THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 57 September 1978 Number 7, Part 1

SG Undersea Cable System:

Introduction and Development Plan

By R. D. EHRBAR, A. E. FORD, and G. GERBIER

(Manuscript received June 30, 1977)

This paper describes the background of the three-nation SG development and the structure of the organization, and comments on the success with which the mission was carried out.

I. INTRODUCTION

Communications traffic across the North Atlantic has increased around 20 percent per annum for some years, carried partly by cable and partly by satellite. This growth continued beyond 1970, following the start of service on the then-newest undersea cable facility, an SF system called TAT-5, in March of that year.[1] There was no reason to think that the growth rate would fall significantly in the next decade. Satellites were carrying a significant part of the North Atlantic traffic, and it was generally recognized that both cable and satellite technologies should be carried forward to their maximum potential. This provided the impetus for pursuing the new cable technology designated SG.

In this paper, we do not discuss the merits of the different transmission technologies or predict future deployment. Instead, we devote our attention to the development of the SG undersea cable system, which as

TAT-6 (3400 nmi long) was turned up for initial service on July 27, 1976 between the United States and France.

II. NEW DEVELOPMENT PLANS

An exploratory program was under way at Bell Laboratories, even before the completion of TAT-5, to determine practical objectives for the next undersea system that would logically follow SF and would continue the reduction in cost per channel mile. In the United Kingdom, production of a system with a top frequency of 14 MHz was also under way. This system, developed for short distance cables, was capable of providing a nominal voice capacity of 1840 two-way channels spaced at 3 kHz. Ongoing development to extend the system to intercontinental distances was well advanced.

A tentative plan for deployment of cables in the North Atlantic had been formulated in negotiations among the many parties sharing responsibility for these communications. One plan called for another SF system between the U.S. and France in 1973, and then a new large capacity system (perhaps as many as 4000 channels) from the U.S. to the U.K. at a later date. To further this plan, two projects were set in motion in 1970: (*i*) Application was made to the American Federal Communications Commission (FCC) in August for permission to lay an SF system from the United States to France and (*ii*) plans were formulated within the Bell System for the development of a new, higher capacity system to be called SG.

The FCC ruled against the SF system in mid-1971 as not being in the public interest and recommended that the development of the proposed SG system be expedited so as to be ready for service as rapidly as possible. This put considerable pressure on completing the SG development so that manufacture could be started at an early date, since there was likely to be a cable circuit deficit across the Atlantic between 1973 and the earliest time considered practical for completion of a transatlantic SG (about 1976).

The pressure to provide another transatlantic system at an earlier date was reduced significantly by a decision of the British Post Office (BPO) and the Canadian Overseas Telecommunications Corporation (now Teleglobe, Canada) to install CANTAT-2, producing 1840 3-kHz message channels between Widemouth, Cornwall and Halifax, Nova Scotia.[2] This link was brought into service in 1974.

III. SHARED DEVELOPMENT PLAN

The idea of a shared development approach to the new SG system had been considered in the early planning stage, and the first joint meeting of AT&T, Bell Laboratories, and the BPO was held in June 1970 in London. In view of the development work already being undertaken by

both the BPO and Bell Labs, partnership in the development of the SG system seemed natural. The division of the development work was changed several times. Only two decisions made at the early meetings remained constant: (*i*) Bell System responsibility for repeater development and (*ii*) BPO responsibility for a new 1.7-in. cable development. Originally, it was intended that the BPO would be responsible for the development of the terminal transmission equipment, but in a three-party agreement, the French Ministry for Postes and Telecommunications (FPTT) accepted responsibility for this work. This was a logical step since, by 1971, it had been internationally agreed that the new, high-capacity system TAT-6 would be installed between Green Hill, Rhode Island and St. Hilaire de Riez, France.

IV. DIVISION OF RESPONSIBILITY

While the division of responsibility could not be completely specified, an attempt was made to define it in some detail. Some responsibilities were common to all, while others were assigned to individual groups.

4.1 Common responsibilities

These can be summarized as follows: Development of overall system design parameters and specification, system planning principles and engineering, specification of test equipment and procedures for installation, commissioning, and maintenance of the system.

4.2 Individual responsibilities

AT&T (Bell Laboratories) was given the development responsibility for the following portions of the system: repeaters and devices for repeaters, repeater factory test sets, laying test sets, couplings and junction boxes, ocean-block equalizers, supervisory tone and repeater monitoring test set, shipboard and shore high-voltage power-feeding equipment, and order wire equipments for cable laying and burying.

BPO was given the development responsibility for cable, jointing methods, cable handling procedures, cable fault location test sets, cable factory test sets, and other associated test equipment.

The FPTT was given the development responsibility for the terminal transmission equipment (multiplex and wide-band line) between the supergroup distribution point and the transmission equipment side of the power separating filter, the order wire and associated terminal signaling equipment, and maintenance test sets.

4.3 Formal agreement

Arrangements were formalized in a document titled "Agreement for Shared Development of a High Capacity Submarine Cable System"

signed on May 2, 1973 although, by mutual agreement, work had started in 1971. The document contained descriptions of the various responsibilities, a method for distribution of development costs, means for handling technical information and inventions, and a list of tentative objectives for system characteristics.

V. PROJECT ORGANIZATION

An undersea cable system is a complex transmission system design consisting of many parts and processes, as is shown in the companion articles in this issue. Each part must have individual objectives, and the complete system must function to produce an end result that is both economical and reliable.

For these reasons, a relatively complex organization was required to guide and monitor the joint development work by technical representatives of the three countries working together for a common goal. Overall guidance was provided by a Development Steering Committee (DSC) consisting of two members each from the BPO, FPTT, and AT&T-BTL. The detailed work was carried out by seven working parties, each having from 8 to 10 members.

A second agreement titled "TAT-6 (SG) Cable Construction and Maintenance Agreement" was a vital part of the project. This agreement assumed the use of the SG system for TAT-6 and defined ownership interest, assignment of channels, responsibility for procurement, installation, etc. of this first SG link to be installed. It was signed by the appropriate telecommunications agency in 16 European countries plus AT&T, ITT Worldcom, RCA Globcom, and Western Union International. From this agreement, the TAT-6 General Committee was formed. A subcommittee, responsible for procurement and installation, was the main point of contact with the SG Development Steering Committee.

VI. METHOD OF OPERATION

Meetings of the development working parties were scheduled as required, and the location of the meetings was rotated among the three countries. Prior to each meeting of the Development Steering Committee, reports of the working parties were circulated for review. The report of each DSC meeting consisted of general discussion and agreements and the following specific sections for control of the project:

(*i*) Review of memoranda from the TAT-6 subcommittee B and memoranda to that subcommittee covering outstanding points.
(*ii*) Review of working party chairmen's reports and instructions to working parties.
(*iii*) Report to administrations.
(*iv*) Review and allocation of development costs.

VII. RESULTS

While the joint development seemed to be overly complex at times and introduced the added expense of traveling, it was a successful endeavor. The interchange of technical information contributed to the development of the SG system and was valuable to each organization for its future work.

The technical results of the development program are discussed in the succeeding papers. The decision to install a 3400-nmi transatlantic system as the first project was rather daring. Laying cable in the North Atlantic early in the year proved to be a hazardous adventure, as the weather refused to be cooperative and accounted for considerable delay and apprehension. Many last-minute decisions were required, but nevertheless, except for the fact that the system objective for intermodulation noise was not fully met, the project was successfully accomplished.

The causes of the intermodulation noise have since been established and development work actively pursued to eliminate these problems, both in future SG systems and by the use of special equipment at one terminal of TAT-6. Bell Laboratories intends to continue to improve the cost effectiveness of the system, and it is hoped that future cable system requirements can take advantage of the facility successfully developed as a result of this unique international cooperation.

REFERENCES

1. "SF Submarine Cable System," B.S.T.J. *49*, No. 5 (May–June 1970), pp. 601–889.
2. "CANTAT-2 Submarine Cable System," Post Off. Elec. Eng. J., *67*, Part 3 (October 1974).

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Requirements and Performance

By S. T. BREWER, R. L. EASTON, H. SOULIER, and S. A. TAYLOR

(Manuscript received September 8, 1977)

This paper reviews the major parameters of the SG Undersea Cable System. SG transmission goals and signal statistics were established, which led to a specific system and equalization design. As the SG design evolved, information on cable aging led to the introduction of four shore-controlled equalizers in the SG transatlantic link (TAT-6). Performance delivered by the TAT-6 link has been generally satisfactory, though some excess noise was encountered in the top third of the high band. Equalization results permit the link to handle 4200 two-way voice channels, 200 more than the objective.

I. INTRODUCTION AND BACKGROUND*

1.1 Traffic growth and forecast

Since 1956, when voice communication across the Atlantic by undersea cable began, growth in traffic has continued at a rapid rate.[1,2,4,5] The two upper curves of Fig. 1 indicate the growth of (*i*) total U.S. overseas phone messages and (*ii*) transatlantic phone messages. Data plotted here include traffic by satellite and radio, as well as by cable. Between 1960 and 1976, annual overseas phone messages have increased from 3.7 to 87.7 million. This represents an exponential growth of 22 percent per year. Transatlantic phone messages to Europe have increased from 1.0 million in 1960 to 29.7 million in 1976. Thus, Atlantic basin growth rate is 24 percent per year. Projecting 1976 traffic at growth rates experienced to date yields an expected 520 million total overseas phone messages in

* This paper broadly covers all aspects of the SG system. For greater detail on any particular facet of the system, the reader is referred to the appropriate subsequent article in this issue of the B.S.T.J.

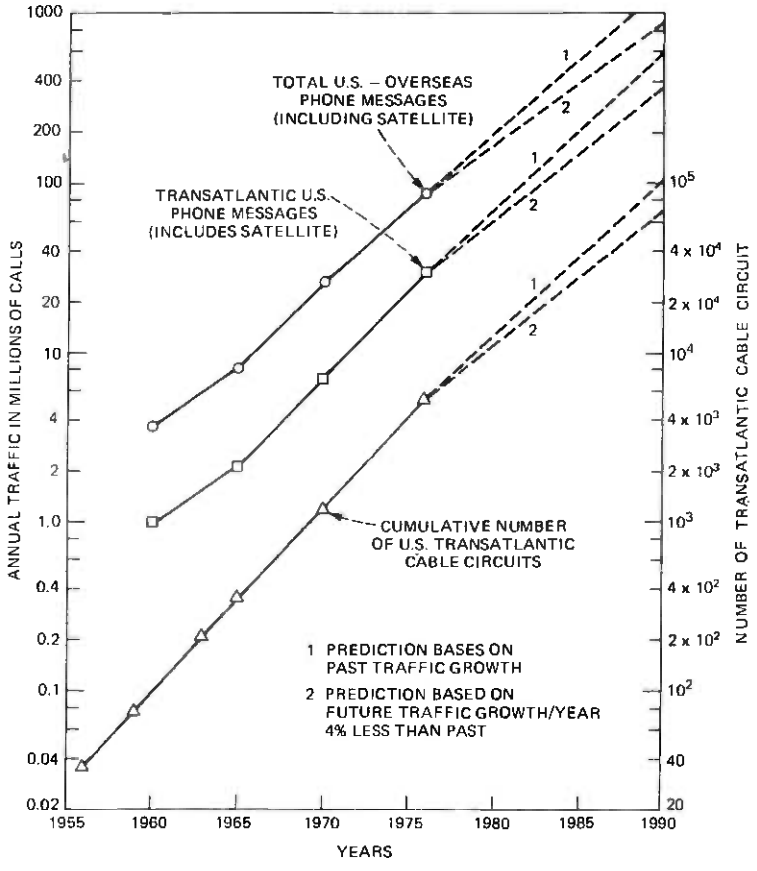


Fig. 1—U.S. overseas and transatlantic growth.

1985, of which 203 million would be transatlantic. These projections are plotted as the upper dashed curves of Fig. 1. If we assume that future growth rates will be 4 percent less than past ones, we get the lower dashed curves of Fig. 1. This more conservative forecast anticipates 385 million total overseas phone messages by 1985, of which 151 million would be transatlantic. The lower cost to the customer that can be expected to result from further economies of advanced technology, plus the convenience of direct subscriber dialing of overseas calls, will be powerful forces to continued rapid traffic growth.

1.2 Transatlantic cable circuit growth

The bottom curve of Fig. 1 shows the transatlantic undersea cable circuits installed to meet traffic needs. Beyond 1976, the circuit projection is at two rates, the past transatlantic traffic-growth rate of 24

Table I — U.S. transatlantic cable circuits

Actual Year	Cumulative Circuits*	Added Circuits	New System
1956	36	36	TAT-1
1959	72	36	TAT-2
1963	212	140	TAT-3
1965	352	140	TAT-4
1970	1172	820	TAT-5
1976	5372	4200	TAT-6

Forecast Year	24%/yr Growth	20%/yr Growth	Additional Cable Circuits Needed Beyond TAT-6 24%/yr	20%/yr
1980	12,700	11,140.	7330.	5770.
1985	37,200	27,720.	31,900.	22,300.
1990	109,200	68,970.	103,800.	63,600.

* Figures are for physical circuits, and do not include a limited number of TASI-derived circuits on installed systems.

percent per year, plus the more conservative rate of 20 percent per year.

Table I shows this growth in tabular form. The lower part of the table indicates the cumulative number of circuits we expect to need up through 1990. If the 24 percent growth rate is sustained, by 1980 another 7330 cable circuits will be needed beyond TAT-6, and by 1985 some 31,900 added cable circuits will be needed.* If we experience the more modest 20 percent per year growth, we will need 5770 additional cable circuits by 1980 and 22,300 new cable circuits by 1985.

With such exponential growth, it is interesting to consider the possible effect speech concentrators such as TASI (Time Assignment Speech Interpolation) could have on future demand. Moderate application of TASI could defer the forecasted circuit needs by one year. Fairly extensive application of TASI could yield a 2-year postponement. After the transient resulting from the introduction of TASI, circuit needs would be expected to grow at the same rate as before.

II. PLANNING THE SG SYSTEM

2.1 Repeater performance

Table II shows key repeater parameters of the SF and SG systems.† Although the SG repeater has five times the top frequency of the earlier SF repeater, it actually has slightly better performance. Better noise figure, linearity, and output power capability result from (*i*) use of silicon

* Future cable circuit needs shown in Table I assume that both cables and satellites will share future growth and that this sharing will continue on the basis of the present cable-satellite facility ratio.

† The appendix following this paper defines various terms and symbols which are used.

Table II — SF and SG repeater parameters

	SF	SG
Nominal top frequency	6.0 MHz	29.5 MHz
At repeater top frequency:		
Insertion gain	40.1 dB	41.0 dB
Noise figure	7.6 dB	3.5 dB
Maximum output power, rms single sine wave	20.0 dBm	23.0 dBm
Loss, amplitude output-to-cable	0.3 dB	1.2 dB
Modulation coefficients*		
M_{2E}	−65 dB	−70 dB
M_{3E}	−95 dB	−113 dB
Nominal impedance	59.4 ohms	50 ohms
Type transistor	Ge mesa	Si planar
Supervisory tone power at repeater output		
Low band	−60 dBm	−50 dBm
High band	−50 dBm	−40 dBm
DC current	136 mA	657 mA
DC potential drop	13.1 V	12 V

* Referenced at power amplifier output.

planar transistors rather than the earlier germanium mesa transistors, (*ii*) use of active rather than passive terminations at the amplifier's input and output ports, and (*iii*) more compact physical structure.

2.2 Channel capacity

The rapid growth in traffic and the state of electronic technology were major factors determining the channel capacity of the SG system. As the start of development, our target was 3500 channels; this goal was later shifted to 4000 channels. The actual U.S.-to-France SG installation satisfactorily equalizes 4200 channels.

2.3 Cable diameter

A separate article describes in detail the SG cable design.[6] Early in the development, studies considered the economics of various possible systems as a function of cable diameter. A larger diameter cable, while more expensive and more demanding of ship capacity, has less loss and therefore saves repeaters and equalizers. Thus, as systems increase in capacity and therefore top frequency, larger cables become more desirable. Our studies showed that cable of 1.7-in. dielectric diameter would yield near-minimum system cost. This size cable also offered the possibility of lower dc resistance, which would reduce terminal voltage.

2.4 Repeater configuration

Figure 2 shows the two repeater configurations considered for SG. Separate high-band and low-band amplifiers avoid the nonlinear sing problem[3] and make design somewhat easier. This configuration also prevents second-order distortion from falling in the high band. On the other hand, the single-amplifier configuration improves reliability by

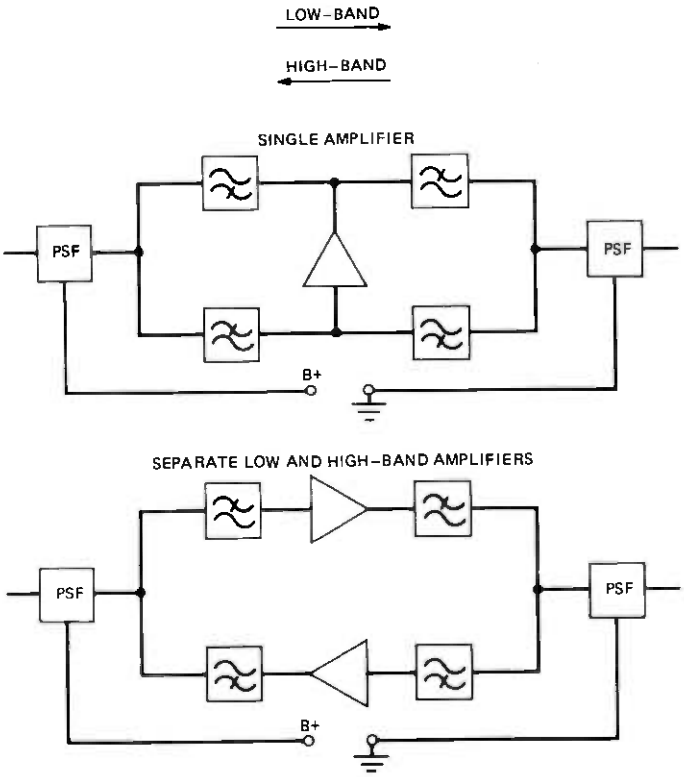


Fig. 2—Possible SG repeater configurations.

reducing the total number of components and devices. Exploratory studies indicated that the noise sing and the other design problems could be solved. These studies led to the choice of the single amplifier repeater configuration.

A side benefit of the single amplifier configuration is the use of intermodulation products from the repeaters to perform fault location and supervision, including detection of a badly modulating repeater. This supervision is done by a frequency-sweeping repeater monitoring set. In addition to the repeater monitoring set, the SG system includes a supervisory oscillator in each repeater, similar to the SF system arrangement.[7] Successive repeaters alternately have low-band and high-band quartz crystal supervisory oscillators by which each repeater may be identified. Since these supervisory oscillators inject known amounts of power, the individual supervisory powers received in the terminal also indicate transmission levels within the system.

III. GENERAL DESCRIPTION

SG is an equivalent four-wire coaxial cable system, with a nominal capacity of 4000 two-way, 3-kHz-spaced channels, and a maximum length capability of 4000 nautical miles (nmi). The TAT-6 physical layout is shown in Fig. 3.

SG frequency allocations are specified in Fig. 4. The nominal low and high bands shown correspond to adequate bandwidth for the 4000 channels and allow for supervisory tone bands, 8-kHz guard bands between supergroups, and three order-wire channels. In the development, transmission was extended as far beyond all band edges as possible without added complexity in the repeaters and without impairing transmission within the nominal band. Practically, this meant that attention had to be paid to transmission in the extended bands of 0.5 to 13.9 MHz and 16.1 to 29.5 MHz. The bandwidths actually achieved on TAT-6 (corresponding to about 4200 two-way channels) are also indicated on Fig. 4.

A noise objective of 1 picowatt per kilometer* applied to the average noise in the 4000 channels that form the basic system, with none of these channels to exceed 2 pW/km.

The ocean-block equalizers (OBEs) equalize the gain deviations present at the time of installation due to the difference between actual repeater gain and cable section loss. The equalizers obtain the gain needed for equalization by shortening to one mile the total cable length between the two repeaters adjoining an equalizer. This makes 80 percent of the gain of a repeater available for equalization. Additional equalization gain is provided in the low band by designing excess gain relative to cable loss into each repeater (low-band gain boost).

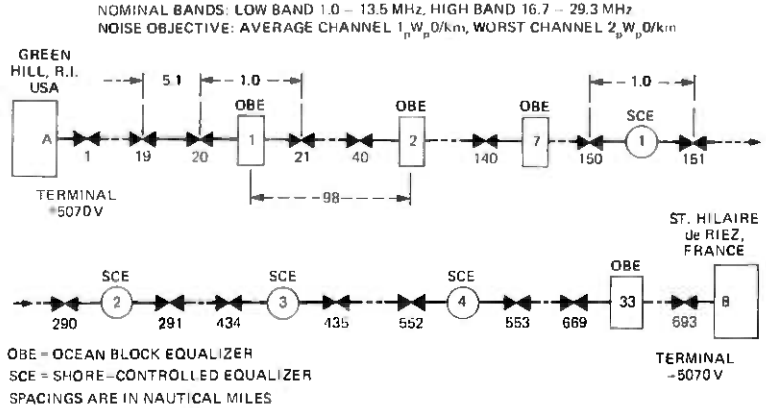


Fig. 3—TAT-6 SG link.

* The 1 pW0p/km channel noise objective corresponds to 38.5 dBrnc0 for TAT-6. Our "ideal" performance target was 36.5 dBrnc0. Thus, 2 dB of noise performance was allowed for the effects of misalignment, etc.

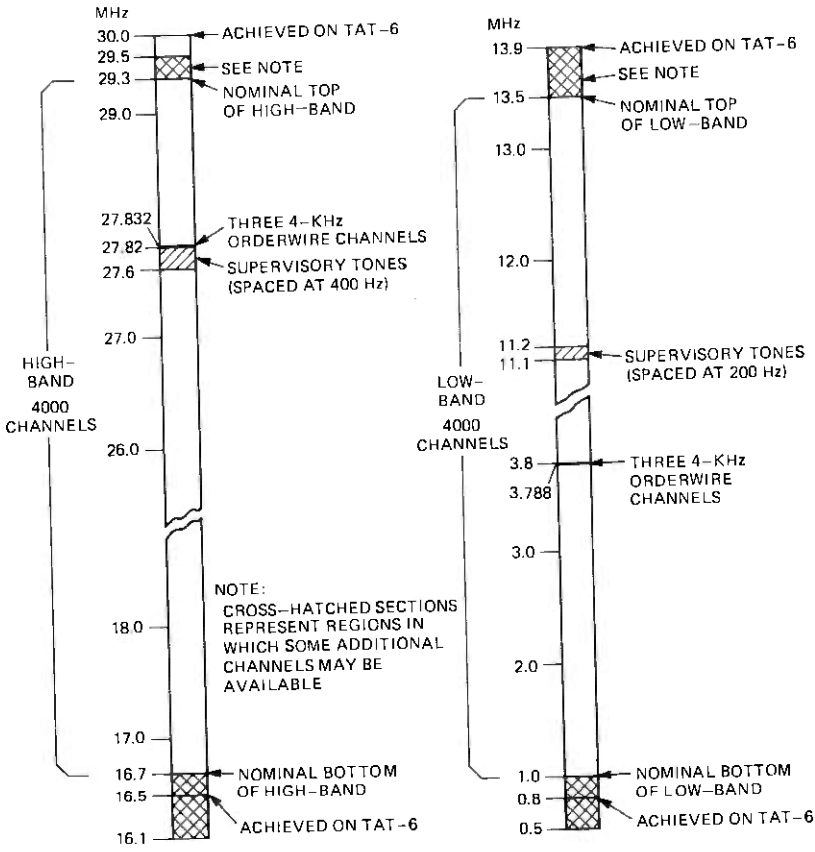


Fig. 4—SG frequency allocation.

Fairly late in the development, problems encountered during the manufacture of cable led to the discovery of phenomena that could cause the attenuation of installed cable to change with time (cable aging). Since the magnitude, frequency characteristic, and even the direction of such cable aging was uncertain, some means of adjusting the equalization along the length of the system subsequent to installation appeared to be required. On a crash basis a shore-controlled equalizer (SCE) was developed. Four such units were included in the TAT-6 link, dividing the system into five approximately equal-length "sectors." The SCEs obtain gain for equalization in the same way as the OBEs, as described above.

3.1 Transmission objectives and computed performance

The average power per channel, P_c, assumed in the SG system design is −13 dBm0. This is a conservative value for present signals and mixtures of voice and data. The −13 dBm0 figure would allow use of speech concentrators on about one-half the channels used to carry voice.

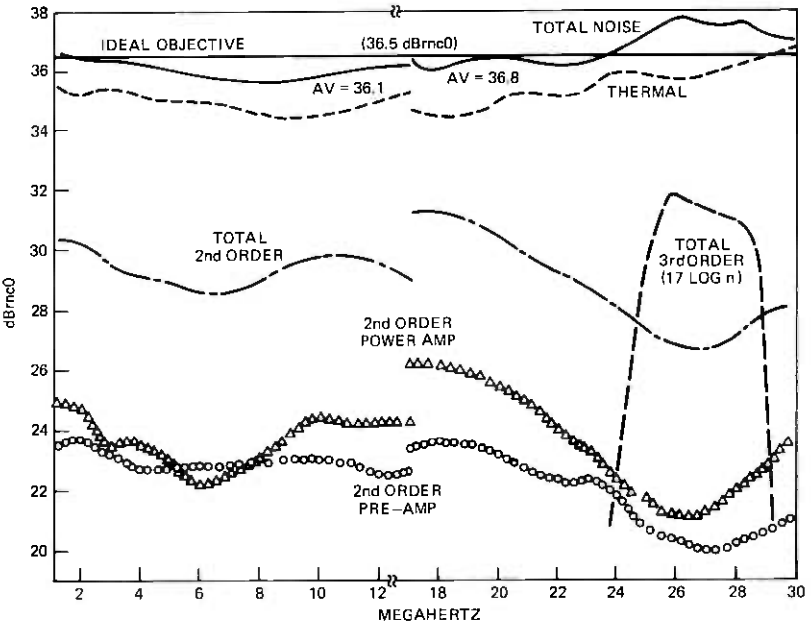


Fig. 5—SG system calculated noise performance (no misalignment, 3500 nmi, 700 repeaters, dBm0/channel = −13, overload margin = 6.2 dB).

Actually, initial convervatism and a continuing decrease in average talker and data volumes will make the −13 dBm0 average channel power adequate in the future even if there should be an increase in power on all voice channels due to the general use of speech concentrators.

A broadband load of 4000 two-way channels carrying the −13 dBm0 load per one-way channel has an average power, P_{av} of 26 dBm0 and a peak value of 39 dBm0.*

The anticipated repeater performance is discussed further in a subsequent section. From this performance, the assumed load, and the number of repeaters necessary to span the Atlantic (about 700), it is possible to compute the expected system noise performance for any assumed transmission levels across the two transmission bands. The goal is to arrive at optimum signal level vs. frequency shaping that corresponds to the minimum flat noise for a particular link. Final shaping is determined by experimentally shaping signals and evaluating the results by system noise power ratio tests.

Figure 5 is the computed noise performance of an ideal system (with only the intentional gain boost appearing as misalignment), corresponding to the transmission levels shown in Fig. 6. The objective for

* Average load = $-13 + 10 \log_{10} (2 \times 4000) = +26$ dBm0. The signal has peaks extending 13 dB higher than this 0.001 percent of the time. This peak-to-average ratio is the multichannel peak factor, k_m.

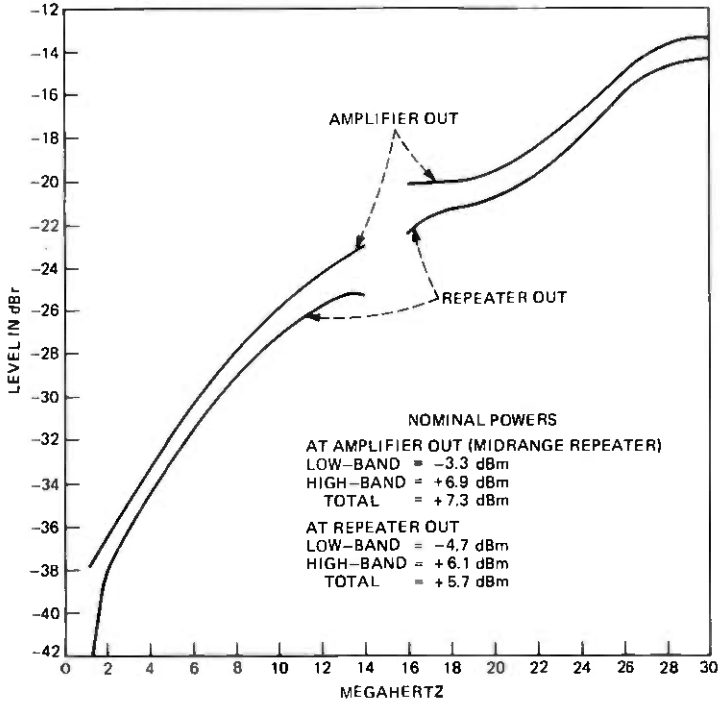


Fig. 6—Nominal transmission levels.

a 3400-nmi transatlantic length on this basis is 36.5 dBrnc0, to allow the 2-dB performance margin mentioned earlier. The results of Fig. 5 show that the estimated noise performance is approximately the desired 36.5 dBrnc0.

As stated in Fig. 5, the overload margin to take care of the highest level repeater is about 6 dB.* A *total* flat misalignment of 12 dB would use up about 1.3 dB of the performance margin. This misalignment is the maximum that can be equalized *equally* at the transmitting and receiving terminals. To avoid overloading any repeaters, additional misalignment would have to be equalized entirely in the transmitting terminal if the misalignment is gain, or entirely in the receiving terminal if the misalignment is loss.† Such nonsymmetric equalization results in considerably greater misalignment penalty than does symmetric equalization.

3.2 Equalization plan

The first level of equalization in the SG Undersea Cable System is to

* That is, the zero-misalignment repeater operates at an average signal power output which is 6 dB below the repeater's capability.

† Ripples across the transmission load that average out (i.e., leave the broadband load unchanged) can always be symmetrically equalized.

match the gain of the repeater to the computed loss of 5.1-nmi of cable under the nominal condition of 2.5°C and 2.5 kilofathoms (kF).

The next level of equalization is provided only in shallow-water repeaters that are subject to seasonal temperature variations. These repeaters have a temperature-sensitive gain characteristic that will compensate for loss changes in the cable resulting from ambient temperature variations.

After a route for a particular installation has been selected, cable sections are ordered from the factories by 29-MHz attenuation, not by length. The idea is to obtain a top-frequency cable section attenuation at sea bottom which corresponds to the mean top-frequency repeater gain. The ordered loss which a particular section is to measure in the factory (at 10°C, 0 fathoms) is computed so that the attenuation of this section will match repeater gain at the depth and temperature corresponding to its position in the system. This procedure thus tends to correct for manufacturing deviations and differences between actual sea-bottom conditions and the nominal conditions (2.5°C, 2.5 kF) assumed in the repeater gain objective.* The accuracy with which this correction can be done depends, of course, on factory measurement accuracy and the accuracy with which temperature coefficients, pressure coefficients, handling or laying effects, and actual seabottom temperatures and depths are known. The coefficients used to predict the seabottom loss are refined, and the accuracy with which the loss is matched to repeater gain can be expected to improve as experience is gained from actual SG installations.

3.2.1 Ocean-block equalizer

In the SG system, a group of 30 repeaters is followed by an ocean-block equalizer (OBE).† Each OBE is adjusted just before being laid and is designed to compensate most of the transmission deviations that have accumulated up to this point in the installation. The OBE contains directional filters that permit the independent equalization of the two transmission bands. The equalization is effected by fixed (mop-up) networks and switchable networks. In addition, each band has a build-out network. The build-out networks are designed so that the total OBE loss is at its nominal value when the switchable shapes are at mid-range.

* Superposed on the procedure described is a small cyclical length perturbation to reduce the systematic addition of interaction ripple due to finite repeater return losses.

† Because of uncertainties associated with first-time installation, the TAT-6 link has an OBE following every 20 repeaters.

This arrangement allows an OBE to provide either system gain or loss around a nominal setting.*

3.2.1.1 Mop-up networks. The mop-up networks are bridged-T designs whose purpose is to equalize deviations known at the time the equalizer is manufactured.† The mop-ups are built from a stockpile of pre-aged components on standardized circuit boards, with maximum utilization of special-purpose computer programs. The aim is to minimize the time required to incorporate new knowledge about cable or repeaters into the OBE characteristics.

It is possible, for example, to incorporate knowledge acquired in the first lay (shipload) of a system in the equalizers being manufactured for the third lay. In fact, if there is sufficient reason, an equalizer can, on an expedited basis, have special mop-ups inserted and be ready for use in the next lay. It should be made clear that different equalizers can contain different mop-up networks. Up to seven mop-up networks may be used in each band. The excess gain available for mop-up networks is 9.5 dB in the low band and 12.5 dB in the high band.

3.2.1.2 Switchable networks. The shapes available in the switched networks are shown in Fig. 7; they are relatively smooth and are related to attenuation versus frequency characteristics associated with the undersea cable. (Figure 7 shows the equalization shapes. The actual bump and $\sqrt{f} - f$ networks have some added $\sqrt{f}$ for physical realizability.) The functional forms are listed in a companion article.[8] Each of seven switchable networks in each band can be independently switched either in or out of the transmission path through the equalizer ($2^7 = 128$ combinations are available in each band). The best combination in a particular equalizer is selected on shipboard just prior to overboarding that equalizer. The choice is based on minimizing the misalignment indicated by transmission measurements made during laying. The measured transmission path extends from the shore terminal through the equalizer whose setting is to be determined.

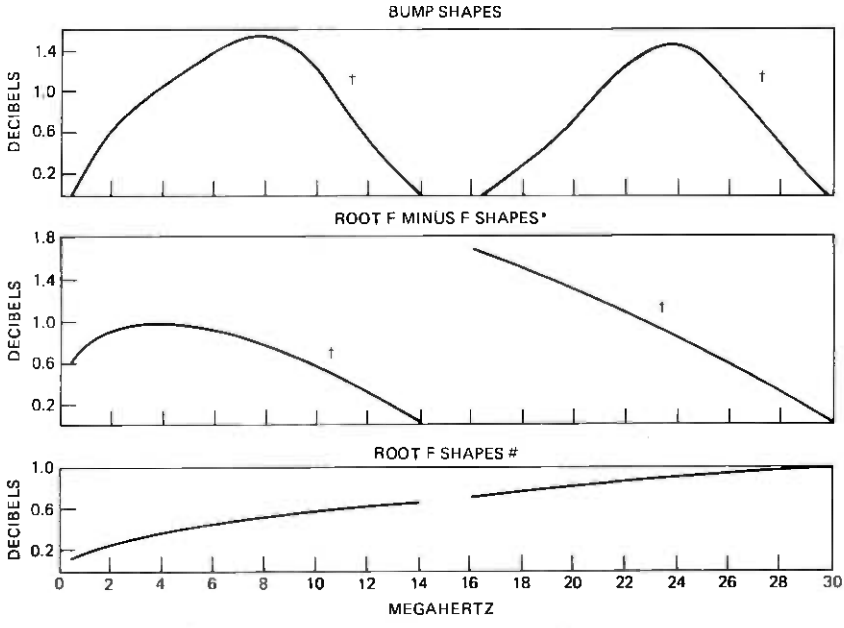
The available amount of shape having a square-root-of-frequency characteristic ($\pm 7.5\sqrt{f/30}$) is somewhat larger than the misalignment allowed in the misalignment allocation discussed later. However, the large noise penalty which results if equalization range is exceeded dictates a comfortable margin.

3.2.2 Shore-controlled equalizer

As mentioned previously, four shore-controlled equalizers (SCE), di-

* The source of any gain, of course, is an adjoining repeater. In effect, the switchable networks may be set so that the total OBE loss is less than the loss of the 4.1 nmi of cable which are omitted from the section containing an OBE.

† Known deviations whose shape is well matched by the broad switchable network shapes can be equalized by switchable network settings during laying. In using the switchable networks to aid in the mop-up function, one must leave enough range to handle deviations between predicted and actual sea-bottom cable transmission.



*ADDITIONAL NETWORKS ARE PROVIDED, WITH 2 TIMES THE LOSSES INDICATED

ADDITIONAL NETWORKS ARE PROVIDED, WITH 2 TIMES, 4 TIMES, AND 8 TIMES THE LOSSES INDICATED

† THE CURVES SHOWN ARE THE DESIRED FUNCTIONAL SHAPES. ACTUAL SHAPES PROVIDED HAVE SMALL AMOUNTS OF $\sqrt{f/30}$ ADDED TO MAKE THEM PHYSICALLY REALIZABLE.

Fig. 7—OBE switchable shapes.

viding the system into five sectors, were included in the TAT-6 link to compensate cable aging. These equalizers contain directional filters, build-out networks, and switchable networks, but do not incorporate the mop-up function. The switchable shapes are the same as those used in the OBE but with about twice the decibel range. The larger range is possible because no gain is allocated to the mop-ups.

The ranges provided by the SCEs in TAT-6 are adequate to handle up to about 75 dB of top-frequency aging in either direction (i.e., gain or loss) and are consistent with the nonlinear sing margins discussed later in this paper. This amount of aging exceeds our forecasts based on modeling of aging to date. System aging is discussed further in Section IX.

The adjustment of SCEs is controlled from the low-band transmit terminal by a pair of control frequencies unique to each equalizer. A first control frequency is used to select the network to be switched in or out, and the second activates the switching. Second harmonics of the transmitted tones are returned to the terminal. By tuning the frequencies of the control tones to maximize the amplitudes of the returned second harmonics, it is possible to center the control frequencies in the pass-

bands of the SCE's crystal filters. Also, level changes of the returned second harmonic are produced by padding switched in by contacts on a deck of the network-select switch. These level changes indicate which SCE network is being addressed.

By the time deep-sea laying commenced on TAT-6, indications were fairly clear that, at least initially, aging would be in the direction of increasing loss. Thus, the OBEs were adjusted to be a little deficient in loss initially, and each SCE was adjusted to compensate this intentional misalignment, which amounted to 2 dB excess gain per sector at the top frequency.

3.2.3 Terminal equalization and protection

The final level of equalization is provided in the shore terminals. A combination of smooth adjustable shapes, custom-designed residual equalizers, and regulation in the receiving multiplex holds transmission levels within acceptable limits.

At the output of the transmitting wideband line, a special circuit switches in extra loss if an excessive broadband signal is sensed. This is done to provide adequate protection against an accidental gross overload which might be capable of inducing gain changes in the undersea repeaters.

3.2.4 Misalignment allocation

In the design of a system, it is necessary to set misalignment objectives consistent with performance objectives and equalization ranges. These misalignment allocations then determine the permissible tolerances for the various contributors to misalignments. Table III shows the SG misalignment objectives, and Table IV shows the unit accuracies required to meet these misalignment objectives.

The initial uniform misalignment is in the form of sharp ripples, whose shapes preclude their equalization by the simple mop-up networks of the OBE. Sharp ripples have little effect on multichannel load and therefore can be equalized equally at the two shore terminals without cutting into the overload margin. Those ripples which are so sharp that they cannot be equalized in the broadband portion of the shore terminals may be eliminated by supergroup equalizers in the receive terminal. In the latter case, positively misaligned channels would be quieter than most, while negatively misaligned channels would be noisier than most.

3.2.5 Noise penalties due to misalignment

In estimating a noise penalty associated with the allocations of Table III, it is assumed that repeaters are fairly uniformly spread across the 5-dB level range of item I. A few repeaters that fall outside this limit on

Table III — Top-frequency* misalignment objectives for 4000-nmi link

I. Initial within-block misalignment (not including effect of OBE)	
A. Maximum difference between highest- and lowest-level repeater in block	5 dB
B. Maximum difference between extreme-level repeaters of this block and extreme-level repeaters in rest of system	5 dB
C. Allocation of above	
1. Repeaters—total in any block	±1 dB/block
a. Systematic (within each block)	±0.02 dB/repeater
b. Random (within each block)	±0.07 dB/repeater
2. Cable	
a. Same in every block (e.g., error in coefficient)	+2.5 dB/block or −2.5 dB/block
b. Different in every block (error in sea-bottom temperature or pressure and manufacturing variations)	±1.5 dB
II. Uniform misalignment along length of system	
A. Initial† (±0.2 dB/block for a maximum-length system with 30-repeater blocks)	±5 dB/system
B. Aging	±5 dB/sector‡

* Allowance at lower frequencies obtained by multiplying top-frequency values by the smaller of 1.5 or $\sqrt{30/f}$ and does not include intentional low-band boost.
† Assumed to be ripply and thus to have negligible effect on overload margin.
‡ The stated range is consistent with the system noise objective. A total equalization range consistent with up to 75 dB of system aging is provided in the SCEs.

Table IV — Accuracy objectives to meet misalignment objectives for a 4000-nmi SG link with 30-repeater ocean blocks

	dB per unit	Percent
I. Repeater		
A. Average deviation from objective within a block (known and equalizable in OBE)*	±0.02/repeater	±0.05
B. Random (within each block) deviation†	±0.07/repeater	±0.15
C. Unequalizable (from Table III, item II A, ±5/800 repeaters = ±0.006 dB/repeater) (Cause: error in knowledge of gain characteristic, especially any component too sharp to be equalized by the OBE's mop-up networks)	±0.006/repeater	±0.015
II. Cable loss deviations		
A. Average sea-bottom loss deviations (causes are errors in the average: temperature coefficient, pressure coefficient, and laying effect) Limits on total error*	±0.08/section	±0.2
B. Random variations (block to block) (causes are: manufacturing variations, errors in seabottom temperature and pressure, and measuring error) Limits on total error*	±0.05/section	±0.1

* Increased by a factor of 1.5 if block length = 20 repeaters.
† Increased by a factor of 1.2 if block length = 20 repeaters.

the low-level side do not have much effect on system noise performance. On the other hand, a few high-level repeaters, appreciably higher than most repeaters, cut into overload margin, and may cost almost decibel for decibel in performance. This must be kept in mind in equalizing the system as it is laid. In other words, if a high-level repeater occurs early in the installation, subsequent equalization should bring repeaters as close as possible to this level. As the installation approaches completion and a level range has been established by the repeaters already laid,

special care should be taken in equalization so that the levels of subsequently laid repeaters do not exceed the maximum of the established range. With these qualifications, a broadband misalignment of 10 dB would increase the ideal noise of Fig. 5 by only 1.25 dB. The ripply uniform misalignment, insofar as it is equalizable in the terminal, could bring the penalty up to 2 dB. The unequalizable portion might degrade a few channels by about 3 dB, improving others by about the same amount.

3.2.6 Nonlinear sing margin

In addition to its effect on noise performance, misalignment is of interest in connection with nonlinear sing margin.

Nonlinear sing is a phenomenon which is possible in equivalent-four-wire systems using a single amplifier for both directions of transmission. Under conditions of gross overload (usually in the presence of large misalignments), these systems can lock themselves up in a so-called noise sing.[3] Under this situation, the nonlinearities result in so much cumulative energy shift from the low to the high band that the energy shifted to the high band can overload that band. The overloaded high band can, in turn, via the common amplifier, produce enough energy shift back to the low band to sustain the whole process.

Detailed computer studies were made using empirical data on repeater behavior under overload for a variety of conditions. It was found that the system would remain stable for superposed misalignments up to 15 to 20 dB.

IV. CABLE

The bulk of any SG system is made up of 1.7-in. core diameter armorless deep-sea cable. The portions on continental shelves are largely composed of single armored cable. The portion from the terminal to the shore as well as the first mile or so out from the shore is shielded and uses a smaller diameter dielectric. The shielding is protection against radio interference in the portions of the system that are exposed to radio fields.

