained without differential coding, and the simulation results for the hard-decision decoder with the same pseudorandom noise sequence. The coding gain obtained from the hard-decision decoder seems consistent with that reported by previous investigators [1]; however, the coding gain contributed by weighted erasure decoding, about 0.5 to 0.6 dB, is less than expected.

Figure 8 shows the actual measurement results. For reference, modem performance with differential encoding was first measured with the modulator and demodulator operating at 1 Mbit/s. The error rate is more than twice the theoretical error rate because of differential coding and modem implementation loss. The performance of the hard-decision decoder, as well as that of the weighted erasure decoder, was measured with the information rate kept at 1 Mbit/s and the channel bit rate set at 2 Mbit/s. In both cases, differential coding was used, and the effect of differential coding on the coding system was the same as that on the uncoded system because the (24, 12) code is transparent to bit inversion. Figure 8 reveals that coding gain achieved by the hard-decision decoder matches the simulation results very closely, that is, about 1.5 dB at 10^{-3}, 1.8 dB at 10^{-4}, and 2.0 dB at 10^{-5}. However, for weighted erasure decoding, a gain of 2.5 dB at 10^{-3}, 2.9 dB at 10^{-4}, and 3.0 dB at 10^{-5} is observed. This measured gain, which is higher than that obtained in the simulation, may be attributed to the capability of soft-decision decoding in reducing modem implementation losses.
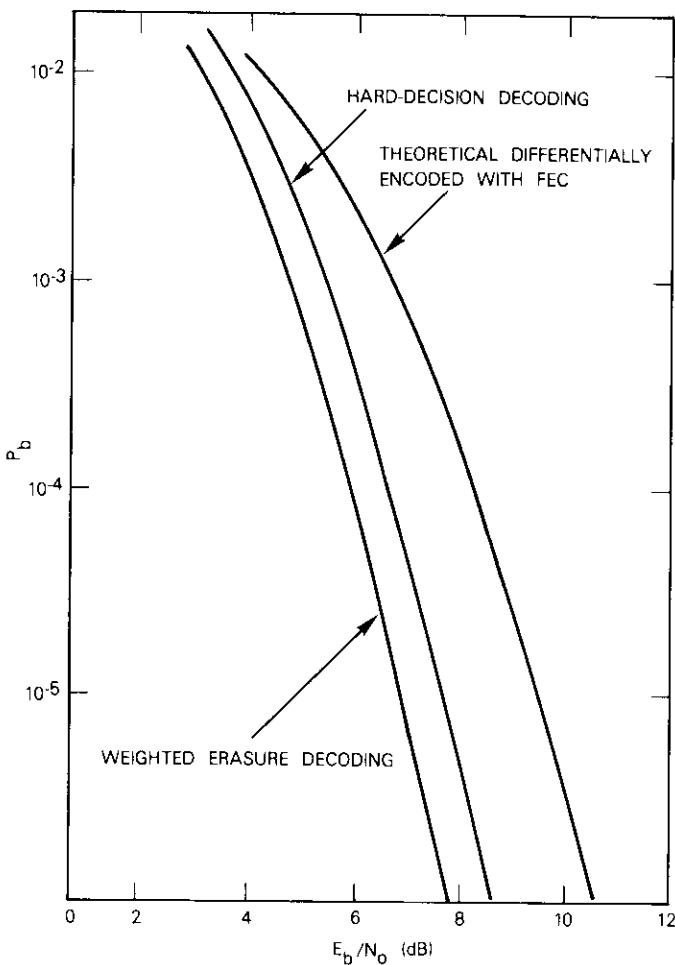
Figure 7. *Performance of Weighted Erasure Decoder With 2-bit Soft-Decision for the (24, 12) Golay Code Obtained from Computer Simulation*

Conclusions

Implementation and laboratory measurements have demonstrated that 2-bit soft-decision weighted erasure decoders can be implemented with simple logic circuits, and that reasonable coding gain can be
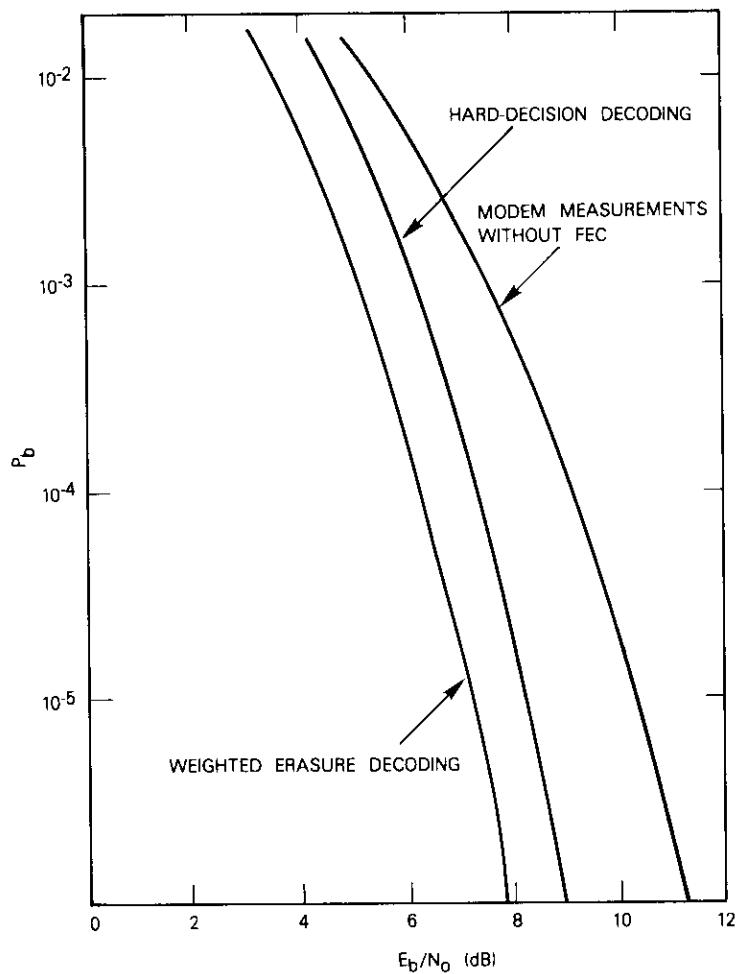
Figure 8. *Measured Performance of Weighted Erasure Decoder with 2-Bit Soft-Decision for the (24, 12) Extended Golay Code*

achieved with a two-fold increase in complexity over hard-decision decoders. When an additional 0.5- to 1.0-dB coding gain is required to meet the performance objectives, the two-fold increase in decoder complexity is usually justified.

Among binary block codes, the BCH codes are the most popular because of their relatively strong error-correcting capability and ease

of decoding. For BCH codes which correct three or more errors, the decoder requires the Berlekamp-Massey algorithm. For double error-correcting BCH codes, list decoding is generally feasible for block lengths up to 255; and for single error-correcting BCH (or Hamming) codes, the error-trapping technique is used. It is a moderate increase in implementation complexity from an error-trapping decoder to a list decoder, and a very significant increase from a list decoder to a Berlekamp-Massey decoder because of the iteration logic involved.

Implementation complexity is also affected by the buffer required (usually slightly more than one block). In this case, a weighted erasure decoder must buffer one block for each of the two hard-decision subdecoders. This is about the same amount of buffer required for a hard-decision decoder twice the block length. Therefore, a weighted erasure decoder for a double error-correction BCH code is much simpler than a hard-decision decoder for a three- or quadruple-error-correction BCH code delivering approximately the same coding gain. There is also a moderate reduction in hardware between a weighted erasure decoder for a single error-correction BCH and a hard-decision decoder for a double error-correction BCH code delivering similar performance.

A further advantage of the weighted erasure decoder using list decoding over the hard-decision decoder with the Berlekamp-Massey algorithm is its capability to operate at high speed. The decoder logic operates at the basic data rate in the list decoder, which is above 10 Mbit/s with the low-power Schottky TTL devices. The input data rate can be even higher if faster logic is used (*e.g.*, 30 Mbit/s with Schottky TTL, or 120 Mbit/s with emitter-coupled logic). This speed is not achievable for a Berlekamp-Massey decoder with the same solid-state technology, without extensive buffering and parallelism.