Out to depths of 640 m (350 fathoms), in addition to using armored cable, the cable and repeaters of TAT-6 are buried approximately 0.6 m (24 in.) beneath the ocean floor. These measures should provide excellent protection against cable breaks due to trawlers and dredges.

The attenuation of the 1.7-in. cable at seabottom conditions (2.5 kilofathoms and 2.5°C) is given approximately by ($\alpha = 1.383\sqrt{f} + 0.0178f$ dB/nmi), where f is in MHz. The square-root term is due to resistive losses in the copper, whereas the linear term is caused by dissipation in the polyethylene. At the top frequency of 30 MHz, total loss is about 8.1 dB/nmi. Of this loss, 0.53 dB, or 6.6 percent, is due to dissipation in the polyethylene.

V. REPEATER CHARACTERISTICS

5.1 Supervisory arrangements and insertion gain

The SG repeater consists of a two-stage preamplifier, a three-stage power amplifier, an output band-limiting filter and shaping network, directional filters, power-separation filters, and a supervisory oscillator. Each supervisory oscillator has its unique frequency, which may be used to identify its associated repeater, and to pinpoint undersea system problems. Normally, repeaters with low-band oscillators alternate with repeaters having high-band oscillators.

The gain objective for the SG repeater is the sum of the loss of 5.1 nmi of cable at 2.5 kilofathoms and 2.5°C and the low-band gain boost. The gain shape is shown in Fig. 8.

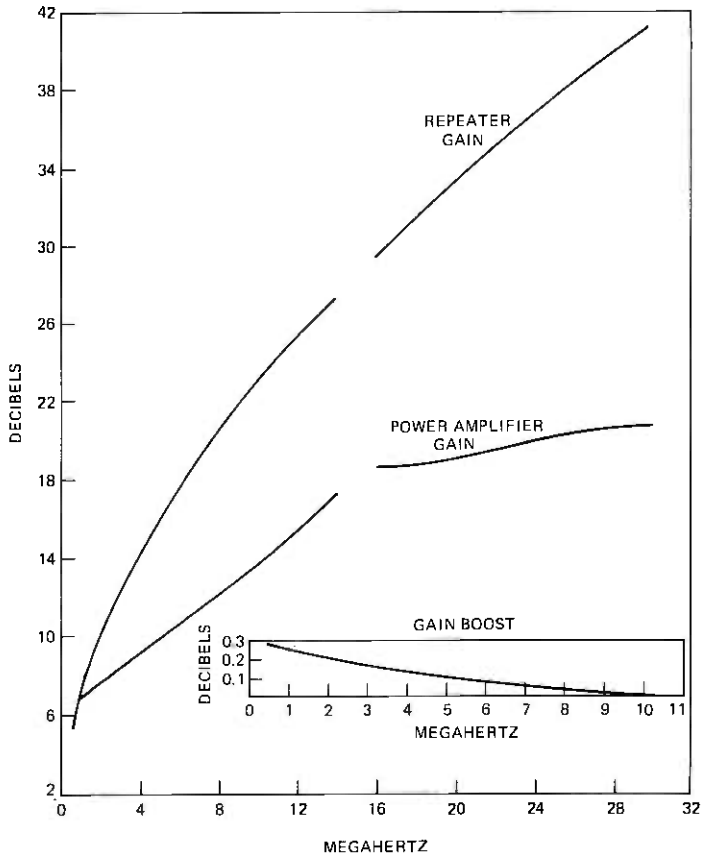


Fig. 8—Repeater and power amplifier gain.

5.2 Noise and modulation performance

Figures 9 through 15 show the values of the various parameters* which, along with repeater gain, determine the noise performance of the system. As was indicated in Fig. 5, the pre-amplifier and power amplifier contribute about equally to the second-order intermodulation noise. The total for the repeater is the voltage sum of the noise produced by the pre-amplifier and power amplifier. The power amplifier is the dominant contributor of third-order noise. The computed contribution of the pre-amplifier raises the total amplifier noise by less than ½ dB, even when the two sources are added on a voltage basis.

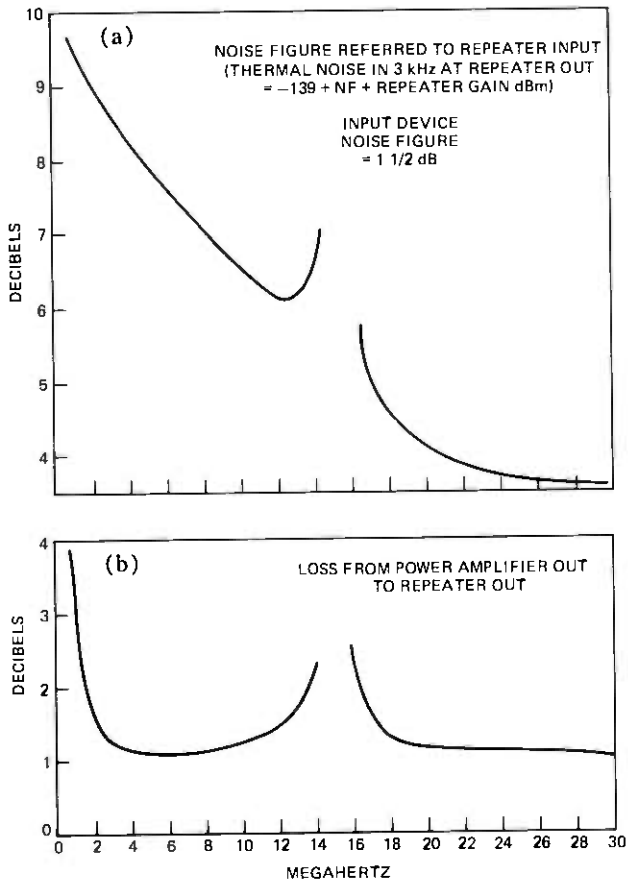


Fig. 9—(a) Repeater noise figure. (b) Output loss.

* These data are the ones determined prior to the installation of TAT-6 and used for the noise computation results of Fig. 5. As discussed later in this paper, the M_{3E} values shown were better than actual repeater performance over the upper portion of the high band.

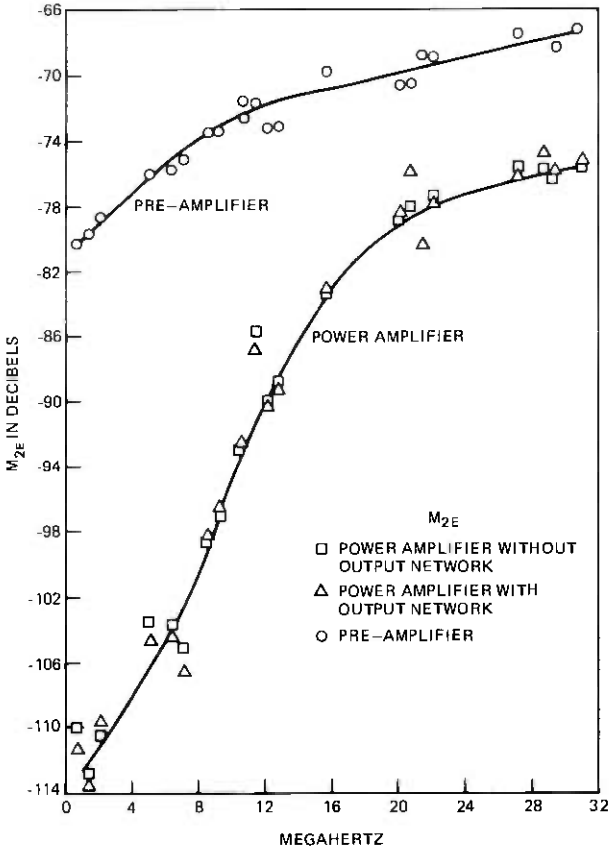


Fig. 10—M_{2E} repeater modulation coefficients (referred to POWER AMP OUT and to PRE-AMP OUT, respectively).

5.3 Load-carrying capacity

Figure 13 shows the behavior of M_{2E} and M_{3E} coefficients as a function of peak repeater output power. Four frequencies are used in the test to give a reasonable peak factor, even though only two or three of the four sine waves are involved in the measured product.

If the repeater behaved according to the classical Taylor series model, the curves displayed on Fig. 13 would be horizontal lines, i.e., intermodulation coefficients would be constant. Actually, the coefficients deviate from this ideal picture as the load is increased, eventually going off in the unfavorable (upward) direction. One can define the load capability in terms of a departure of the coefficients by a few decibels from their low-level values. One can conclude from Fig. 13 that the peak power capability of the SG repeater is approximately 26 dBm.

To obtain the results shown in Fig. 14, the circuit is loaded with frequency-shaped thermal noise except for a slot at 27 MHz. In this slot,

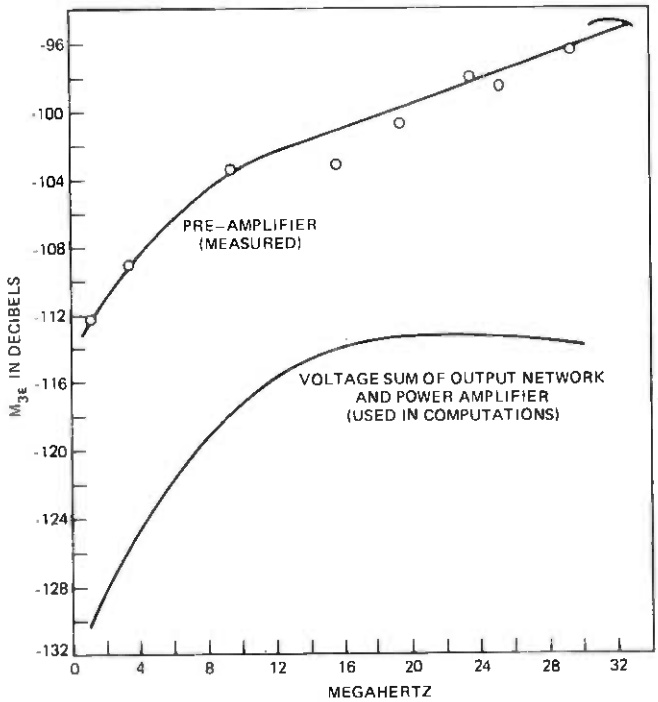


Fig. 11—M_{3E} repeater modulation coefficients (referred to POWER AMP OUT and to PRE-AMP OUT, respectively).

the intermodulation noise of a single repeater is fed into a 1.74-kHz bandwidth detector and then into a multi-channel amplitude analyzer.

The resulting amplitude distributions are plotted for a number of repeater output powers in Fig. 14. As the curve shows, at +14 dBm, very slight departures from a Gaussian distribution are noted. By the time the test signal has been raised to +15 and +16 dBm, gross departure from a Gaussian distribution is observed. Although the results are not plotted in Fig. 14, for test signals of +13 dBm and lower, the amplitude distribution of intermodulation noise is indistinguishable from a Gaussian distribution.

The test results shown on Fig. 14 prove to be a very sensitive means of determining the repeater's load capacity. Note that a random noise test signal of +13 dBm corresponds to a peak power of +26 dBm or an rms sine wave power of +23 dBm. Thus, the load capability determined from Fig. 14 agrees with that determined from Fig. 13 and the value used in the system layout shown in Table II.

Finally, Fig. 15 presents the limits on gross overload; exceeding these would risk inducing repeater gain change. These last limiting powers, which exceed considerably the load capacities defined by Figs. 13 and

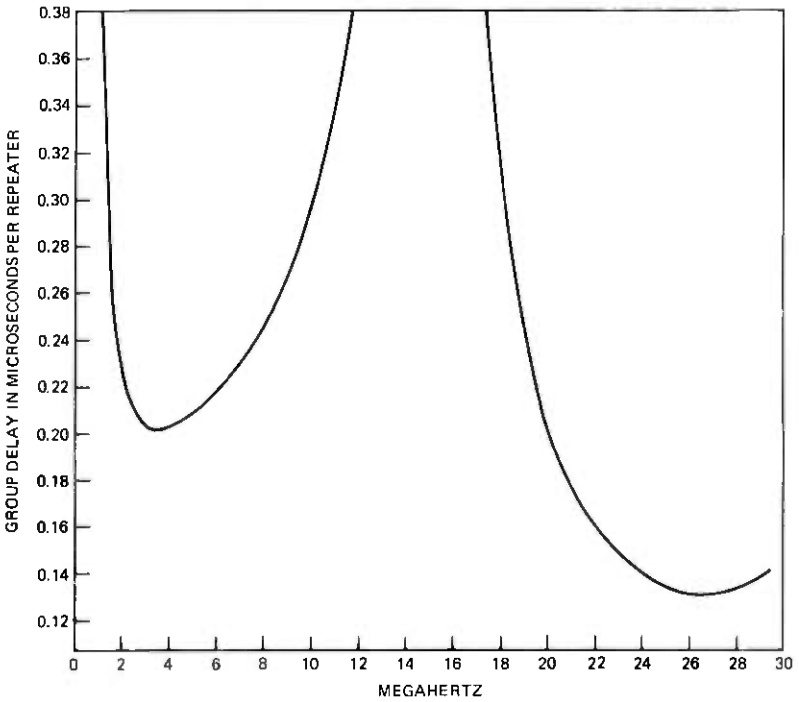


Fig. 12—Repeater group delay.

14, form the basis for adjusting the SG terminal's undersea system protection circuit, which was referred to in Section 3.2.3 and is covered in greater detail in Section VII.

VI. RELIABILITY OBJECTIVES

Newer generations of undersea cable systems pose ever more challenging problems. Between one generation and the next, the number of repeaters for a given length approximately doubles. At the same time, the complexity of each repeater increases. Despite these trends, the newer generation goal is to achieve approximately the same system reliability as that of earlier systems.

In the SG development, reliability was enhanced by a number of measures. Throughout development, mechanical design proceeded hand-in-hand with electrical design. (This was necessary for both performance and reliability reasons.) Highest quality raw materials were used. Inspection extended all the way from raw materials to the finished product. The 30-MHz top frequency required components and devices which were close to the state of the art. Each such new item was accompanied by its own carefully thought-out reliability testing and inspection program tailored to the new art.

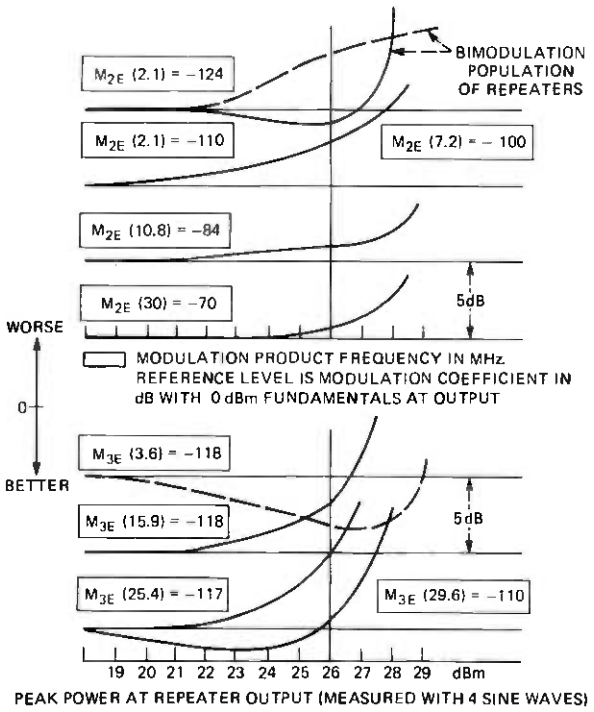


Fig. 13—Repeater excess modulation vs load power.

Early in the project, studies allocated reliability targets to various devices and components. Table V compares the device and component objectives of the SG system with reliability achieved in earlier generations of undersea cable systems. Note that the design objectives of the SG system correspond approximately to those which have been achieved in field experience on the earlier systems.

Achievement of the SG reliability targets would result in a repeater electronics reliability of 21 FITS. On the basis of experience, one might expect an approximately equal number of failures associated with other portions of the undersea link. If this proved to be the case for SG, the TAT-6 link would experience a mean system rate of 0.26 failures per year.

At the time of writing, TAT-6 has functioned reliably with no interruption since it went into service in July 1976.

VII. SHORE TERMINALS

Functionally, the shore terminal consists of: (*i*) terminal transmission

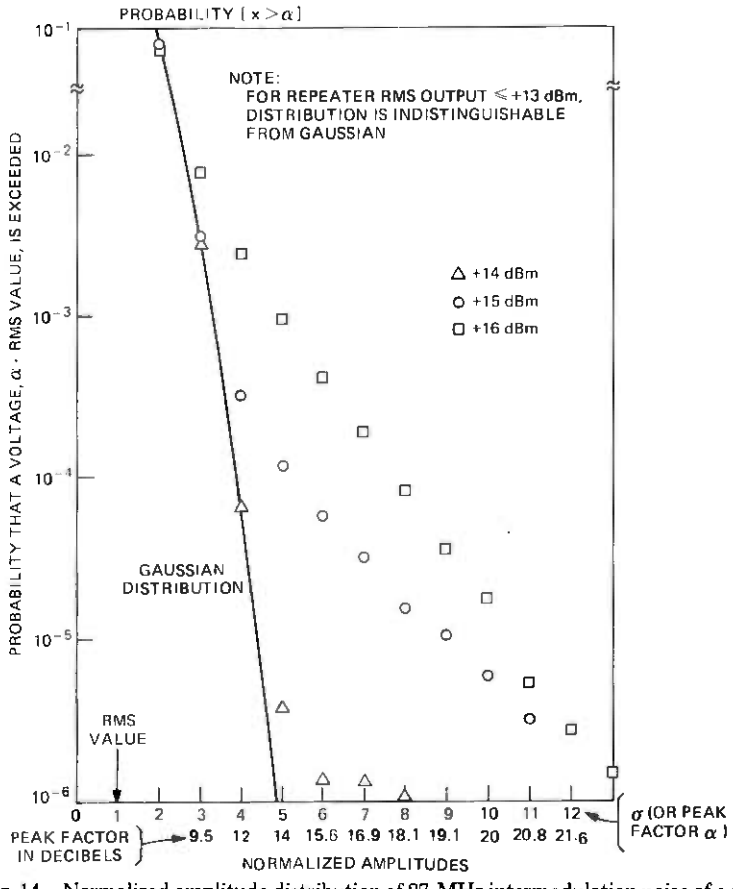


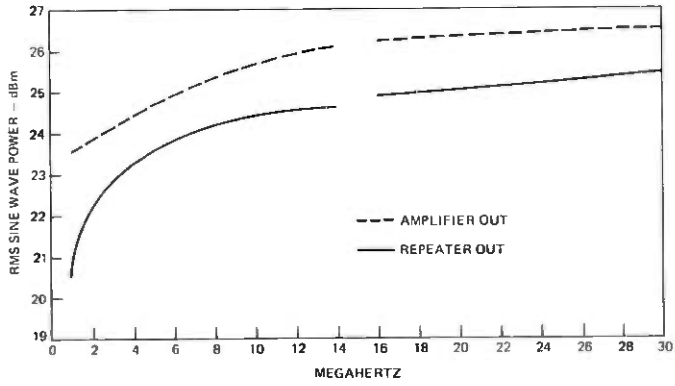
Fig. 14—Normalized amplitude distribution of 27-MHz intermodulation noise of a single repeater as a load-capacity defining phenomenon.

equipment made up of the multiplex, wideband line, and pilot monitoring equipment, (*ii*) power feed equipment for the undersea repeaters, and (*iii*) fault location measuring equipment.

From a transmission point of view, the SG system is defined from transmit supergroup input to receive supergroup output. The multiplex (and its associated carrier supply) consists of supergroup and hypergroup* frequency translation. Terminal wideband lines provide equalization capability to optimize noise performance and provide end-to-end flat transmission.

Signal limiting is provided at the input of each transmitting super-

* In SG, the hypergroup consists of an assemblage of 10 supergroups. Hypergroup 1, an exception, contains 14 supergroups.



NOTES:
1. RMS BROADBAND POWER OF A MULTI-CHANNEL SIGNAL AT AMPLIFIER OUT SHOULD BE LIMITED TO + 18 dBm.
2. EXCEEDING LEVELS SHOW IN GRAPHS CAN RESULT IN REPEATER GAIN CHANGE

Fig. 15—Maximum permissible single sine-wave power.

Table V — Reliability of active and passive components

Component Type	Component Reliabilities in FITS*	
	SG Design Objectives	Field Experience†
Active device (transistor or electron tube)	1.0	0.9
Diodes	0.5	0.0
Passive components	0.05	0.06

* FITS are defined as the number of failures in 10^9 component hours.
† Basis is failures observed on the earlier generations of U.S. systems designated SB, SD, and SF.

group. In addition, a protection arrangement in the transmit wideband line introduces added loss if necessary to avoid gain change in undersea repeaters due to gross broadband or single-frequency overload. (Figure 15, referred to earlier, shows these repeater limits.) This protection is adequate, even in the presence of positive misalignment.

The power feed equipment provides a current-regulated source of 657 ± 0.3 mA to power the undersea repeaters. It is capable of providing up to 7500 volts and shuts down in the presence of voltages and currents capable of damaging the system.

Fault location equipment consists of general purpose dc and low-frequency test sets, and a repeater monitoring set. The repeater monitoring set is capable of measuring: (*i*) repeater supervisory tones and (*ii*) "echoes" produced by high-level test signals acting on repeater nonlinearities.

VIII. ACHIEVED PERFORMANCE

8.1 Transmission equalization

Initial misalignments encountered during installation of TAT-6 were generally within the allocations of the system plan. Figure 16 shows the within-block misalignment (i.e., net gain or loss from the first repeater to the last repeater in an ocean block, ignoring the intentional low-band gain boost) averaged over all the blocks. Comparison of the two curves of Fig. 16 shows the effect of OBEs in reducing misalignment. The broad component of pre-OBE misalignment was less than ±2 dB in most blocks and was about ±4 dB in the worst block. The ripply structure was somewhat larger than expected and hard to predict since it was mostly a result of reflection at the repeater termination. It is a function of the exact length of the pigtail (flexible coaxial lead) between repeater and cable termination and is different for a pigtail surrounded by water as compared to one in air. The actual pigtail length depends on the relative orientation of repeater and termination aboard ship at the time of splicing. The ripple is very small per termination (and therefore hard to measure) but becomes important when many terminations occur in series (1400 in a system).

The solid curve of Fig. 17 shows the transmission of the undersea system from first to last repeater. The dashed curve in the figure shows transmission from a flat level point in one terminal to that in the other. Comparison of the two curves shows the effect of the terminal broadband equalization. Beyond this, there is still equalization at the receive supergroup level, where required, to flatten out the transmission.

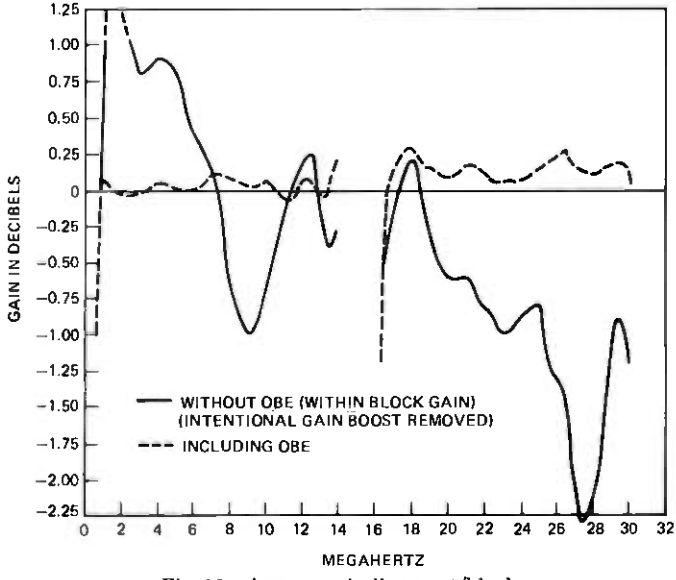


Fig. 16—Average misalignment/block.

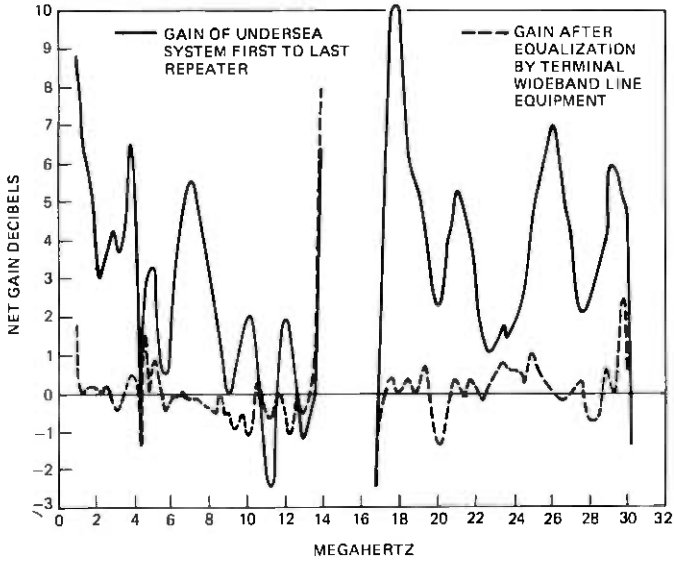


Fig. 17—Gain of TAT-6 before and after equalization by terminal wideband line equipment.

Figure 18 shows some level diagrams along the length of the system at some sample frequencies to illustrate SG equalization as a function of length. Most of the diagrams show behavior at ripple maxima and minima, with only the cases at 10.5 and 22 MHz representing behavior near zero crossings. The sawtooth pattern at 2 MHz is due to the gain boost. The data in these figures are obtained from laying measurements.

8.2 Noise performance

Noise performance of a new system is measured and optimized during commissioning[9] by means of noise loading tests. Commissioning refers to the large number of tests and adjustments required to carry a system from the time of the final splice to the time of initial commercial service. In these tests, the multichannel load is simulated by a noise "signal," and the system noise is measured in selected frequency slots, which can be cleared of signal by inserting band-stop filters in the transmitting terminal. The stop filters are sufficiently narrow to have negligible effect on the total load. A signal-to-noise ratio is determined by the ratio of the received power in the selected slot without and with the insertion of the corresponding band-stop filter (this ratio is referred to as the noise power ratio, or NPR). Transmitting levels are adjusted to give the best achievable signal-to-noise ratios.

Initial noise loading tests on TAT-6 were made with the noise power

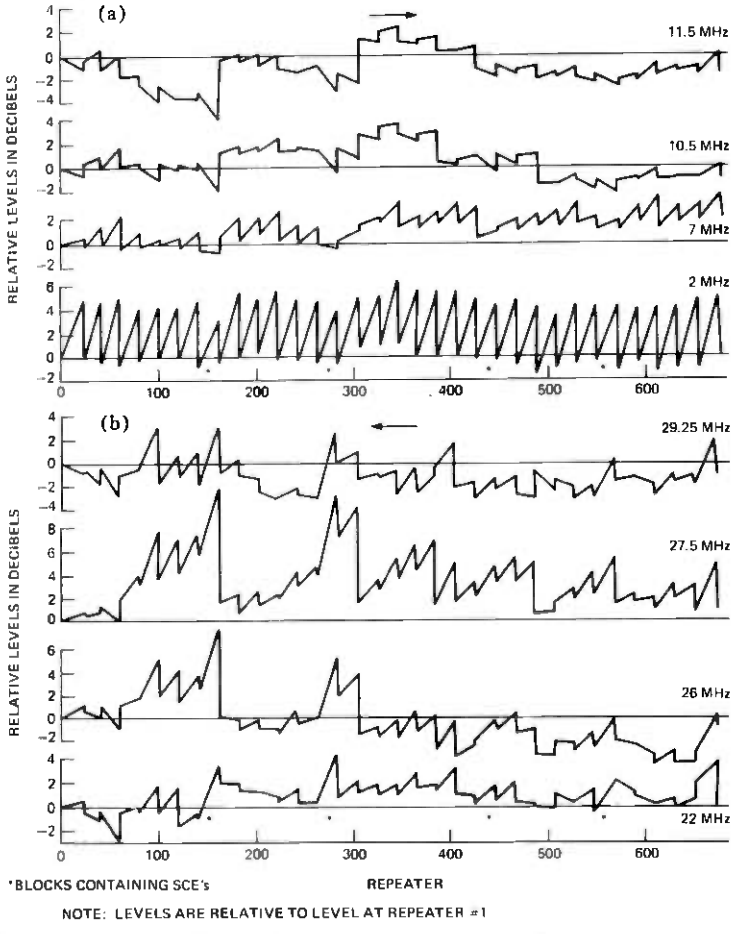


Fig. 18—Levels along length of TAT-6 at selected frequencies. (a) Low band. (b) High band.

density shaped vs frequency in such a way as to realize near-optimum noise performance if the repeater behaved in the expected manner. Actually, in these tests, the low band and the lower two-thirds of the high band came out largely as expected. However, the top one-third of the high band was noisier than anticipated. After considerable time was spent verifying that this was due neither to a measuring problem nor to a few repeaters suffering from some kind of a fault condition, it became apparent that this was an intrinsic property of the system.

Figure 19 shows the noise performance achieved after level re-optimization, as compared to that originally expected. Figure 20 shows optimized levels corresponding to this result, also compared to those expected. The results of Fig. 19 were obtained by measuring thermal noise

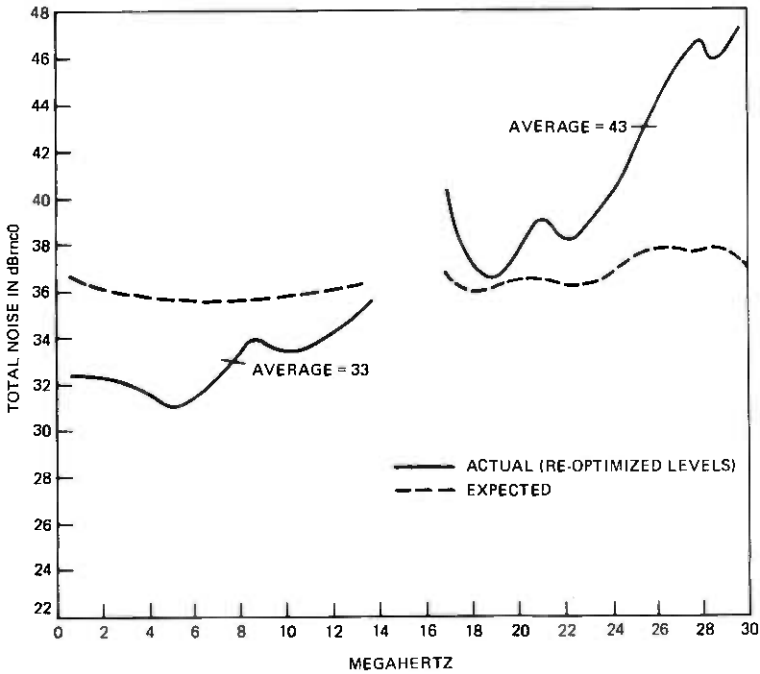


Fig. 19—System noise performance.

and calculating the increase due to modulation noise. The modulation noise for this calculation was obtained from noise loading measurements, interpolating between measured points where necessary. System noise performance shown on Fig. 19 was better than anticipated at lower frequencies. This was because optimized levels at the higher frequencies were lower than expected, thus reducing the *A-B* intermodulation products falling in the low band.

8.2.1 Nature of excess noise

The portion of the frequency spectrum in which the excess noise occurred suggested that the source of excess noise was third (or higher)-order intermodulation. This was confirmed by loading the high band only. Since this band is less than one octave wide, no second-order intermodulation can contribute to noise within it.

The NPR curves of Fig. 21 show that, whereas one would expect the curve at higher loads to have a slope of −2,* the system shows a much

* Third-order products or noise are expected to increase 3 dB for each 1-dB increase in signal. Thus, the signal-to-third-order noise ratio should decrease 2 dB per 1 dB increase in signal.

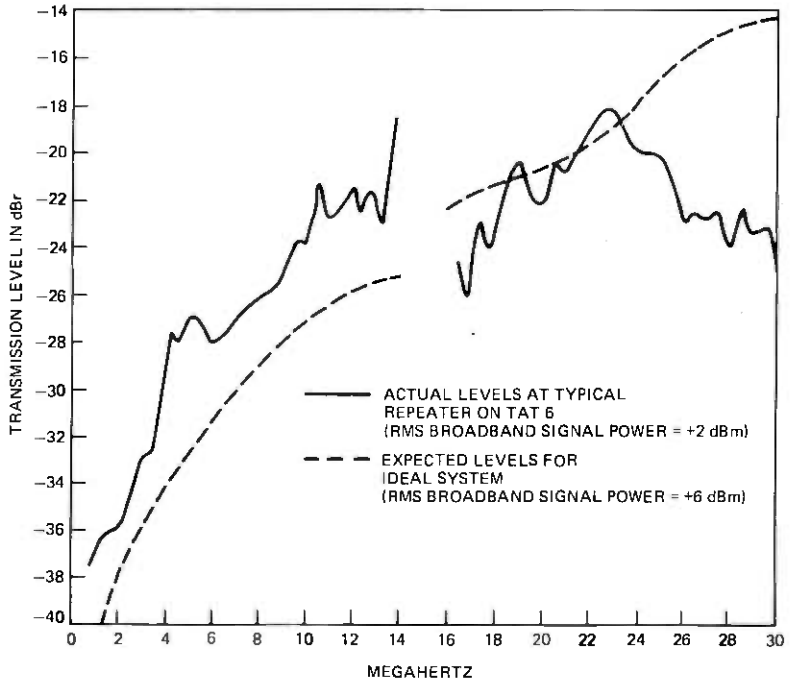


Fig. 20—Transmission levels at repeater out.

shallower slope. This is clearly indicative of anomalous behavior and provides the beginning of an explanation of why this behavior had not been clear from measurements on single repeaters or from testing an assembly of ten repeaters during SG development. The dashed curves show what was expected from the assumed constant values of M_{3E}, but using the "actual" levels of Fig. 20.

Third-order intermodulation noise is only of system significance compared to second order and thermal noise because, in the absence of delay distortion, it adds on a voltage (systematic) basis while the other two noises add on a power (rms) basis. Thus, by normal measuring techniques, the only way to see third-order intermodulation noise produced by noise loading of one or a few repeaters is to load with a power much higher than normal for a system, and extrapolate down to the operating point, assuming the classic 2-for-1 slope. The results of Fig. 21 show that a much shallower slope prevails, so that such an extrapolation would be much too optimistic.

The usual way to compute intermodulation performance is to measure intermodulation products produced by two or three discrete fundamental frequencies. Techniques then exist to estimate performance with a real load from the modulation (M) coefficients determined from the discrete tone measurements.[10] However, these techniques assume that

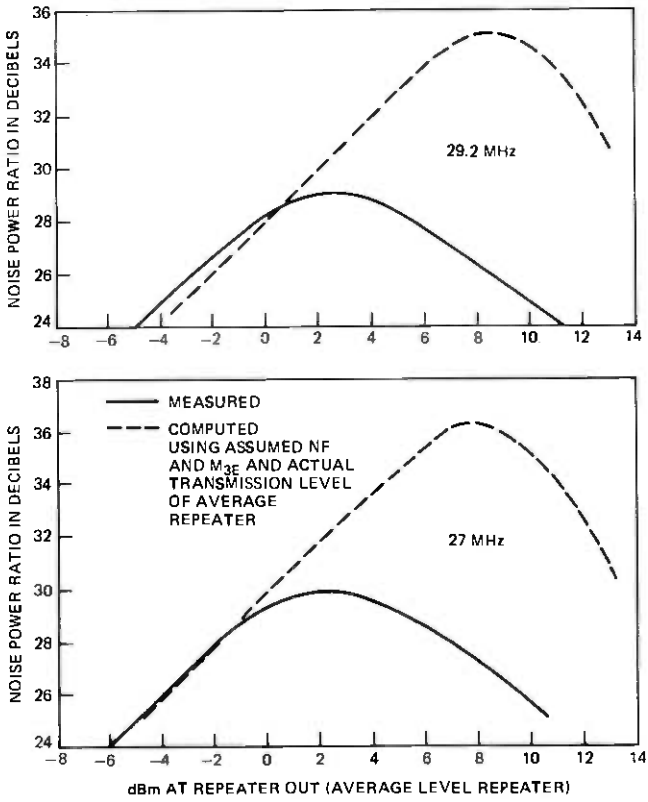


Fig. 21—NPR (signal-to-noise ratios) on system—highband only loaded.

the discrete tone intermodulation is a function only of product frequency and not of the frequencies of the fundamentals. This assumption had been substantially true in earlier systems.

In the development and manufacture of SG, a set of fundamental frequencies believed to be typical had been selected to measure intermodulation. The behavior of products generated by these fundamentals was normal, and the values obtained were those shown in Fig. 11. These results led to computed noise performance consistent with the objective, as shown in Fig. 5.

After the discovery of the third-order intermodulation problem, extensive additional measurements were made. It was found that fundamentals existed which produced products of a much higher level than those produced by the standard fundamentals used during development and manufacture. Furthermore, these products did not vary 3 dB for each decibel of change in the power of fundamental power (i.e., M coefficients were not fixed, but were a function of power). Finally, whereas third-order modulation from the standard fundamentals was

slightly better at sea-bottom temperature as compared to room temperature, these new fundamentals generated much more intermodulation at the cooler sea-bottom temperature than at room temperature.

What made the problem much worse was that many of these "bad" fundamentals were in the portion of the spectrum where there was almost no delay distortion (see Fig. 12). This meant that the intermodulation products of highest amplitude were the very ones that added in-phase from repeater to repeater. Figure 22 shows some M_{3E} coefficients determined from discrete tone measurements on the system using fundamentals very close to the product frequencies. These can be compared to the "nominal" −113 dB M_{3E} and to the M_{3E} coefficients of Fig. 11. The values in Fig. 11 which were measured with "standard" fundamentals had been thought to be the correct ones.

Figure 23 shows the dependence of M_{3E} on source frequencies. These modulation coefficients were measured on a single repeater at sea-bottom temperature at a particular power level and for a particular product frequency. Instead of the usual situation in which one might expect at most a few decibel variation over the whole range of possible source frequencies (up to the overload power), these "isomods" show gross changes as a function of fundamental frequencies.

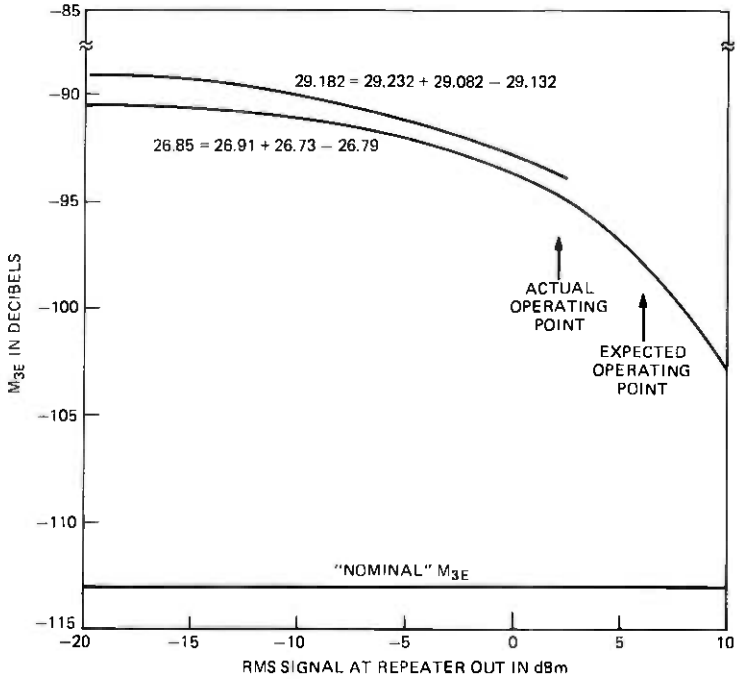


Fig. 22—M_{3E} determined from system measurements using fundamentals close to product frequencies.

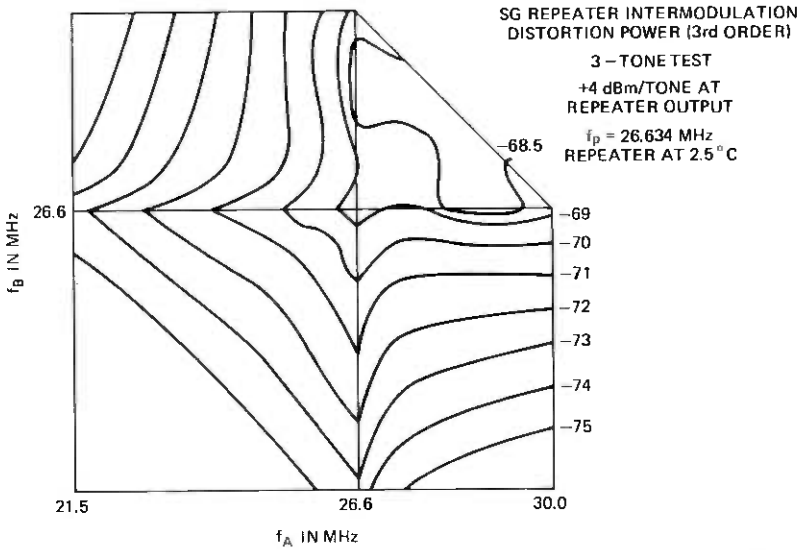


Fig. 23—Three-tone ($A + B + C$) isomod contours for SG repeater with surge protection diodes (sea-bottom temperature).

Thus, the much-worse-than-anticipated values of M_{3E} at the anticipated operating level were the source of the problem. It was then further aggravated by the shallow slope; that is, signal-to-noise ratio improved at a much more gradual rate than 2 for 1 as levels were reduced to diminish the third-order noise. This behavior forced us to reduce levels even more, degrading the signal-to-thermal noise ratio in the process.

In the top third of the high band, the achieved noise performance is poorer than the worst-channel objective by as much as 5 dB, and the average noise in the high band exceeds the average-channel noise objective by about 4.5 dB. These levels of noise do not render circuits unusable, but do represent an undesirable service degradation.

8.2.2 TAT-6 modulation improvement

Our program of modulation improvement involves placing equipment at the Green Hill, Rhode Island terminal to cancel third-order distortion in the upper portion of the high band. Transmission and delay equalizers will duplicate the electrical conditions at the middle of each ocean block of the undersea system. Then at each such virtual ocean-block point, a distortion generator will introduce distortion to balance out third-order distortion of the corresponding ocean block of the undersea link.

Field trials at Green Hill with a single distortion generator have achieved significant noise cancellation. These were feasibility trials performed by noise-loading a single supergroup at a time in the part of the high band where delay is relatively flat. Indications are a 12- to 15-dB reduction in third-order intermodulation can be realized, which will allow the TAT-6 link to meet its original performance objectives.

The Green Hill installation will include duplicate distortion canceling chains. Normal terminal monitoring and automatic switching arrangements will assure service continuity.

8.2.3 Intermodulation improvement of future SG links

Studies of the third-order modulation concluded that the reverse-biased output protector diodes were a major source of anomalous distortion. These diodes protect the power amplifier output stage against surge damage.

Independent of the modulation problem, Bell Laboratories had examined a double-epitaxial transistor design. The deeper second epitaxial layer makes these transistors highly resistant to second-breakdown damage. Using double-epitaxial transistors in the output stage eliminates entirely the need for the reverse-biased diodes. Laboratory tests indicate that the new combination is actually more rugged than the earlier arrangement.

Studies of intermodulation further indicated that interface states at the boundary between silicon and silicon-oxide layers of the transistor were a second source of the anomalous third-order distortion. This boundary, or interface, is produced by the transistor's silicon oxide (plus silicon nitride) passivation. Experiments suggested that shifting the crystalline orientation of the transistor starting material from ⟨111⟩ to ⟨100⟩ would greatly reduce the number of interface states, and hence the amount of anomalous distortion. Future SG links will have improved performance based on such a change in the output stage of the power amplifier. The first two stages of the power amplifier will retain transistors with the ⟨111⟩ orientation, since experiments established that a *small* amount of anomalous distortion will cancel some of the normal distortion, leading to an amplifier which is more linear than one using ⟨100⟩ transistors in every stage.