It is concluded that the weighted erasure decoder with list decoding may have great potential in low-cost, high-speed applications.

Acknowledgment

References

[1] J. M. Wozencraft and I. M. Jacobs, *Principle of Communications Engineering*, New York: John Wiley and Sons, Inc., 1965.

[2] A. J. Viterbi, "Error Bounds For Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, IT-13, No. 2, April 1967.

[3] G. D. Forney, "Generalized Minimum Distance Decoding," *IEEE Transactions on Information Theory*, IT-12, No. 2, April 1966, pp. 125–131.

[4] D. Chase, "A Class of Algorithms for Decoding Block Codes with Channel Measurement Information," *IEEE Transactions on Information Theory*, IT-18, No. 1, January 1972, pp. 170–181.

[5] E. J. Weldon, "Decoding Binary Block Codes on Q-ary Output Channels," *IEEE Transactions on Information Theory*, IT-17, November 1971, pp. 713–718.

[6] J. M. Goathal, "A Metric for Maximum Likelihood Decoding on Binary Input, Q-ary Output Channels," *Proc. IEEE*, 1972 Canadian Communications and EHV Conference, Montreal, Canada, November 9–10, 1972, pp. 42–43.

[7] E. J. Weldon, "A Comparison of an Interleaved Golay Code and a Three Dimensional Product Code," Final Report, USNELC, Contract N0095368M, August 1968.

[8] T. Kasami, "A Decoding Procedure for Multiple-Error-Correcting Cyclic Codes," *IEEE Transactions on Information Theory*, IT-10, pp. 134–139, April 1964.

[9] S. Lin, *An Introduction To Error-Correcting Codes*, Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Lin-Nan Lee received a B.S.E.E. degree (1970) from National Taiwan University, an M.S. (1973), and a Ph.D. in electrical engineering (1976) from the University of Notre Dame. He is a staff scientist in the Communications Systems Analysis Department of COMSAT Laboratories. Before joining COMSAT in 1977 he was a communications officer in the Chinese Air Force, a research assistant at the University of Notre Dame, and an employee of the Linkabit Corporation involved in the design and development of packet communications protocols for the satellite networks. He is a member of IEEE and Sigma Xi.

Ming-I Weng received a B.S. in 1971 and an M.S. in 1973 in electronic engineering from National Chaio Tung University. Since 1975 he has been with Telecommunications Laboratories, where he is engaged in the design and implementation of digital and data communications equipment. From October 1978 to September 1979, he was on leave with the Communications Systems Analysis Department, COMSAT Laboratories.

CTR Note

Ship-to-shore tests of a MARISAT 56-kbit/s data service

T. O. CALVIT

(Manuscript received May 15, 1980)

Introduction

Since the launching of the MARISAT satellites in 1976, the system has provided reliable teleprinter and voice grade service, including facsimile and 1200/2400-bit/s data [1]. It presently serves ships in three oceans, and experience indicates rapid growth in both utilization and the number of ships equipped for service.

Long before the MARISAT system became fully operational, it was recognized that the satellites could provide inexpensive, ship-to-shore, high-speed, data transmission service. The operational concept for such service at 56 kbit/s was disclosed in 1977. During the latter part of November 1978, simplex 56-kbit/s data transmission tests (over 130 hours in about two weeks) were conducted from a ship in the North Sea to the MARISAT shore station in Southbury, Connecticut. The test results effectively demonstrated that a 56-kbit/s ship-to-shore data transmission service could be implemented for small ships with a high degree of reliability and availability, and, with minimal additional equipment, to a standard MARISAT ship terminal.

Mr. Calvit is Program Director of SATCOL, Programs and Systems Development Department, Systems Technology Services Division.