IX. SYSTEM AGING

In its two years of service (since commissioning), the system has aged in the direction of increased loss by about 25 dB at 30 MHz with a shape that is approximately linear with frequency. The incorporation of shore-controlled equalizers is thus seen to have been well worth while. Work is currently going on to identify the physical mechanisms responsible for the transmission change. Until a particular physical mechanism or mechanisms are proven to be the cause, predictions of aging over the life of the system must be made with caution. On a pure curve-fitting basis, log or exponential functions of time appear to fit the observed data. These functions would predict end-of-life top frequency aging as large as 50 dB or as low as 30 dB.

Suppose end-of-life total aging proves to be 35 dB. This is 7 dB per

SCE sector. Since the TAT-6 link was pre-equalized by 2 dB per sector, as described in Section 3.2.2, we would be left with a net loss per sector of 5 dB, which is within the system misalignment allocations.

X. EQUALIZATION AND PERFORMANCE

The initial equalization results achieved on TAT-6 are satisfactory and in keeping with misalignment allocations. Equalization knowledge gained in laying TAT-6 will allow the use of 30-repeater ocean blocks in future systems, still keeping misalignments within the desired range.

Regarding reliability, definitive statements require further operational experience. However, the TAT-6 link has been operating continuously since commissioning in July 1976. We fully expect that the superior job of burying the shore ends with Sea Plow IV, plus the great care in each step of design and of manufacture, will result in meeting our reliability targets.

Although less than six months elapsed between the start of design and initial shipment of shore-controlled equalizers, the four SCEs in the TAT-6 link are performing satisfactorily. These have been reset once since commissioning, to compensate cable aging.

It now appears that there will be enough range in the SCEs to compensate aging throughout TAT-6's lifetime. By keeping misalignment within bounds, and thus minimizing the noise penalty associated with cable aging, the SCEs are contributing importantly to the performance of TAT-6.

APPENDIX

Glossary

A.1 Terms for system description

REPEATER SECTION A length of cable with its associated repeater. Nominal deep-sea repeater sections for SG are 5.1 nmi long.

OCEAN-BLOCK EQUALIZER (OBE) Passive unit designed to reduce the residual gain or loss at the time the system is laid, which results from discrepancies between the gain of repeaters and the loss of the associated cable sections.

OCEAN BLOCK A number of repeater sections and their associated OBE. For TAT-6, ocean blocks were 20 repeater sections (98 nmi) long. Future SG links will have 30-repeater (149-nmi) ocean blocks.

SHORE-CONTROLLED EQUALIZER (SCE) Undersea unit placed following a number of ocean blocks. The basic purpose of the SCE is to compensate long-term changes in transmission occurring after installation, such as those due to cable aging.

SECTOR, OR SCE SECTOR An SCE and its associated ocean blocks. In TAT-6 there are seven ocean blocks per SCE; hence, sectors are about 680 nmi long.

A.2 Terms for signal description

ZERO LEVEL POINT (0 dBr) Used to provide a reference for expressing transmission level.

TRANSMISSION LEVEL The net gain or loss in decibels from a reference point (the zero level point) to the point in question. For net gains, the transmission level is positive; for net losses, it is negative. Transmission level may be expressed as $\pm X$ dBr.

CHANNEL LOAD, P_c The average signal power in a one-way voice channel. This power is expressed in dBm0, decibels relative to one milliwatt at the zero-level point. The SG average channel load was taken as -13 dBm0.

BROADBAND LOAD, P_{av} The total average power in a wide-band signal. Where this signal is composed of N two-way voice channels combined with equal weighting, $P_{av} = P_c + 10 \log_{10} 2N$ dBm0.

MULTI-CHANNEL PEAK FACTOR, k_m The ratio, expressed in dB, between the 0.001 percent peak power and the average power of the multi-channel signal. When the number of channels is very large, the signal's k_m approaches that of Gaussian noise. Thus, for SG the multi-channel peak factor = 13 dB.

PEAK FACTOR, α The *ratio* of peak voltage to rms voltage. A peak factor of 5 would indicate voltage five times larger than the rms voltage. To express the peak factor as a power ratio in decibels, we take 20 log α.

A.3 Channel noise terms

CHANNEL NOISE The background interfering power in the bandwidth of a single channel.

WEIGHTING The function vs frequency which allows for the fact that different frequencies interfere with voice communication to different degrees.

PSOPHOMETRIC WEIGHTING, (p) A noise weighting curve adopted as international standard by the CCITT.

C-MESSAGE WEIGHTING, (c) Noise weighting currently in use in the Bell System domestic plant.

REFERENCE NOISE, rn One picowatt, or -90 dBm before applying any weighting.

A.4 Equalization terms

EQUALIZATION The process of adjusting system elements (typically equalizers) so that transmission is kept within prescribed limits at all frequencies and points in the system.

MISALIGNMENT (dB) The extent to which equalization departs from the ideal. Positive misalignment represents gain, while negative

represents loss. Misalignment can be absolute (departure from a nominal level), or relative (departure of level from levels at other units along the system). In laying an undersea system, one is concerned with relative misalignment, since all undersea levels may be shifted up or down by terminal adjustments after the system is laid.

UNIFORM MISALIGNMENT Misalignment which increases at a constant rate going along the system.

A.5 Cable terms

NAUTICAL MILE (nmi) When used for cable, refers to a cable nautical mile, 1855.3 meters, or 6087 feet.

DISSIPATION FACTOR, tan δ The tangent of the polyethylene dielectric loss angle. For high quality dielectrics, tan $\delta \doteq \delta$, and the dissipation factor is given in microradians at a particular frequency.

TEMPERATURE COEFFICIENT, $\Delta\alpha_t$ The change in loss of cable with change in temperature. Usually given in percent per degree Celsius at specified frequency and reference temperature and pressure.

PRESSURE COEFFICIENT, $\Delta\alpha_p$ The change in loss of cable with change in pressure. Usually given in percent change per kilofathom at specified frequency and reference temperature and pressure.

A.6 Repeater terms

NOISE FIGURE, NF* The number of decibels by which the output thermal noise power of a repeater, when referred to the input by the repeater's available power gain, exceeds the ideal value given by kTB (k = Boltzman's's constant = 1.3806×10^{-23} joules/degree Kelvin, T = temperature in degrees Kelvin, and B = bandwidth in hertz). At room temperature of 17°C, or 290°K, this limiting or ideal value of noise is given by P = (−174 + 10 log B) dBm.

MODULATION COEFFICIENT, M_{2E} An equivalent second harmonic coefficient which, acting in a repeater whose nonlinearity can be modeled by a Taylor series, would give the observed A-B products. If $M_{\alpha-\beta}$ is the A-B coefficient, $M_{2E} = M_{\alpha-\beta} - 6$ dB. For an M_{2E} of −70 dB, application of α and β frequencies, each with a power of 0 dBm at the point at which the modulation coefficient is defined, would result in a power at frequency $(\alpha - \beta)$ given by: $P_{\alpha-\beta} = -70 + 6 = -64$ dBm.

MODULATION COEFFICIENT, M_{3E} An equivalent third-harmonic coefficient which, acting in a repeater whose nonlinearity can be modeled by a Taylor series, would give the observed $A + B - C$ product. If $M_{\alpha+\beta-\gamma}$ is the $A + B - C$ coefficient, $M_{3E} = M_{\alpha+\beta-\gamma} - 15.6$ dB. For an M_{3E} of −113 dB, application of α, β, and γ frequencies, each with power of 0 dBm at the point at which the modulation coefficient is defined, would result in an $\alpha + \beta - \gamma$ power given by $P_{\alpha+\beta-\gamma} = -113 + 15.6 = -97.4$ dBm.

* Reference 10 discusses noise figures in detail.

MAXIMUM SINE WAVE OUTPUT POWER, P_R The maximum power the repeater can handle before some specified impairment or distortion occurs.

REFERENCES

1. Special Transatlantic Cable issue, B.S.T.J., *36*, No. 1 (January 1957), pp. 1–326.
2. Special SD Submarine Cable System issue, B.S.T.J., *43*, No 4, Part 1 (July 1964), pp. 1155–1467.
3. C. D. Anderson "Overload Stability Problem in Submarine Cable Systems," B.S.T.J., *48*, No. 6 (July–August 1969), pp. 1853–1864.
4. "SF Submarine Cable System," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 601–798.
5. R. L. Easton, "Undersea Cable Systems—A Survey," IEEE Communications Society, *13*, No. 5 (September 1975).
6. G. E. Morse, S. Ayers, R. F. Gleason, and J. R. Stauffer, "SG Undersea Cable System: Cable and Coupling Design," B.S.T.J., this issue, pp. 2435–2469.
7. R. G. Buus, J. J. Kassig, and P. A. Yeisley, "Repeater and Equalizer Design," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 631–651.
8. C. D. Anderson, W. E. Hower, J. J. Kassig, V. M. Krygowski, R. L. Lynch, G. A. Reinold, and P. A. Yeisley, "SG Undersea Cable System: Repeater and Equalizer Design and Manufacture," B.S.T.J., this issue, pp. 2355–2402.
9. D. N. Harper, B. O. Larson, and M. Laurette, "SG Undersea Cable System: Commissioning: Final System Alignment and Evaluation," B.S.T.J., this issue, pp. 2547–2564.
10. Members of Technical Staff, *Transmission Systems for Communication*, 4th ed., Murray Hill, N.J.: Bell Laboratories, 1970.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Repeater and Equalizer Design and Manufacture

By C. D. ANDERSON, W. E. HOWER, J. J. KASSIG, V. M. KRYGOWSKI, R. L. LYNCH, G. A. REINOLD, and P. A. YEISLEY

(Manuscript received May 30, 1978)

Linearity and reproducibility of undersea electronics play key roles in determining the achievable performance of an undersea cable system. This paper describes the characteristics of undersea electronics and the measures taken in design and manufacture to achieve the desired performance and reliability, giving special emphasis to new features and applications. It describes in detail the performance achieved in the recently completed SG system linking the U.S. and France.

I. INTRODUCTION

Exploratory work on undersea amplifiers and devices in the late 1960s produced models of repeaters with the performance required for a 3500-channel, 4000-nmi system with a top frequency of 27.5 MHz. The specific development of the 30-MHz, 4000-channel, SG system was started early in 1971. The first SG installation, TAT-6, which connects Green Hill, R.I., with St. Hilaire, France, was completed in July 1976, yielding 4200 channels.

The fivefold increase in bandwidth of SG over SF[1] was made technically possible by advances in the design of reliable, ultralinear, silicon microwave transistors.[6] Manufacturing the necessary numbers of undersea bodies economically and on schedule required substantial innovations in the basic mechanical construction and assembly techniques of the electronic units.

II. GENERAL OBJECTIVES

2.1 Noise and load

The SG system was designed to meet the international objectives of 1 pWp0/km (38.6 dBrnc0 for 3500 nmi), average, with a signal load of −13 dBm0 per 3-kHz spaced channel. The "worst channel" could be 2 pWp0/km. Once the noise, load, channel capacity, and cable diameter are specified, and the achievable amplifier noise figure and modulation coefficients are known, the repeater gain (and spacing) can be determined from the following approximate relationship:

$$\text{s/n} = -1.8 - \frac{1}{3}\left[M_{A+B-C} + 10\log(\overline{n_p}) + 20\log(n) \right] - \frac{2}{3}\left[10\log(KTB) + N_F + G_R + 10\log(n) \right], \quad (1)$$

where

s/n = signal-to-noise ratio in decibels,
M_{A+B-C} = repeater third-order modulation coefficient for 3-tone products of the $A + B - C$ type (−95.0 dBm),
n = number of repeaters in the system (690),
$\overline{n_p}$ = equivalent number of third-order intermodulation products which add in phase from repeater to repeater (376,000),
KTB = thermal noise = −139 dBm at 300°K in 3 kHz,
N_F = repeater noise figure (3.5 dB),
G_R = repeater insertion gain (41.0 dB).

The numbers in parentheses indicate the approximate 29.5-MHz values. The channel noise and load objectives (38.6 dBrnc0 and −13 dBm0, respectively) translate to a system s/n ratio of 36.4 dB. Equation (1) evaluated at 29.5 MHz yields an s/n ratio of 36.4 dB. More detailed analysis showed that proper signal level shaping would significantly reduce the noise at lower frequencies. It thus appeared that we could meet the noise objectives.

2.2 Transmission

Because optimum repeater noise performance can be obtained only over a narrow range of signal power, it is necessary to accurately match the repeater gain to the cable section loss. The errors between the two are compensated for in the ocean-block equalizers (OBEs) which follow every 20 or 30 repeaters. The allocation of repeater, equalizer, and cable loss deviations are detailed in another article.[2] For a 20-repeater block (TAT-6), the top frequency loss deviation allowances are ±0.03-dB systematic and ±0.1 dB random for repeaters and ±0.1 dB total for OBEs.

The deviation allocations are increased at lower frequencies up to a maximum of $1\frac{1}{2}$ times the top-frequency value, in proportion to $30/f$, where f is in MHz.

Available gain for the equalizers is obtained by reducing the length of cable by 4.1 nmi, that is, from 5.1 nmi for repeater sections to 1.0 nmi for equalizer sections. Additional low-frequency gain range is derived from a repeater gain boost of the form:

$$\begin{aligned} \text{repeater gain boost (dB)} &= 0.00218(11.5\text{-}f)^2 \quad f \leq 11.5 \\ &= 0 \qquad\qquad\qquad\quad f > 11.5, \end{aligned} \tag{2}$$

where f is in MHz.

Cable loss variations due to the seasonal temperature changes on the continental shelves are equalized by temperature-controlled repeaters. The gain of these repeaters is controlled by an ambient-temperature-sensing thermistor in the amplifier feedback circuit. The objective was to equalize at least $\frac{2}{3}$ of the shallow-water loss variation, leaving only a smooth residual which could be readily equalized in the terminals.

To meet noise objectives over the 20-year minimum life of the system requires means of equalizing any reasonably likely transmission change which occurs during that period. Shore-controlled equalizers (SCEs) were introduced when cable "aging" (change in attenuation with time) was predicted to be of serious consequence.[3] Each SCE has a controllable loss range of ±16 dB at 30 MHz with 2-dB step sizes.

2.3 Mechanical

To ensure that the repeaters and equalizers survive the shocks associated with laying and recovery, they are designed to withstand:

(*i*) A 60-g shock peak for 60-ms duration between half-amplitude points along each of the three principal axes.
(*ii*) A vibration limit of 1.27-cm sinusoidal peak-to-peak displacement from 5 to 11 Hz.
(*iii*) A sinusoidal vibration at 3-g maximum from 11 to 500 Hz.

The underwater electronics are designed to operate over a temperature range of 0° to 30°C (32° to 86°F). The allowable storage temperature range is: −18° to 57°C (0° to 135°F).

The maximum permissible leak rate of a sealed pressure housing under helium pressure test is 1×10^{-7} standard cc/s at 844 kg/cm^2 (12,000 psi). This is the pressure at a depth of 8050 m (26,000 ft). Individual components such as seals are allowed a maximum leak rate of 5×10^{-8} standard cc/s.

Table I — New materials, processes, and techniques

Materials	
Aluminum	AA 356.0 and AA-6061
Bronze	CA-725
Thermoplastic	Glass-reinformed polybutylene terephthalate (PBT)
Stainless steel	Types 305, 310, and 384

Processes	Techniques
Back extrusion	Printed-wiring boards
Electron beam welding	Vacuum drying
Low-pressure permanent-mold castings	

III. MECHANICAL DESIGN

3.1 New techniques

The physical design of the SG underwater hardware (Fig. 1) embodies some principles proven in the past and some which are new to undersea use. Those materials, processes, and techniques newly exploited in undersea cable use are listed in Table I.

3.2 Pressure hull

The cylindrical copper beryllium housing is back-extruded rather than forward-extruded. This change has led to a housing which costs about the same as those used for the earlier SD and SF systems in current dollars. The back-extruded housing more than satisfies the corrosion resistance, grain size, porosity, and strength requirements.

The tungsten inert gas (TIG) method of welding the cover to the housing has been replaced by electron beam (eB) welding. A high-intensity (about 1.4 MW/cm^2) stream of electrons is focused to a controlled beam size upon the weld seam in a vacuum. The kinetic energy of the beam is converted into heat upon impact with the copper beryllium, causing it to melt, vaporize, and then condense on the cooled material surrounding the melt trough. The weld progresses continuously as the cover and housing rotate under the beam.

The weld depth and melt zone geometry have been optimized for minimum stress by rigidly controlling the beam current, diameter, accelerating voltage, focal distance, work distance, and welding speed. Weld depths are typically 3.8 mm. In the rare case of a process or apparatus failure during assembly, the eB process permits up to two repair welds. Such a failure with the TIG process required a machining operation which destroyed either housing or cover.

3.3 Electronics unit

The most significant departure from previous practice was the elimination of hermetic sealing of the electronics block of the repeater and equalizer units. Analyses had shown that by taking advantage of the

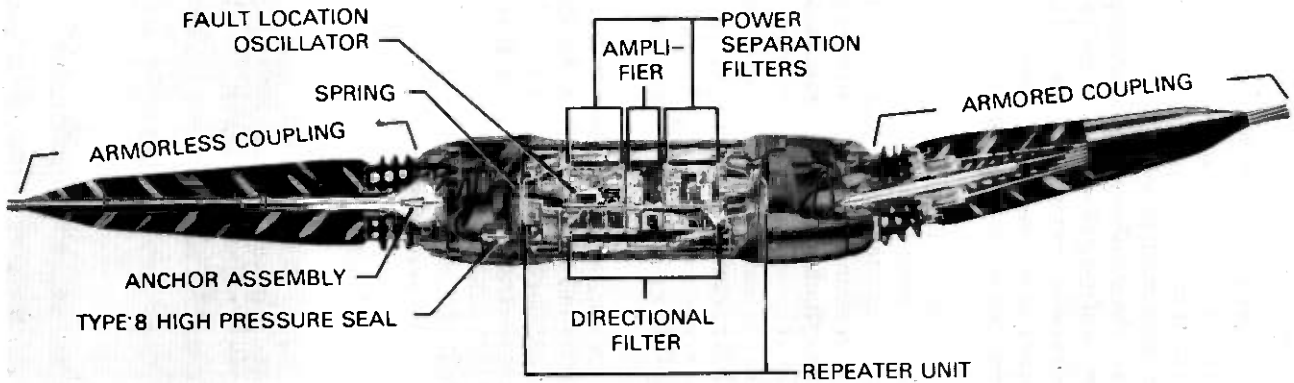


Fig. 1—SG repeater.

water absorption properties of the plastic materials within the electronics unit, the inner assembly (Fig. 2) can be left open inside the pressure housing without danger of moisture condensation. This greatly simplified the design of the inner unit. It allowed the cylinder to be extruded, which simplified the epoxy coating procedure and subsequent assembly with the main frame. The main frame and associated structures were cast by the low-pressure permanent-mold process (LPPM).

The frequency range of the SG system necessitated close coordination of the electrical and physical designs of the electronics units to achieve the needed performance and to minimize unit-to-unit variations. In the repeater, assembly procedures allowed adjustment of over 30 inductors. Five of these adjustments are made after final assembly of all the networks.

Each individual circuit was mounted on its own aluminum frame as a separate module (Fig. 3). Functional partitioning allowed networks to be assembled, tested, adjusted, and "locked" individually. Each module was mounted on a main frame which in turn was slid into the epoxy-coated aluminum cylinder. The main frame was then locked to the cylinder and the ends covered with plastic end caps.

The modular type construction using cast frames provided low thermal impedance (about 1°C/W) between the transistor mounting studs and the housing. This type of construction also minimizes corona due to the high voltages (up to 7 kV) by maintaining isolation, proper spacing, and smooth surfaces.

Glass epoxy double- and single-sided circuit boards, 1.588 mm (0.0625 in.) and 2.381 mm (0.0935 in.) thick, supplied the structural rigidity needed for mounting components and highly repeatable circuit paths to minimize variations in electrical characteristics. Path widths were set at 2.54 mm (0.100 in.) with a minimum spacing of 1.27 mm (0.050 in.). Dielectric constant, insulation resistance, peel strength, plating thickness, and breakdown voltage were required to meet stringent limits to insure high reliability.

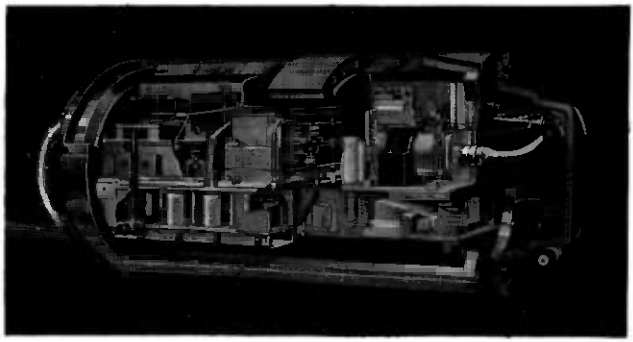


Fig. 2—Electronics unit, inner assembly.

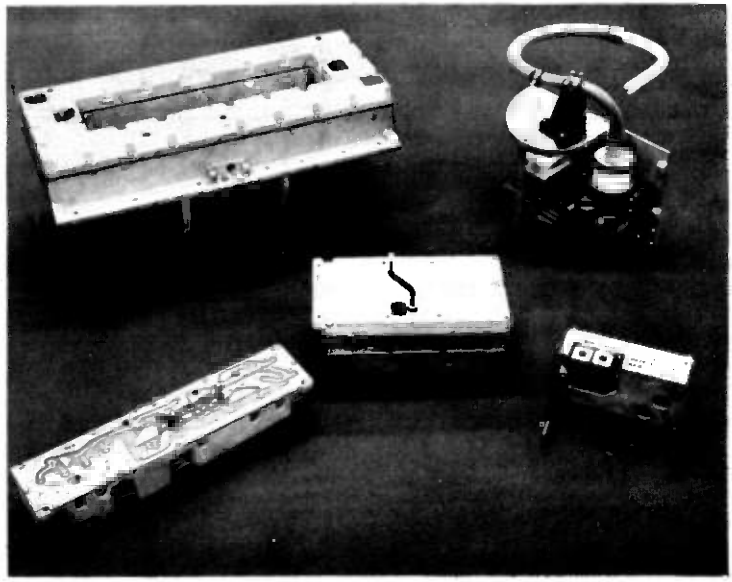


Fig. 3—Circuit modules. Top left, directional filter (DF); top right, ground separation filter (GSF); bottom left, output network; center, amplifier; bottom right, oscillator.

Because of the high potential of the electronic units, it was desirable to make the end caps from insulating material. A glass-filled polybutylene terephthalate (PBT) was selected because of its excellent mechanical and electrical properties. In addition, its low shrink factor and good flow characteristics make it a good material for injection molding. PBT has a melting temperature greater than 216°C (420°F) and therefore allows for soldering on or near its surface. The material has also been used for electrical standoffs and insulating covers.

Although the decision to use a nonhermetically sealed electronics unit permitted the exploitation of new materials and processes, it required a new vacuum drying procedure which considered the outgassing and water absorption properties of all the materials in the unit. A drying cycle and a closure procedure have been devised which predict a dew point of less than −15°C (+5°F) after 20 years at sea bottom at an external pressure of 844 kg/cm^2 (12,000 psi).

IV. REPEATER ELECTRICAL DESIGN

4.1 Configuration

As described in a companion article,[2] the repeater gain is provided by a pair of tandem amplifiers common to both transmission bands.

Table II contains a summary of the repeater electrical performance.

Table II — SG repeater electrical performance summary

Frequency (mHz)	Repeater Gain (dB)	Amplifier Gain (dB)	Network Losses (dB)	Repeater Noise Figure (dB)	Modulation Coefficients* M_{2E}* (dB)	Modulation Coefficients* M_{3E}* (dB)	Repeater Return Loss A end (dB)	Repeater Return Loss B end (dB)
1	7.48	10.56	3.08	8.7	−112	†	34	25
2	10.37	12.13	1.76	8.0	−116	†	36	24
4	14.53	15.94	1.41	7.7	−110	†	33	22
8	20.50	22.07	1.57	7.0	−100	†	25	30
12	25.34	27.67	2.33	5.8	−84	†	22	17
13.75	27.20	30.64	3.44	5.8	−79	†	22	22
16.50	29.96	33.00	3.04	5.2	−77	−113	27	19
20	33.23	34.99	1.76	4.0	−73	−113	24	26
25	37.46	39.09	1.63	3.6	−71	−113	21	23
30	41.37	42.89	1.52	3.5	−70	−113	24	20

* Defined at power amplifier output. The value of M_{3E} is that determined before installation of TAT-6. As discussed later in Section VII, this value is better than the actual repeater performance.

† M_{3E} is not a significant parameter in the low band.

4.2 Passive networks

4.2.1 Ground-separation filter

The ground-separation filters (GSFs) shown in Fig. 4 connect the cable to the directional filters. They were designed as high-pass filters. Coaxial capacitors (0.013 μF) provide a low-impedance ground path for signal frequencies, while blocking the high dc voltage (up to 7 kV) which exists between the repeater unit (repeater ground) and the housing (sea ground). To prevent coupling between the repeater input and output, this impedance should ideally be zero. However, at low frequencies the reactance of the high-voltage capacitor becomes large, while at high frequencies inductive parasites dominate. The coaxial design virtually eliminates the inductance problem in the capacitor. A coaxial choke with a longitudinal inductance of 250 μH reduces the unbalanced current in the GSF and thereby increases the isolation between the repeater input and output. Adequate GSF loop loss* was achieved without building up the capacitance between the inner unit and the housing.

Powder core inductors carry the 657-mA dc line current to the amplifier biasing circuits. Each inductor along with the adjacent 0.013-μF coaxial and low-voltage ceramic capacitors forms a full-section, 50-ohm, high-pass filter with a cutoff frequency of about 300 kHz.

4.2.2 Directional filters

The directional filter is a 50-ohm, four-port network with two high-pass and two low-pass filters. They are of conventional design using oppositely phased transformers in the two low-pass portions to help

* GSF loop loss refers to stray transmission around the amplifier via the GSFs and the repeater's power path.

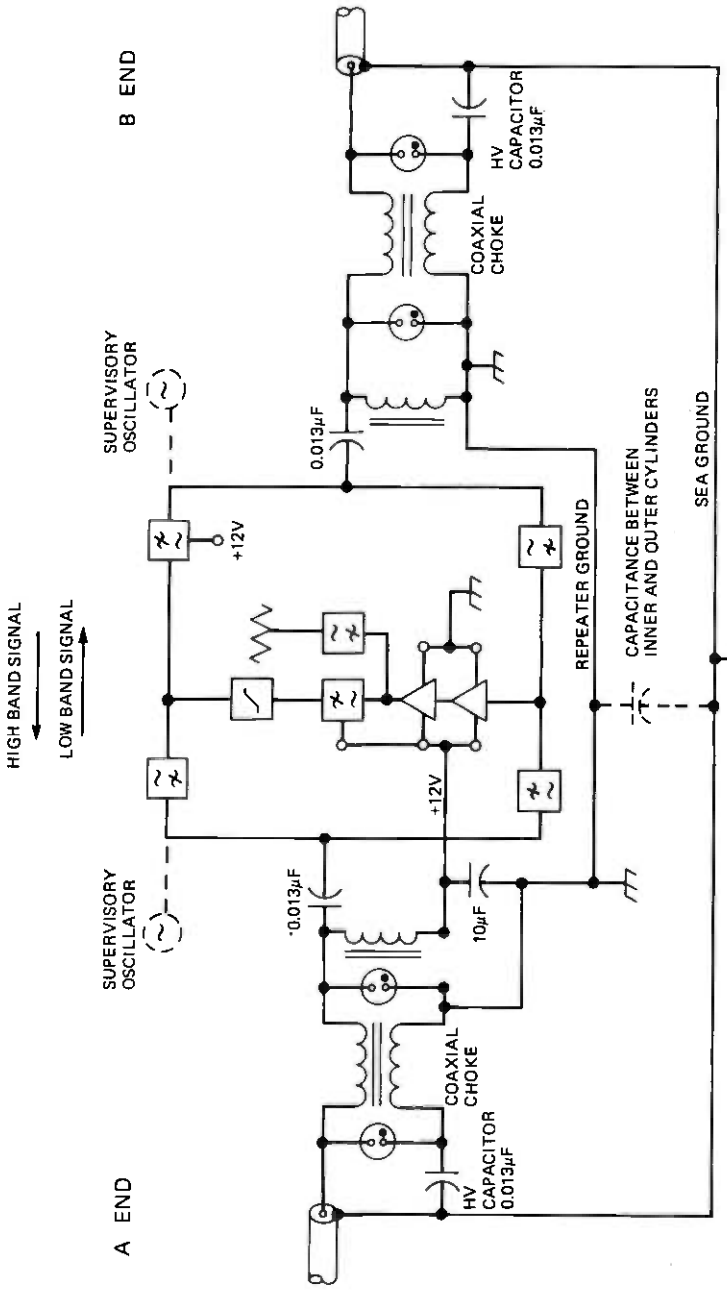


Fig. 4—SG repeater block diagram showing GSF circuit.

increase directional filter loop loss.† Figure 5 shows the rectangular arrangement and the individual shielded inductor compartments that provide ideal isolation between both sections and ports. All components are mounted on a single board and connected with printed wiring. The individual sections are tuned "in place" after all filter wiring is complete.

4.2.3 Output network

The output network, shown in Fig. 6, contains surge protection diodes, a low-pass bandlimiting filter, a low-frequency loss equalizer, and a high-pass terminating filter. The filter pair was designed on a constant-R basis with a 3-dB crossover at 33.5 MHz. The low-pass filter increases the high-frequency loop loss around the amplifier and improves the noise sing margin[4] by suppressing the transmission of above-band noise. The high-pass section properly terminates the power amplifier at high frequencies, thereby maintaining a uniform and controlled stability margin.

A constant-R bridged-T equalizing network helps shape the repeater

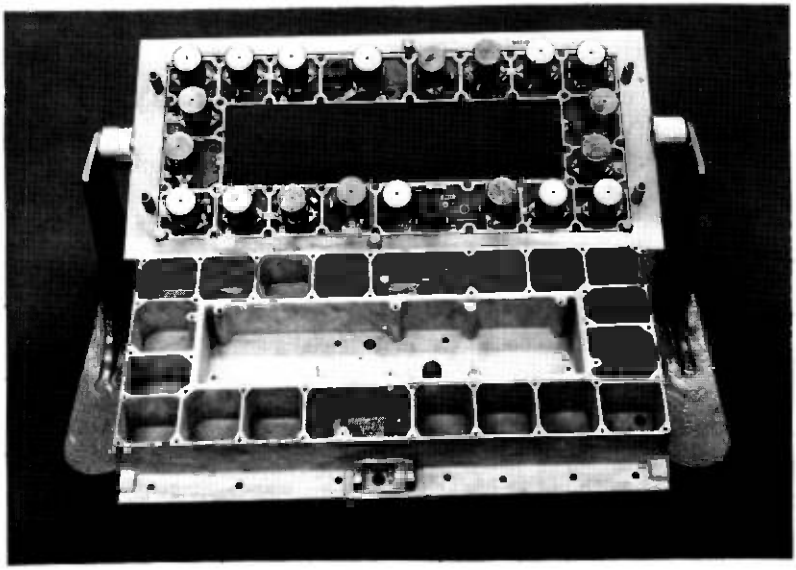


Fig. 5—Directional filter.

† Directional filter loop loss refers to the stray transmission around the amplifier via the tandem-paired filters, both high pass/low pass and low pass/high pass. Providing a phase reversal in one of the two paths increases the total loss because transmission via one path tends to cancel that via the other.

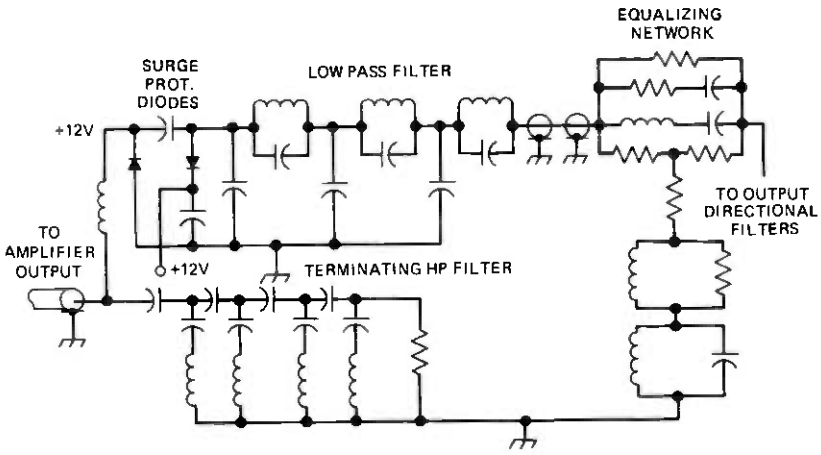


Fig. 6—Output network schematic.

gain and increases the output return loss and the external loop loss at low frequencies.

A pair of oppositely poled, reverse-biased diodes, whose capacitance is absorbed in the low-pass filter, provides surge protection for the power amplifier.

4.3 Amplifier

4.3.1 Configuration

The amplifier (Fig. 7) consists of a two-stage preamplifier (PRA) and a three-stage power amplifier (PWA). Both PRA stages are common emitter while the PWA uses a common-base output stage, driven from a Darlington pair via a current step-up transformer. The gain is shaped with feedback. Shunt input, series output is used in the PRA, and shunt input, shunt output in the PWA. Feedback with hybrid-connected autotransformers actively terminate the amplifier input and output ports. The PRA, which has a high output impedance, is connected to the low-input impedance PWA with a resistor. This arrangement reduces the impedance interaction effect between the two at their respective loop gain and phase crossover frequencies and minimizes stability problems.

4.3.2 Bias

The repeaters are biased at 657 ± 0.3 mA with an average voltage of 11.9 V (at 2.5°C). Nominal operating points of individual transistors are listed in Table III. The input stage of the PRA is biased from the emitter of the second stage. During power turn-up, the first stage does not begin to conduct until the second-stage current reaches about 50 mA.

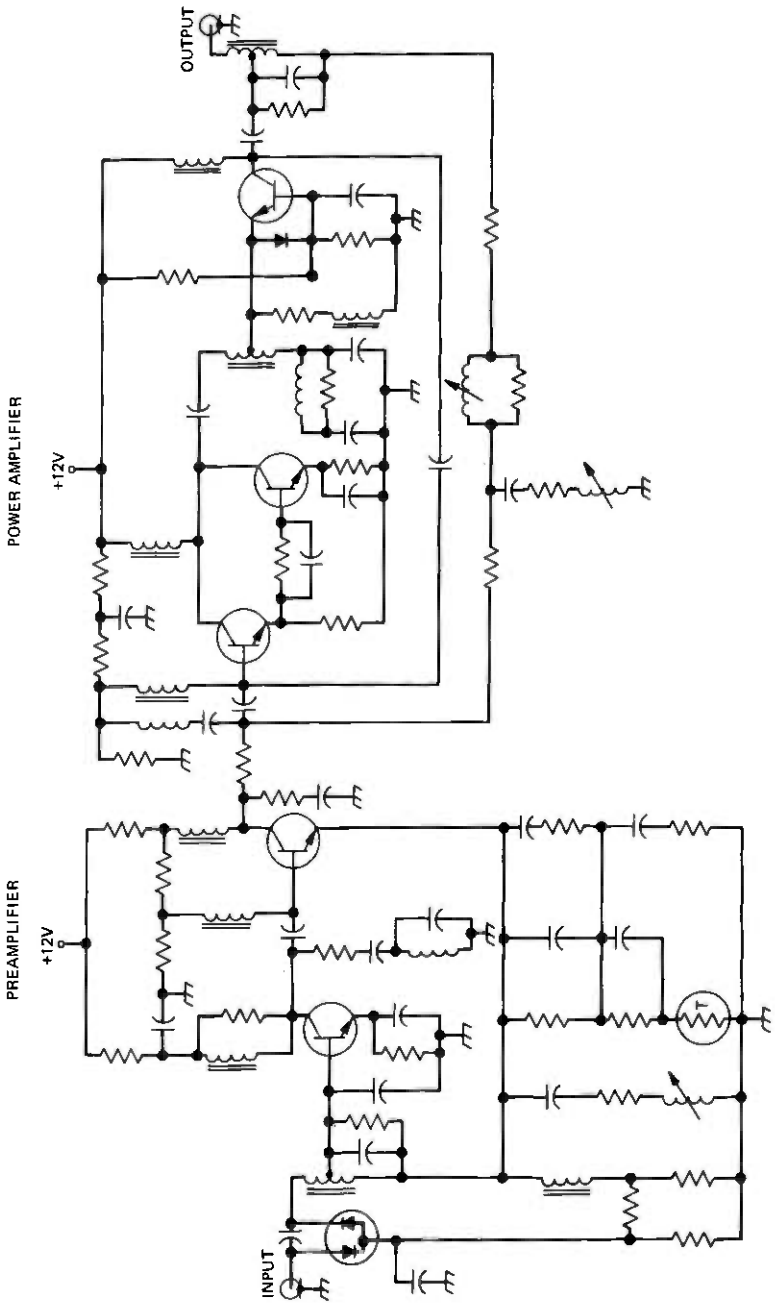


Fig. 7—SG amplifier schematic.

Table III — Nominal transistor operating points (2.5°C)

	Stage	Collector-Emitter Voltage (V)	Collector Current (mA)
Preamplifier	1	9.44	49.8
	2	9.24	153.1
Power amplifier	3	9.16	50.2
	4	10.05	147.2
	5	9.99	147.2
Oscillator		8.05	9.2
Subtotal			556.7
Bleeder current (mA)			100.3
Total (mA)			657.0

Staggering the currents at which the two stages begin to develop gain avoids oscillation which can occur during power turn-on of feedback amplifiers. The dc-coupled Darlington pair in the PWA operates similarly.

4.3.3 Gain

The amplifier gain objective equals the sum of:

(*i*) The loss of 5.1 nmi of cable at 2.5°C, 2.5 kilofathoms.
(*ii*) The loss of the repeater passive networks.
(*iii*) The low-band gain boost.

The amplifier gain, 43 dB at 30 MHz, is divided approximately equally between the PRA and the PWA. It is controlled by the feedback circuits in both amplifiers. The repeater gain, which is further shaped by the bridged-T equalizer in the output network, was initially matched to within +0.03 dB of the objective.* This necessitated extensive characterization of each component and the printed wiring boards. Component tolerances were assigned based upon both sensitivity studies and adjustment capability. The amplifier contains three adjustable inductors. Two are adjusted during the final repeater network testing stage along with an inductor in the output network and two in the directional filters.†

The shapes obtained with these final adjustments closely match the shapes of the three most significant sources of gain deviation:

(*i*) Directional filter and output network loss variations.
(*ii*) Miller capacitance of Darlington pair.
(*iii*) Alpha sensitivity of the second stage in the Darlington pair.

The effectiveness of the gain control and adjustment techniques is

* The low-frequency gain was later increased slightly to obtain more channels below 1 MHz at the expense of increasing the low-frequency gain deviation somewhat.

† This operation is discussed in Section 6.3.5.

indicated in Table IV, which is based upon the entire TAT-6 repeater production.

There are two basic types of SG amplifiers, fixed-gain and temperature controlled (TC). The latter, which are new to undersea cables, are used in those repeaters to be laid in shallow water areas where bottom temperatures change seasonally. The gain of TC repeaters is designed to increase with temperature, as does cable loss. Gain is controlled by an ambient temperature-sensing parallel-disk thermistor in the PRA feedback circuit. Figure 8 shows the repeater gain and the cable-loss change for a temperature rise from 2.5° to 10.1°C. Over this range, the repeater compensates for at least $\frac{2}{3}$ of the cable loss change. The TC repeater is designed to have the same gain at 2.5°C as the fixed-gain repeater.

4.3.4 Feedback

The achievable feedback is limited by both conventional and noise-sing stability considerations. As shown in Fig. 9, the nominal PRA feed-

Table IV — Repeater gain objective and average* deviation (dB) at 2.5°C

Low Band				High Band			
Freq (MHz)	dB Gain Objective	Deviation without Termina-tions	Deviation† with Termina-tions	Freq (MHz)	dB Gain Objective	Deviation without Termina-tions	Deviation† with Termina-tions
0.75	6.404	−0.018	−0.018	16.50	29.950	−0.006	−0.017
1.00	7.346	0.092	0.092	16.75	30.190	0.001	0.026
1.50	8.935	0.094	0.094	17.00	30.430	−0.006	0.026
2.00	10.281	0.060	0.060	17.25	30.667	−0.005	0.026
2.50	11.469	0.044	0.044	17.50	30.904	−0.001	0.025
3.00	12.546	0.036	0.036	17.75	31.138	0.008	0.023
3.50	13.540	0.032	0.032	18.00	31.373	0.015	0.010
4.00	14.470	0.037	0.036	18.50	31.835	0.027	0.002
5.00	16.176	0.027	0.029	19.00	32.293	0.027	0.002
6.00	17.730	0.007	0.010	19.50	32.745	0.026	0.012
7.00	19.170	−0.019	−0.020	20.00	33.193	0.020	0.020
8.00	20.520	−0.047	−0.051	21.00	34.075	0.008	0.016
9.00	21.797	−0.063	−0.057	22.00	34.940	0.001	0.008
10.00	23.015	−0.044	0.046	23.00	35.788	−0.002	0.006
10.50	23.604	−0.030	−0.028	24.00	36.622	−0.007	0.006
11.00	24.182	−0.018	−0.019	25.00	37.441	−0.006	−0.003
11.50	24.751	−0.009	−0.003	26.00	38.248	−0.002	−0.029
12.00	25.309	−0.001	−0.014	27.00	39.042	0.008	−0.032
12.50	25.859	−0.009	−0.003	27.50	39.435	0.012	0.024
12.75	26.130	−0.015	−0.020	28.00	39.824	0.018	−0.001
13.00	26.398	−0.015	−0.030	28.50	40.211	0.023	0.026
13.25	26.665	−0.013	−0.024	29.00	40.596	0.024	0.045
13.50	26.929	−0.007	−0.007	29.50	40.978	0.016	0.048
13.75	27.192	0.001	0.009	30.00	41.357	−0.013	0.012

* Based upon all the TAT-6 repeaters.
† Refers to armorless terminations.

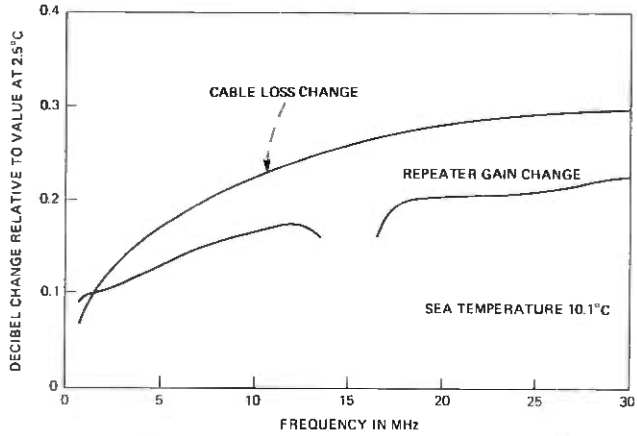


Fig. 8—Gain tracking of temperature-controlled repeater.

back ranges from 32 dB at 5 MHz to 11 dB at 30 MHz, while the PWA feedback is maximum at 1.5 MHz and decreases to 12 dB at 30 MHz. The repeater is stable for all combinations of open and short circuit terminations, for all possible combinations of transistors up to at least 45°C and at reduced currents, which occur during power turn-up. The

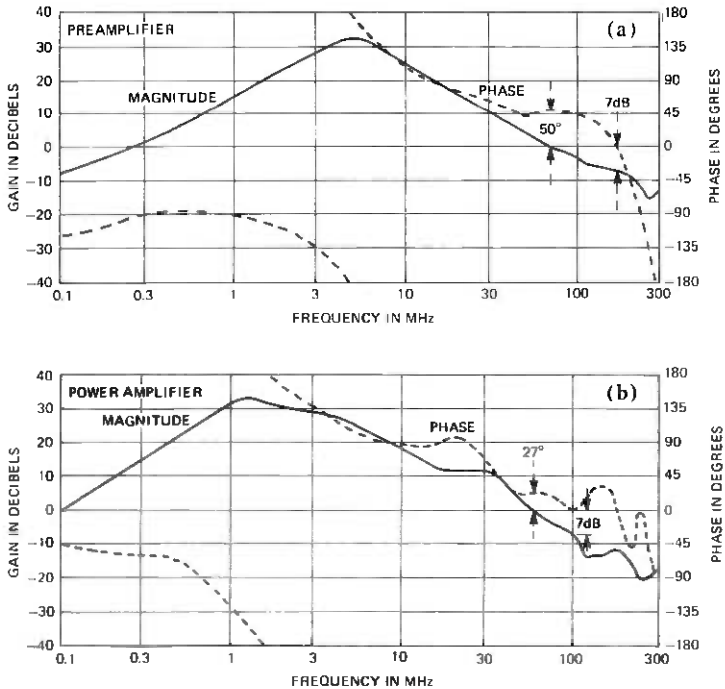


Fig. 9—Amplifier feedback.

noise-sing margin of the system is strongly influenced by the phase of both the inband and aboveband loop feedback of the PWA. Maintaining the phase characteristics near 90° across the high band and as large as possible aboveband resulted in significant improvement in the noise-sing margin.

4.3.5 Thermal noise and intermodulation distortion

The nominal repeater noise figure and intermodulation coefficients are shown in Table II. A 30-MHz noise figure of 3.5 dB was achieved by using a low-loss active termination and by selecting low noise-figure transistors for the input stage.

Second-order distortion coefficients range from −70 dBm at 30 MHz to −112 dBm at 1 MHz. The tandem amplifier configuration results in some fundamental frequency dependence in the coefficients. Because of the shaped gain in both amplifiers, the PRA second-order distortion contribution is significant for low-frequency fundamentals.

Third-order distortion products of the $A + B - C$ type are the most important because of their number and their tendency to add in-phase from repeater to repeater. Measurements of many different products of this type during development and manufacture of the SG repeater indicated a value for M_{3E} of approximately −113 dB in the high band.* As discussed further in Section 7.2, the characterization of third-order distortion was misleading because a source of distortion existed that did not obey the usual rules.

4.3.6 Overload protection

Reverse current flow across the emitter-base junction of a silicon planar transistor can cause a change in the current gain. Although these junctions are forward-biased in the amplifier, abnormally large signal voltages can, in the output stage, produce voltages in the emitter-base breakdown region. Such abnormal signals can result from certain failures in the terminal equipment, or in the land facilities feeding the system, or from improper test procedures. The output stage of the PWA is protected against signal overload by a diode connected across the emitter-base junction. Additional protection is provided in the shore terminals.[5]

4.4 Supervisory oscillator

A single-transistor, crystal-controlled oscillator injects a unique frequency at the input of each repeater. These supervisory tones, which fall

* Defined at amplifier output.

alternately in the high and low bands, are used for monitoring transmission levels and localizing faults on the system. A diode limiter and a thermistor-controlled attenuator maintain a stable tone power at the repeater output of −50 and −40 dBm in the low and high bands, respectively. The quartz crystals are circular, plano-convex, AT cut. They resonate in the fifth overtone thickness shear mode. Crystal characteristics are listed in Table V.

4.5 Transistors

The 82-type transistor[6] used in the SG amplifier features a common design for all stages. It is a silicon planar epitaxial type with an f_T of 2.7 GHz designed to operate for 20 years at 1.5 W and a maximum junction temperature of 65°C with a maximum α_0 change of 0.0005. (Sample aging measurements to date indicate no discernible aging.)

Each transistor's small signal characteristics were determined at 10 V, 150 mA, and 9 V, 50 mA, the two nominal operating points. Two-port transmission parameters at 15 frequencies from 100 kHz to 400 MHz were computer-processed by an "on-line" circuit-modeling program to yield element values in the equivalent circuit of Fig. 10. Thus, each set of 60 complex transmission parameters was reduced to nine circuit elements. Noise figure and intermodulation coefficients at the two bias points were also measured. A transistor assignment program optimized the usage of a given lot of transistors by assigning each transistor to a stage in the repeater on the basis of its so-called "noise numbers." Five noise numbers were computed for each transistor. These numbers were proportional to the system noise which would be contributed by the transistor in each possible use.

The input stage noise number was proportional to the 30-MHz noise figure, while the output-stage noise number was dominated by the transistor third-order modulation coefficient. The program minimized the sum of the noise numbers, while yielding a required number of complete transistor sets. There were no grouping requirements based upon stability or gain deviation considerations. Table VI lists the median ac parameters of all transistors used in the TAT-6 link.

Table V — Crystal characteristics

	Low Band	High Band
Frequency band (MHz)	11.120–11.216	27.628–27.820
Spacing between tones (Hz)	200	400
Accuracy from nominal (Hz) (over 20 years and 0° to 45°C)	−118, +40	−293, +98
Resistance-min-max (ohms)	20–100	10–50
Shunt capacitance (pF)	4.25	8.4
Inductance (mH)	500	40
Q	350K to 1750K	140K to 700K

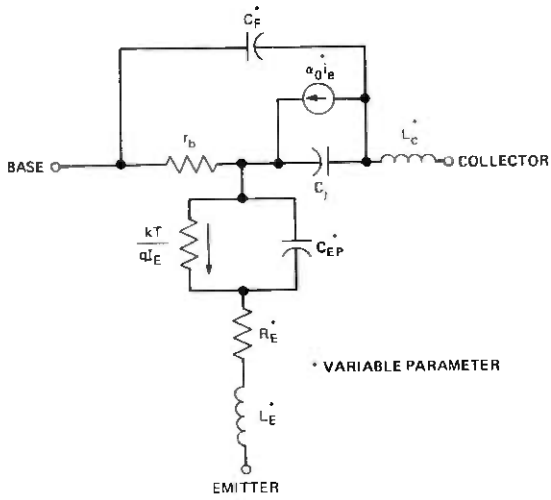


Fig. 10—Circuit model of 82-type transistor.

Table VI — Median transistor parameters

Transistor Type	Function	Bias Volts	Bias mA	Median Parameter Values* α_0	R_E (ohms)	C_F (pF)	C_{EP} (pF)	Stage Placement Criteria
82A	Ampl. stage 1	9.4	50	0.9920	0.173	3.38	128	Noise figure and α_0
82B	Ampl. stage 2	9.2	153	0.9892	0.192	3.29	376	α_0
82C	Ampl. stage 3	9.2	50	0.9885	0.166	3.29	149	C_F
82D	Ampl. stage 4	10.0	147	0.9854	0.189	3.20	419	M_2 and α_0
82E	Ampl. stage 5	10.0	147	0.9881	0.188	3.05	379	M_3 and α_0
82F	Oscillator	8.0	9					$0.9780 < \alpha_0 < 0.9938$
82G	Oscillator	8.0	9					$0.9780 < \alpha_0 < 0.9890$

* Refers to TAT-6 production. The median 30-MHz noise figure for the 82A was 1.87 dB. The median 28.9-MHz M_3 for the 82E was 93.1 dB.

4.6 Surge protection

Two gas-filled electron tubes in each GSF provide the first level of protection against the high-energy surges which accompany faults in the dc power path. These tubes break down at a voltage in the range of 75 to 600 v, depending on the rate of change of the incident surge voltage. Although the gas tubes prevent damage to the passive components, the transistors require a second level of protection which is furnished by reverse-biased diodes (Table VII).

Table VII — Repeater surge protection diodes

Location	No. of Diodes	No. of Diode Packages	Reverse Bias, Volts	Protection Function
Ground separation filter	3	3	12.0	Power path
Preamplifier	2	1	0.86	Amplifier input
Directional filter	2	1	4.0	Amplifier output
Output network	2	2	12.0	Amplifier output

Three Zener diodes in the GSF share equally (within 10 A) peak current surges of up to 160 A caused by the cable discharge, while limiting the voltage across the dc power path to the range −3 to +19 V.

The PWA output transistor proved to be very susceptible to damage from the surge's low-frequency components which pass through the low-band output directional filter. A pair of 4-V reverse-biased diodes placed across the directional filter reduced the surge and also improved the noise-sing stability margin by compressing gain at abnormally high low-band signal power.

Surge protection was probably the most elusive problem encountered in realizing an acceptable repeater design. The results of an extensive testing program in which hundreds of devices were repeatedly surged indicated that the average output stage transistor would survive 50 worst-case surges before failure. Achieving this degree of protection cost what appeared at the time to be a 3- to 4-dB increase in repeater third-order distortion. As further discussed in Section 7.3, the actual degradation was much greater.

4.7 Corona noise

Because of the high voltage present in the repeaters and equalizers, it is possible that static discharges (corona noise) could either excessively degrade system noise performance or lead to an eventual component failure. The requirements adopted were based upon the possibility of insulation damage and failure and are about 20 dB more stringent than current requirements based on group band (48-kHz) data.

Each repeater and equalizer is backfilled with 4.4 atmospheres (50 psig) of dry nitrogen and tested for 21 hours at 7 kV. Broadband detectors, 0.1 to 30 MHz, record all corona "pops" which exceed 1 mV at the output of the repeater. Up to 12 counts exceeding this threshold are allowed in a 21-hour period.

4.8 Passive components

A summary of the types and values of SG repeater passive components[5] is included in Table VIII.

All resistors are the tantalum film type on alumina substrates and have thermo-compression bonded leads.

NPO ceramic capacitors have replaced mica and paper types in all except high-voltage applications. The paper-castor oil high-voltage capacitors are of radically new design, featuring a coaxial structure to reduce parasitic inductance. Tantalum electrolytic capacitors are used for coupling and bypassing purposes where large values are needed.

Ferrite-core chokes, smaller air-core coils for filter networks, and a cavity-tuned inductor are all innovations from previous designs. Hybrid feedback at the amplifier's input and output ports is implemented by means of tightly controlled autotransformers wound on ferrite cores. An autotransformer is also used in the power amplifier to obtain a 2:1 current gain between the Darlington pair driver and the common-base output stage. In the ground-separation filters, coaxial longitudinal chokes help provide the necessary isolation between the amplifier input and output.

4.9 Reliability

As discussed in a companion article,[2] the design objectives for SG component reliability are 0.05, 0.5, and 1.0 FITs (failures in 10^9 device-hours) for passive components, diodes, and transistors, respectively. These levels of reliability have been achieved on previous systems.

Using the preceding reliability objectives and the repeater component count of Table IX results in a repeater failure rate of 21 FITs. (This excludes the oscillator, since its failure would not degrade system performance.)

4.10 Terminations

The cable-repeater terminations, described in detail in a companion article,[3] have a significant effect on system transmission. Although their actual insertion loss is small, about 0.13 dB per repeater at 30 MHz, they contribute a loss ripple due to impedance interactions with the repeater and cable. Reflections occur at each end of the "pigtail," the flexible coaxial transmission lead between the cable and repeaters within the termination. Because the insertion phase of the directional filter and output network varies rapidly with frequency (typically 360 degrees across the high band), the phase of the repeater reflection coefficient across the high band changes by about two revolutions. The reflection coefficient at the cable end of the termination results from excess capacitance in the transition area, 6 pF in the armorless and 3 pF in the armored terminations. The interactions between these two reflections, separated by the electrically long pigtail, causes a rapidly varying loss ripple whose magnitude approaches ±0.035 dB in the high band (see Table IV).

Table VIII — SG repeater passive components

Resistors:

Type	Values (ohms)	Tolerance (percent)	Dissipation (Watts)	Temperature Coefficient (PPM/°C avg)	20-Year Stability (max. percent)
Metal film	3–5000	±0.1 min.	1/16, 1/8	0–40	+0.05

Capacitors:

Type	Values	Tolerance (percent)	Max. VDC	Temperature Coefficient (PPM/°C)	20-Year Stability (max. percent)
Coaxial H-V	0.013 μF	±3	8000	−40	+0.5
Ceramic	5 PF − 0.047 μF	±1 min.	15	±30	+0.02
Tantalum ESR	1–10 μF 2 ohms at 2 MHz	±7	15		

Inductors:

Type	Values (μH)	Tolerance (percent)	Max. DC Current (mA)	Temperature Coefficient (PPM/°C)	20-Year Stability (max. percent)
Air core	0.1–44.3	±1 to ± 10	—	—	—
Torroidal powder core	12.9, 16.25	±3	700	−200	+0.5
Air core slug-adjusted	0.05–5.4	±2 (min. adj)	—	—	—

Transformers:

Type	Remarks
Longitudinal choke	Wound with 50-ohm coaxial cable, 250-μH minimum L to longitudinal currents, max. current 700 mA
Phase reversing	1:1 and 1:−1, 50 ohms, 250 μH min. L, ferrite core
Autotransformers	5:20 ohms and 33.3:50 + 100:150 ohms, ferrite core

Thermistors:

Type	Values (ohms)	Tolerance (percent)	Temperature Coefficient (ohms/ohms/°C)	20-Year Stability (max. percent)
Parallel disks	13.83 at 11.2°C	±1	−0.044 at 25°C	+2
Individual disk	29.0–34.0 at 25°C			

V. EQUALIZER ELECTRICAL DESIGN

For the first SG installation (TAT-6), two types of undersea equalizers were developed, a conventional ocean-block equalizer (OBE) and a special shore-controlled equalizer (SCE). The location in the system of these equalizers and their functions in the overall equalization plan are described in a companion article.[2]

The primary job of an OBE is to equalize the accumulated misalignment at the end of an ocean block. Such misalignment arises from design and manufacturing deviations in both repeaters and cable and from

Table IX — SG repeater component count

Component Type	Ground Separation Filters	Directional Filters	Output Network	Preamplifier	Power Amplifier	Oscillator HB	Oscillator LB	Subtotal
Resistors		4	8	19 or 21	17	9	9	59
Capacitors:								
Ceramic	4	36	19	12	8	6	6	85
Paper	2							2
Tantalum	1	1		6	8		2	16 or 18
Thermistor*				1				1
Inductors:								
Air core		20	11	2	4	3	3	40
Ferrite				4	4	1	1	9
Dust core	2							2
Transformers	2	2		1	2			7
Diodes	2	2		2	1	2	2	9
Transistors				2	3	1	1	6
Gas tubes	4							4
Crystals						1	1	1
Subtotal	17	65	38	51	47	23	25	†

* Thermistor is used only in shallow-water temperature-controlled repeaters. It replaces two resistors.

† Total components: repeater with HB oscillator: 241; repeater with LB oscillator: 243.

uncertainties associated with predicting the overall transmission of the system in the ocean-bottom environment. The OBE is adjusted to its final setting aboard ship shortly before it is overboarded.

The primary job of the SCE is to compensate for transmission changes occurring in one sector over the life of the system. (For TAT-6, a sector is approximately one-fifth of the system length.) From the standpoint of the transmission path, both the OBE and SCE are passive. Gain for equalization is obtained by shortening the cable length between the two repeaters adjacent to an equalizer to a total of one mile (one-half mile on either side of the equalizer). In the lower part of the low band, additional equalization gain is provided by the repeater gain boost.

5.1 Ocean-block equalizer

The SG ocean-block equalizer design retained those features of the SF system OBE that proved so successful in equalizing that system, viz:

(*i*) Independent equalization in the two transmission bands.
(*ii*) Switchable networks—a set of 14 networks (7 per band), identical in all OBEs, that can be individually switched "in" or "out" of the transmission path.
(*iii*) Mop-up networks—a series of up to seven tandem networks in each band that can be designed and fabricated late in the manufacturing cycle of an equalizer. These could be different for each OBE.

The switchable and mop-up equalization modes accomplish the primary OBE function and are discussed in detail in separate sections.

Figure 11 shows the OBE. Since there are no active electronics in the OBE, a power-separation filter (PSF) at each end bypasses the high-voltage direct current. This arrangement places the internal chassis ground at the same potential as the outer pressure housing (sea ground). (The repeater and SCE differ in having the internal chassis at center-conductor potential.)

Separation of the two transmission bands within the OBE is achieved through use of two directional filters, each consisting of a low-pass and high-pass filter paralleled at one end to form a three-port network.

A build-out network in each band centers the switchable network adjustment range around the nominal OBE loss, while preserving the maximum possible flat loss for mop-up network design. Nominal OBE total loss corresponds to that of 4.1 nautical miles of 1.7-in. SG cable at sea bottom plus a loss equal to the per-block repeater gain boost.

A coaxial transmission test lead exits the OBE housing through a high-pressure seal. This lead is used during tests of the assembled shipload of cable and repeaters prior to laying and during measurements between ship and shore while laying.

Within the OBE, the test lead terminates in a symmetrical 3-port splitting pad that can be switched "in" or "out" of the transmission path. After final adjustment on shipboard, the pad is switched "out." This disconnects the test lead and connects transmission straight through. Finally, the end of the external test lead is overmolded.

5.1.1 Switchable networks

The insertion loss functions of the 14 switchable networks are given in Table X. Each network is realized by a constant resistance (50-ohm) structure. Three classes of shapes are identifiable, "root-f," "root-f minus f," and a broad "bump," whose peak loss is near the center of each band. (The latter two shapes have small amounts of root-f loss added to make them physically realizable.) These three shapes are plotted in a companion article.[2] The frequency characteristics of 16 of the total of 128 settings obtainable in the low band are shown in Fig. 12. The switchable networks are capable of equalizing only smooth, relatively broad components of misalignment. (The more ripply components are handled by the mop-up networks.)

Of the three classes of switchable network shapes, the one with by far the largest range is root-f. This is because repeater gain and cable loss, as well as changes in cable loss due to temperature and pressure effects, are approximately proportional to the square root of frequency. It follows, therefore, that residual differences in these parameters, including those due to inaccuracies in predicting sea-bottom temperature or depth

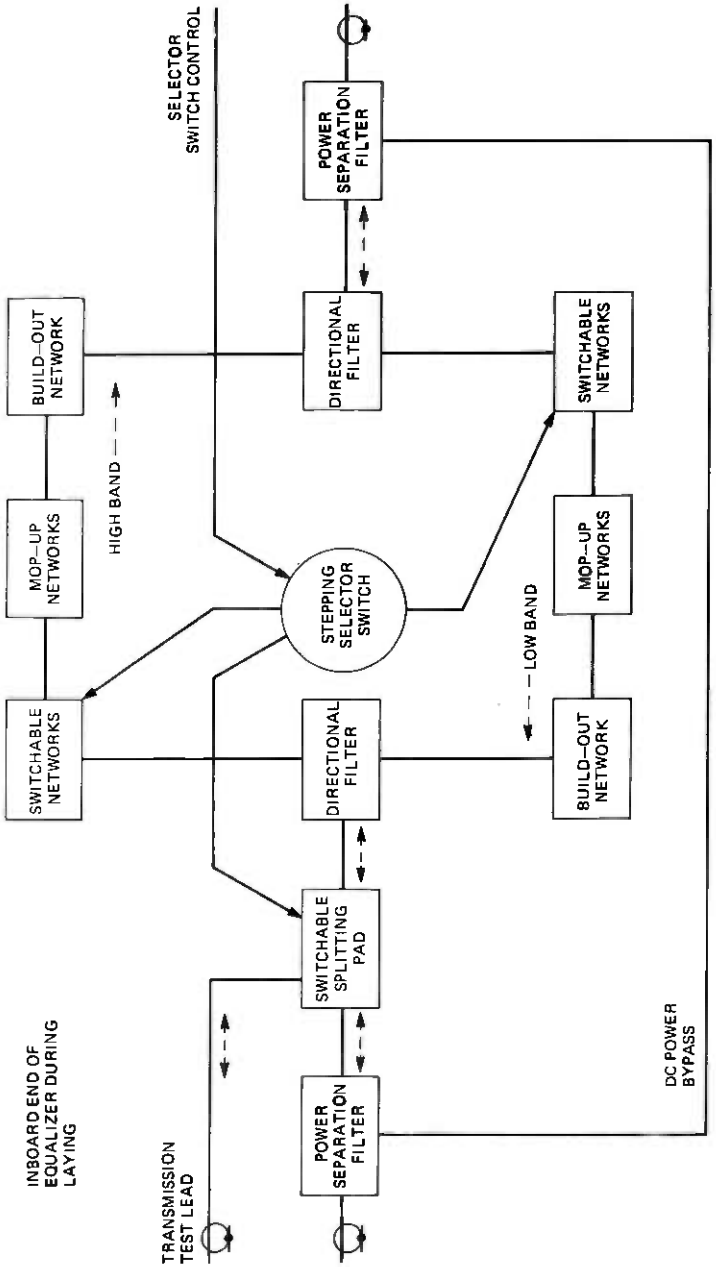


Fig. 11—Block diagram of ocean-block equalizer.

Table X — Ocean-block equalizer switchable loss shapes (dB)

Low Band	High Band
$8\sqrt{f/30}$	$8\sqrt{f/30}$
$4\sqrt{f/30}$	$4\sqrt{f/30}$
$2\sqrt{f/30}$	$2\sqrt{f/30}$
$1\sqrt{f/30}$	$1\sqrt{f/30}$
$8(\sqrt{f/14} - f/14) + 2\sqrt{f/30}$	$17.26(\sqrt{f/30} - f/30) + 0.5\sqrt{f/30}$
$4(\sqrt{f/14} - f/14) + \sqrt{f/30}$	$8.63(\sqrt{f/30} - f/30) + 0.5\sqrt{f/30}$
1.5-dB bump + $\sqrt{f/30}$	1.5-dB dump + $\sqrt{f/30}$

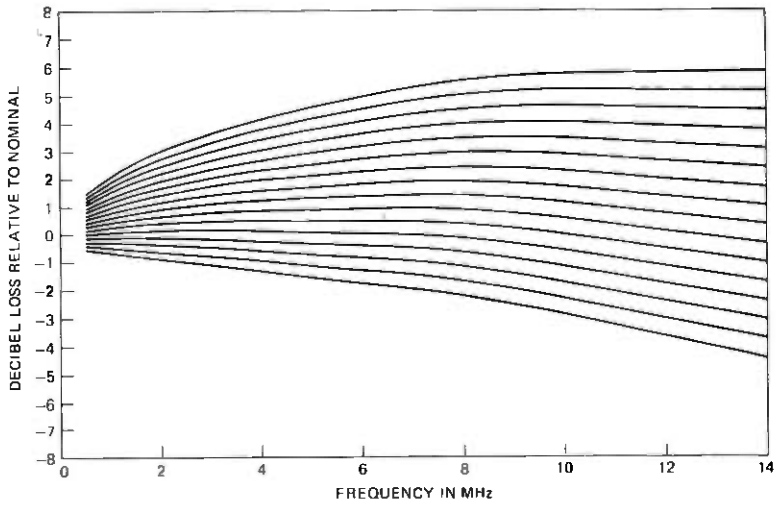


Fig. 12—One-eighth of the possible switchable network settings obtainable in the low band (nominal loss is equivalent to half the total switchable network loss).

produce transmission changes which are roughly proportional to root-f.

The rationale behind giving the "root-f-minus-f" shape the second largest range is more subtle. Consider the following: Cable loss per unit length, α, can be expressed quite precisely as a function of frequency, f, by

$$\alpha = A\sqrt{f} + BDf, \tag{3}$$

where A and B are independent of frequency and D, the dissipation factor, is a measure of the loss due to the cable dielectric. Variations in nominal α for manufactured cable that are due to variations in A are quite closely compensated for over the entire transmission band when each cable section is cut individually to a specified top frequency loss. Variations in B or D, on the other hand, are compensated for correctly only at the cutting frequency, because one is attempting to correct an

attenuation error that is directly proportional to f* by varying a loss that is closely proportional to $\sqrt{f}$. (For SG cable, at 30 MHz the first term in eq. (3) is 94 percent or more of the total.) The resulting misalignment is compensated for reasonably well in the OBE by the root-f-minus-f switchable shape.

The third switchable shape, the broad bump, is the only one that is not specifically cause-associated. However, the root-f shape has its greatest effect toward the top of each band, and the root-f-minus-f shape has its greatest effect near the bottom. Thus, the bump, whose effect is primarily in the middle of each band, provides a general-purpose equalization characteristic to the switchable networks.

Associated with each switchable network (and the test-lead splitting pad) is a small magnetic latching double-transfer relay. These relays are set to one of two possible states by means of a stepping selector switch. The stepping selector is operated by a series of pulses of either polarity applied to the switch control lead that enters the OBE housing at the end opposite the transmission test lead (Fig. 11). The selector is a single-pole, 16-position switch, each position associated with a particular relay and network, or with a reference condition. Each pulse applied to the switch control lead accomplishes two functions. First, it advances the selector one position (independent of pulse polarity). Second, a positive pulse switches the network (relay) associated with the new position "out," whereas a negative pulse switches the network into the transmission path.

5.1.2 Mop-up networks

In contrast to the switchable mode, two features of the mop-up networks worth noting are:

(*i*) Their ability to provide fine-grain (i.e., ripply) compensation.
(*ii*) The relatively short interval required between design of these networks and completion of manufacture of the OBE.

To achieve these features, each mop-up network uses a modular-design, constant-resistance (50-ohm) bridged-T configuration constructed from a predetermined stockpile of components and standardized hardware.† Parasitic circuit effects of components and hardware have been well-characterized in advance, because there is no time for the usual engineering development to assess and correct second-order effects. Specially written computer programs are used to expedite design and documentation effort. Network performance is computed taking into account the effects of parasites and discrete component values, and elec-

* This is not strictly true for variations in D which can be frequency-dependent, but the point is still valid.

† This ensures that aged undersea cable quality components are available simultaneously with a network design.

trical test specifications are generated for both individual networks and tandem assemblies. The OBE is designed so that the complete equalizer, less mop-up networks and high-pressure housing, can be constructed and tested before mop-up design information is available.

An individual bridged-T network can provide a positive bump shape (loss vs frequency), a negative bump (dip), simple low- or high-pass shapes, or flat-loss. Networks can be connected in tandem (with minimum interaction due to the constant resistance design) to obtain combinations of these shapes. In practice, most mop-up networks are the bumps or dips, whose circuit configuration is shown in Fig. 13. For a fixed characteristic impedance, R_0, such simple shapes can be completely specified by peak (bump or dip) amplitude, center frequency, and stiffness (which can be defined as the ratio of upper half-loss frequency to center frequency).

The range chosen for the SG mop-up component stockpile is shown graphically in Fig. 14. The symmetry of this figure results from the nature of the bridged-T for bump and dip shapes which calls for a pair of LC resonance values that intersect equidistant from the horizontal R_0 line along a vertical line corresponding to the center frequency. Similarly, the pair of shunt and series resistance values falls equidistant above and below the R_0 line in Fig. 14. In some cases, when a network design calls for component values outside this range, a flat-loss network transformation can be applied which moves the required values inside the stockpile range. The resulting transformed network then has a loss which is the sum of the desired loss shape and additional flat loss.

The designer of mop-up networks must "spend" his flat loss wisely, because the available flat-loss is limited to 9.5 dB in the low band and 12.5 dB in the high band. The number of mop-up networks in a particular OBE could be less than the seven per band allowable from physical

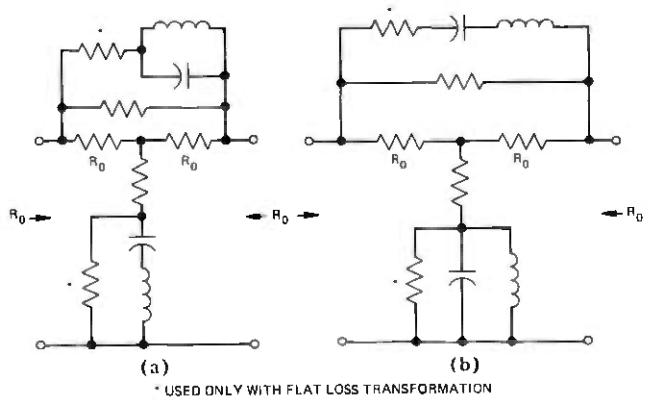


Fig. 13—Mop-up network configurations. (a) Loss dip. (b) Loss bump.

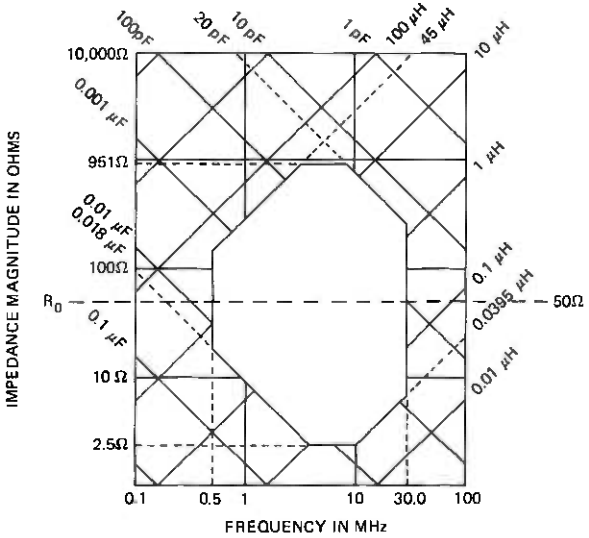


Fig. 14—Stockpile range.

Table XI — Passive OBE components

Component	Description	Value	Tolerance (percent)	Use
Resistor	Wirewound, vitreous enamel, power resistor	14.0 ohms	±1	Selector switch circuit
Capacitor	HV oil-impregnated, paper dielectric, housed in ceramic cylinder with metal end caps	0.02 μF	±3	PSF(A) and (B) assemblies
Inductor	Duolateral-wound over plastic form capable of continuous operation with up to 0.7 amperes through windings	1.0 mH	±5	PSF(A) and (B) assemblies
Selector-switch	18-point, unidirectional switch			
Relay	"Crystal can" size, hermetically sealed, magnetic-latching relay containing two transfer break-before-make contacts with No. 1 contact metal. Relay has an H-shaped spring-supported armature-magnet assembly and an isolated, hermetically sealed coil chamber.			

constraints because of the limited amount of flat loss which is available.

5.1.3 Components

With the few exceptions noted in Table XI, all components used in the ocean-block equalizer are identical in type with those used in the repeater, which have already been described.

5.2 Shore-controlled equalizer

Because of uncertainty in the long-term stability of cable loss, plans for the earlier SD and SF systems had included shore-controlled equalizers (SCEs). In both cases, however, development effort on the SCEs ceased when adequate cable loss stability was assured. Based on this experience, the SG system development began assuming SCEs would not be required. In mid-1975, six months prior to the beginning of the first deep sea lay of TAT-6, cable loss changes noted in factory measurements[3] clearly indicated the necessity for installing SCEs. Development was started in August 1975, less than six months before the date the first unit was due to be shipped. This extremely short time interval required the use of readily available components.

5.2.1 SCE realization

The SCE has much in common with the OBE: directional filters, switched network configuration, and certain hardware. The OBE stepping-selector switch is replaced by an instrument-type 50-mW dc motor and gear train furnished by CIT.* A modified repeater GSF replaced the PSF of the OBE, and a final gear reducer and wafer switch assembly were procured from commercial sources. Special filters, amplifiers, and rectifiers using available and aged SG components were designed to control the electromechanical switching operation.

5.2.2 Method of control

The SCE uses the same type of magnetic latching relays as the OBE for switching networks in or out of the transmission paths. Relay states in each SCE are controlled by two unique frequencies spaced 400 Hz apart in the 11.464- to 11.472-MHz band. As shown in Fig. 15, the two signals transmitted from the A terminal are selectively amplified and rectified in the SCE to produce two dc control voltages. The output of the motor channel powers the motor which turns the selector switch; the output of the relay channel, when present, serves as an "activate" signal for the latching relays as they are sequentially accessed by the selector switch. Position identification information is transmitted back to the A terminal by amplitude modulating the second harmonic of the motor channel frequency. Two adjoining selector positions are associated with each relay and network. One position corresponds to switching the network in, and the second corresponds to removing the network.

* Compagnie Industrielle des Télécommunications.

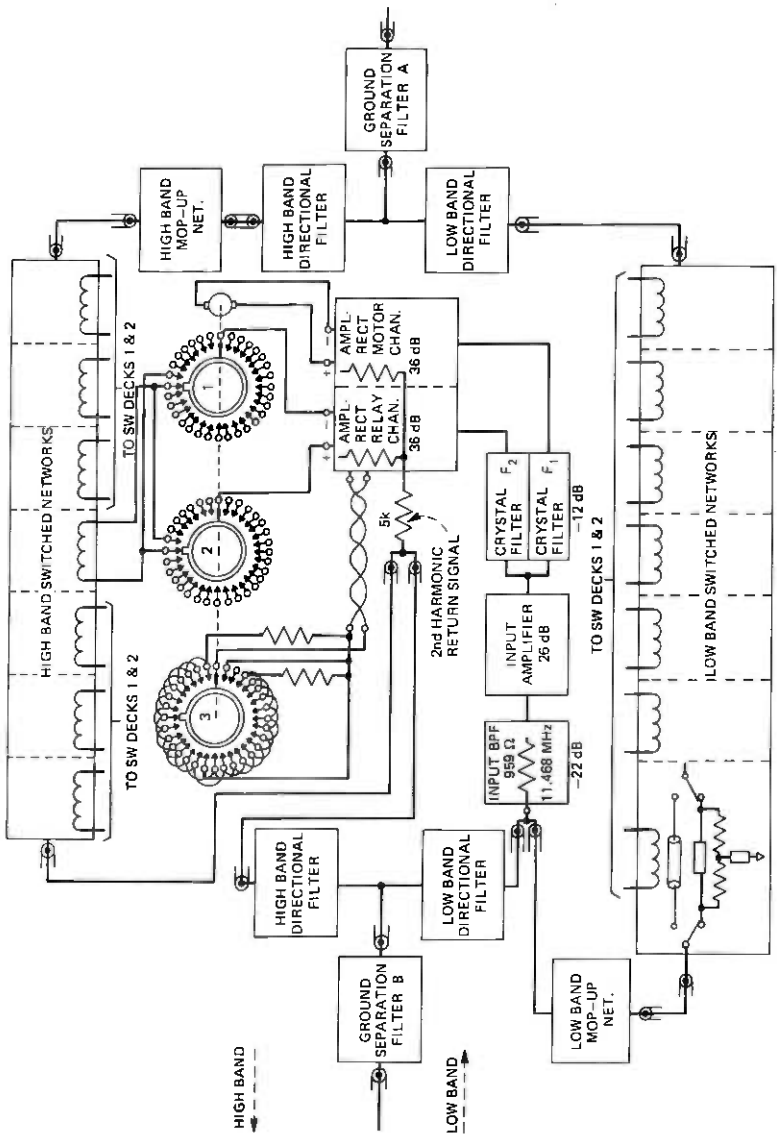


Fig. 15—Block diagram of SG shore-controlled equalizer.

Figure 16 is a chart recording of the second harmonic signal as received during an SCE-adjusting operation. The annotations indicate the loss shapes of the seven switchable networks in each band. Individual networks are switched in or out by transmitting a short burst (2–3 s) of the relay channel tone during the interval, typically 12 s, that the network coil is connected to the relay channel rectifier. By monitoring the level change of a supervisory tone which goes through the SCE, one can verify immediately that the network has switched.

Figure 17 shows the typical characteristics of the two control channels. The SCE is normally left in the index position after adjustment. In this position, the relay channel rectifier is connected to a diode rather than a relay coil. As the relay channel signal level is slowly increased, the second harmonic return increases smoothly until an abrupt step of about 20 dB occurs, indicating that the rectifiers have begun to conduct. This point serves as a reference level; a signal 12 dB above the reference will produce sufficient power to switch a relay. The motor channel characteristics are similar. The control tones are tuned to the precise center frequencies of their respective crystal filters by monitoring the returned second harmonic signal.

5.2.3 Control system design

The control system of the SCE is a modified version of that used in the French S-25 system designed by Compagnie Industrielle des Télécommunications. The success of the accelerated SCE project is largely due to the cooperation of CIT in sharing their knowledge and experience and in furnishing the critical motor-gear train assemblies.

A number of measures prevent false operation of the SCE. The control frequency band, 11.464 to 11.472 MHz, is a space between supergroups normally used for noise monitoring. A band elimination filter at the A terminal blocks undesired incoming signals in this band. The operating level for the control tones was chosen sufficiently high, up to −6 dBm at repeater output, to preclude the possibility of inadvertent operation. The second harmonic return signal exceeds the system background distortion by at least 20 dB. From the normal "rest" position of the controls, *both* F_1 and F_2 must be applied to change the SCE setting.

5.2.3.1 Filters. The input bandpass filter, centered at 11.468 MHz, is an LC type with a 4-dB bandwidth of 0.5 MHz and is designed to operate between 1000- and 50-ohm impedances. Together with the 959-ohm isolation resistor, it has a through loss of 22 dB and adds less than 0.2-dB bridging loss to the low-band transmission path.

Each crystal filter contains a single crystal of the same design used in supervisory oscillators. The crystal, with a series resistance of 20 to 36 ohms, operates in series between impedances of typically 5 ohms. Each filter provides at least 34 dB of adjacent channel suppression (400 Hz away) while adding 12 dB of flat loss.

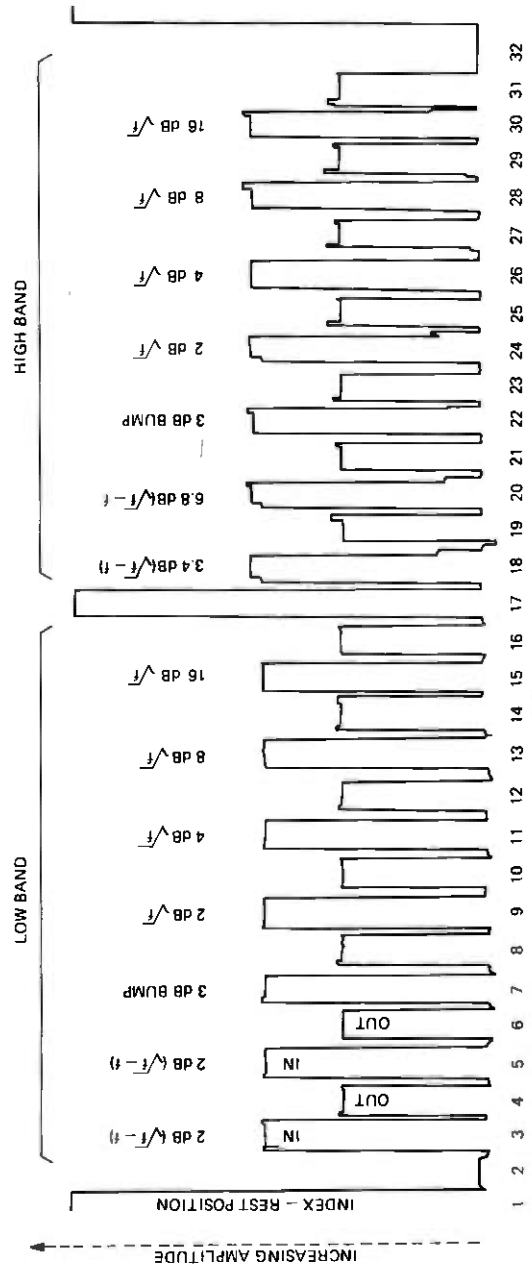


Fig. 16—Shore-controlled equalizer—second harmonic signal, $2F_1$, returned to shore during adjustment.

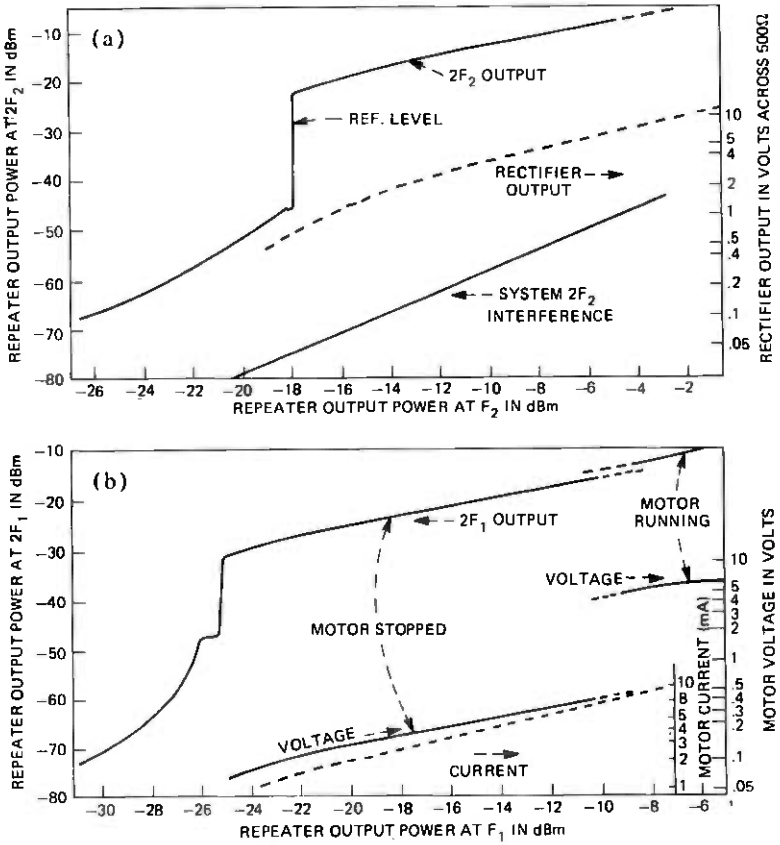


Fig. 17—(a) SCE relay-activate channel characteristics. (b) SCE motor channel characteristics.

*5.2.3.2 **Amplifiers.*** The three amplifiers are identical except for gain and terminations. Each has two common emitter stages with gain controlled by feedback and each is biased at 12 V, 219 mA. The input amplifier has 26-dB gain and 50-ohm input and output impedances. Each power amplifier provides a nominal gain of 36 dB. The power amplifier input impedances are low, as required by the crystal filters, and the outputs are tuned to operate efficiently into a voltage doubler-type rectifier. The actual gain of each power amplifier is tailored, by choice of feedback resistor, to accommodate the loss of its associated crystal filter. The gains and losses are arranged so that each rectifier will deliver 50 mW (5.0 V, 10 mA) to its load (motor or relay) when the signal power at the SCE input is −8 dBm. Each amplifier rectifier will deliver a maximum output power of at least 200 mW.

*5.2.3.3 **Electromechanical.*** A precision-instrument-type, permanent magnet, dc motor drives an attached 2000:1 ratio gear train. A 15:1 worm

gear which connects to the wafer switch assembly provides the final reduction. Each of three identical switch wafers has a single pole and 32 positions, 31 of which are active. Two of the wafers connect to the relay coils. Each coil requires two positions, one for switching the network in and one for switching the network out. The third wafer is used as a four-level loss switcher which amplitude-modulates the power of the second harmonic return signal as the switch slowly turns. This return signal informs station personnel of the select switch position.

5.2.4 Future improvements

Since completion of TAT-6, a solid-state circuit has been developed to replace the motor-switch assembly. The remaining components in the SCE are virtually unchanged.

5.2.5 Switchable range

The switchable loss shapes available in the SCE are listed in Table XII. The networks are similar to those in the OBE except that the range and step sizes are twice as large. The SCE has no mop-up network capability.

VI. MANUFACTURING

6.1 Schedules and production rate

Repeater and equalizer final assembly and much of their manufacturing were done at a special factory in Clark, N.J. as with earlier repeaters for SD and SF.

Authorization to proceed with production was received in February 1972, with the first repeater scheduled for shipment in December 1974. Prior to the start of manufacturing, it was necessary to design new manufacturing and test facilities required to handle the SG hardware and the increased frequency range.