Service description

It is technically feasible to introduce a new MARISAT service which can provide ship-to-shore data transmission at 56 kbit/s [2]. For example, data transmitted from a ship in either the Atlantic or Pacific Ocean could be instantly available to a data processing center anywhere within the continental United States. Even relatively small ships, such as those engaged in seismic research, could easily accomodate the necessary shipboard equipment. Such service would be available within those ocean areas which have 10° or higher elevation angles toward the Atlantic or Pacific MARISAT satellite locations (15°W longitude and 176.5°E longitude, respectively).

The services would consist of two communications channels:

a. A ship-to-satellite to shore station simplex channel capable of transmitting data at rates up to 56 kbit/s.

b. A protocol and control channel from shore station to satellite to ship. The use of both a 50-baud TTY channel and a 1200-bit/s data over voice channel has been considered for the control function, and either appears feasible; however, the lower control channel bit rate would require a proportionately larger buffer memory both on the ship and at the shore station to preserve an acceptable throughput efficiency [2] of better than 80 percent.

Based upon current work, these services could be provided using special modem, codec, and control components to augment existing MARISAT ship terminal and shore station equipment. An available MARISAT ship terminal using rate 1/2 convolutional encoding and soft-decision decoding will provide high link availability and bit-error rates (BERs) better than 10^{-6} at user elevation angles of 10° or higher.

Tests at sea

Feasibility tests were conducted in the North Sea aboard the *M/V J. E. Jonsson,* a 168-ft (51 m), 1,008 gross ton, seismic exploration vessel operated by Geophysical Service Inc. The ship is equipped with a Scientific-Atlanta Model SA-3055 MARISAT terminal. Modulation and forward-error-correction (FEC) equipment elements were added to the existing terminal on the ship and at the shore station. Rate 1/2 convolutional encoding and a QPSK modulator were used on the ship. The input data stream to the coder/modulator was provided by a BER tester.

An existing high-speed data modem, originally used for laboratory tests at 240 kbit/s, was reconfigured to operate at an effective data rate of 56 kbit/s, and the package was mechanically modified for the anticipated shipboard environment. The equipment was tested at COMSAT Laboratories with the MARISAT simulator and then with the MARISAT satellite for demonstration in a static environment.

TEST EQUIPMENT

Figure 1 is a block diagram of the equipment configuration used in the shipboard tests. A BER test set generated a pseudorandom sequence at 56 kbit/s, and rate 1/2 encoding provided forward error correction. The encoded data stream at 112 kbit/s entered the QPSK modulator at an IF frequency of 90.8 MHz which was compatible with the MARISAT terminal used. This signal was up-converted to 1640.8 MHz in the terminal and transmitted via the Atlantic MARISAT satellite to the shore station at Southbury, Connecticut.

At Southbury, the received signal was filtered and demodulated to a digital data stream at 112 kbit/s. This data stream was decoded to detect and correct data errors and provide an output at 56 kbit/s. The test set at the ground station checked each received bit against an identical internally generated pseudorandom sequence. Each bit error was counted, and the error statistics were recorded on a paper tape printer.

TEST OPERATIONS AND RESULTS

The shipboard test equipment was located in a rack adjacent to the MARISAT terminal. A coaxial switch was used to select either normal MARISAT voice or test data as input to the terminal. The request for start-up of operations was coordinated through the operator at Southbury using normal MARISAT telephone service. Support personnel at the shore station provided the required circuit patch and monitored the operation of the shore-based test equipment. Tests were initiated to run for several hours at a time depending on ship motion conditions, circuit availability, and support personnel schedules.