Because of the mid-1976 service date of TAT-6 and because the system called for over 800 repeaters and equalizers, it was necessary to achieve

Table XII — Shore-controlled equalizer switchable loss shapes (dB)

Low Band	High Band
$16\sqrt{f/30}$	$16\sqrt{f/30}$
$8\sqrt{f/30}$	$8\sqrt{f/30}$
$4\sqrt{f/30}$	$4\sqrt{f/30}$
$2\sqrt{f/30}$	$2\sqrt{f/30}$
$8(\sqrt{f/14} - f/14) + 2\sqrt{f/30}$	$34.52(\sqrt{f/30} - f/30) + 2\sqrt{f/30}$
$8(\sqrt{f/14} - f/14) + 2\sqrt{f/30}$	$17.26(\sqrt{f/30} - f/30) + 0.5\sqrt{f/30}$
3-dB bump + $\sqrt{f/30}$	3-dB bump + $\sqrt{f/30}$

production capacity of 16 units per week. Concurrently, it was necessary to maintain capacity for other systems at a rate of three units per week.

6.2 Environmental conditions

All critical assembly operations are performed in "clean" rooms where temperature, humidity, and airborne contaminants are closely monitored. Objectives are less than 50,000 dust particles above 0.5 microns per cubic foot, temperatures of 22.8° ± 1.1°C (75° ± 2°F), and relative humidity of less than 40 percent. In the paper capacitor winding room, relative humidity is controlled to less than 20 percent. Dust counts on three sizes of particles (0.5, 1.0, and 2.0 microns) are monitored weekly for each clean room. Positive air pressures are maintained and monitored in all rooms and associated vestibules to prevent entry of contaminants.

6.3 Product development and manufacturing facilities

The design of the SG system required improved manufacturing and test facilities. The purpose of these facilities was to enable the manufacture, aging, and testing of new components such as ceramic capacitors, to provide new transmission test facilities, and to provide new and improved facilities for closure type operations. Development work on facilities began at Clark, N.J. in February 1971.

6.3.1 Aluminum castings

The large number of repeaters and the elimination of the hermeticity requirement for the electronics unit required that more economical procedures for making aluminum castings be investigated.[7] The investigation revealed that the castings should be made from the low-pressure permanent mold (LPPM) process and the cylinder should be extruded. Expensive machining operations were virtually eliminated with no sacrifices in quality. In addition, the separate, one-piece cylinder simplified the epoxy coating procedure and subsequent assembly with the main frame, which simply slides into the surrounding cylinder, where it is locked in place.

6.3.2 Ceramic capacitors

Ceramic capacitors, new to undersea cable application, were substantially different from their predecessor,[8] the mica capacitors, both in physical appearance and electrical characteristics (Fig. 18). Their use in SG circuits was mandated by the need for smaller encapsulated units having negligible parasitic inductance. Replacement of mica with ceramic also reduced costs and avoided a serious procurement problem, that of purchasing top quality mica.

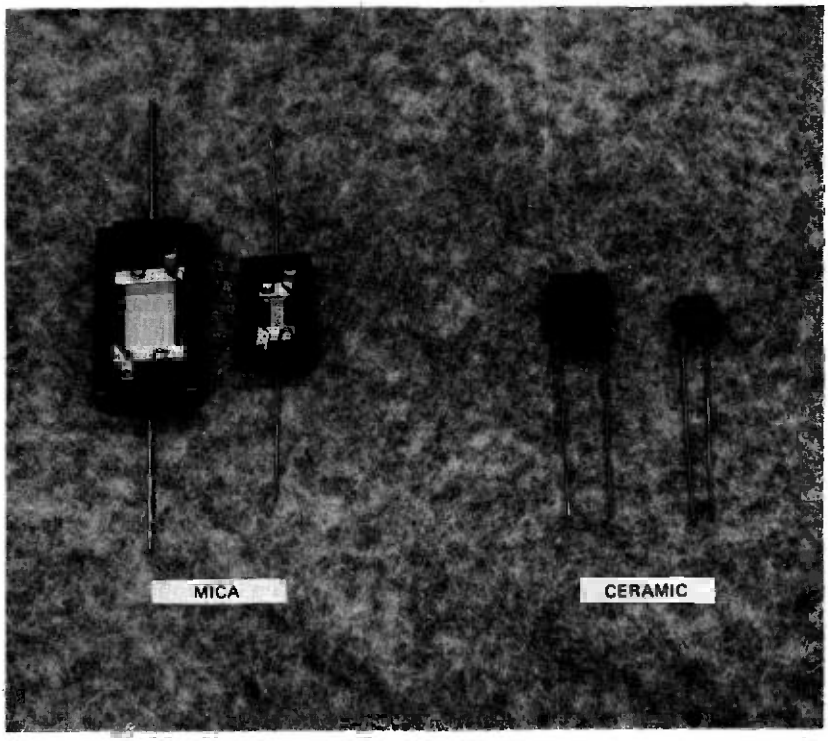


Fig. 18—Comparison of ceramic and mica capacitors.

Although ceramic capacitors are purchased from commercial suppliers, prior to manufacture of electronics, complete in-house qualification procedures had to be specified and implemented. Principal tests performed included insulation resistance, dielectric strength, capacitance, and conductance along with 17 weeks of voltage conditioning aging (see Fig. 19).

Because of its physical design, the ceramic capacitor was readily adaptable to automatic testing methods. Special test processing fixtures were provided onto which 25 capacitors were attached for processing through the full complement of tests. Flexibility was designed into the fixtures, enabling the capacitors to be tested individually for capacitance, conductance, and insulation resistance, or as a group of 25 as is required during aging and breakdown testing. Because of stringent tolerance requirements, 0.1 percent in some cases, test set accuracies and repeatability has to be in the order of 0.01 percent. Special design features such as gold-plated contact points on fixtures and shielding to eliminate stray parasitics were necessary.

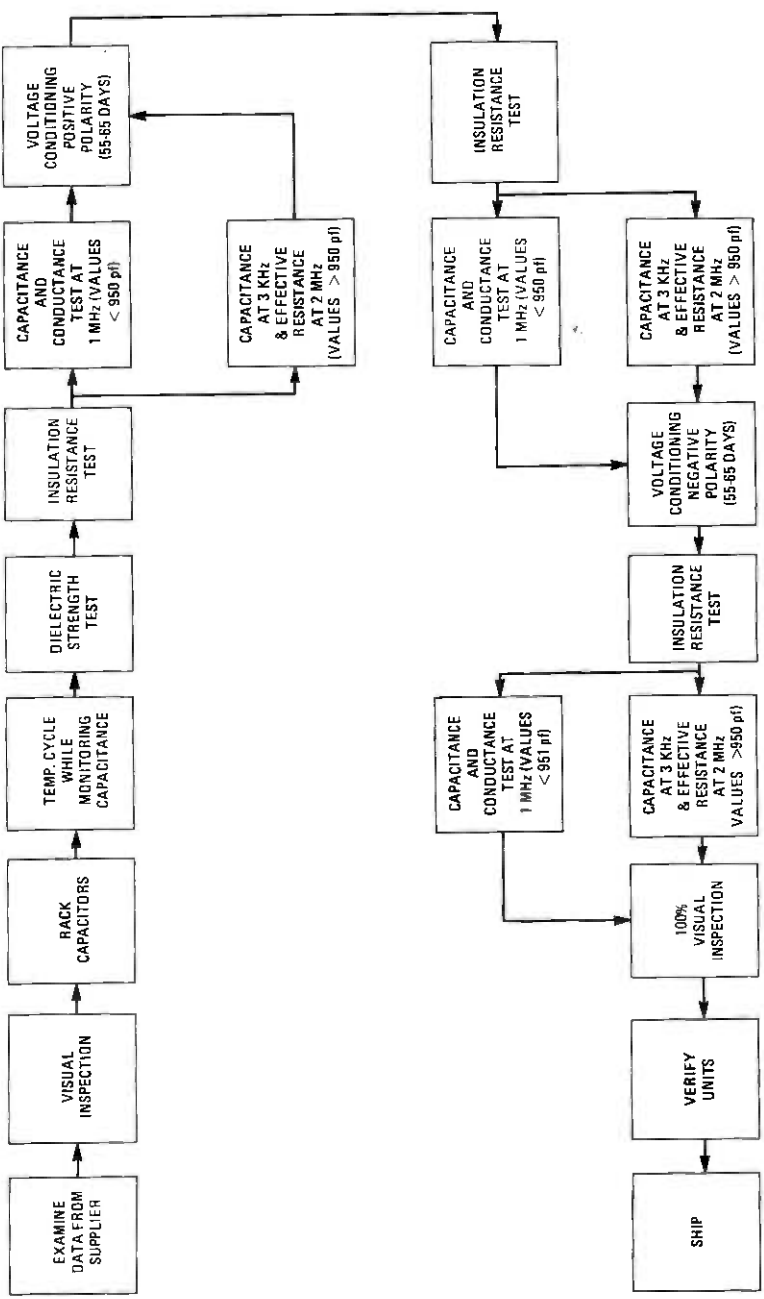


Fig. 19—Ceramic capacitor process flow chart.

6.3.3 Printed-wiring-board assemblies

From a manufacturing viewpoint, one of the significant changes in physical design was the use of printed-wiring boards (PWB). PWB requirements and assembly techniques were developed with specific attention to component mounting procedures, fixturing, and soldering techniques. Overlays were provided to assist the technician in proper location of components. In addition, training programs were developed to acquaint the operators and inspectors with component identification, and to identify various component and board defects which could degrade reliability.

Detailed investigations were undertaken to establish improved board manufacturing processes and requirements. Plating procedures were revised and improved at printed-wiring board manufacturing locations. New visual standards to aid in judging dewetting of solder-coated circuit paths were also developed.

6.3.4 Seals

The function of the high-pressure seal is to provide a dc power and signal transmission path into the high-pressure housing. It must withstand high dc voltage and full sea pressure, while maintaining a satisfactory coaxial transmission path. The seal chosen for the SG repeater and equalizer was a modified version of the 8-type seal manufactured at Western Electric's Burlington, N.C. plant. Its predecessor, the 3-type seal used in SF, did not have suitable transmission properties in the SG high band.

The primary sealing mechanism in both seals is the thin polyethylene gasket adjacent to the ceramic. In the 3-type, the ceramic is backed up by a large metal disk which provides satisfactory mechanical performance but introduces excessive capacitance with respect to the 50-ohm transmission impedance. The metal disk was eliminated in the design of the 8-type seal and a larger ceramic cylinder was used. This produced a suitable mechanical seal and at the same time reduced the capacitance to provide proper transmission impedance.

The 8-type seal (Fig. 20) was completely molded in one operation, including attachment of both pigtail leads, rather than three molding operations needed for the 3-type seal. This resulted in reduced cost and increased production capacity from each molding press.

The most severe problems encountered in developing the 8-type seal involved the feed-through assembly. The copper caps on the ends of the ceramic are provided to control electric field strengths. Initially, these were brazed to the ceramic to provide hermeticity in the assembly, but forces generated in the molding operation were sufficient to break the ceramic adjacent to the braze joint, destroying the hermeticity. This problem was overcome by making the conductor in two parts with the

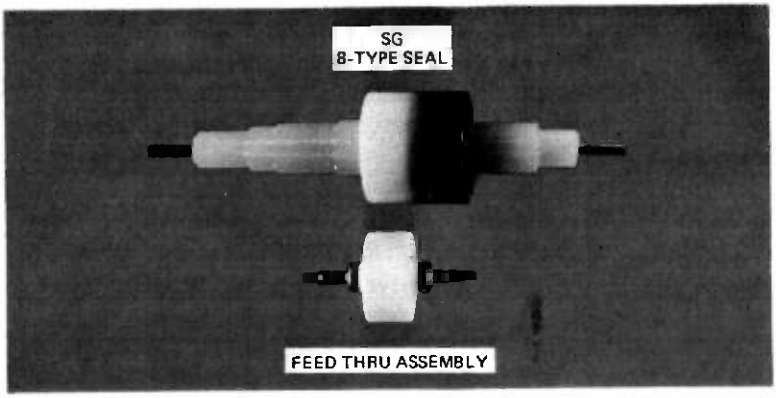


Fig. 20—High-pressure seal.

cap configuration integrally formed on each. These are assembled to the ceramic part (with small fluoroplastic washers between the ceramic and the caps) and screwed together with a coupling sleeve within the hole in the ceramic. A small slug of solder is included in this assembly. The unit is then heated to melt the solder into the threads and the threads are torqued hand tight. Upon cooling, the conductor parts are fused together. Differential thermal shrinkage compresses the fluoroplastic to form a hermetically sealed assembly.

This design change eliminated the need for metallizing the ceramic and brazing the caps to the ceramic, reduced the cost of the feed-through assembly, and produced a higher quality seal.

6.3.5 Subassembly and final repeater testing

Early in the planning, it was decided that the bulk of the transmission testing would be performed by a computer-operated transmission measuring set[9] (COTMS). Duplicate sets were provided. Choice of the COTMS was based on: (*i*) its ability to handle a measuring sequence automatically, including recording and interpreting results, and (*ii*) its precision, repeatability, and precision up to the prescribed top frequency. The COTMS has an accuracy with measurement-averaging of within 0.001 dB. Full utilization of the COTMS included time-sharing of the main measurement unit via remote stations, and formulation of programs specifically tailored to the product needs.

Software packages provided specific instructions to the tester so that type of unit, operation to be performed, and serial number could be double-checked prior to the start of testing. In addition, the software packages identified the proper sequence of test frequencies, checked test results at each frequency against specified limits, and stored a test history for each unit. CRT display units were used as input and output devices

to provide communications between the tester and the computer-monitored test facility. Data were collected on disk and later transferred to magnetic tape for analysis and storage. With the above programming and proper use of remote stations, a test capacity in excess of 20 repeaters per week was attained.

In addition to the COTMS facilities, a precision-adjusting test facility known as MINI-COTMS was also needed. Although called a MINI-COTMS, it bears little resemblance to the COTMS described above. Application of this facility was confined to repeater network and equalizer network tuning.

In the repeater network, five inductors are available for final tuning. Nominal gain for each of 118 discrete in-band test frequencies is stored in computer memory, while deviations from nominal for each frequency are visually displayed on the MINI-COTMS. A special program and adjusting procedure had been formalized which enabled the tester to minimize deviations at most in-band frequencies. The true value of this facility is readily apparent during adjustment, since some of the inductor adjustments interact with one or more others. In this situation, it is a great advantage to have the entire range of frequencies displayed on the CRT output. (Insertion loss measurements are displayed with a precision of ±0.015 dB.) The CRT output, which includes the RSS value of the gain deviation, is observed by the tester while he is performing the final tuning operation. Maximum deviations allowable are of the order of ±0.05 dB with the number of peaks being minimized. Hard copy output is obtained in two forms: values printed to the thousandth of a decibel and a plotted form. Optimum adjustment and use of results to apply corrective action led to minimum deviations among repeaters, which in turn led to a more readily equalizable system and more voice channels.

Equalizer adjustment procedures are similar to the procedures used to tune repeater networks but on a larger scale, with two parameters (insertion loss and return loss) being monitored. Over 70 inductors per equalizer are adjusted on this facility, with a maximum of 14 appearing on any one subassembly. Without a facility having the interactive capability and accuracy of MINI-COTMS, equalizer adjustment would have been virtually impossible.

Other facilities designed for network testing were the crystal oscillator test set, accurate to within 0.1 Hz in the 27-MHz frequency range, a noise and modulation test set, and coarse adjustment (±0.1 dB) test facilities.

6.3.6 Corona testing

Repeaters have been designed to eliminate corona. However, disturbing "pops," or electrical discharges, can be generated under certain conditions. "Pops" above certain voltage levels can cause errors in data

transmission, degrade the quality of voice transmission, and deteriorate insulation.

For SG, extensive effort was spent on developing corona detection equipment and trouble-shooting techniques. A new wide-band detector was designed with the capability of detecting all possible wave shapes with uniform sensitivity across the 30-MHz transmission band. A 2.44-m × 3.66-m shielded room was constructed to provide for high production levels and stringent shielding requirements and to house the product during test. The capability of testing ground separation filters in a pressurized state was also provided. Experience obtained on previous projects proved that pressurizing tends to retard corona impulses. This is an acceptable procedure since final repeaters are pressurized with 4.4 atm (50 psig) of dry nitrogen. In actual production, 12 fixtures were provided and used with a majority of the ground separation filter testing performed with 1.68 atm (10 psig) of dry nitrogen applied to the product.

6.3.7 Closure

Operations associated with closure of the repeater required some completely new facilities as well as an upgrading of existing facilities. Among the more significant changes were: introducing eB (electron beam) welding for joining the high pressure end covers to the copper beryllium housing, instituting a low-pressure test for checking the prime sealing surface of the seal, and modifying vacuum-drying facilities to enable this operation to be performed in the assembled repeater stage of production.

Seals were tested at low pressure to check the mating of the prime sealing surface of the high-pressure seal to the surface of the high-pressure end cover. In performing this test, an O ring and seal are assembled into a high-pressure cover under a small axial load. This relatively minor load, compared to the load when subjected to ocean-bottom pressures, permits the prime sealing surfaces to be tested without fear of the seal bulging to form a side or secondary seal. Helium at 7.8 atm (100 psig) is applied to the seal for two hours. The leak rate of helium past the O ring and prime sealing surfaces, plus migration through the polyethylene body of the seal or along the interface of the internal feed-through, is monitored with a mass spectrometer.

Because of the elimination of a sealed inner unit, the vacuum-drying operation had to be performed in the final stage of manufacture (unit sealed in its copper beryllium housing). This meant that existing facilities had be to modified to:

(*i*) Include heating blankets that reduced the heat losses and were capable of heating the large copper beryllium housing to a temperature of 57°C (135°F).

(*ii*) Handle a 500-pound (227 kg) repeater.

6.4 Quality and verification of product

Manufacture of submarine cable repeaters and equalizers places emphasis on ensuring long-term stability and system reliability. To achieve this end, factors such as training personnel, maintenance of a clean work environment, and adherence to detailed processing specifications were stressed. A series of multiple inspections includes audits of procedures, methods, drawings, and product to assure conformance. These are all vitally important. For TAT-6, over 12,000 items of component performance data were transmitted on a daily basis to the data center for analysis, subsequent statistical manipulation, and finally storage.

As previously mentioned, transmission tests on the electronic subassemblies were performed and verified by the COTMS test facility. These measurements were immediately compared against specified limits, thereby providing instant feedback as to the acceptability of the unit tested. This instant feedback feature was important since it enabled engineering to investigate the cause of unusual test results while a unit was still in the original "fixtured" condition. Data were analyzed for trends so that impending out-of-limit conditions could be detected and compensating adjustments made before serious problems developed. Statistical data were also used as a basis for analyzing out-of-limit conditions and determining proper resolution.

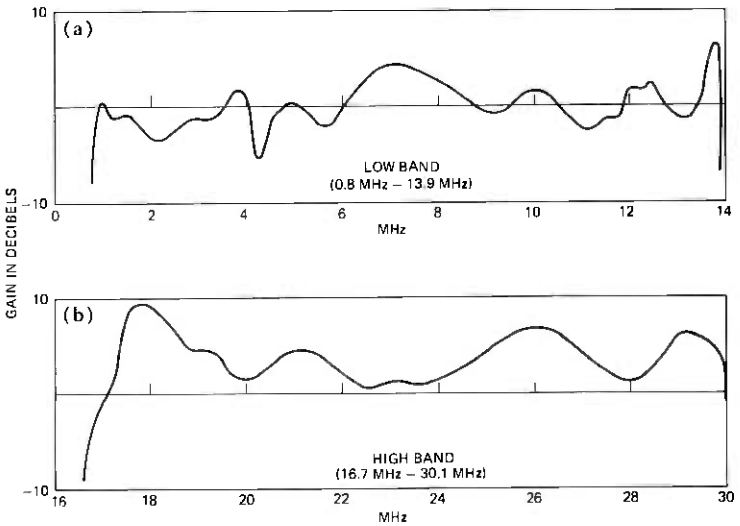


Fig. 21—TAT-6 undersea system misalignment.

Although the aging and testing program is time-consuming, it provides maximum assurance that system components will operate for their intended life.

Equally important with aging and test programs are the procedures which ensure that physical design requirements have been met and that accurate data have been transmitted. It is also necessary to insure that all operations, inspections, and deviations have been properly recorded. For these purposes, multiple inspection procedures are used. First, all operations, inspections, and tests are documented and verified by the signature of the person performing the operation. Second, product examiners verify that prior operations have been performed satisfactorily through entries on data sheets. Third, before any apparatus item (component or subassembly) can be shipped, inspection must verify that all operations have been properly performed.

All data pertinent to the assembly and test of a repeater likewise are compiled in a data book prepared for that specific repeater. This data book is verified for completeness and conformance by a Western Electric Auditor of Manufacturing Practices (AMP). The pertinent data (primarily electrical) are also checked by a resident Bell Laboratories engineer.

As an additional check of quality, the AMPS perform an independent random check of product previously accepted by product examiners.

The Quality Assurance Organization, independent of Western Electric product engineering, is responsible for preparing and performing quality surveys in collaboration with their Bell Laboratories counterparts. One such Bell Laboratories QA engineer is resident at Clark. QA surveys cover detailed investigations of compliance with specified procedures, methods, and drawings in all manufacturing areas.

VII. TAT-6 EXPERIENCE

7.1 Undersea system misalignment

The net misalignment of the TAT-6 undersea system is shown in Fig. 21. The low and high bands contain 53 and 54 supergroups, respectively, including the frequency bands occupied by the order wire and supervisory tones. As indicated in a companion article,[2] the repeater level variation at some frequencies exceeds the end-to-end misalignment. The unpredicted repeater gain deviations arose largely from reflections between the repeater and the cable terminations. The design of the OBE mop-up networks was based upon transmission information obtained from measurements and computer simulation of prototype cable terminations. Evidently, the measured samples of terminations were not representative. With the completion of the first deep sea lay, it was obvious that a special OBE mop-up design would be required following each lay to maintain the misalignment within acceptable limits. Manufacture

of the first such special OBE was completed within a period of two weeks following availability of new mop-up network designs. The OBE was installed near the beginning of the second lay.

7.2 Noise performance

The measured thermal and second-order noise performance of TAT-6 was both satisfactory and consistent with predictions. Third-order intermodulation noise in the top part of the high band, however, exceeded our predictions by as much as 18 dB. The signal-to-noise ratios obtained by noise loading tests during commissioning are shown in Fig. 22. To minimize the total noise at high frequencies, repeater levels were lowered roughly 6 dB. Both the measured and the predicted 29.182-MHz signal-to-noise ratios are indicated in Fig. 23. Third-order intermodulation distortion normally increases 3 dB per 1-dB increase in signal level. On TAT-6, we observed a ratio that varied between 3:1 at low signal levels to 1.8:1 at high signal levels. Narrow-band noise loading tests (2 to 10 supergroups) indicated that most of the third-order distortion was generated by signals close to the frequency of the measured channel.

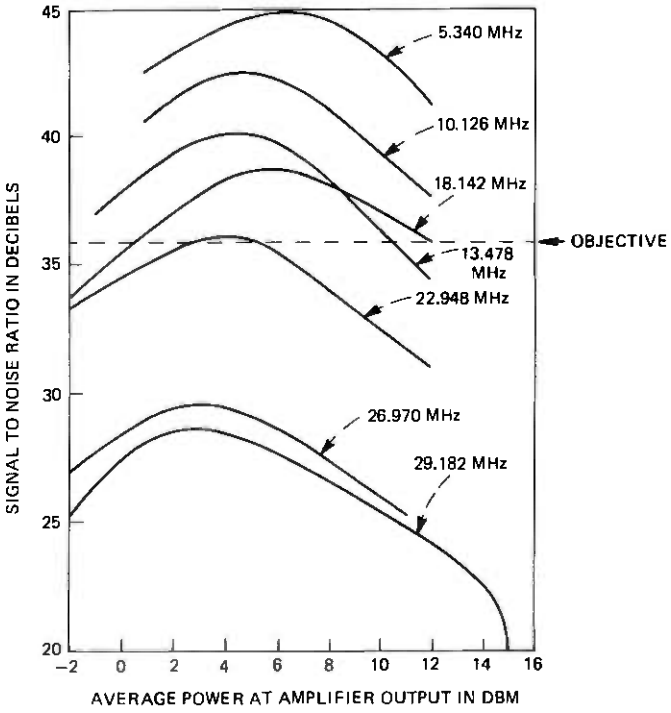


Fig. 22—TAT-6 measured signal-to-noise ratio.

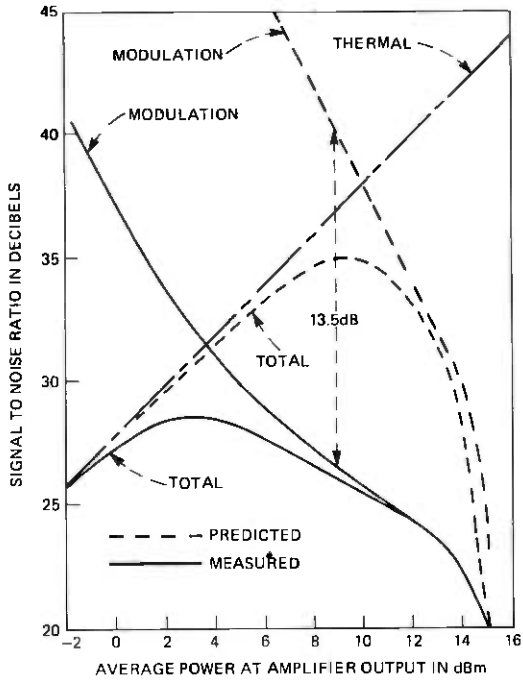


Fig. 23—TAT-6 measured and predicted 29.182-MHz signal-to-noise ratio.

7.2.1 Characterization of M_{3E}

Calculations of system intermodulation noise were based upon average repeater levels, repeater delay characteristics, and repeater modulation coefficients, each being a function of frequency. The modulation coefficients M_{2E} and M_{3E} were characterized in the same manner that had been successfully used in past systems.

A repeater is loaded with three or more fixed fundamental tones of frequencies A, B, and C. Various product frequencies of the type $A \pm B$ and $A \pm B \pm C$ are measured. In the simplest case when the amplifier output stage is the dominant source of distortion, the coefficients M_{2E} and M_{3E} depend only on the product frequency and not on the specific fundamental frequencies. If there is significant dependence on fundamental frequencies, one must consider the effects of signal shaping and delay distortion in choosing the single most appropriate value of a particular modulation coefficient for use in calculating system noise.

In developing the SG repeater, the following fundamental frequencies were used to evaluate repeater intermodulation: 5.8, 14.5, 15.2, 16.6, 21.7, and 27.5 MHz. Nine different third-order and 17 different second-order product frequencies were measured in characterizing the repeater distortion coefficients. In addition, the distortion was measured at two different power levels, +5 and +12 dBm per tone.

Discovery of the excessive third-order intermodulation noise on TAT-6 triggered an intensive search for the source. Instead of measuring repeater distortion with fixed fundamental frequencies, fixed pairs of bandstop-bandpass product-frequency filters and continuously variable fundamental frequencies were used. A contour plot of the measured 3-tone ($A + B - C$) distortion at 26.6 MHz and 2.5°C is shown in Fig. 24 with A and B as the independent variables. (Frequency C depends on A, B, and the product frequency.) Frequency triplets that produce equal distortion are connected with "isomodulation contours." The distortion increases with the frequency of each fundamental and the product. Also the distortion tends to peak for products whose fundamentals are close to the product frequency. The latter are the same products which tend to add in phase from repeater to repeater and thus dominate on long systems. These "significant" distortion products increase in power by about 6 dB between room and sea-bottom temperatures, while those which were measured during development and manufacture decrease about 1 dB over the same temperature range.

7.2.2 Source of excess third-order distortion

Because of the frequency, power level, and temperature dependence of repeater third-order distortion, the initial characterization was inadequate and misleading. Improved measurement techniques indicated that the two surge-protection diodes in the output network increased the repeater distortion by about 12 dB. Distortion caused by these diodes cannot be explained in terms of conventional behavior involving nonlinear resistance and capacitance. The distortion can, however, be explained semi-quantitatively in terms of nonlinear MOS (metal-oxide-semiconductor) capacitance and conductance effects. In addition, tests have shown that when the reverse-biased output network diodes are replaced with reverse-biased MOS capacitors, the resulting distortion contours have characteristics similar to those of Fig. 24. The distortion, while of third-order type, is actually produced by second-order interaction; second-order products are generated and modulated again with the fundamental signals. At the time of this writing, it is not clear in detail how the MOS-type nonlinearity is produced in the surge-protection diodes. Figure 25a is a block diagram of a nonlinear network whose calculated isomodulation contours have the same characteristics, Fig. 25b, as those measured on an SG repeater.

7.2.3 Future noise improvement on TAT-6

Following the successful analytical modeling of the repeater third-order distortion, an experimental post-distorter network was designed, built, and tested. The distorter, consisting of diodes and linear networks, achieved 15 to 20 dB of third-order distortion reduction on a single repeater over a 10-dB dynamic range.

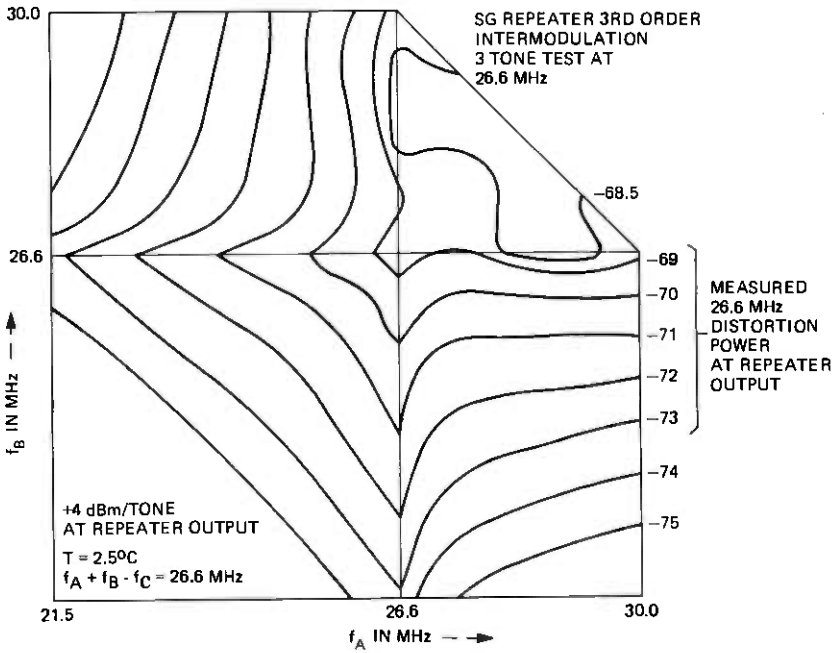


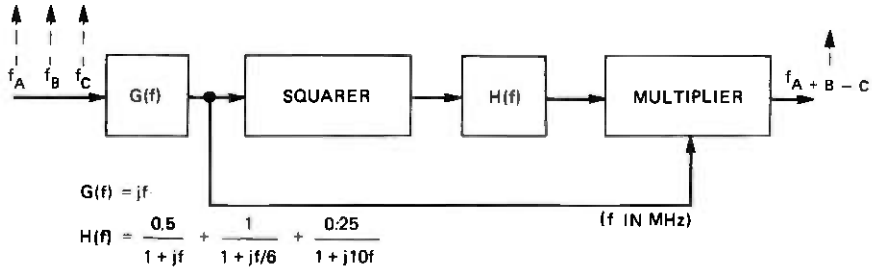
Fig. 24—Isomodulation contours of SG repeater with surge protection diodes.

It was further successfully tested as a possible noise canceler on TAT-6 by noise-loading single supergroups in the portion of the high band where delay distortion is relatively small.

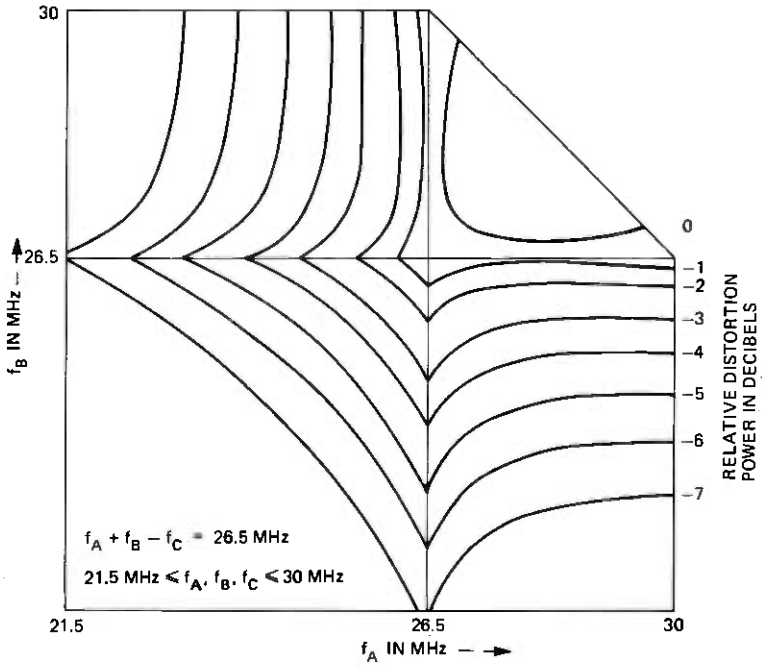
A post-distorter is presently being constructed for the three top hypergroups of TAT-6. It consists of 34 tandem distorter networks connected by loss equalizers, phase equalizers, and amplifiers. Each distorter network will cancel the distortion produced by one block of 20 repeaters. A 13-dB suppression of third-order distortion is expected, which should reduce total noise by the 6 dB required to meet the original system design objectives. The loss equalizers in the post-distorter are adjustable to track the system loss change with time.

7.2.4 Noise Improvement on future SG systems

Prior to completion of TAT-6, the design of the 82-type transistor had been modified to improve its ability to withstand surges. The new "double epitaxial" design features a two-step collector impurity profile which increases the collector-emitter breakdown voltage while maintaining exceptional linearity. This improvement allowed removal of the two surge-protection diodes in the output network and one gas tube in each GSF. Since the semiconductor diodes contributed nearly all the excess third-order intermodulation noise, most of the necessary im-



(a)



(b)

Fig. 25—(a) A fundamental frequency dependent third-order distorter. (b) Computed 26.5-MHz response of distorter.

provement in repeater linearity was achieved before the problem was discovered.

Further investigation of transistor third-order distortion revealed that it had two characteristics of the diode distortion, unusual power level and fundamental-frequency dependence. It was discovered that these characteristics could be essentially eliminated by changing the crystal orientation of the transistor from the ⟨111⟩ axis to the ⟨100⟩ axis. It

appears that the MOS capacitance associated with metallization and the oxide passivation layer is the source of the unusual distortion. Third-order distortion measurements ($A + B - C$ type products) on samples of MOS capacitors have shown that capacitors built on a ⟨111⟩ crystal orientation generate 13- to 17-dB more distortion than those built on ⟨100⟩ crystals. This difference in distortion is probably proportional to the difference in interface energy state densities for the two orientations.[10] Semiconductor distortion studies are being continued in an effort to more fully understand and characterize the physical mechanisms which produced this unforeseen behavior. The results of these studies will be published when they have been completed.

REFERENCES

1. "SF Submarine Cable System," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 601–798.
2. S. T. Brewer, R. L. Easton, H. Soulier, and S. A. Taylor, "SG Undersea Cable System: Requirements and Performance," B.S.T.J., this issue, pp. 2319–2354.
3. G. E. Morse, S. Ayers, R. F. Gleason, and J. R. Stauffer, "SG Undersea Cable System: Cable and Coupling Design," B.S.T.J., this issue, pp. 2435–2469.
4. C. D. Anderson, "Overload Stability Problem in Submarine Cable Systems," B.S.T.J., *48*, No. 6 (July–August 1969), pp. 1853–1864.
5. M. Brouant, C. Chalhoub, P. Delage, D. N. Harper, H. Soulier, and R. L. Lynch, "SG Undersea Cable System: Terminal Transmission Equipment," B.S.T.J., this issue, pp. 2471–2496.
6. W. M. Fox, W. H. Yocum, P. R. Munk, and E. L. Sartori, "SG Undersea Cable System: "Semiconductor Devices and Passive Components," B.S.T.J., this issue, pp. 2405–2434.
7. J. V. Milos and P. A. Yeisley, "Manufacturing Aluminum Castings and Extrusions for use in SG Submarine Cable Repeaters," Western Electric Engineer, July 1975.
8. A. T. Chapman et al., "Manufacture of Submarine Cable Repeaters and Ocean Block Equalizers," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 663–681.
9. W. J. Geldart, et al., "A 50 Hz–250 mHz Computer-Operated Transmission Measuring Set," B.S.T.J., *48*, No. 5 (May–June 1969), pp. 1339–1381.
10. P. V. Gray and D. M. Brown, "Density of SiO_2-Si Interface States," Appl. Phys. Lett., 8 (1966), pp. 31–33.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Semiconductor Devices and Passive Components

By W. M. FOX, W. H. YOCOM, P. R. MUNK, and E. F. SARTORI

(Manuscript received September 28, 1977)

In this paper we describe the active devices and passive components used in the SG undersea cable system for the TAT-6 link between Green Hill, Rhode Island, and St. Hilaire de Riez, France. We explain reasons for component choice and present information about design, performance, screening, and reliability.

I. INTRODUCTION

Traditionally, in high-reliability communication systems such as undersea telephone cables, well-established component and device designs and manufacturing processes are used. But with the development of the SG undersea cable system, this conservative approach has been challenged by the need for devices of higher performance. This requirement resulted, in some cases, in sophisticated device designs that border on "state-of-the-art" concepts. For example, the increase in channel capacity or bandwidth requires high-frequency transistors incorporating design and processing features that are relatively new. Nevertheless, new features were carefully evaluated so that they would meet the system reliability requirements. In general, reliability is the governing criterion for the selection of components and devices for the undersea cable system. This concept and how it applies to passive components and active devices are covered in the following subsections: II, transistors; III, diodes; IV, gas tubes; V, resistors; VI, capacitors; and VII, inductors and transformers. Each subsection is complete for each component or device and covers: (*i*) choice and requirements, (*ii*) design and construction, (*iii*) characterization and screening, and (*iv*) aging.

II. TRANSISTORS

2.1 Requirements

Although there are seven transistor codes, six in each repeater, all are the same basic design. This, of course, is a very desirable arrangement for simplicity and economy of manufacture. The basic development code is a passivated npn silicon planar transistor and is divided into: (*i*) the 82A for the input stage, selected chiefly for low noise; (*ii*) the 82B and 82C as intermediate-stage devices selected for high gain and low capacitance, respectively; (*iii*) the 82D and 82E as driver and output transistors, which are optimized for low second- and low third-order modulation coefficients, respectively; and (*iv*) the 82F and 82G for low- and high-band oscillators. Table I shows assignment of the transistor codes to the various functions and their most critical characteristics.

2.2 Design

The design of the transistor draws on the epitaxial, planar, and noble contact metallurgy technology commonly used in contemporary discrete and integrated devices.

2.2.1 Thermal considerations

Being a power high-frequency transistor, the transistor's active element is in essence an array of conventional, small, high-frequency devices electrically connected in parallel on the same silicon chip. Packing a large number of these small devices close together (to keep capacitance low) naturally creates a thermal problem. And where both reliability and low distortion performance are important, this problem becomes critical. For a long service life, the transistor must operate at as low a temperature as possible, and at the same time all parts of the active element must be

Table I — Location and characteristics of transistors

Transistor Code	Circuit Designation	Function	Operating Voltage	Stage Placement Criteria (Median)
82A	Q201	Input stage	V_{CE} = 9.2V I_C = 50mA	NF = 1.87 a_o = 0.992
82B	Q202	Intermediate stage	V_{CE} = 10V I_C = 150mA	a_o = 0.0990
82C	Q301	Intermediate stage	V_{CE} = 9.2V I_C = 50mA	C_F = 3.20
82D	Q302	Driver stage	V_{CE} = 10V I_C = 150mA	M_{2E} = −54.2 a_o = 0.986
82E	Q303	Output stage	V_{CE} = 10V I_C = 150mA	M_{3E} = −93.1 a_o = 0.988
82F	Q601	Low band oscillator	V_{CE} = 10V I_C = 10mA	$0.978 < a_o < 0.98$
82G	Q501	High band oscillator	V_{CE} = 10V I_C = 10mA	

uniform in temperature so that low levels of distortion are achieved. This is accomplished by grouping the emitters into four quadrants of nine emitters each, with appropriate spacing between (see Fig. 1). A thermal profile of the 82-type transistor indicates the temperature uniformity across the active element is within a range of 3°C.

2.2.2 Impurity distribution

A unique feature of the transistor manufacture is that each individual silicon wafer of starting material is carefully profiled for the impurity distribution of its epitaxial layer. Poon[1] showed that intermodulation distortion could be reduced by properly shaping the transistor doping profile in the collector region. In practice, optimum low-distortion performance is obtained by grading the epitaxial layer so the concentration of impurity atoms gradually increases in the epitaxial layer in a direction approaching the substrate (see Fig. 2). In processing, the epitaxial profile seen in Fig. 2 is readily achieved by a combination of processing temperature and an appropriately shaped profile at the start. During sub-

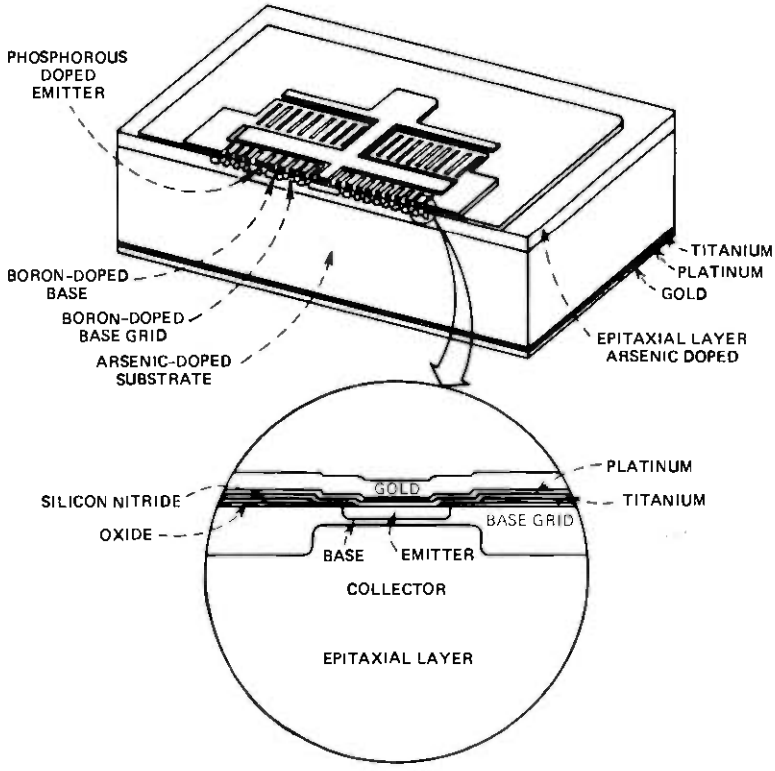


Fig. 1—Cross section of transistor chip. Insert shows enlarged view of a single emitter.

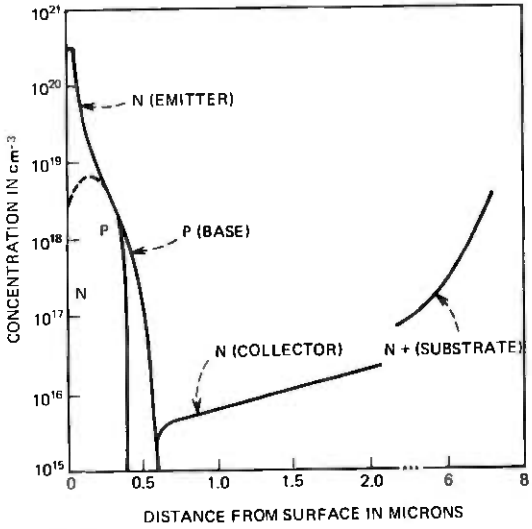


Fig. 2—Transistor impurity profile, with concentration as a function of depth from the surface.

sequent high temperature operations, a redistribution of impurity atoms occurs in just the right amount to produce the desired collector impurity distribution. Intermodulation coefficients represented by M_{3E} below −100 dBm have been realized in this manner. However, early transistor models which were optimized for low distortion by using epitaxial layers 5 μm thick were found to be vulnerable to transient surges. Therefore, to improve transistor surge capability, a 7-μm thickness of the epitaxial layer was found to be necessary. An inherent problem associated with silicon microwave transistors is that the device may go into secondary breakdown in some modes of operation. Use of the thicker epitaxial layer tends to shift the onset of secondary breakdown to higher voltages. This compromise in epitaxial layer thickness resulted in a slight penalty in the third-order modulation coefficient.

Emitter and base impurity distributions shown in Fig. 2 are achieved by straightforward diffusion processes adjusted to meet the transistor's electrical requirements.

Finally, the silicon substrate through which the collector current flows is made thin and is heavily doped with arsenic. Consequently, low collector series resistance, low collector contact resistance, and low thermal resistance are achieved.

2.2.3 Active element

The active element (Fig. 1) consists of 36 emitter stripes. Each emitter is 2.5 μm wide and 65 μm long and is separated by 12.5 μm. The number and size of the emitters were chosen to provide the necessary power

dissipation, low intermodulation distortion, and a flat gain-current characteristic over the operating current range.

A band of deep diffused boron surrounds each emitter and forms a p^+ grid-like structure over the entire active element. This diffusion serves as both a gettering diffusion and as a means of lowering the base resistance between the emitter stripes and at the base contact itself. A combination of a relatively large base contact area and a high surface concentration provides a low base contact resistance (r_b''), necessary for low noise performance. Similarly, a combination of very narrow emitter stripes and stripe spacing and the boron grid provide a low base resistance (r_b'). The active base region is also boron-diffused and is tailored to provide a gain range of $45 < h_{FE} < 160$, an emitter breakdown voltage greater than 5 volts, and a base layer width of 0.23 μm. In summary, the above provides the necessary vertical and horizontal geometry for a gain bandwidth cutoff frequency (f_T) of 2.7 GHz.

The contact metallization used is the Bell System standard titanium, platinum, and gold system. Emitter and base areas are contacted by overlays which lead to wire-bonding pads outside the active element. Collector contact is made through a metallized layer on the back side of the chip.

The titanium-platinum-gold metallization makes use of a new sputter-etch process. This process was first devised by Herb and Labuda[2] and later applied to the transistor manufacture as the first production device sometallized. Even though this appears inconsistent with the policy of using only well-established processes for undersea cable devices, early results were so encouraging it was decided to proceed with the process. In practice, this process is remarkably uniform in quality and manufacturing yields when compared to the older plating process.

Briefly, the process consists of depositing uniform layers of Ti, Pt, and Au; masking the contact areas with nickel plate; and then rf sputter-etching away those areas not wanted. The mechanism depends on the relative etch rates of the various materials used.

2.2.4 Device stabilization

Traces of metallic impurities are always present in semiconductor starting materials and sometimes are introduced in subsequent processing steps. They are known to degrade electrical properties of semiconductor junctions and upon their removal the junction characteristics are restored. Any scheme which removes or immobilizes these impurities is considered a gettering process.

Two such processes are successfully used in the transistor processing. The first makes use of a high concentration, deep, and highly stressed diffusion region in the transistor's active element. As mentioned earlier, a deep boron-diffused grid-like structure is formed in the base area to

create a p^+ region in the proximity of each emitter stripe. A diffusion such as this is known to have two important properties: (*i*) the enhanced solubility of metallic impurities in a p type diffusion, and (*ii*) the high density of dislocations associated with this type of boron diffusion.[3] Both these effects, as well as their close proximity to the emitters, result in a very effective gettering mechanism. Junction leakage currents have been reduced several orders of magnitude to a few hundredths of a nanoampere. Furthermore, the transistor exhibits a very flat gain-current characteristic at low currents (Fig. 3) which can be attributed to the reduction in the number of recombination sites in the gettered base region.

The second gettering process is used on both transistors and diodes, which contain silicon dioxide and nitride layers. Because stability of surfaces is so important in establishing reliable device operation, these layers are, in addition to other functions, used where appropriate in passivation techniques for controlling semiconductor surface potentials. However, silicon dioxide is known to be porous to alkali metals, particularly sodium ions. These ions can migrate to the silicon surface and alter the electrical characteristics of the device. Every effort possible is exerted in manufacture to eliminate these contaminants but, as a precautionary step, the device is protected by a silicon nitride coat which is placed over the oxide as a permanent seal. In addition, immediately before the nitride coat is applied, the underlying layers of oxides are given a gettering treatment to rid them of possible ionic contamination.

The final step in device stabilization is a high temperature anneal of each completed transistor wafer to relieve any stresses introduced during the metallization operations.

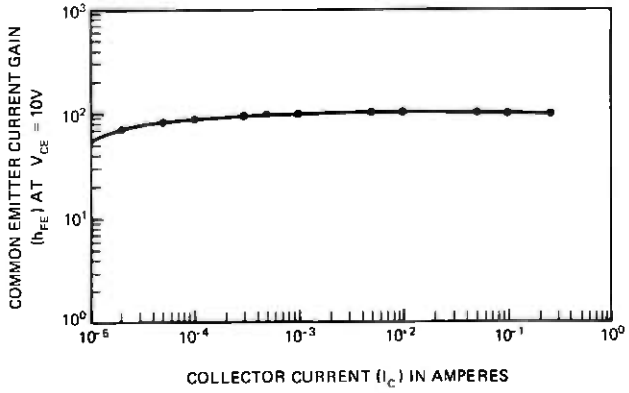


Fig. 3—Common emitter current gain characteristics, with transistor gain as a function of collector current.

2.2.5 Encapsulation

Because operating temperature is so important a factor in establishing device reliability and performance, a significant amount of effort is invested in the design of low thermal impedance packages. Furthermore, these packages are required to remain vacuum tight over substantial excursions of temperature and time. They must consist of materials that do not introduce excessive stress on the fragile semiconductor chip.

The structure chosen for the transistors and diodes is a hermetically sealed encapsulation which provides alumina ceramic isolation for low parasitic capacitance and inductance, stud mounting, and low thermal resistance. The transistor combination of the active element and package has a thermal derating factor of less than 24°C per watt. This is consistent with the thermal objectives of the system, whose goals are based on a maximum operating junction temperature of 65°C. Although the package was developed specifically for the transistor, it is also used for diodes to simplify manufacture.

The encapsulation features are also based on providing the transistor or diode with the best possible environment during its lifetime, even with the device active element itself equipped with a passivated design. Charge accumulation is enhanced by the presence of water vapor and electrolysis contributes to degradation of internal structural members; therefore, all semiconductor devices are hermetically sealed. Also high temperature baking of piece parts and cold-weld closures minimize the chance of residual gas evolution inside the enclosure after the seal is made.

Figure 4 is an illustration of the package. This encapsulation consists of two subassemblies (a header assembly and a stud can assembly). These assemblies are joined together by a cold-welding process after the semiconductor chip is bonded to the header and the connections to the feed-through leads have been made.

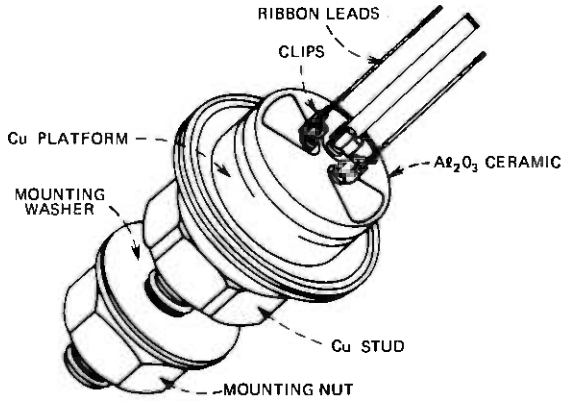


Fig. 4—Semiconductor device encapsulation. Diodes and transistors are hermetically sealed in this package.

The header assembly consists of a number of separate piece parts brazed together. The welding flange of the header assembly is fabricated from oxygen-free high-conductivity copper using cold forming operations. The molybdenum platform, upon which the semiconductor chip is mounted, is punched from special material which will not delaminate during fabrication. Ceramic insulators consist of an aluminum oxide disc metallized with a mixture of molybdenum and manganese, over which a thin layer of copper is plated. Feed-through terminals are made of Kovar lead wire and Kovar washers. All parts are brazed together simultaneously using special brazing alloys, and the final assembly is then finished with layers of nickel and gold plate.

Can assemblies are also made of separate piece parts. The stud is cold formed of an alloy of copper selected to meet the torque requirements necessary to ensure good thermal contact to the repeater housing. The can is formed from oxygen-free copper and is ultrasonically welded to the stud. The assembly is finished with a nickel and a gold plate.

To ensure reliability and integrity for the package, a number of tests and examinations are performed during assembly and upon the completed package. Some are destructive sampling tests and designed to determine failure limits. An example is hydrostatic pressure testing until rupture occurs. Another test is a lead push test, in which force is applied axially to the end of the lead of the header until it is pushed through the ceramic.

As part of the final screening of devices, the package is again scrutinized for mechanical defects by exposing the devices to acceleration tests, shock tests, and particle tests as well as detailed visual examination.

2.3 Characterization

Electrical characterization of underseas cable semiconductor devices is, of necessity, a detailed and exacting process. The high system performance objectives and the rigorous reliability goals require testing procedures not usually applied to ordinary devices. For example, the reliability objectives for the transistor demand that a change in common emitter current gain (h_{FE}) be no more than 2.5 percent over the service life of the transistor. This is equivalent to a change in small signal current gain (a_o) of less than 0.0005 over a 20-year period. Therefore, these devices must not only be designed to meet such stability requirements, but must be characterized to such a degree that it can be determined that such minute changes can be detected. Toward this end a battery of tests is performed, some of which are repeated many times during the screening program.

Characterization is in two parts: dc tests primarily determine the reliability qualities of the device, RF tests show mostly the circuit perfor-

mance capabilities. In addition to conventional dc tests, several of which are performed at multiple bias points, special tests such as pulse tests and surge capability tests are conducted. Transistor current gain measurements are made at several levels of collector current on each individual device.

Two tests are made at use conditions (50 and 150 mA), one test is made at a level very sensitive to transistor changes (5 mA) to monitor aging and selection, and another is made at high current (250 mA), which is a good indicator of whether all emitters are properly functioning. Each candidate transistor must be able to meet the gain-current characteristic shown in Fig. 3.

Another example of special testing is the test placed on the transistor's emitter junction. The quality of the emitter junction is particularly critical because of its exposure in service to noise overload which may momentarily push the emitter into reverse conduction, its exposure to residual surges which act similarly on the device and stem from cable fault, and its exposure to inverse bias which may occur temporarily during fault monitoring. Therefore, the shape of the emitter reverse voltage-current characteristic is critical and must be determined for each device. Table II shows typical dc characteristics of the transistor.

RF testing consists of four noise figure measurements, ten intermodulation coefficient measurements, and "on-line" transistor modeling of each device. On-line modeling is transmission characterization performed on a computer-operated test system which produces a circuit model for each transistor. The S parameters of each transistor are measured at the two use conditions and at 15 frequencies from 0.1 MHz to 436.8 MHz. These data are automatically fed into a computer and by means of a circuit-oriented model (Fig. 5) reduced to six basic parameters. The model is constructed so that it fixes the values of those transistor elements which are not likely to vary widely in manufacture. Fixed values shown in Fig. 5 are calculated from the geometry, impurity concentration, etc., for a typical transistor. The variable elements are (*i*) small signal current gain a_o, (*ii*) capacitance associated with a_o cut-off frequency *Cep*, (*iii*) collector/base capacitance less inner capacitance C_F, (*iv*) total emitter resistance R_e, (*v*) total emitter inductance, L_E, and

Table II — Some typical DC characteristics

I_{EBO} (V_{EB} = 2V)	<0.35nA
I_{EBO} (V_{EB} = 3V)	<10nA
I_{EBO} (V_{EB} = 4V)	<235nA
I_{CBO} (V_{CB} = 10V)	<0.07nA
I_{CEO} (V_{CE} = 10V)	<0.09nA
BV_{CBO} (I_C = 100 μA)	>60V
BV_{EBO} (I_E = 100 μA)	>5V
V_{CB}, V_{EB} (I_E = −100mA)	<0.91V
h_{FE} (I_C = 150mA, V_{CE} = 10V)	$50 < h_{FE} < 160$

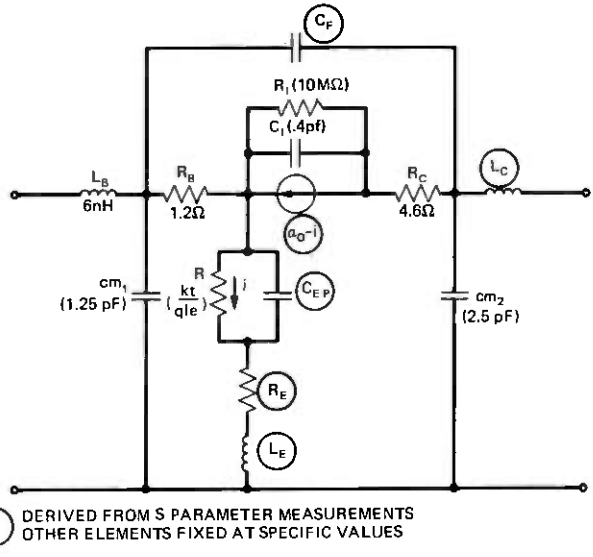


Fig. 5—Circuit-oriented model used to establish transistors' behavior in a repeater set. Circled elements are varied to match data. Other elements are calculated from geometry and impurity distribution.

(*vi*) total collector inductance L_c. Distributions of all RF parameters are compiled, and each transistor is assessed with respect to its relative position in the total population. It is then assigned to a specific repeater set where it has a unique and optimized function. Table III illustrates how tightly the model parameters are controlled for TAT-6.

2.4 Screening and aging

Although, in essence, mechanical screening begins at the very onset of manufacture, a series of mechanical evaluations is performed on all completed devices to confirm the integrity of the device's structure. Electrical screening begins with environmental stresses such as channel aging, overstress power aging, and high temperature stress, and ends with long-term aging.

Channel aging is done on a 100-percent basis and serves to eliminate inferior devices that, despite all the precautions taken in manufacture, may still have some surface charge accumulation. This is accomplished by holding each device at elevated temperature with its junction at reverse bias for a period of one week.

Accelerated power aging is a shakedown test. Each transistor is power-overstressed at a level three times its operating level (4.5 W) for a period of one week. This test is well within the capability of the transistor structure and enables detection of devices with marginal quality and incipient failure modes.

Table III — 82 type transistor—AC model parameter data

	82A*		82B†	
	-2σ	$+2\sigma$	-2σ	$+2\sigma$
a_o	0.9894	0.9932	0.9842	0.9926
$R_e(\Omega)$	0.1530	0.1962	0.1729	0.2156
C_F(pF)	3.035	3.863	3.011	3.679
C_{EP}(pF)	111.2	151.3	292.3	462.0
L_E(nH)	6.535	6.833	6.565	6.891
L_C(nH)	6.024	6.689	6.314	6.892
	82C*		82D†	
	-2σ	$+2\sigma$	-2σ	$+2\sigma$
a_o	0.9818	0.9926	0.9808	0.9900
$R_e(\Omega)$	0.1455	0.1956	0.1712	0.2176
C_F(pF)	3.131	3.452	3.015	3.529
C_{EP}(pF)	120.2	202.3	340.3	486.6
L_E(nH)	6.543	6.868	6.554	6.892
L_C(nH)	6.010	6.674	6.348	6.848
		82E†		
		-2σ		$+2\sigma$
a_o		0.9814		0.9925
$R_e(\Omega)$		0.1669		0.2155
C_F(pF)		3.085		3.745
C_{EP}(pF)		317.8		466.4
L_E(nH)		6.566		6.889
L_C(nH)		6.318		6.814

* (V_{CE} = 9.2V, I_C = 50 mA).
† (V_{CE} = 10V, I_C = 150mA).

The high-temperature stress technique is a sampling test based on the observation that semiconductor device life is related to the magnitude of the stressing temperatures. A sample of every lot of transistors is stressed at three different time intervals and three junction temperatures. Distributions of failure based on the reliability goals (Δh_{FE} < 2.5% in 20 years) are then made for each time and temperature combination. From these data, the commonly known Arrhenius Failure Distribution (Fig. 6) is made which provides a convenient method of determining the failure rate for any time-temperature combination. Inspection of these acceleration life curves reveals that the sigma of the failure distribution remains constant at various stress levels; therefore, we may extrapolate the curves in Fig. 6 to long-time, low-stress-level conditions. Note in Fig. 6 that the intersection of the 20-year objective with a 65°C use condition (these being the reliability objectives) is coincident with a value of sigma of the transistor failure distribution, which is extremely small. This establishes the confidence level or percentage of failure to be expected. With six transistors in each repeater of the 700-repeater system, these extrapolations indicate that in 20 years one would anticipate less than one transistor failure in the entire system.

Finally, the devices are placed on six-month long-term age at maxi-

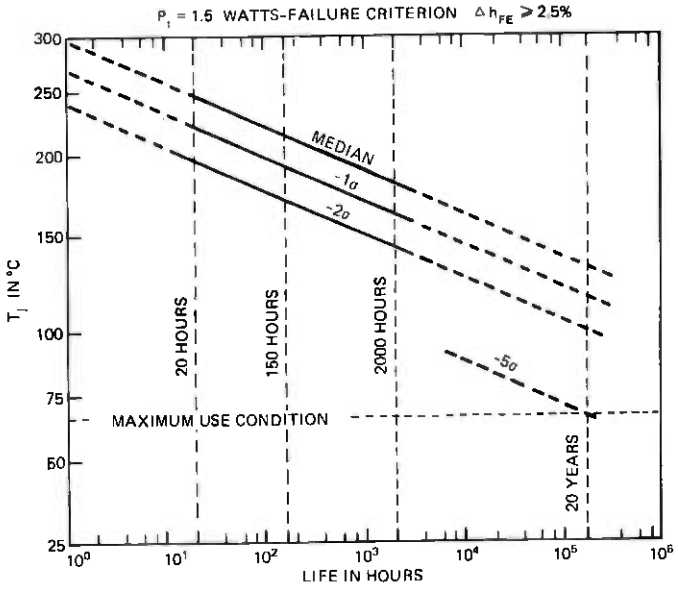


Fig. 6—Pattern of accelerated aging for transistors.

mum operating conditions. During this time, devices are tested at regular intervals, the results of which are observed for trends. These data are statistically extrapolated over a 20-year period, with the objective of assuring a 20-year life in the repeater. Figure 7 is a graph of how a group of SG transistors age. The graph contains a plot of the median as well as the standard deviation in the population. After initial stabilization, no further trends are evident.

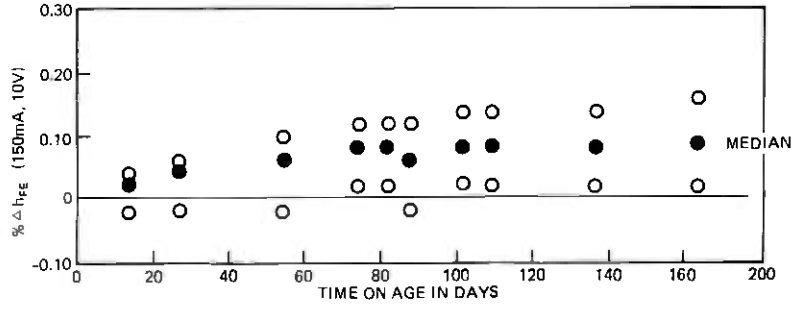


Fig. 7—Stability of a group of transistors during long-term aging.

2.5 Transistor improvements for future systems

Upon completion of transistor manufacture for TAT-6, several important design improvements were made to the 82-type transistor. The transistor's surge capability was significantly upgraded by fabricating the device in double epitaxial silicon, where the normally lightly doped thin epitaxial layer is separated from the heavily doped substrate by a thick epitaxial layer of intermediate doping level. Use of the double epitaxial layers shifts the onset of secondary breakdown to higher power levels while still maintaining good linearity and high frequency performance. A compromise in epitaxial thickness as reported in Section 2.2.2 is no longer necessary. Furthermore, unexpected high distortion near the upper end of the transmission band was markedly reduced by changing the orientation of the starting silicon material from ⟨111⟩ to ⟨100⟩. Experimental devices indicated interface surface effects were the cause of this anomalous behavior. Moreover, a recent study by Olson[8] shows surface effects can contribute to intermodulation distortion.

Another important change concerns the semiconductor device encapsulation. Gold-plated piece parts that make up the package have been replaced by unplated nickel parts at a considerable cost savings with no penalties in reliability or integrity of the encapsulation. And finally, additional high-temperature annealing of the assembled transistors prior to final sealing has further enhanced the transistor's stability. The slight change in gain during the initial period of aging under operational conditions (Fig. 7) is no longer evident, so that now the transistor is essentially stable from the very beginning.

III. DIODES

3.1 Requirements

A total of seven diode codes are used to obtain the required protection for each repeater. They consist of three basic types, two of which are similar to those used in the SF system.

The 522 is a silicon planar diode used for signal overload protection and is new for the SG system.

The 523B,C,D,E, and F are silicon mesa-type diodes. Specifically, the 523B acts as an amplitude control for the oscillator, the 523C furnishes input protection for the preamplifier, and the 523E is used for signal path surge protection and is located in the directional filter.

The 524A is also a silicon mesa-type diode and is used as a voltage limiter for power path surge protection in the ground separation filter. They are used in a set of three in parallel, packaged as the 532A. In the event of a cable break or large transient through the repeater, these diodes must regulate the peak voltage across the repeater during the transient. Table IV shows the diode complement.

Table IV—Location and characteristics of diodes

Diode Code	Circuit Designation	Function	Operating Voltage	Critical Characteristic
522A	CR-301	Output transistor signal overload protector	−0.8V	C < 10 pf series resistance 10 to 13 Ω
523B	CR-501	High/low band oscillator amplitude limiters	0V	max. C/pkg. = 20 pf
523C	CR-201	Signal path surge protectors	−0.85 (each chip)	max. C/pkg. = 8 pf chip Δ = <1 pf
523D	CR-401	Signal path surge protectors	−12V	max. C/pkg. = 4.8 to 5.8 pf
523E	CR-101	Signal path surge protectors	−4V (each chip)	max. C/pkg. = 11.6 to 12.6 pf
523F	CR-402	Signal path surge protectors	−12V (one chip)	max. C = 4.8 to 5.8 pf
532A	CR-701	Power path surge protector	−12V	forward characteristic matched to within 0.1V @ 75A/set of 3 V_{BR} = 15V @ 10mA

3.2 Design

New for the SG system, the 522A employs a planar technology similar to that of the transistor, including silicon nitride passivation. The design is tailored to provide protection during signal overload to the emitter-base junction of the output transistor of the power amplifier. Signal overload can result in reversing the bias of the emitter-base junction, causing it to go into breakdown. This is known to degrade stability characteristics of the transistor. Diode characteristics which will provide this protection during signal overload and which would not alter the current during normal signal levels are seen in Fig. 8.

To meet these requirements, a small gold-doped silicon chip was designed. Small area keeps the capacitance low, and gold doping controls the slope of the forward V-I curve after forward conduction begins. The diode junction is formed by boron diffusion and protected by a silicon nitride layer. To achieve the required forward V-I characteristic, a thin layer of gold is evaporated on the back of the wafer and diffused into the silicon. This lowers the lifetime of the minority carriers and increases the series resistance of the diode. Contact metallization is provided to the chip by a nickel-gold overlay.

The chips for the 523 and 524 diode types are similar in design to 468A and 467A used in the SF system[4] and include a silicon dioxide passivation layer over the junction area.

The 522A chip is mounted in the same SG package as the transistor. The 523 type chips are also in the same package but are mounted in series, two to a package. The 524A is mounted in a slightly modified one-lead version of the basic SG package and then mounted on a heat sink as a matched group of three diodes (Fig. 9).

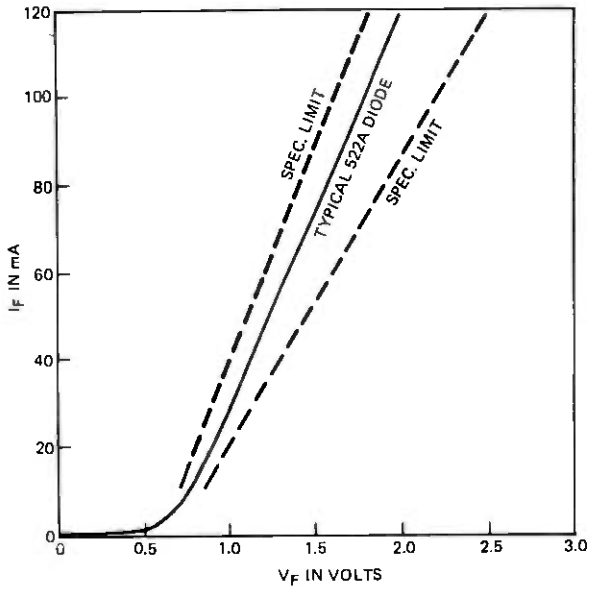


Fig. 8—Forward characteristics of a 522A diode.

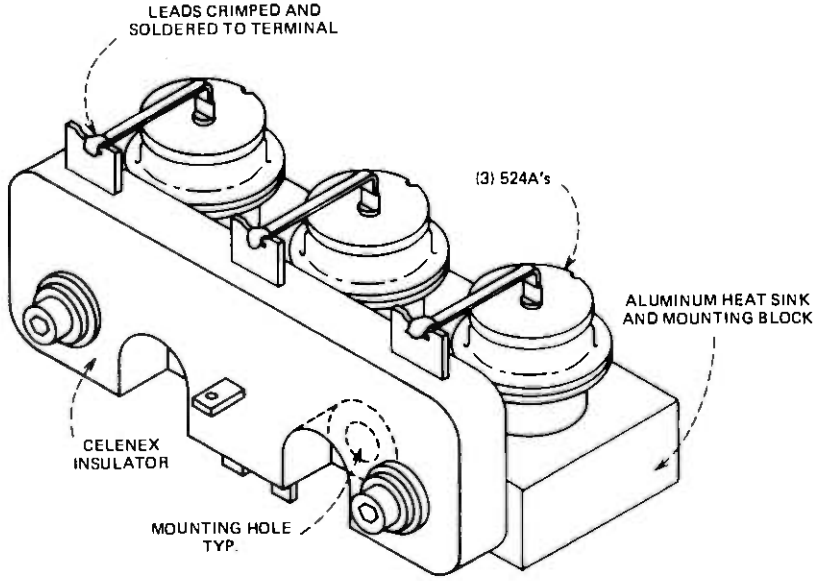


Fig. 9—Matched diodes mounted on a heat sink and connected in parallel.

3.3 Characterization

A battery of tests is performed, some of which are repeated many times during the course of the screening program. In addition to the normal diode tests of saturation current, forward and reverse voltages, surge

tests, etc., specific tests are performed on the various codes. For instance, the diode characteristic curve in Fig. 8 is ensured for each 522A by measuring the forward voltages and currents at a number of operating levels. The 523 family of diodes which contains two chips is characterized for each chip and the combination of the two chips, and then fitted to the specific protective functions (see Table IV).

Finally, the 524A which is the basic building block for the 532A surge protector is characterized by measuring the individual diode breakdown voltages during a 75A surge and then matching them to within 0.1 V for the three diodes set.

3.4 Screening and aging

Screening consists of channel age tests, accelerated age, and operational aging at use conditions. The channel age test consists of holding the diode at reverse bias for 16 hours at 150°C and at 80 percent of the breakdown voltage. The accelerated age test consists of holding the diodes at high temperature for 144 hours. Devices that behave differently from the main populations are discarded, and the remaining diodes are aged for six months at maximum use conditions. Those diodes that exhibit the most stable characteristics are then selected for use in the repeater.

The aging characteristics of SG diodes are similar to those of the SF system, in which many years of accumulated data indicate unprecedented reliability.[5]

3.5 Device manufacture

The manufacture of semiconductor components makes use of the principles devised for the devices used in earlier undersea cable systems. To reduce the possibility of unforeseen sources of device instability, all manufacturing operations are meticulously performed under stringent conditions. All the crucial operations are performed in a clean room environment. Raw materials and piece parts are inspected to assure conformance to the design intent. Thorough records of all inspections are kept, to create an atmosphere of quality consciousness and to provide diagnostic information in the event of deviant behavior of devices or device lots. Control charts are used to monitor the uniformity of the processes, and inspections take place at every step in the manufacture to prevent introduction of incipient failures. Pedigree data on each device is stored so that any particular device is traceable from its material point of origin throughout its history in the manufacturing shop.

In the SG program, many new and sophisticated pieces of apparatus were used to improve precision and control. Furnace operations such as oxidation, diffusion, and annealing are all performed in an automatically controlled furnace complex. This system achieves with fine pre-

cision the processing repeatability and uniformity required for device manufacture. Process variables such as gas flow, temperature, insertion and withdrawal times, speed, and other related parameters no longer involve human variability. Vacuum deposition stations used in metallization operations are similarly automated. Wafer separation is accomplished by laser-beam cutting. Finally, a specially constructed, scanning electron-beam microscope is used to inspect each individual device before it is sealed in its encapsulation.

IV. GAS TUBES

Primary signal path surge protection is accomplished by four WE 469A gas tubes similar to those used in early undersea cable systems.[6]

V. RESISTORS

5.1 Choice

Tantalum nitride film resistors were chosen for the SG system because of their stability, the predictability of their aging characteristics, and their low parasitic effects. In addition, their performance in the Bell System as well as in the SF undersea cable system[7] has been excellent.

5.2 Requirements

Resistors are used in bias networks, current limiting, feedback networks, wave shaping, equalizers, impedance matching, and level adjustment. Tolerances range from ±15 percent for some of the low-valued resistors to ±0.1 percent for others in critical applications.

Two sizes of resistors are required. Type 265 resistors are rated at $\frac{1}{8}$ W but are operated at a maximum of $\frac{1}{16}$ W, at which power the film temperature is such as to guarantee end-of-life tolerances well below those required by the system. Type 266 resistors are used in the repeater output stage and are rated at $\frac{1}{2}$ W but are operated at a maximum of 0.3 W to keep end-of-life tolerances well within system requirements. There are fifty-nine 265-type resistors and fifteen 266-type resistors in the repeater. A typical ocean-block equalizer may contain one hundred nine 265-type resistors, of which twenty-two are used for "mopping up." Resistor patterns range from 0.6 square to 250 squares. Although the SF system used resistors formed on sapphire (the 243-type), considerable experience since has demonstrated that resistor films on bare alumina substrates possess more than sufficient stability for the SG application. In addition, the use of alumina substrates amounts to a considerable cost saving.

5.3 Construction

Resistors range in value from 3 to 10,000 ohms and are fabricated using as-sputtered sheet resistances ranging from 4.5 to 25 ohms/square.

Both 265- and 266-type resistors are made by sputtering the appropriate tantalum nitride film onto 9.52 × 11.43 cm ($3\frac{3}{4}'' \times 4\frac{1}{2}''$) ceramic substrates, followed by evaporated layers of titanium and palladium, over which is plated 3 μm of gold. Resistors are patterned photolithographically and are laser-scribed into "mini" substrates 2.54 cm (1 inch) square. These squares contain ten 265- or four 266-type resistors which are then pre-anodized, heat-treated, and then final-anodized to value. Then the squares are sawed apart into discrete resistors, and leads are attached by thermal compression bonding. A pre-stamped polyvinylidene fluoride sleeve is shrunk over the resistor for protection. The stamped information on the sleeve includes the sputter run number, the substrate number in the run, the "mini" number, and the individual resistor number from the "mini," thus allowing complete traceability for each individual resistor. Complete resistors are shown in Fig. 10.

The resistors are assembled by gold-to-gold thermocompression bonding, thus precluding whiskers and other possible intermetallics.

5.4 Screening and reliability

The resistors are heat-treated in "mini" form for two weeks at 150°C to determine resistor film quality. Then assembly of the resistors is

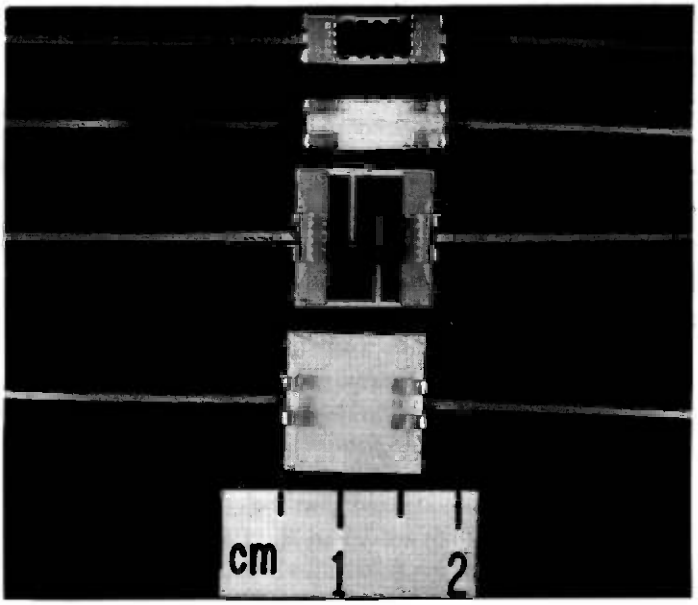


Fig. 10—Thin film resistors.

completed, and they are shipped to the repeater assembly location where each resistor is visually examined, a sample of the product is voltage-surge tested, then all resistors are baked 24 hours at 105°C, followed by five temperature cycles from −40° to +85°C. After that, each resistor sustains a short-time overload of 6.25 times rated power for 5 minutes. This is followed by 1,500 hours at rated power to screen out any resistors with anomalous behavior and, after a final resistance check and noise measurement, the resistors are ready for assembly. Overall yield after screening and testing is 75 percent.

VI. CAPACITORS

6.1 High-voltage paper capacitors

As described elsewhere in this publication, the coaxial cable supplies dc power at constant current to operate each repeater, as well as carrying the composite carrier signals being transmitted in both directions. To separate the dc power from the carrier signals, power separation filters are included on each side of the repeater amplifiers (see Fig. 11). Sufficient dc power is tapped off to power the repeater, and the dc current is returned to the cable. The major source of destructive high frequency feedback around the repeater results from the series inductance in the path between the "sea" ground and the internal repeater ground. It is possible to eliminate this inductance by making the signal path a completely coaxial structure in which the outer conductor comprising a coaxial capacitor is used to "connect" the two grounds (see Fig. 11).

6.1.1 Choice of capacitor materials

The coaxial capacitor must be able to stand the full cable voltage (7 kV) plus any anomalous spikes generated during fault conditions. Consequently, castor-oil impregnated kraft paper capacitors were chosen for this service, as they have been for earlier submarine cable systems, with excellent results.

6.1.2 Requirements

The maximum voltage sustained by these capacitors at the shore ends of the system is 7 kV. Capacitance value is 13 nF ±3 percent. The smallest physical size together with the required coaxial structure and voltage rating dictated the unit shown in Fig. 12. Notice the tube with expansion bellows to allow for the expansion of the castor-oil-impregnant, particularly during the high-temperature processing which the repeater undergoes. Coaxial construction plus near-perfect axial electrical symmetry reduces series inductance to 20 nH.

6.1.3 Construction

As can be seen in Fig. 13, this capacitor is actually two capacitors in

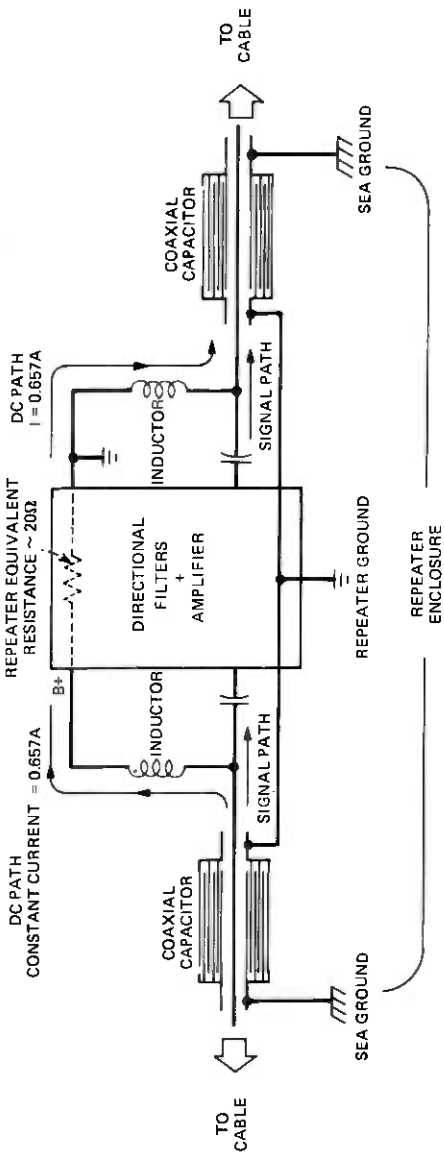


Fig. 11—Diagram showing how coaxial capacitors are used to "connect" the "sea" ground and the repeater ground.

Fig. 12—Completed coaxial capacitor used in ground separation filter. Note expansion bellows.

series. Notice how the conducting foils are interleaved so that there is a "floating" foil not attached to the external connections. The capacitor is wound on a hollow alumina ceramic tube with metal ferrules soldered to it at both ends. The floating foil plus the connected foils are separated by 5 layers of 25-μm (1-mil) paper. Foils are approximately 6-μm ($\frac{1}{4}$-mil) thick aluminum. This combination of papers and foils is wound to the appropriate diameter (and capacitance value) with conductor "flags" attached to the outside foils at positions that are axially symmetrical and 90 degrees apart around each end of the capacitor. This symmetry is necessary to keep the series inductance of the structure as low as possible. The capacitor is completed by enclosing the paper-foil unit in a ceramic housing which is soldered in place. One ferrule is equipped with the expansion bellows plus a filling tube where the oil enters the unit. After a vacuum bake, the capacitor is impregnated with pure castor oil and sealed. All external metal parts are nickel + gold plated for solderability and to eliminate corrosion. Nickel is included as a diffusion barrier.

Other high-voltage capacitors used in the equalizers are constructed in a similar manner, but are not in a coaxial configuration.

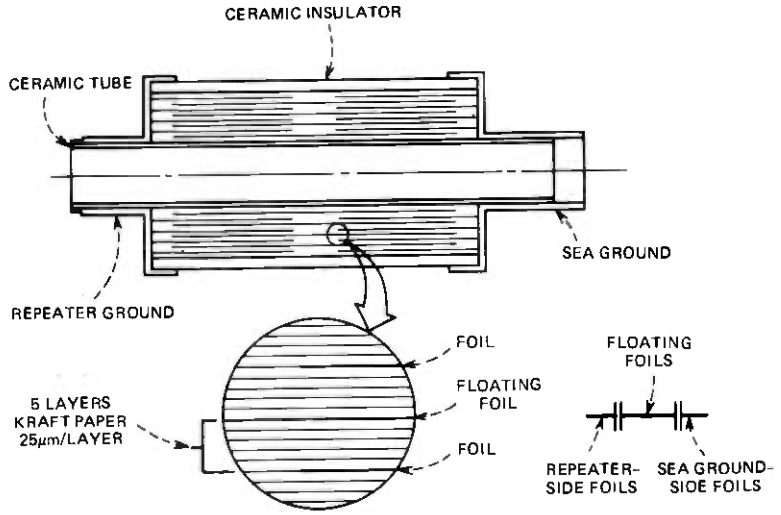


Fig. 13—Diagram showing cross section of coaxial capacitor (expansion bellows not shown).

6.1.4 Qualification and testing

Every roll of capacitor paper purchased is qualified for use by testing capacitors made from the paper. These test capacitors are wound and impregnated with castor oil, tested for 25 days at 66°C and more than twice-rated voltage. Capacitance and insulation resistance shifts are checked as well as catastrophic failure.

Ten percent of all manufactured product is sacrificed for a similar test, and all product is screened for 6 months at sea-bottom temperature (4°C) and at higher than rated voltage. During these tests, all significant electrical parameters must remain within tight limits.

Capacitance tolerance for this application is ±3 percent, with a total shift during life of perhaps 0.5 percent. Capacitors are initially stabilized by temperature cycling between −18°C and +66°C, and capacitance shifts are again held to tight limits. Insulation resistance and effective series resistance at 100 kHz (a measure of internal, stable, electrical continuity) are both held to absolute values; the change in those parameters during processing and aging is also held to tight limits.

6.2 Ceramic capacitors

6.2.1 Choice

Previous submarine cable systems used mica capacitors because of their excellent trouble-free history of providing stable, reliable operation at sea-bottom temperatures. The parasitic performance of mica capacitors at SG frequencies, however, precludes their use. Ceramic capacitors were chosen instead, although their long-term reliability had to be es-

tablished as part of the program. Results of testing and screening to date have shown this choice to have been a good one.

6.2.2 Requirements

As Western Electric does not manufacture ceramic capacitors, the required units are purchased from outside suppliers under specification KS-21013. Required values range from 3 pF to 0.047 μF, with 80 percent of the values less than 1000 pF. Tolerances range from ± 5 percent to ± 0.2 percent, with the majority of applications requiring ± 1 percent. Ceramic capacitors are available in a range of temperature coefficients but, for this application, NPO (zero temperature coefficient) capacitors are used because their capacitance stability with temperature averages better than ± 0.02 percent. There are no applications where the less temperature stable, high-K dielectric capacitors would suffice, in spite of their smaller size.

6.2.3 Construction

Ceramic capacitors use a thin (37 μm), modified titanium dioxide ceramic as a foundation onto which are screen-printed palladium electrodes. The ceramic sheets are piled up and punched into stacks which are pressed and fired. The resulting "chip" with alternating electrodes emerging from each end is coated on the ends with a fritted silver, which is then fired, thus connecting alternating electrodes together (see the cross section in Fig. 14). The completed chip is measured for capacitance and other parameters and then sorted into a "post office" of graded values, which are then paired to give the correct total value. Leads are soldered to the paired chip which is then mounted into a molded diallyl phthalate case and encapsulated with a silicone rubber potting material (see Fig. 15). After testing, capacitors are finally mounted into the appropriate printed wiring board.

The copper wire leads are nickel- and gold-plated to keep them cor-

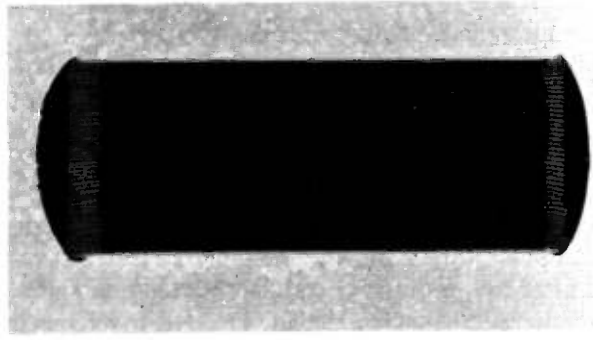


Fig. 14—Cross section of ceramic capacitor.