During the test period, a wide range of environmental and ship operational conditions was encountered. The elevation angle from ship to satellite was about 17° as the ship operated in Storfjorden Fjord in Norway (62.5°N, 7°E), and approximately 23° when the ship was in Moray Firth in Scotland (58°N, 3°W). The typical link budget of Table 1 shows that with expected atmospheric and propagation losses, an
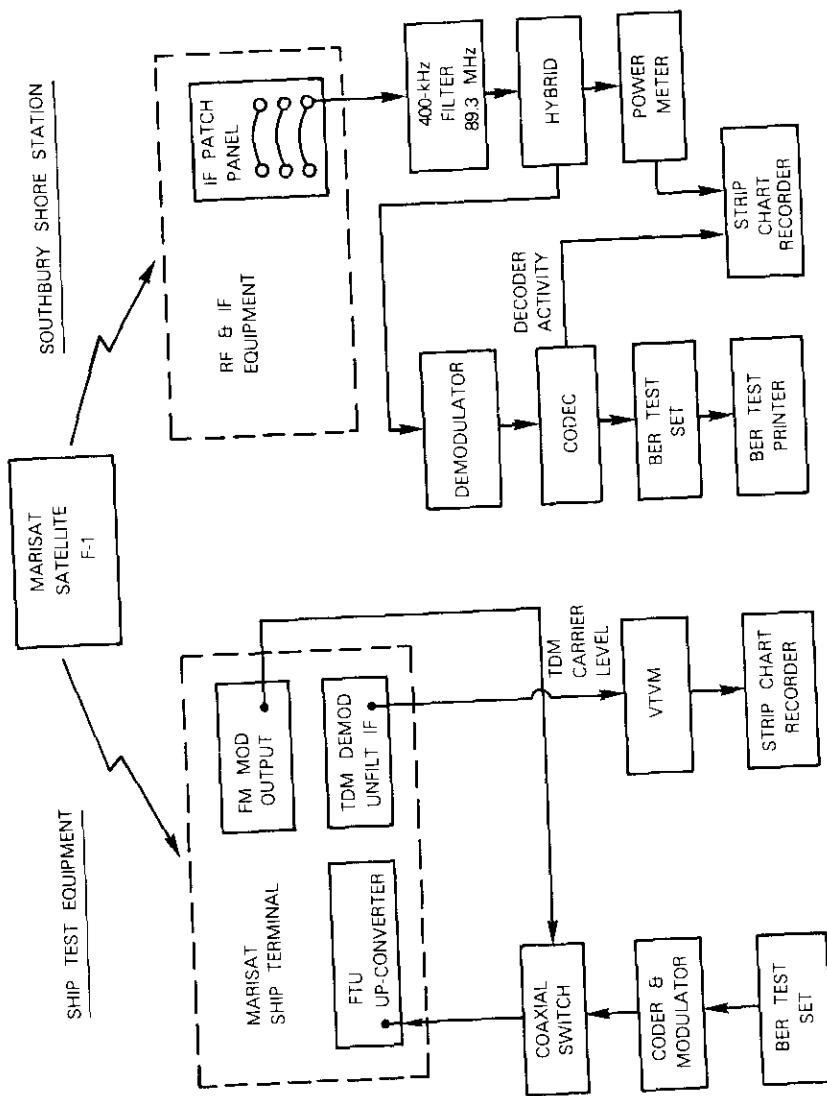
Figure 1. *Test Equipment Configuration*

TABLE 1.   TYPICAL LINK BUDGET

Up-Link		
Ship Transmitter Power	37.2	W
Diplexer/Feed Loss	0.6	dB
Ship Antenna Gain	23.5	dB
Ship e.i.r.p.	38.6	dBW
Free Space Loss	188.6	dB
Losses		
Wet Radome	0.5	dB
Polarization Coupling	0.2	dB
Atmospheric Absorption	0.4	dB
Satellite G/T	−15.9	dB/K
C/N_o		
Without Losses	62.7	dB-Hz
With Losses	61.6	dB-Hz
Flux Density		
Without Losses	−124.3	dBW/m^2
With Losses	−125.4	dBW/m^2
Saturation Flux Density	−100.6	dBW/m^2
Input Backoff		
Without Losses	23.7	dB
With Losses	24.8	dB
Down-Link		
Satellite Saturation e.i.r.p.	20.5	dBW
Carrier Output Backoff		
Without Losses	17.5	dB
With Losses	18.5	dB
Carrier e.i.r.p.		
Without Losses	+3.0	dBW
With Losses	+2.0	dBW
Free Space Loss	196.9	dB
Losses		
Atmospheric Absorption	0.3	dB
Rain Attenuation	0.5	dB
Polarization Coupling Loss	0.4	dB
Rain-Induced Sky Noise	1.2	dB
Shore Station G/T	33.0	dB/K
C/N_o		
Without Losses	67.7	dB-Hz
With Losses	64.3	dB-Hz
C/N_o		
Without Fade	81.7	dB-Hz
With Fade	83.3	dB-Hz