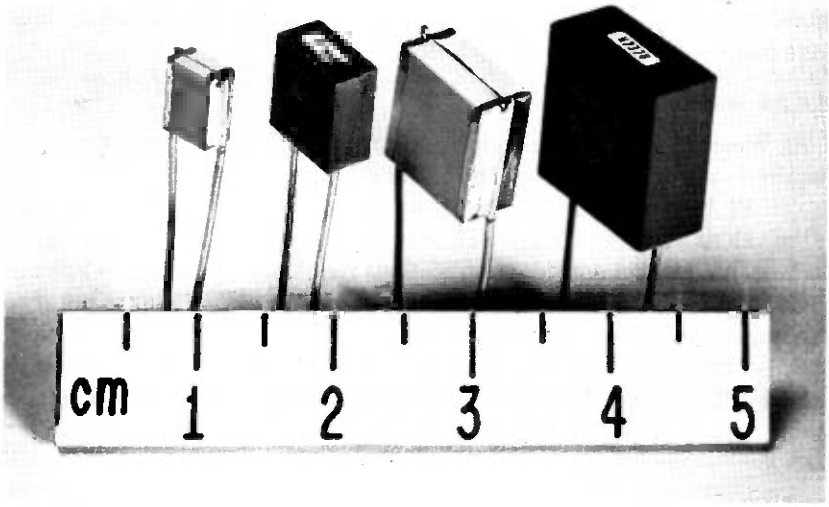


Fig. 15—Ceramic capacitors.

rosion-free and solderable. The nickel is used as a diffusion barrier between the copper and gold.

6.2.4 Screening procedures

With the application of voltage, ceramic dielectric undergoes an electrochemical degradation which eventually results in breakdown. The preliminary accelerated life testing of 3500 capacitors of the type to be used under a variety of conditions led to the conclusion that under sea-bottom conditions their average life would be over 100,000 years, but that it would be necessary to screen out atypical capacitors that might fail much earlier.

All capacitors undergo a rigorous testing program by the supplier, and a sample of each lot is subjected to destructive electrical and mechanical tests to determine the capability of the lot.

After delivery to Western, the capacitors are subjected to additional testing. To screen out any which may have an intermittent internal connection, each capacitor is cycled over the entire operating and storage temperature range (−20° to 60°C) twice while it is continuously monitored for capacitance. The capacitors are also life-tested at voltages ranging, according to dielectric thickness, from 150 to 450 V dc. Electrical parameters are monitored and compared before test, after two months of test, and after four months of test. Both fixed and statistical limits are placed on the values and the changes in value, to screen out any units which could be peculiar.

6.2.5 Reliability

In 96,702 capacitors screened, 8 short-circuited on the four-month life test. Assuming that the failure rate is proportional to the third power of the applied voltage, they would short-circuit at the rate of 0.005 FIT* in the system.

6.3 Solid tantalum capacitors

6.3.1 Introduction

Solid tantalum capacitors were used extensively for the first time in the SF submarine cable system. They are particularly useful because of their large capacitance density. Twenty solid tantalum capacitors are in each SG repeater: fifteen 1-μF and five 10-μF units (see Fig. 16). They are used exclusively as coupling and bypass elements, since their characteristics do not allow their use as elements in frequency-sensitive networks.

6.3.2 Screening procedures

All capacitors are measured for capacitance value, series resistance, and leakage current as they enter the Western Electric, Clark, New Jersey shop. Then, at specified periods during nine months of screening, they are measured for the same parameters four more times. Screening

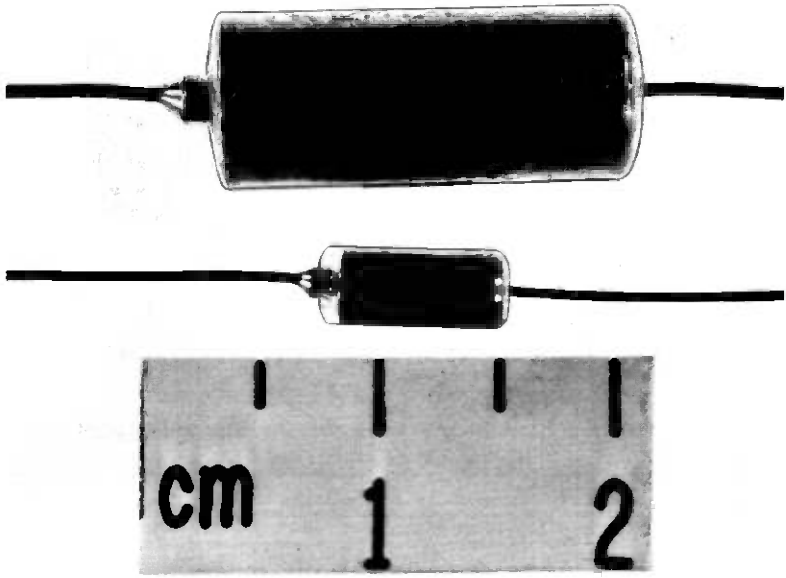


Fig. 16—Tantalum capacitors.

* Number of failures in 10^9 hours.

consists of temperature cycling, pulsing, surge testing, and life testing for three 60-day periods at elevated temperature and voltage.

After these tests, any unit displaying a ±0.5-percent change in capacitance is rejected. The absolute differences in successive readings of series resistance and leakage current are limited to fixed maximum values. In addition, the results are looked at statistically, so that any units showing absolute differences outside 3.5σ for either parameter are rejected.

VII. INDUCTORS AND TRANSFORMERS

7.1 Transformers

Transformer designs for undersea cable systems are influenced profoundly by the ratio of highest-to-lowest transmitted frequency. The SB, SD, and SF systems all maintained a ratio of about 12:1. In the SG system, however, this ratio has been increased to 50:1. The 30-MHz top frequency requires smaller designs to reduce parasitics, but sizes should not be so small that the low-frequency characteristics are degraded or the reliability is compromised.

Figure 17 shows the signal transformer designs. The 2678B/C is used either as an input or output transformer, while the 2678A is used for interstage coupling. All are "transmission line" transformers, and the input and output units are designed as hybrids. All use the same toroidal

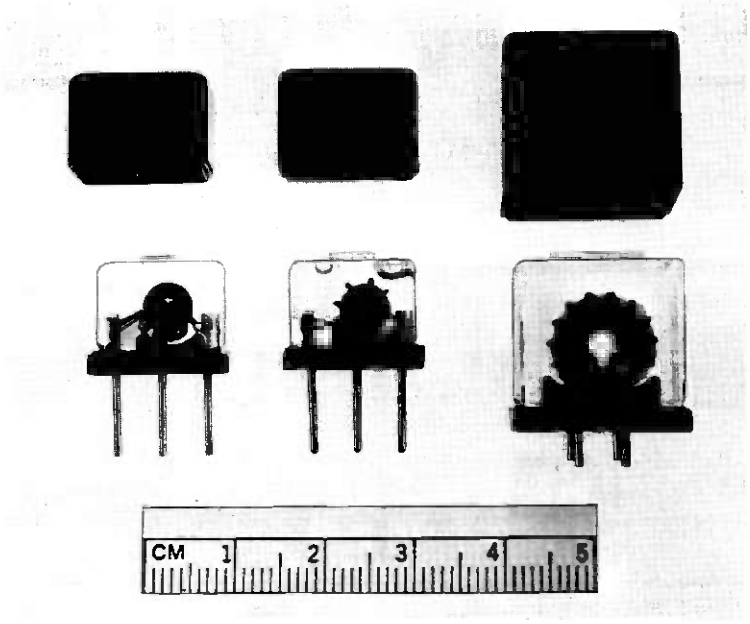


Fig. 17—Signal transformers.

core material, a newly developed manganese-zinc ferrite. The electrical integrity of these units must exist well outside the system's bandwidth, where overall gain and phase of the feedback amplifier continue to be important.

The 2679A transformers are used in such a way that their parasitic capacitance and short-circuit inductance become elements in the directional filter. Consequently, this puts a premium on the reproducibility and stability of these components. This is achieved by making the winding from a pair of wires, twisted under controlled tension at a specified number of turns per inch.

Figure 18 shows the 2680A longitudinal choke whose purpose is to suppress longitudinal currents while remaining transparent to the transverse signal voltages with no more than a 2-percent reflection coefficient. Its self-inductance is part of the power separation network; as such, it must carry the 0.657-A cable current. It is designed to withstand the 8-kV surge which could appear between the windings in the event of an accidental cable cut near shore. This combination of re-

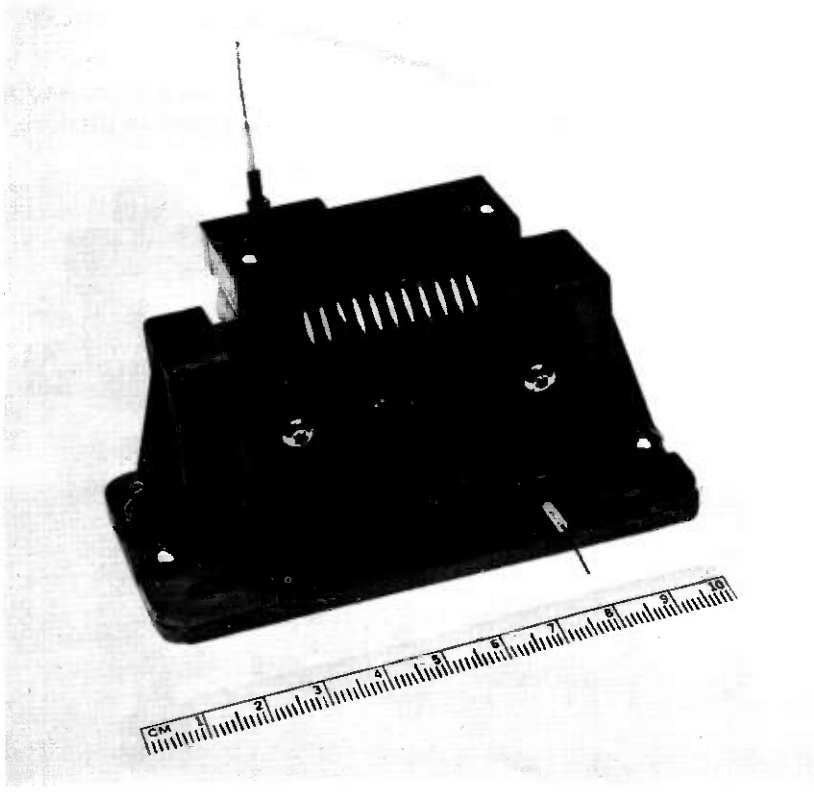


Fig. 18—The 2680A longitudinal choke.

quirements has resulted in the development of a special coaxial cable which comprises the winding and in the use of manganese-zinc ferrite core material. The compromise between permeability and coercive force appears to be appropriate.

7.2 Inductors

Eight inductor structures were developed and are shown in Fig. 19. Of the 68 codes using these designs, 27 are adjustable and form a stockpile of continuous inductance (0.0395 to 45 μH) for adjusting the mop-up filters. The remaining codes range up to 1000 μH. Wherever possible, designs have used the highly reliable procedures and materials developed for earlier systems. A new glass-bonded mica formulation with lower dielectric constant has been developed to lower distributed capacitance. New completely shielded designs have been necessary because of the tighter packing of components. Lower inductance units feature precisely spaced solenoidal windings, which are wound under tension and varnished in place. Higher inductance values are achieved by using varnish-coated, multiple-pie, duolateral windings.

The 1714A unit, shown in Figs. 19 and 20, is the most crucial inductor in the group. It measures 3.85 μH and is adjustable over a range of ± 2.5 percent. It is used to control repeater gain by varying β in the feedback amplifier. The length-to-diameter ratio is optimized to produce the

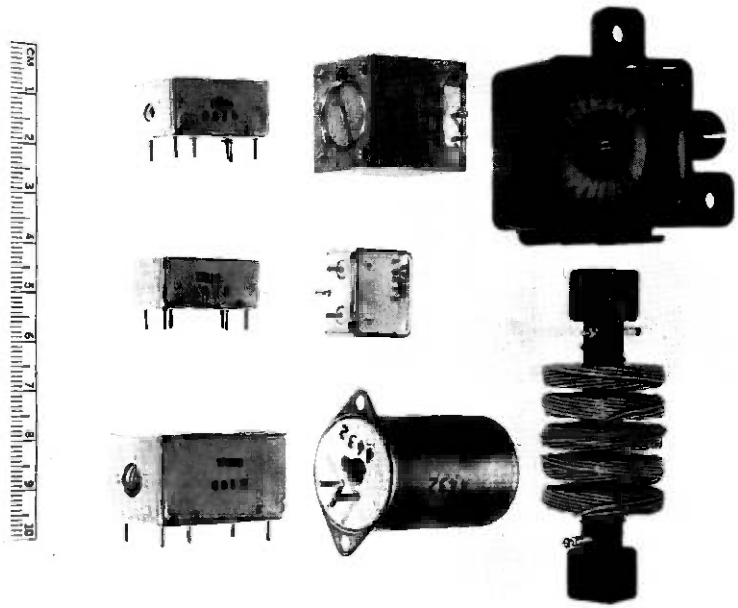


Fig. 19—Inductors.

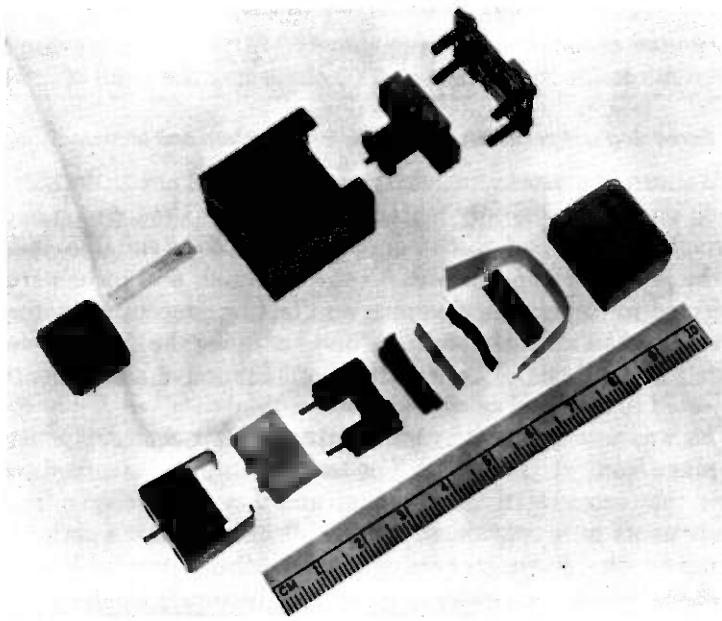


Fig. 20—Exploded view of 1710B and 1714A inductors.

highest possible principal resonance consistent with the attainment of high impedance at 200 MHz. Shielding is achieved by enclosing the winding in a brass block. Inductance is varied by turning a hollow brass plug in and out of the cavity to modify the inductor's magnetic field.

The 1710A and 1710B transistor feed chokes are 56 μH with ± 5 and ± 2 percent tolerances, respectively. Besides some initial adjustment capability, their first resonance (parallel resonance) must be at a high frequency and must exhibit high impedance. Also, the in-band third-harmonic products must not exceed -115 dBm when both fundamentals are at $+10$ dBm. These requirements have resulted in the design shown exploded in Fig. 20. The magnetic circuit consists of a small ferrite bar, separated from the ferrite winding spool by a mica shim to give a precise air gap. This ferrite bar is moved relative to the core to give the correct value of inductance. The position of the ferrite bar is maintained securely by spring-loaded clamps until the bar can be bonded in place.

The 1719A inductor shown in Fig. 19 is designed to produce 1 mH with the 0.657-A dc cable current through the winding. Because the unit must be corona-free to ground at 7,000 V dc, a magnetic core is not practical. As this inductor is in shunt with the signal path, it must maintain a specified minimum impedance up to 60 MHz. This is accomplished by winding the inductor as a duolateral coil of five sections. The *shunt* resonance (high impedance point) due to the overall inductance tuned by the effective capacitance is broad and occurs at 3 MHz. The first *series*

resonance due to magnetic coupling between the winding pies and the distributed capacitances occurs above 60 MHz, thus achieving the required minimum impedance of 1000 ohms over the band of interest.

7.3 Screening and reliability testing for transformers and inductors

To ensure that the inductive components would not be degraded after a long period of use, only materials were considered which had been thoroughly tested and which were determined to be suitable candidates for the proposed applications. All raw materials and piece parts were subjected to inspection, cleaning, and testing procedures before they were used. As a means of verifying and justifying the choice of material as well as methods of construction, all inductive components were subjected to artificial aging and stabilization.

This was accomplished by temperature cycling each component over the range from −18 to +66°C. The temperature cycles served two purposes: they accelerated aging changes and they established a trend line. Components were considered acceptable only if, after a series of temperature cycles, parameter changes were within permissible limits and the changes after each series of cycles were becoming smaller so that the extrapolated end-of-life changes would also be within the accepted limit.

REFERENCES

1. H. C. Poon, "Implications of Transistor Frequency Dependence," IEEE Trans. Elec. Devices, *ED-21*, No. 1 (January 1974), pp. 110–112.
2. G. K. Herb and E. F. Labuda, "Semiconductor Device Fabrication Using Nickel to Mask Cathodic Etching," U.S. Patent 3,808,108, applied for December 1971, issued April 1974.
3. M. R. Poponiak, W. A. Keenan, and R. O. Schwenker, "Gettering of Biopolar Devices with Diffused Isolation Regions," H. Huff and R. Burgess, eds., *Semiconductor Silicon* (1973), pp. 701–708.
4. A. J. Wahl, W. McMahon, N. G. Lesh, and W. J. Thompson, "Transistors, Diodes, and Components," B.S.T.J., *49*, No. 4 (May–June 1970), pp. 683–698.
5. A. J. Wahl, "Ten Years of Power Aging of the Same Group of Undersea Cable Semiconductor Devices," B.S.T.J., *56*, No. 6 (July–August 1977), pp. 987–1005.
6. V. L. Holdaway, W. Van Haste, and E. J. Walsh, "Electron Tubes for the SD Submarine Cable System," B.S.T.J., *43*, No. 4 (July 1964), pp. 1759–1782.
7. Ref. 4, pp. 683–698.
8. H. M. Olson, unpublished work.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Cable and Coupling Design

By G. E. MORSE, S. AYERS, R. F. GLEASON, and J. R. STAUFFER

(Manuscript received May 26, 1978)

This paper describes 1.7-in. SG submarine cable and cable terminations and compares their mechanical and electrical performance with the design objectives. At 30 MHz, transmission loss in the cable is very sensitive to variation in dielectric loss. Precautions required to prevent excessive dielectric loss change during cable manufacture and subsequent system life are discussed. Sea-bed temperature and pressure coefficients are reviewed in the light of TAT-6 experience, and an estimate is made of possible attenuation change over a 20-year interval.

I. INTRODUCTION

As discussed in the preceding papers, the British Post Office (BPO) had primary responsibility for cable development for the SG system and Bell Telephone Laboratories (BTL) was responsible for the repeater-to-cable coupling design.

A number of factors ruled out the use of 1.5-in. diameter-over-dielectric (DOD) SF[1] cable for the SG system; not least of these was its high loss. The channel capacity chosen for SG required an upper frequency of approximately 30 MHz. At this frequency, the attenuation of SF cable for a transatlantic system would have reached a formidable 32,600 dB but, by using a low-loss dielectric at an increased diameter, 1.7 in. DOD SG cable reduced this total by 5000 dB. This saved 123 repeaters and reduced the system power feed voltage.

However, cable attenuation remained far higher than on any previous system. This was reflected by: an increased sensitivity to manufacturing tolerances, significant attenuation changes with time (cable aging), and increased importance of laying effects and sea-bed temperature and

pressure coefficients. Of particular concern was the control and characterization of dielectric loss behavior. Surprisingly, the adoption of a low-loss dielectric seems to have compounded the difficulties and eventually came to dominate the cable development. This may be understood in part when it is appreciated that a change in dielectric loss-angle of 1 microradian on all the cable would cause a change of 26 dB at top frequency on TAT-6.

II. DESIGN OBJECTIVES

Many factors influence the design of a new cable. These include system costs, the development schedule, manufacturing capability, considerations of shipboard handling and maintenance, and compatibility with the repeater, its housing, and termination.

Early considerations of the possible lines of development indicated that, in the available time, a series of objectives based largely on modifications to the SF design would be the best approach. Economic studies showed that, for an upper frequency of 30 MHz, a more expensive cable with lower attenuation would reduce the total system cost in addition to increasing system reliability by reducing the number of repeaters.

The following improvements to the SF design were therefore set as general objectives.

(*i*) Reduce the cable attenuation at 30 MHz by the use of a larger cable and a lower-loss dielectric.

(*ii*) Reduce the dc resistance of the inner conductor by some 50 percent to offset the increased line current required by the SG repeater.

(*iii*) Increase the strength of the deep water cable to exceed, by at least 20,000 lb_f (88.96 kN) the weight in water of 3000 fathoms (5486 m) of cable plus three repeater bodies.

(*iv*) Improve the handling performance in terms of the minimum safe bending radius and the number of reverse bends before the inception of a crack in the outer conductor.

Work in conjunction with Imperial Chemical Industries Ltd. (ICI) and Standard Telephones and Cables Ltd. (STC) showed that a dielectric material with a loss angle of 47 μrad at 30 MHz could be extruded satisfactorily at a diameter of 1.7 in. (43.18 mm), the cable dimension giving minimum system cost. The inner conductor retained the SF geometry with the diameter increased to 0.478 in. (12.14 mm) to meet the other essential requirements. This dimension gave the cable the convenient impedance of 50 ohms, which is very near to the value for minimum attenuation and provides a value of attenuation negligibly greater than the minimum.

With these basic parameters fixed, the more detailed objectives were formulated, and development concentrated on providing a detailed

specification for all stages of material supply, cable manufacture, and testing.

III. DESCRIPTION OF CABLE

A number of different cable designs are required to meet the wide range of laying stresses and operational risks that a long-distance, deep-water system must survive. To satisfy these requirements, five versions of 1.7 in. (43.18 mm) cable were developed: an armorless cable for deep water use, a light armored transition cable, a heavy single-armored cable, a heavy double-armored cable, and a screened cable.

3.1 Armorless deep sea cable

The main system cable (Fig. 1) is an armorless design protected by a polyethylene sheath and provided with an internal strength member. It is suitable for use in deep water where there is little risk of damage by fishing trawls, clam dredgers, and ships' anchors, or of disturbance and abrasion by ocean currents on a rocky sea bed. Cable to this design constituted 93 percent of TAT-6.

3.1.1 Inner conductor assembly

The strength member of the composite inner conductor consists of a 41-wire, high-tensile steel strand in which three layers of wires are arranged, all with a left-hand lay, around a central king wire. The design, shown in Fig. 2, is essentially a scaled-up version of the SD/SF strand with the wire sizes increased to give a nominal diameter of 0.4305 in. (10.93 mm). The previous distribution of bending stresses was retained at the increased diameter by increasing the lay length to 9.10 ± 0.35 in. (231.1 ± 8.9 mm). If steel wire with a tensile strength of 300,000 lb_f/-

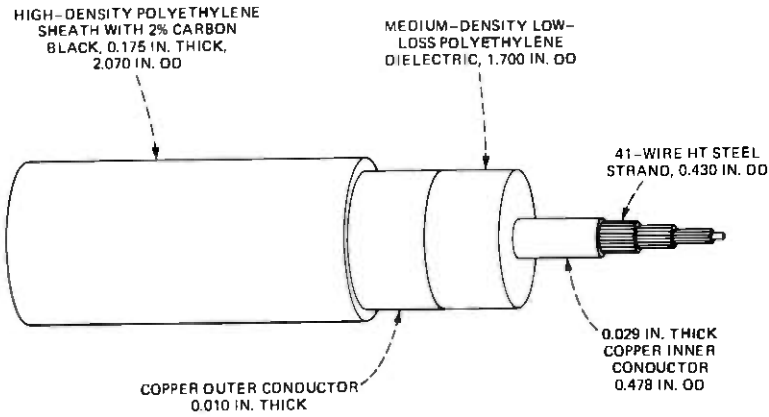


Fig. 1—Armorless SG ocean cable.

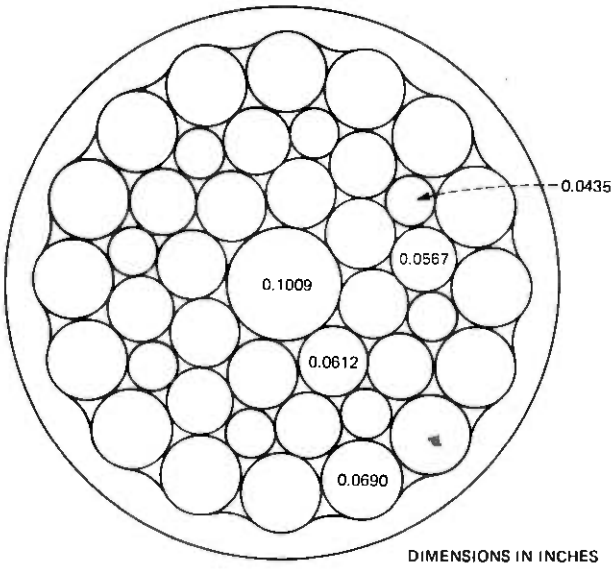


Fig. 2—Cross section of inner conductor. The tolerance on the wires is ±0.0005 in., except for the center wire, where it is ±0.0010 in. The wire is high tensile steel, 291,200 to 336,000 lb_f/in^2 (2008 to 2317 MPa).

in^2 (2068.5 MPa) is used, the strand has a strength of 37,000 lb_f (164.6 kN). Based on an estimated weight in water of 3500 lb_f (15.57 kN) per nautical mile, this gives the cable a strength-to-weight modulus* of 10.6 nmi (19.67 km). The immersed weight of a repeater body is 550 lb (2446 N) so when three bodies are supported at a depth of 3000 fathoms the excess strength of the cable is nearly 25,000 lb_f (111 kN). This exceeds the design target by 25 percent without including the strength contribution of the other materials outside the steel strand.

The inner conductor is formed from a 0.029-in. (0.737 mm) thick, fully annealed, high-conductivity copper tape. It is wrapped around the strand, in tandem with the strand-forming operation, to produce an oversize tube. The tape edges are closed in a longitudinal butt joint which is seam-welded with an inert gas arc. The conductor is completed by drawing the tube tightly onto the strand to give a finished diameter of 0.478 ± 0.001 in. (12.14 ± 0.025 mm). The previous practice of specifying oxygen-free high conductivity copper was relaxed to allow the use of a cheaper high conductivity electrolytic tough pitch copper. To maintain weld quality and reliability, a minimum purity of 99.9 percent and maximum oxygen content of 0.05 percent were specified. A further

* The modulus of a cable is equal to the strength divided by the weight per nautical mile in sea water.

control was placed on the maximum size (0.0014 in., 0.036 mm) and dispersion of oxide grains. The minimum conductivity allowed was 100.8 percent as defined by the International Annealed Copper Standard (IACS).

To ensure that the copper-to-steel interlayer shear strength is sufficient to transfer laying and recovery stresses to the strength member, the copper is heavily swaged into the interstices of the strand. The thickness of copper immediately over the wires is reduced to 0.024 in. (0.61 mm), but the effective thickness of the conductor is held at 0.029 in. (0.737 mm) by suitable adjustment of the difference between tape feed rate and line speed. As a result of cold forming, the conductor tape is work-hardened, and its conductivity is reduced on an average by 1.2 percent. This is later offset by approximately 0.35-percent increase due to a partial anneal during core extrusion. In the design model, a final conductivity of 99.5 percent IACS was assumed. This value was set low to ensure that sufficient repeaters were ordered. In TAT-6, the mass conductivity, determined from 119 samples of finished inner conductor, was 100.2 percent IACS. This result assumes a value of 8.89 g/cm^3 for the density of copper. If account is taken of the true density, then the volume conductivity would appear to have been slightly higher at about 100.4 percent IACS. The design value for the inner conductor resistance was 0.911 ohms per nautical mile at 10°C. Cable manufactured for TAT-6 in fact came out slightly higher at 0.931 ohms per nautical mile because the conductivity of the steel was lower than expected.

3.1.2 Dielectric

As system bandwidths are increased, the dielectric loss becomes more significant. Its contribution to cable attenuation is proportional to the product of loss angle and frequency and is independent of cable size, whereas the contribution from the conductor loss is inversely proportional to cable diameter and, once skin effect is established, is substantially proportional to the square root of frequency. The effect of dielectric loss at high frequencies is to increase cable attenuation and accentuate the deviation from a root frequency law, particularly in large diameter cables, thus complicating the compensation of manufacturing deviations by length adjustment.

In cable used for SF at 6 MHz, the dielectric was responsible for only 3 percent of the total loss at the top of the band, but using this polyethylene in SG cable increases the proportion to 10 percent at 30 MHz. To gain a worthwhile improvement, it was necessary to halve the dielectric loss while maintaining the other essential properties. Two UK suppliers, ICI and Bakelite Xylonite Ltd (BXL), met this requirement during the development stage and UBE Industries of Japan was approved during TAT-6 cable production.

Table I — Loss angle requirements

Frequency (MHz)	Loss Angle in Microradians at 23°C	Change from 23° to 3°C
30	47 ± 6	11 ± 2
	Loss angle expressed as a proportion of the loss angle at 30 MHz, 23°C	
1	0.62 ± 0.14	
6	0.77 + 0.10 −0.14	

The selected dielectric is a high molecular weight polyethylene characterized by a low contamination level, a precisely controlled permittivity, and a low dielectric loss. The reduction in loss was achieved by raising the density to 0.930 g/cm^3 (thus reducing the amount of amorphous material contributing to the loss), reducing the concentration of polar groups, and controlling more closely other impurities and adventitious contamination. Although approximately 3 μrad of the reduced loss angle was offset at 30 MHz by an increase in permittivity caused by the increase in density, the shift from SF to SG polyethylene is responsible for a net reduction in TAT-6 cable loss of 1300 dB at 30 MHz. The nominal* SG polyethylene loss angle is shown in Table I.

The tolerance at 1 and 6 MHz was expressed as a proportion of the 30-MHz value to control the shape of deviations in the linear frequency loss term while, at the same time, allowing the material suppliers a reasonable margin. A control on the loss-angle change with temperature was specified to limit the uncertainty in sea-bed loss at 30 MHz and to avoid a multiplicity of cable coefficients.

To allow for some small variation among sources of material, the quality assurance control was related to the nominal values of each material as determined during the type approval process. However, in approving alternative polyethylenes, the need to process them under similar conditions was appreciated; hence, close controls were placed on those properties affecting extrusion characteristics. Thermal oxidation was controlled by the low-loss antioxidant used in SF cable—Ethyl Antioxidant 330 [1, 3, 5 trimethyl-2, 4, 6 tri (3, 5-ditertbutyl 4-hydroxy-benzyl) benzene]. As a further protection against thermal oxidation, the granule feed hopper to the extruder screw was flooded with nitrogen. The increased density and crystallinity of the low-loss material led to a reduction in its stress crack resistance, but the level achieved was considered acceptable in view of the protection provided by the sheath and outer conductor and the absence of any record of a service failure due to stress cracking.

* The nominal loss angle of SG polyethylene at 30 MHz is based here on the value obtained with the Bell Laboratories Murray Hill test set. The Bell Laboratories set was used as the reference against which the dielectric test sets at the polyethylene suppliers and the two cable factories were calibrated by a careful exchange of plaques.

The polyethylene is pressure-extruded over the inner conductor and cooled in a series of water troughs which successively reduce the core temperature to the factory ambient (further details are given in Section V). The rate of cooling is arranged to prevent retraction of the core from the inner conductor. By this means, the interlayer adhesion required to transfer longitudinal laying stresses to the inner strength member is maintained. During development, the interlayer shear strength over a 3-in. length of conductor typically varied between 500 and 1000 lb_f (2225 and 4450 N) and occasionally reached 2000 lb_f (8900 N). However, under production conditions, the adhesion achieved over the same length varied between 250 and 500 lb_f (1112 and 2225 N). The required adhesion was 225 lb_f (1000 N). After extrusion, the dielectric is shaved to a diameter of 1.700 ± 0.003 in. (43.18 ± 0.076 mm) at a temperature of 20°C. The eccentricity of the inner conductor is limited to 0.020 in. (0.51 mm).

3.1.3 Outer conductor and sheath

The outer conductor is fabricated from an annealed copper tape, 0.010 ± 0.0005 in. (0.254 ± 0.0127 mm) thick by 5.630 ± 0.010 inch (143.0 ± 0.254 mm) wide, having essentially the same properties as the inner conductor tape. It is wrapped tightly around the shaved core at a temperature of 20 ± 1°C with the edges overlapped longitudinally by approximately 0.25 in. (6.35 mm). The maximum stretch permitted in the forming process is 0.5 percent. To provide the conductor with support and protection, it is oversheathed in a tandem operation with high-density polyethylene to a diameter of 2.07 ± 0.03 in. (52.58 ± 0.76 mm). To ensure adequate bending performance and avoid alignment problems when jointing, the minimum sheath thickness was limited to 0.150 in. (3.81 mm) and maximum eccentricity to 0.020 in. (0.51 mm).

In the design model, a conductivity of 100.5 percent IACS was assumed for the outer conductor, but during TAT-6 production an average mass conductivity of 100.8 percent IACS was measured on 121 samples. This figure becomes 101.0 percent IACS if a correction is made for the error in assumed density. The average dc resistance for the entire production run was 0.836 ohms per nautical mile at 10°C.

During development, the bending performance of three outer conductor thicknesses, 0.008, 0.010, and 0.012 in. (0.203, 0.254, and 0.305 mm), was examined in combination with three sheath thicknesses, 0.125, 0.150, and 0.165 inch (3.175, 3.81, and 4.19 mm). A wide variation was found among cable samples having the same outer conductor and sheath thickness, which made comparison difficult. In general, the resistance to buckling and cracking improved as the conductor and sheath thicknesses increased, but the thicker the conductor the more pronounced the effect of the sheath. Each combination was assessed from flex tests but the extent of bending damage was very dependent on the test rig

used. There was also a marked improvement in bending performance when the bending diameter was increased from 9 to 10 ft (2.74 to 3.05 m). These differences were considered sufficient to justify setting 10 ft (3.05 m) as the bending limit for 1.7-in. (43.18 mm) cable.

The performance of the 0.008-in. outer conductor was very marginal at a bending diameter of 10 ft (3.05 m) and it failed to survive 50 reverse bends in every test at 9 ft (2.74 m) irrespective of sheath thickness. The selected combination of 0.010/0.165 in. (0.254/4.19 mm) was also slightly marginal, but the risk was thought to be acceptable provided the minimum sheath thickness did not fall below 0.150 in. (3.81 mm).

The sheath material was a high molecular weight, high-density ethylene copolymer, containing 2.6 percent by weight of carbon black. The density specified was slightly higher than that required in SF cable. This increased the flexural modulus which increased the hoop stress applied to the outer conductor by the sheath, thus improving the reverse bend performance. Four sources of sheath material were approved, although only three were used in TAT-6. Towards the end of production, the inclusion of carbon black as protection against actinic radiation was discontinued except for cables ordered as repair lengths. Ironically, the removal of carbon black brought a substantial improvement in bending performance. In continuous loop tests, where cable is driven in a figure eight between two sheaves, the onset of cracking at 6-ft (1.83 m) diameter changed from 20 to 40 reverse bends and at 10 feet (3.05 m), cracking did not occur before 100 reverse bends. Clearly, the use of a natural sheath material opens up the prospect of a useful cost saving in future systems from the use of a thinner conductor/sheath combination. However, those responsible would be well advised to retain the original dimensions for repair lengths containing carbon black. In the final design, the weight in sea water was 1.6 tons* per nautical mile and in air 5.7 tons per nautical mile. The strength was 16.5 ton_f and the modulus 10.5 nmi (19.5 km).

3.1.4 Jointing

Shipboard and factory joints were developed along separate lines to take account of differences in the conditions and equipment available for their preparation. For example, while both share a common standard for tensile strength and bending performance, completion of a shipboard joint is required in the shortest possible time.

Both methods use a swaged steel ferrule to make a joint in the inner conductor assembly, and each type of joint has a minimum strength of 15 ton_f. In the factory, this is achieved using a 6.5-in. (165.1 mm) long ferrule pressed to a mean diameter of 0.82 in. (20.8 mm) and in a repair by swaging a 4.5-in. (114.3 mm) long ferrule to a diameter of 0.87 in. (22.1 mm). With these dimensions, there is little to choose between them

* 1 ton = 2240 lbs, or 1016 kg.

electrically. The factory ferrule has a return loss at 30 MHz of 20 dB compared with 21.5 dB for the shipboard version.

The dielectric is restored using an injection molding process, and both versions must be capable of withstanding 200 kV dc for 1 minute in type approval tests. All core joints are examined radiographically for voids, cracks, and inclusions. Each joint made in the factory must also withstand a production test of 100 kV dc for 1 minute without failure. In a shipboard joint, the outer conductor is connected through by the insertion of a split copper wire braid soft-soldered to the parent outer conductor. While this technique has a superior bending performance to the SF outer conductor restoration, lingering doubts regarding its long-term electrical stability have prevented its adoption for factory use. To improve the flexibility of the outer conductor restoration, instead of spot-brazing a split copper tube directly to the outer conductor as in SF, three helical copper tapes are spot-brazed to the split tube and to the outer conductor. This provides a semiflexible connection which, in a complete joint, is capable of withstanding 50 reverse bends at a diameter of 10 ft (3.05 m) without loss of continuity. In the factory, the joint is completed by fusing a split section of sheath to the cable and sealing it with a longitudinal seam weld. The hoop stress of the sheath is restored by a tight, close-wound binding of $\frac{1}{16}$ in. (1.59 mm) galvanized iron wire for a distance of 5 ft (1.52 m) over the joint. For the shipboard joint, the sheath is restored by a polyethylene sleeve slipped on the cable and located over split sheath sections placed as packing over the braid. The ends of the sleeve are fused to the sheath in an injection mold and reinforced with a galvanized wire binding.

3.1.5 Pre-formed dead ends for holding armorless cable (stoppers)

When holding armorless cable during a repair, it is necessary to grip the cable by a method that transfers the load from the inner strength member to the dead end without causing damage to the cable. This is arranged using a dead end manufactured by Preformed Line Products (GB), Ltd, Andover, England. In essence, thirteen 0.212-in. diameter high tensile steel wires are formed into a helical tape 26 ft long. The wires are fixed together with a vinyl adhesive and coated on the inside of the helix with aluminum oxide. By reversing the lay of the helix at the middle of the tape, the ends may be folded into a hairpin which, when wrapped around a cable, leaves a convenient loop for the attachment of a shackle.

A single dead end is capable of holding armorless cable at tensions up to 18 ton_f but, if cable is held at the bow of a ship for long periods, tension, bending, and vibration fatigue may cause premature outer conductor failure well below this value. Since a splice may take several hours to complete, the use of two dead ends in tandem, linked by an equalizing sling to spread the load, is essential.

With careful protection of the cable at the eye of each dead end, it is possible to hold SG cable for 12 hours under a cyclic load of 1 to 7 tonf with a period of 6 seconds. However, if the tension range is increased to 1 to 9 tonf, the time to develop fatigue cracks in the outer conductor is halved.

3.2 Armored cable

In shallow water, additional protection is required to prevent cable failure due to abrasion or disturbance by fishing trawls and ships' anchors. This may be achieved by the provision of a suitable armor or by cable burial depending on the circumstances. Since the TAT-6 route was suited to cable burial using a plow, it was planned to maximize protection by laying heavy single-armored cable to a depth of 1000 fathoms (1829 m) and by burying the cable at depths less than 500 fathoms (914 m). Clearly, the description "shallow water" is something of a misnomer, but the need for protection is real. Fish are now caught at 800 fathoms (1460 m), and a trawl fault has occurred at a depth of 735 fathoms (1340 m). In TAT-6, the plow was used to a maximum sea-bed depth of 350 fathoms (640 m), while the transition to armorless cable was made at 750 fathoms (1372 m). However, a short length of single-armored cable was successfully laid at 1100 fathoms (2012 m) when the depth briefly increased after the transfer from armorless to armored cable on completion of the deep sea lay.

When low attenuation is not a prime requirement, operational constraints during laying and repair favor the choice of a small diameter coaxial for the armored cable. However, during the SG development the feasibility of armoring the main system cable was demonstrated and on a sea trial this cable was laid and recovered without serious difficulty. Indeed, the feasibility of using a heavy double-armored 1.7-in. cable was also demonstrated. The 1.7-in. cable was selected because its greater strength and weight provided increased protection from trawler attack and also because its use reduced the number of repeaters required.

3.2.1 Single-armored cable

To reduce the risk of an insulated break in the inner conductor and the attendant difficulty in location, it is customary in the design of armored cables to provide a ductile conductor that more than matches the elongation characteristic of the armor. In SG, the continued use of solid copper was prohibitively expensive, but there was insufficient time to develop an alternative* structure with high elongation before TAT-6 was laid. As a result, the TAT-6 armored cable used the high tensile steel strand inner conductor assembly of the armorless cable. It was therefore

* See possible alternatives in Section 3.2.3.

necessary to select the armor material with some care. The use of Grade-65 (British Standard abbreviation) medium tensile steel wire armor gives a residual strength* well below the minimum strength of the strand at a strain equal to the yield strain of the strand. In practice, the strand yields at approximately 1 percent strain when the armor has a residual strength of 11 ton_f. Since the minimum strength of the strand is 16.5 ton_f, its failure under load will cause a cable break and exposure of the inner conductor to the sea. The armor has a minimum strength of 68 ton_f; therefore, taking into account the contribution from the strand, the effective strength of the cable is 73.5 ton_f. In proving trials, the inner conductor integrity was demonstrated at tensions up to 79 ton_f with failure occurring at approximately 0.8 percent strain.

The single-armor design is shown in Fig. 3. It comprises 25 wires of 0.291 ± 0.004 in. (7.391 ± 0.102 mm) diameter, medium tensile galvanized steel applied with a left-hand lay of 36.5 ± 1.0 in. (927 ± 25 mm) over a bedding of 75-pound (34 kg) jute. The armor is coated with bitumen and bound with two servings of 3-ply, 28-pound (12.7 kg) jute applied in opposite directions and coated with a bituminous compound to a finished diameter of 3.05 in. (77.5 mm). The cable specification in fact allowed the use of an approved alternative to jute and, since the completion of TAT-6, synthetic yarns have replaced it as the preferred bedding and serving material.

The single-armored SG cable has a weight in sea water of 15 tons per nautical mile and a weight in air of 24 tons per nautical mile. It is clearly not a cable that a small repair ship could handle easily but, provided the serving remains intact, it is relatively docile and can be coiled in a tank to a diameter of 12 ft (3.66 m) and drummed without difficulty at a diameter of 10 ft (3.05 m).

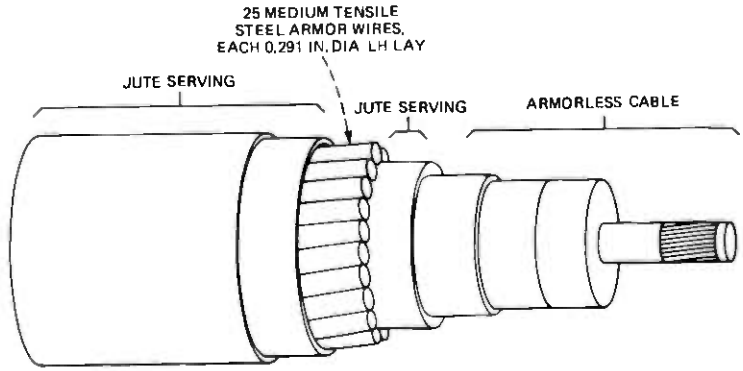


Fig. 3—Single-armored SG cable.

* Residual strength is the difference between breaking strength and the strength at a particular value of strain.