TABLE 1.   TYPICAL LINK BUDGET (CONTINUED)

Total Link		
C/N_o		
Without Losses	61.5	dB-Hz
With Losses	59.7	dB-Hz
E_b/N_o		
Without Losses	14.0	dB
With Losses	12.2	dB

E_b/N_o of 12.2 to 14.0 dB could be provided. Earlier laboratory tests indicated that soft-decision decoding and QPSK modulation could achieve a BER of better than 1×10^{-5} for an E_b/N_o above 5 to 6 dB. Experimental results indicated agreement with calculated performance.

Over a 2-week test period, there were approximately 136 experimental hours. Tests were initiated whenever shipboard motion permitted work activities (characteristically, for continuous ship rolling motion of less than 10°). A summary of the test results presented in Table 2 shows five separate sequential periods corresponding to locations along the route of the vessel. Figure 2 illustrates the test route with vessel locations marked at 6-hr periodic time intervals. During test periods, link availability was defined as that percentage of time, after initial carrier acquisition, when carrier lock was maintained and data were being received. According to this definition, the minimum time for loss of acquisition was 1 second because of bit error recording equipment limitations. The BER is defined as the ratio of bits in error to the total number of bits transmitted during a period of availability.

The first test extended over 2 days at dockside in Middlesbrough, England. During a total of 29.7 hr of test data, the average BER measured was 2.1×10^{-8}; the errors occurred in bursts of about 40 bits or less. An error rate of 1.5×10^{-10} was observed during one 13-hr period. These BERS compare favorably with terrestrial and other satellite communications circuits.

Soon after the journey from Middlesbrough to Alesund, Norway, began, a storm was encountered. High winds and waves produced peak ship rolls in excess of 20° during the test; however, 32.5 hr of data were recorded during the 2.5 days required to cross the North Sea. Circuit availability dropped to 97.2 percent and the average BER observed was 1.8×10^{-7}. Sometimes tests had to be completely suspended because of extreme ship motion when the vessel was rolling beyond 30°.

TABLE 2.   SUMMARY OF TEST RESULTS

Dates	Location	Total Test Hours	Availability (Percent)	BER	Sea Environment and Vessel Motion
11/15–11/16	Dockside Middlesbrough	29.7	100.00	2.1×10^{-8}	Quiescent
11/17–11/19	Enroute England To Norway	32.5	97.2	1.8×10^{-7}	High Wind and Waves Ship Roll to 20°
11/20–11/23	Operating In Storfjorden Fjord	17.8	98.4	7.4×10^{-8}	Calm Waters Negligible Ship Motion
11/23–11/24	Enroute Norway To Scotland	6.2	92.4	5.3×10^{-8}	Very Heavy Seas Winds to 40 knots Waves to 10 ft (3 m) Pitch Angles as large as 20°
11/25–11/29	Operations In Moray Firth	50.2	95.6	1.6×10^{-6}	Cloudy and Rainy Some Heavy Seas Rapid Pitch Motions Peaking at About 12° Occasional Roll to 10°

During operation in the Storfjorden Fjord near Alesund over a 3.5-day period, both the circuit availability and BERs improved. The waters were relatively calm in the fjord; however, some signal degradation was noted when the vessel periodically made rapid heading changes. Very heavy seas, high winds and waves, and peak pitch angles to 20° were observed on the return voyage from Alesund to northern Scotland. Only 6 hr of data were collected during the 45 hr required to cross the North Sea because of environmental conditions. The measured circuit availability was the lowest recorded during the test. The antenna stabilization system on the Jonsson terminal was designed to operate at pitch angles of less than ±7.5° and with pitch motion periods of 2π seconds or greater. These specifications were frequently exceeded during this part of the voyage; however, the BER was better than expected for the prevailing conditions.

The final tests, performed in Moray Firth off the coast of northern Scotland, involved the greatest number of test hours and operation in
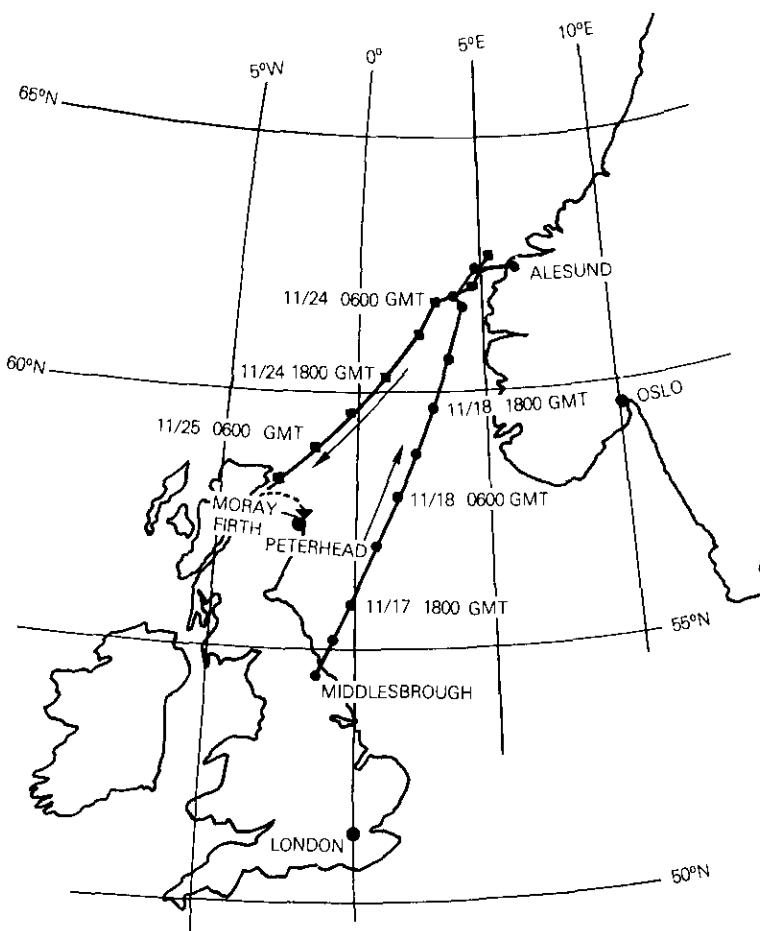
Figure 2. *Ship Test Route*

a variety of weather and sea-state conditions. Rapid pitch motion, peaking at about 12°, caused the highest average BERS observed during the tests. However, this 1.6×10^{-6} BER will allow adequate throughput efficiency if the proper protocol and data block sizes are chosen.

Conclusion

It has been demonstrated that 56-kbit/s data transmission is technically and operationally feasible between seismic vessels and shore

facilities. The shipboard test allowed performance evaluation over a wide range of operational extremes for the seismic vessel. Measured BERS were better than 1×10^{-7} for much of the available test time. Circuit availability above 97 percent was typical for most of the test. The performance data exceeded expected results and show that the MARISAT link is suitable for data transmission. The relatively trouble-free operation of the equipment indicates the potential for good system reliability.

References

[1] D. W. Lipke et al., "MARISAT—A Maritime Satellite Communications System," *COMSAT Technical Review*, Vol. 7, No. 2, Fall 1977, pp. 351–391.
[2] T. O. Calvit, "A MARISAT 56 kbps Ship-to-Shore Data Service," Presented at the 1979 San Francisco RTCM Assembly Meeting, April 23–26, 1979.
[3] T. O. Calvit and L. B. Heitman, "High-Speed Satellite Data Transmission of Maritime Seismic Data," AIAA 8th Communications Satellite System Conference, Orlando, Florida, April 20–24, 1980.