3.2.2 Transition cable

Since the static weight in water of single-armored cable at 1000 fathoms (1828 m) approaches very nearly the strength of the armorless cable, a light-armored transition cable is required at a transfer from armored to armorless cable. In the absence of previous experience of laying heavy-armored cable at this depth, it was considered advisable to provide a double-armored transition cable with a neutral torque characteristic. The design strikes an optimum strength-to-weight modulus of 2.6 nmi (4.82 km) for the armorless cable with respect to the transition cable and for the transition cable with respect to the single-armored cable. (That is, the armorless cable has strength to support the weight in water of 2.6 nmi of transition cable and, similarly, the transition cable has strength to support the weight in water of 2.6 nmi of single-armored cable.)

The transition cable is shown in Fig. 4. Each layer of armor has 48 Grade-65 galvanized steel wires coated with polyvinyl chloride to space the wires uniformly around the cable. In the first layer, the wires are 0.085 ± 0.001 in. (2.159 ± 0.025 mm) diameter applied with a left-hand lay of 42 ± 1 in. (1067 ± 25 mm) and in the second layer they are 0.078 ± 0.001 in. (1.981 ± 0.025 mm) diameter applied with a right-hand lay of 54 ± 1 in. (1372 ± 25 mm) over an intermediate serving comprising two opposing layers of 3 × 17 pound (7.7 kg) jute coated with bitumen. The cable is completed with an outer serving of two opposing layers of 3 × 28 pound (12.7 kg) jute flooded with a bituminous compound to an overall diameter of 3.17 in. (80.5 mm). Its weight in sea water is 5.9 tons per nautical mile, and its weight in air is 15.5 tons per nautical mile.

The armor strength varies from 21 to 25.8 ton_f, depending on the quality of steel, but because the inner conductor fails at 1-percent strain, only about 88 percent of the potential strength is realized. The combined strength of the armor and strand therefore varies from 35 to 39.2 ton_f.

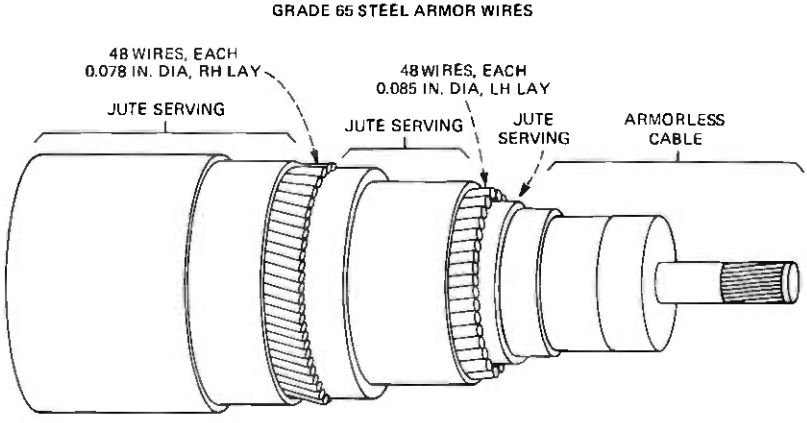


Fig. 4—Transition cable.

In proving trials, failure occurred at 39 ton_f with an extension of 0.84 percent. No rotation of the armor was observed.

3.2.3 Double-armored cable

As stated previously, the feasibility of laying a double-armored SG cable was demonstrated but such a cable was not adopted for TAT-6. The bedding and first armor pass is as for the single-armored cable. Then, with coatings of bituminous compound under and over each layer, two intermediate servings of 3-ply 17-lb (7.7 kg) jute are specified with opposing lays. This is followed by a second armor pass of 33 wires, 0.291 ± 0.004 in. (7.391 ± 0.102 mm) diameter, Grade-65 medium tensile galvanized steel, applied with a left-hand lay of 48 ± 1 in. (1219 ± 25 mm). The second pass is served with two opposing layers of 3-ply 28-lb (12.7 kg) jute coated in bitumous compound to ensure that the serving adheres firmly to the armor. The finished cable has a weight of 32.8 and 46.8 tons per nautical mile in sea water and air respectively and an overall diameter of 3.8 in. (96.5 mm). It has a tensile strength of 162 ton_f.

The tremendous strength of this cable rules out the use of a composite inner conductor with a high tensile steel strand if an insulated inner conductor break is to be avoided. However, preliminary trials of a 19-wire mild steel strand and of copper-clad aluminum rod indicate that either could be used, the former with virtually no further development.

3.3 Screened cable

At shore ends and in land sections, additional screening is required against electromagnetic interference in the low-frequency band. In SG, the primary concern is with the radio broadcast band, 0.55 to 1.6 MHz. At these frequencies, adequate protection is given by increasing the outer conductor thickness to 0.030 in. (0.762 mm). This was shown to be practicable towards the end of SG development and may be used in future systems, but in TAT-6 a modified version of the SF screened cable[1] was used. The cable impedance was converted to 50 ohms by increasing the inner conductor diameter to 0.281 in. (7.137 mm). Corrosion-prone neoprene-covered armor wire was replaced by a conventional armor, protected with tar and jute servings.

IV. ELECTRICAL CHARACTERIZATION

To enable the system to be equalized, the level of uncertainty in the predicted attenuation of laid cable must be reduced to that which can be corrected by the ocean-block equalizer (OBE). While many factors contribute to uncertainty, some tend to be random while others are systematic. The random factors include manufacturing variations and measurement error. An error in the temperature or pressure coefficient, on the other hand, will cause a systematic deviation from the predicted

sea-bed loss. In the OBE, provision was made for the correction of cable misalignment amounting to ±2.5 dB systematic and ±1.5 dB random per ocean block at 30 MHz. In a 20-repeater block, this corresponds to a total of ±0.48 percent of cable loss.

Since cable and repeater development proceeded simultaneously, predictions of cable loss and laying coefficients were required in advance of the manufacture of experimental cable sections. The initial estimates were provided from theoretical consideration of the physical properties and dimensional behavior of cable material based on experience with SD and SF systems and supplemented as necessary by experimental data of dielectric loss at frequencies up to 30 MHz. Uncertainties in sea-bed estimates were narrowed in the light of experience gained from sea trials and refined again as experimental data led to a better understanding of the results. The objective was the characterization of cable typical of large-scale production rather than the accurate description of experimental lengths.

4.1 The effect of variation in manufacture

Strict control of material properties, cable dimensions, and manufacturing processes is exercised through the cable specification, in order to limit the variation in cable loss. Table II shows the potential magnitude of variations due to raw material and dimensional tolerances. In the case of conductivity, no advantage is to be gained from an upper limit, so the anticipated variation is shown. A similar situation exists with the

Table II — Effect of variation in material properties and cable dimensions on cable attenuation

Parameter	Tolerance	Effect on Attenuation in Percent		
		1 MHz	6 MHz	30 MHz
Inner conductor:				
Diameter	±0.001 in. (±0.025 mm)	±0.00	±0.00	±0.00
Thickness	±0.0005 in. (±0.013 mm)	±0.00	±0.00	±0.00
Conductivity	±0.4%	∓0.15	∓0.15	∓0.15
Core insulation:				
Diameter	±0.003 in. (±0.076 mm)	∓0.18	∓0.18	∓0.17
Relative permittivity	±0.004	±0.09	±0.09	±0.09
Loss angle: 30 MHz	±6 μR	—	—	±0.57
6 MHz	+ 5.5 μR	—	+0.24	—
	−7.4 μR	—	−0.33	—
1 MHz	±7.4 μR	±0.14	—	—
Outer conductor:				
Air gap	±0.0006 in. (±0.015 mm)	∓0.11	∓0.10	∓0.10
Thickness	±0.0005 in. (±0.013 mm)	±0.00	±0.00	±0.00
Conductivity	±0.4%	∓0.04	∓0.04	∓0.04

air gap* between the core and outer conductor. Although it cannot be quantified by the specification, every attempt is made to control and regulate its size. It is perhaps worth commenting that the limits on diameter and permittivity variation were met in every cable section manufactured for TAT-6 (the former by a fair margin), and that the polyethylene core material met the loss angle requirement in more than 97 percent of the cable sections.

4.2 Control of deviation from design

With one exception, the effect of individual tolerances is small, but when combined they amount to a total deviation of ±1.2 percent at 30 MHz. Half of this is due to the dielectric loss, but since its effect at low frequencies is far less, equalization at 30 MHz by length adjustment will cause a significant deviation from design in the low band. This was thought to be sufficient justification for a further control on the final characteristic. When the first experimental cable lengths were produced, another reason emerged.

Attenuation measurements on trials cables revealed an enhanced loss which had every appearance of being due to the dielectric but which proved difficult to confirm from dielectric loss measurements on samples prepared from extruded core. The apparent† increase in loss angle was 7 to 8 μrad at 30 MHz. Since this had a significant impact on attenuation, it was incorporated into the design model. It became necessary therefore to provide a means of checking that the effect remained constant during production. This was achieved by extracting the apparent loss angle of the cable at 30 MHz from an examination of the ratio of the losses at 30 and 1 MHz using the following expression:

Apparent loss angle at 30 MHz = $57 + 188(R - 5.756)$ μrad,

where

$$R = \frac{\text{measured insertion loss (dB) at 30 MHz}}{\text{measured insertion loss (dB) at 1 MHz}}.$$

It can be shown from consideration of the transmission equations that the ratio R is proportional to the loss angle at 30 MHz but relatively independent of it at 1 MHz. It is also virtually unaffected by deviations in loss which are "cable shape." Combining errors from all sources, including test equipment, the technique is considered accurate to ±1.7 microradians at 30 MHz. By smoothing the measured data, it is possible

* The outer conductor does not conform precisely to the surface of the dielectric; the term "air gap" refers to the space between the outer surface of the dielectric and the inner surface of the outer conductor. The thickness of the air gap is the average value of the separation between these surfaces.

† The apparent increase in loss angle is now thought to be due in part to a calibration error of 5 μrad in the reference dielectric test set. It is hypothesized that the remainder is due to oxidation or contamination of the polyethylene during core extrusion.

to reduce this uncertainty to ±1 microradian. Given that the dielectric as supplied remains within tolerance, then the apparent loss angle should not vary by more than ±8 μrad from a nominal value of 62 μrad at 10°C. This proved to be the case, but it was soon found that a tolerance of this magnitude allowed large variations in the change of dielectric loss between raw material and finished cable resulting from differences in manufacturing conditions. Since this had an adverse effect on subsequent loss stability, the control was placed instead on the processing change. This was implemented by restricting the change between raw material measured at 23°C and finished cable measured at 10°C to 14 ± 4 microradians. To give an early warning of approaching trouble so that immediate remedial action could be taken, the use of moving range control charts showing the dielectric process change was made mandatory toward the end of TAT-6 production.

Having extracted a value for the dielectric loss at 30 MHz, the combined deviation of the cable shape components of cable loss can be estimated and controlled. A tolerance of ±0.5 percent was specified which encompassed the total production deviation from the mean characteristic at each factory.

It is evident from Table II that small variations in the air gap could be responsible for a sizeable "laying effect," i.e., the amount by which the cable misses its sea-bed design loss. To check that the application of the outer conductor was under control, one section in 10 was required to be repanned (removed from one storage pan and coiled into another) after the factory transmission tests and then remeasured. A large increase in loss as a result of handling the cable would point to a slack outer conductor and the need for adjustment of the forming mill.

In the final design loss for factory measurements at 10°C, shown in Fig. 5, an air gap of 0.0020 in. (0.051 mm) was assumed. In repanning operations, the air gap collapsed on the average by 0.0007 in. (0.018 mm) after a single turnover and by a further 0.0002 in. (0.005 mm) following a second turnover. However, the degree of scatter was fairly large and appeared to be related to previous storage conditions and the time elapsed before turnover.

4.3 Sea-bed coefficients

Laying coefficients to allow correction for temperature, pressure, and handling effects were obtained from 18 nmi (33.4 km) of prototype SG cable manufactured by STC. The trials cable was produced in four equal lengths, half with ICI and half with BXL low-loss polyethylene. The temperature coefficient was determined from transmission measurements of individual and paired lengths at five equally spaced temperatures varying from 22° to 4°C.

At high frequencies, the cable temperature coefficient is increasingly

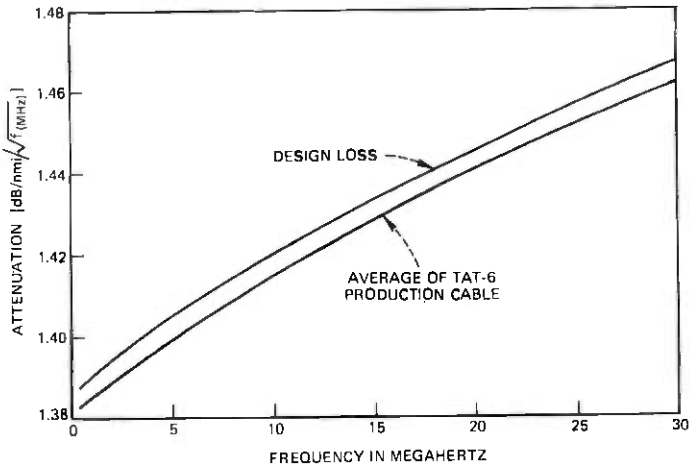


Fig. 5—Attenuation at 10°C and 0 fathom. An air gap of 0.002 in. between the outer conductor and core was assumed in the calculation.

dependent on the dielectric loss, but there was little evidence of difference among the SG dielectrics. The mean of the trials cable coefficients is compared in Fig. 6 with the final design coefficient. This is split into two parts to allow for laboratory evidence which suggested that the rate of thermal expansion of the dielectric differed over the temperature ranges 2.5° to 10°C and 10° to 20°C. However, a measured temperature coefficient was liable to apparently substantial but quite illusory modification by small variations in outer conductor behavior. If the direction of temperature change was from low to high, then the outer conductor did not move until the air gap closed. This caused an apparent increase in the temperature coefficient by overriding part of the dielectric contribution. An apparently nonlinear coefficient was readily caused by excessive delay between two measurements at successive temperatures. This allowed additional collapse of the outer conductor as the sheath very gradually contracted. A similar effect could be produced by successively overshooting and overcorrecting several times before reaching a stable temperature. The time constant of these effects was far longer than that required for dc resistance stabilization.

4.4 Sea trial

Having established the temperature coefficient, a sea trial was conducted in June 1973 by Cable Ship *Alert* to determine the pressure coefficient of the armorless cable and to assess the handling characteristics of the armored cable. The two 9-nmi (16.7 km) lengths of armorless cable were laid at depths of 1500 and 2500 fathoms (2743 and 4572 m) in the form of a hairpin, keeping both ends on board to facilitate trans-

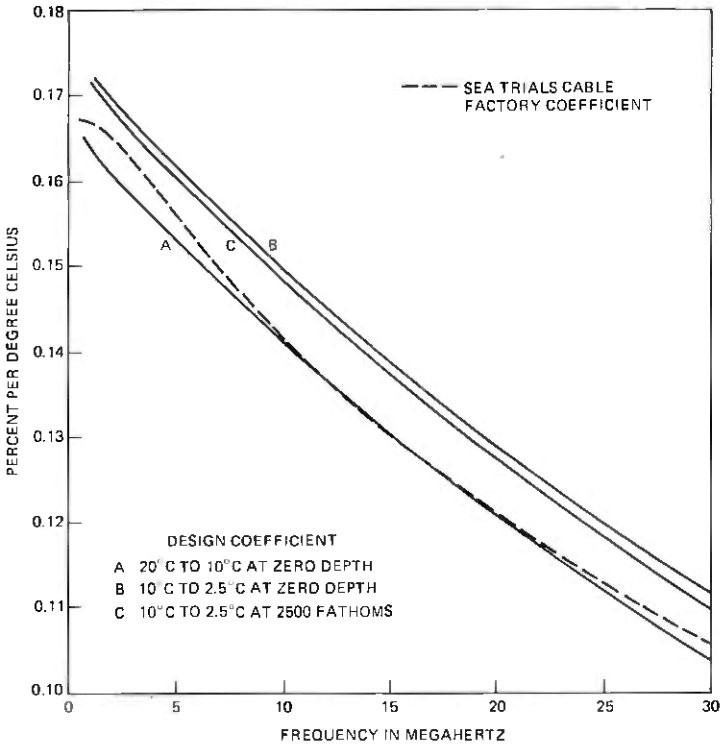


Fig. 6—Attenuation temperature coefficient for 1.7-in. SG cable.

mission tests. The design coefficient shown in Fig. 7 was based on the two points obtained from each cable measurement. The depth coefficient increased with frequency due to the increased influence of the dielectric. The coefficient obtained from the ICI cable was smaller than that from the BXL and, at 1 MHz, was 0.12 percent lower than the computed value for SF cable. It was therefore decided to bias the SG design coefficient towards the BXL characteristic with a value slightly above the SF design at 1 MHz.

The curve in Fig. 7 labeled "Mean Sea Trials Cable Coefficient" was calculated using additional depth information gained from downrunner* measurements. These were taken when the cable just reached the sea bed at each depth. The sea-trial cable coefficient is the average of a linear fit through the four points thus obtained for each length. Treated in this way, the ICI and the BXL data confirm the previous ICI two-point coefficient. The data also indicate the presence of a large air gap in the BXL cable which is consistent with the attenuation change seen during handling operations.

* The downrunner is the portion of cable between the ship and the sea bed.

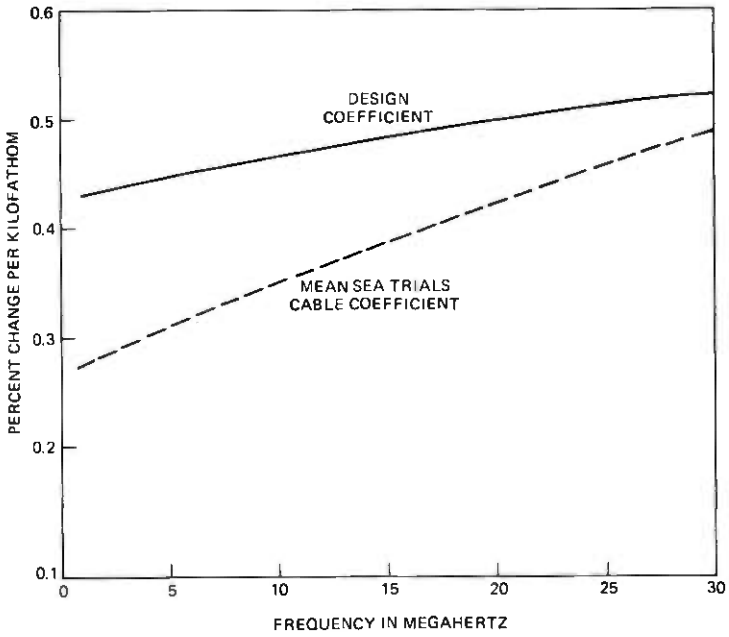


Fig. 7—Attenuation pressure coefficient for 1.7-in SG cable.

The essentially non-depth-dependent component of the pressure coefficient due to closing the air-gap under the outer conductor was modeled from air-gap behavior determined from laboratory measurement and sea trials results. The design value was set at 0.37 percent flat with frequency.

After the sea trials, both cables were streamed (laid out on the sea bottom), one at each depth, with a view to recovery at a later date to assess their stability over the intervening time period. This was undertaken in June 1975 by Cable Ship *John W. Mackay*. During the two-year interval, the cable which had been manufactured under conditions closest to those approved for TAT-6 production had increased in loss by the equivalent of 1 μrad at 30 MHz.

V. ANOMALOUS DIELECTRIC LOSS INVESTIGATION DURING TAT-6 CABLE PRODUCTION

When cable production started at CDL*, it was found that some freshly produced lengths exhibited excessively high apparent loss angles at 30 MHz, 20 to 30 μrad above nominal. It was quickly discovered that the effect was related to extruding core into an initial trough of water

* Cables de Lyon, Calais, France.

at 105°C† instead of the then accepted value of 95° to 97°C, as shown in Fig. 8. The higher temperature had been chosen to maintain adhesion and freedom from voids at increased extrusion line speeds. As a result of the investigation described below, the subsequent extruded core was chilled in a first water trough having a temperature of 92°C.

The high apparent loss angle of the affected cable continued to increase with time in the factory, appearing to reach a peak value after three to six months and then showing signs of slowly decreasing. When the temperature of the water in the first trough was reduced to 97°C, cable was produced within specification but some changes in apparent loss angle with time still remained (see Fig. 9). This resulted in lengthy investigations into the effects of water on polyethylene.

Freshly produced core, extruded into water at 97°C, was rushed from the factory to the laboratory and water measurements were made on three concentric layers. The outer layers were found to contain up to 200 parts per million (ppm) of water, while the other layers contained close

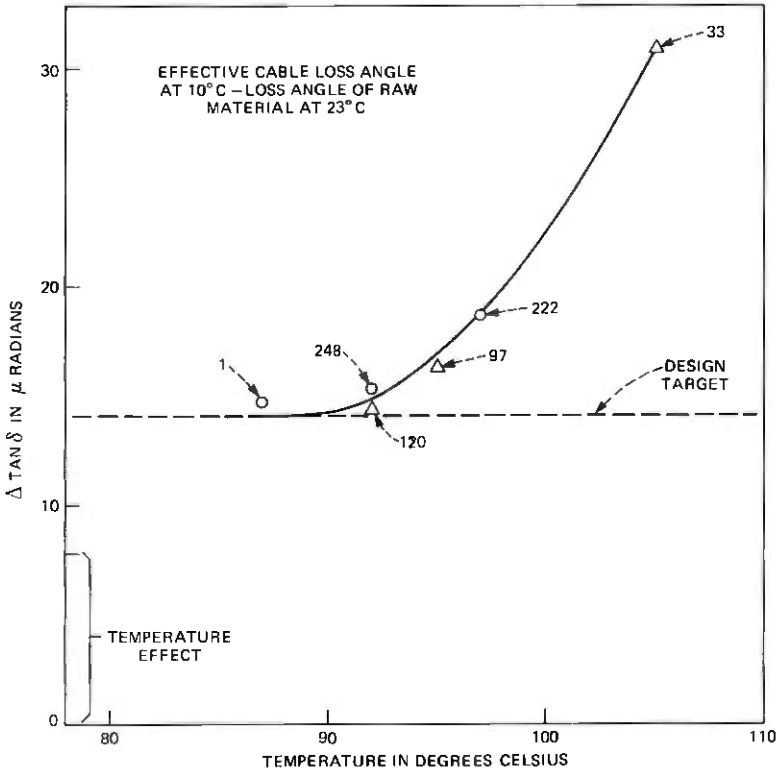


Fig. 8—Core extrusion cooling temperature versus loss angle change. Number of cable sections represented by each plot is shown on the curve.

† Pressurized cooling troughs allow temperatures above the atmospheric pressure boiling point.

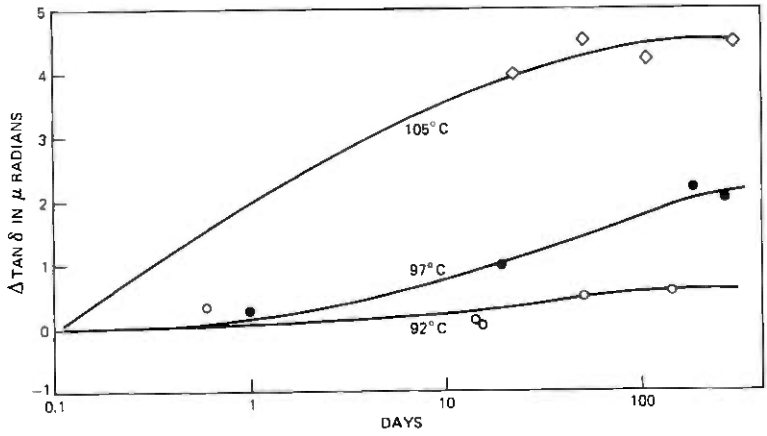


Fig. 9—Typical factory aging characteristic for cable core cooled at different temperatures. Change in apparent cable loss angle with time at 30 MHz and 10°C.

to the normal saturation levels of about 10 ppm at room temperature. Plaques were also molded from material from these three layers and the loss angles determined at several frequencies. Plaques from the outer layers of the core initially showed enhanced loss angles with an excess of 20 to 30 μrad at 30 MHz, but the values decreased quite rapidly with time, recovering to expected values within 24 hours.

5.1 Water measurements

As a result of this initial work, techniques were developed for examining the quantity and nature of the water trapped in the core. A cutter was produced which could remove a cylindrical plug from a cable core without affecting the distribution of water. If six of these plugs were used and sliced into thin disks, the radial distribution of the water in the core could be measured using a DuPont moisture analyzer to an accuracy of ±5 percent or ±2 ppm for small quantities of water. The distribution of water through a fresh core of SG polyethylene is given in Fig. 10. Most of the water is concentrated in the outer layers and is far in excess of normal saturation levels. It was suspected that the excess water was clustered in the form of microdroplets in small voids less than 1 μm in size. A Perkin-Elmer scanning calorimeter was used to confirm this theory of clustering. By cooling samples to −55°C and then monitoring them while gradually heating to 30°C, a peak in heat capacity was observed at 0°C caused by the melting of ice crystals formed where the water had clustered. The quantity of clustered water was calculated from the enthalpy of fusion of frozen water. Attempts to see the microdroplets under a microscope proved unsuccessful, although clusters 2 μm in diameter were detected when plaques were subjected to high-pressure steam and quenched in water at 23°C. This supports the suggestion that

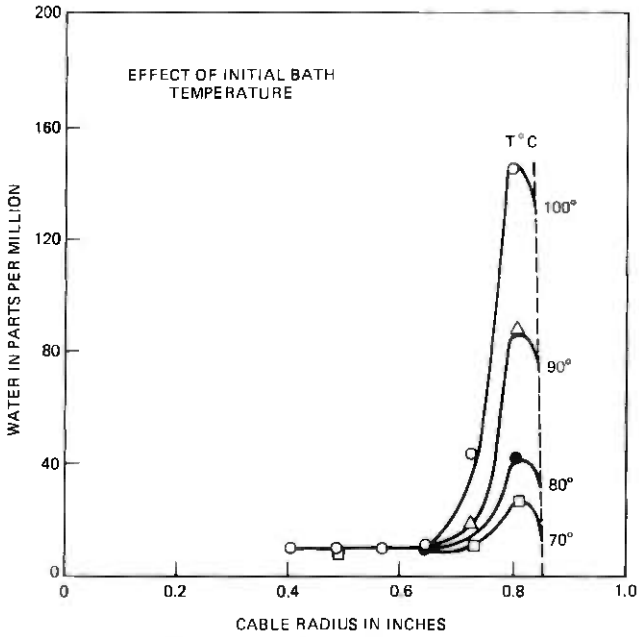


Fig. 10—Distribution of water through an SG cable core.

the microdroplets produced under normal, less severe conditions are smaller than 1 μm.

The techniques described above were used to determine the quantity of both the total and clustered water in samples saturated at temperatures between 4° and 100°C, with samples saturated above 23°C, then quenched in water at 23°C (see Fig. 11). Within the limits of detection (10 ppm), no clustered water melting at 0°C was observed in samples saturated at or below 50°C.

5.2 Exploration of water trapped during core extrusion

The polyethylene core is extruded at about 190° to 200°C and is then cooled by traversing a series of water troughs starting at temperatures of 80° to 100°C and descending in temperature to 20°C. When the molten core enters the first trough, water vapor diffuses into the polyethylene to produce a concentration of 100 to 200 ppm near the surface, reducing to about 10 ppm at a depth of 0.2 in. As the temperature of the polyethylene drops below approximately 105°C, crystallization starts and it is hypothesized that noncompatible materials (polar groups, antioxidant, and water) are swept ahead of the crystallizing surfaces. When the temperature decreases further, the clustering of these impurities triggers the precipitation of globules of water near the outer surface of the core. The globules contain traces of water-soluble salts and antiox-

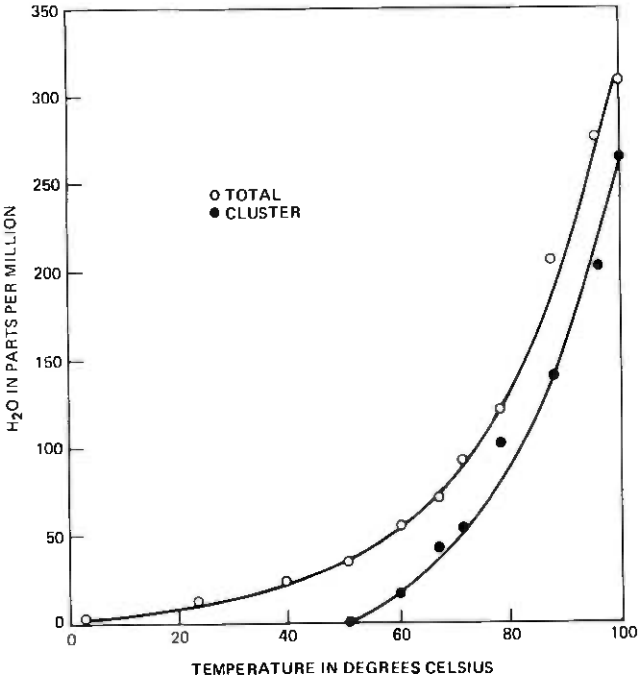


Fig. 11—Saturation content in SG cable.

idant and attract to the polyethylene-water interface any polar groups in the vicinity. Thus, the water droplet conditions its surroundings and produces some small water-filled voids with hydrophilic surfaces. The diffusion coefficient drops sharply with temperature, and very little of the precipitated water can escape from the cable before it reaches room temperature. When the core is exposed to the atmosphere prior to sheathing, the trapped water slowly dries out and if it is left for longer than one week in the air, it will usually recover to a low water content. However, many of the voids will be left with hydrophilic surfaces so that, if exposed to water on the sea bed, they can refill, causing changes in the dielectric loss.

The standard procedure of shaving 0.070 in. (1.78 mm) from the outside of the core removes the more heavily contaminated material but is insufficiently deep to remove all the damaged polyethylene.

5.3 Effects of water on dielectric loss

To determine the effects of water on the dielectric loss of the polyethylene core, numerous experiments have been carried out both on cable and on plaques simulating cable core. It was found that, although water in the vapor phase has a small effect on the dielectric loss below the microwave range, microdroplets as small as 0.1 μm in diameter can have

considerable influence on the loss in the megahertz region. They act as small regions of high permittivity and finite resistivity dispersed in a low-loss dielectric medium. The dielectric response of such a system is governed by relaxation equations and in the simplest case will exhibit a single Debye-type peak whose frequency is mainly dependent on the permittivity and resistivity of the small regions. In the case of the microdroplets, the resistivity of the water will be modified by the presence of soluble impurities such that the peak occurs in the megahertz range. The general behavior is known as the Maxwell-Wagner effect. In practice, there may also be a small increase in loss in the megahertz region due to water in the vapor phase causing an enhancement of the gamma process,[2] which is centered around 1 GHz in polyethylene.

Typical behavior is illustrated in the results of the following experiment, shown in Fig. 12. To simulate the conditions in the outer layer of the core immediately after extrusion and cooling, a series of plaques (2 mm thick) was molded at 180°C. The molten plaques, attached to metal backing sheets, were then dropped into baths of water at 95° to 70°C for 30 minutes. The dielectric loss was measured at frequencies up to 30 MHz as a function of drying time at 23°C and 35-percent relative humidity. The effect of the water on the polyethylene was found by normalizing the times and plotting the excess loss over that for a normal molded plaque cooled in air. A typical example simulating a first cooling trough at 95°C is given in Fig. 12; the initial effect is large, with a peak

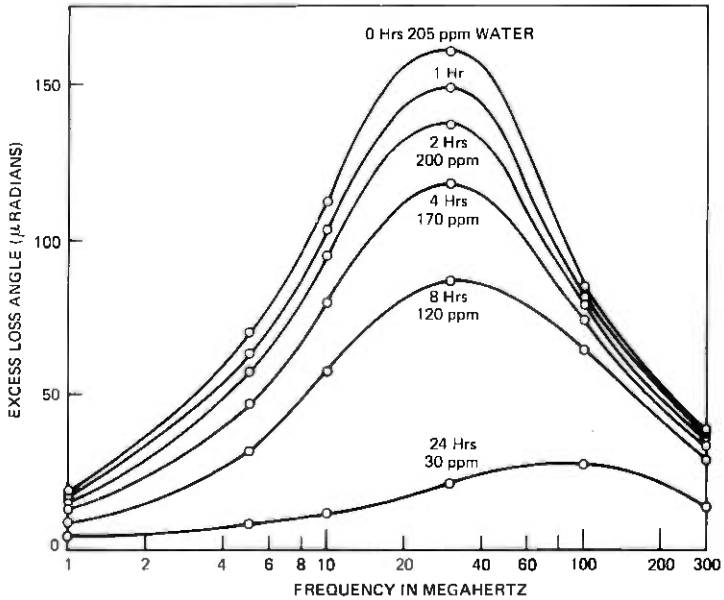


Fig. 12—Increase in loss angle caused by immersing molten plaques in water at 95°C for 30 minutes, plotted as a function of drying time.

excess loss of approximately 160 μrad at 3 MHz for a water content of 245 ppm. The effect decays quite fast so that, after eight hours, the peak excess loss is down to approximately 85 μrad; after twenty-four hours, the peak has shifted up to 8 MHz and is approximately 25 μrad high for a water content of 30 ppm. The experiments on plaques highlight the effect of clustered water on polyethylene, but luckily the influence on the overall attenuation of finished cable is heavily diluted by a number of factors.

The behavior of plaques represents only the outer layers of the core. The inner layers are free from clusters and contain water to the normal saturation level of about 10 ppm. This level remains substantially constant while the outer layer dries down. So whereas the effect of water on the dielectric loss of the outer layer of freshly extruded core is initially very large, the rest of the core is scarcely affected, and the dielectric loss integrated over the whole cross section shows only a small increase. As a rough rule of thumb, the overall change in dielectric loss integrated over the entire SG cable core is about $\frac{1}{10}$ the excess loss of a plaque simulating its outer layer. Taking as an example the results of the plaque experiment given in Fig. 12, the plaque initially exhibited an excess loss of 37 microradians at 30 MHz, decreasing to 12 microradians after 24 hours. The dielectric loss of an equivalent SG core would be increased initially by 3.7 microradians; this would then drop to 1.2 microradians.

In terms of overall cable attenuation, the effect is further diluted by the fact that, at low frequencies, the dielectric loss contributes only a very small proportion of the total loss; even at 30 MHz it contributes only 6 percent. Even so, a 1-μrad change at 30 MHz is equivalent to 26-dB change in an overall system of 3400 miles which, if it occurred on the sea bed, would require that the system be re-equalized at several points along its length.

VI. CABLE AGING

The effects of water introduced during the extrusion process and the changes in loss observed on cable lengths monitored in the factory raised questions about the long-term loss stability of the cable in service. Prior to laying the cable, there were only conjectures of the likely behavior on the sea bed but, in spite of inadequate evidence, it was essential to make some estimate of the possible changes in loss over the 20-year life of the system so that an adequate equalization strategy could be formulated. The only piece of evidence available from measurements on sample cable was that a piece of SG sea-trials cable, laid at 1500 fathoms for 2 years, had aged by 1 μrad at 30 MHz. Using this information and results based on other work, it was predicted that the dielectric loss of the SG cable system would age by −1 to +3 μrad in 20 years. (That is, the uncertainty embraced possibilities ranging from a 1-μrad *decrease* to a 3-μrad *increase* in 20 years.)

As predicted, the TAT-6 system is slowly aging on the sea bed: 346 days after commissioning, the phase delay had increased by about 3.5 μs and 583 days after commissioning the attenuation of the deep-water sections had increased by 21.6 dB at 27.5 MHz. The change of attenuation with frequency is approximately linear. The Requirements and Performance paper in this issue of the B.S.T.J. discusses the time behavior of the observed aging and the use of shore-controlled equalizers to compensate for it.

It is hypothesized that water is entering the TAT-6 cable on the sea bed at the repeater terminations and is traveling along the cable causing compression of the core and expansion of the sheath. The resulting sea-water film between the core and outer conductor acts as a lossy dielectric in series with the polyethylene. This, together with the dimensional changes of the cable structure, will cause an increase in attenuation and a change in phase delay.

It is further hypothesized that, in addition to aging from this cause, water in contact with the core will gradually diffuse into the polyethylene and refill some of the dried microvoids. The rate of diffusion will be very slow at deep-sea pressure and temperature. To confirm that polyethylene with dried cluster sites will reabsorb water, laboratory experiments were carried out on a series of plaques similar to those described in Section 5.3. The plaques were allowed to dry under laboratory conditions for a week (simulating the average time between extruding and sheathing the core), by which time their water contents were less than 10 ppm. They were then exposed to approximately 98-percent humidity, and the resulting changes in loss angle as a function of exposure time were plotted for frequencies up to 30 MHz. The results for a plaque initially conditioned in 95°C water are given in Fig. 13. After exposure for 28 days, the plaque had about 60 ppm of water in it. The experiments are still continuing.

The results have been expressed in Fig. 14 in terms of overall attenuation for a complete cable length of 3400 nmi. The curves illustrate how aging from this cause initially might appear linear with frequency and subsequently change shape. In the case of the cable on the sea bed, the effect of exposure to water will proceed at a considerably slower rate than that observed in the experiments. The rate will be limited by the construction of the cable, the longer diffusion paths, and the effects of increased pressure and reduced temperature. Attempts have been made to account for these effects; using data from experiments on short cable samples, they predict similar shapes but about $\frac{1}{3}$ the magnitude of the aging shown. Reabsorption of water into cluster sites will not change the phase delay significantly.

While neither of the two aging mechanisms described above can individually account for all aspects of the observed cable changes, com-

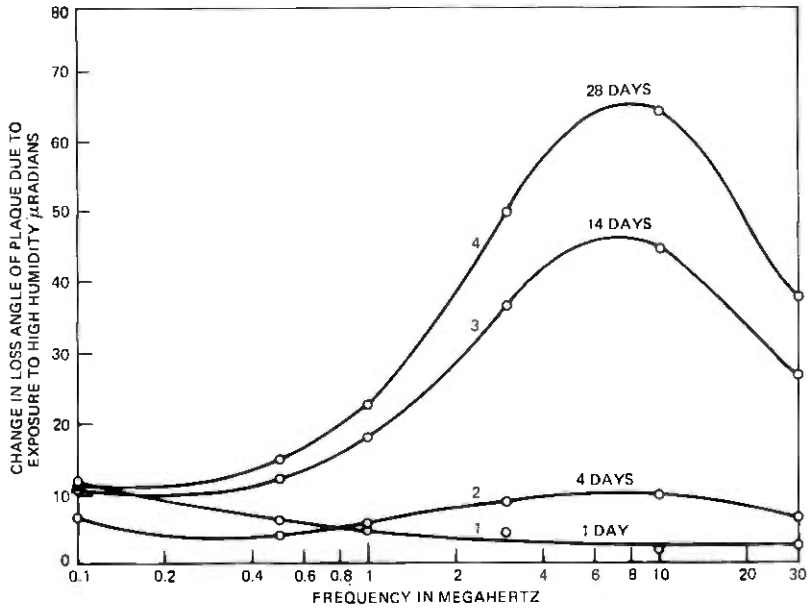


Fig. 13—Change in loss angle of plaque on exposure to high humidity after drying for one week.

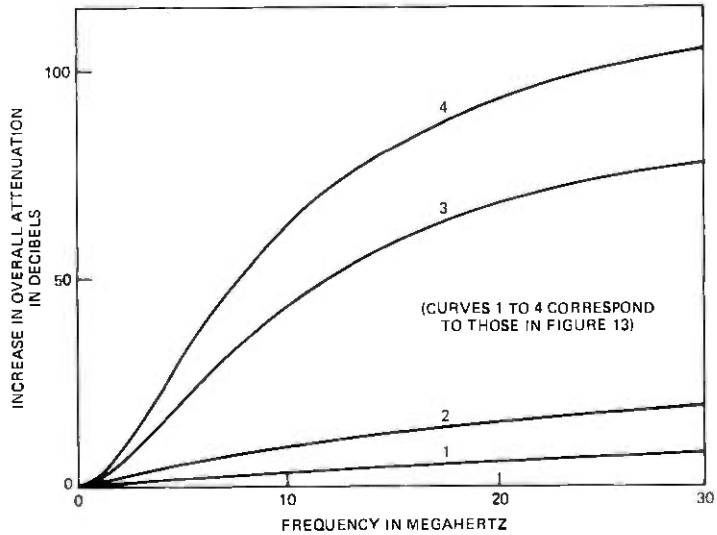


Fig. 14—Calculated attenuation changes in a 3400-nmi 1.7-in. SG cable assuming dielectric changes due to water.

binations of the two mechanisms can be made to fit the data, including the rate of loss change with time. However, the long-term aging prediction cannot be confirmed without better knowledge of (*i*) the relative movements of core and outer conductor on the sea bed and (*ii*) the effects

of reabsorption of water on the dielectric loss under deep-sea conditions.

VII. LAYING EFFECT

Laying effect is the difference between predicted and actual sea bed transmission losses at the time of lay. Predictions are based on cable and repeater factory measurements that are corrected for estimated sea-bed conditions. A non-zero laying effect can result from errors in factory or shipboard measurements, from cable aging during the period between these measurements; from errors in cable temperature, pressure, and handling coefficients, and from differences between predicted and actual sea-bed conditions at the time of lay. It can also be caused by imperfect knowledge of repeater and equalizer termination losses and impedance discontinuities at the cable-pigtail and the repeater-pigtail interfaces.

In TAT-6, the observed laying effect was remarkably consistent among the several cable lays. It was possible to differentiate between shallow and deep water lays and between cable from different manufacturers. A typical result is shown in Fig. 15 for the first deep water lay. Laying effect was essentially zero at 30 MHz but reached + 0.45 percent at the bottom of the low band. This may be interpreted as a constant cable shape component of + 0.45 percent, which is offset at 30 MHz by a dielectric loss angle at sea bottom which is 4 μrad higher than predicted.*

The average cable shape and linear-loss components of the deep-water laying effect were, from Table III, +0.42 percent and + 4.4 μrad, respectively. Because these errors cancelled at 30 MHz, the resulting misalignment remained well within the equalization range of the OBE but their magnitude is sufficient to warrant an explanation. The difference between the design and the mean sea trials cable pressure coefficient, shown in Fig. 7, suggests a starting point. At 1 MHz, this is +0.16

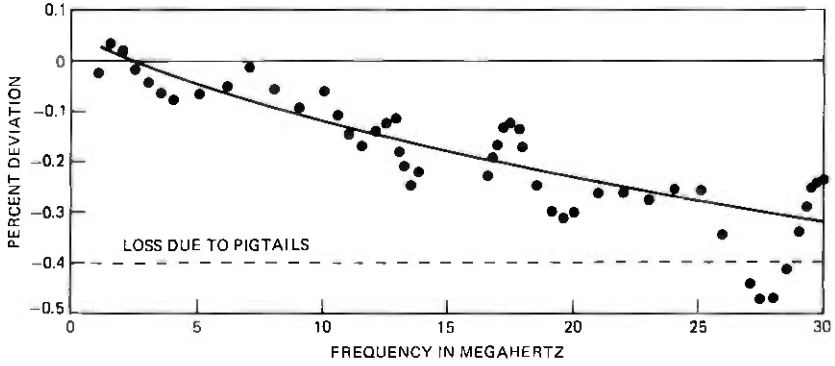


Fig. 15—Lay 1 (deep water) laying effect including 0.4 percent pigtail loss.

* Positive laying effect corresponds to gain relative to prediction.

Table III — TAT-6 cable laying effect

Lay	Cable Length and Supplier (nmi)	Mean Annual Sea-Bed Tempera-ture (°C)	Mean Sea-Bed Depth (kF)	Cable Shape Laying Effect (%)	Linear Loss Laying Effect at 30 MHz (μrad)
Green Hill burial	111 STC/SWC*	10.4†	0.045	+0.30	+1.5
Lay 1	605 STC	2.6	2.19	+0.45	+4.0
Lay 2	634 CDL	2.3	2.67	+0.34	+4.0
Lay 3	640 STC	2.3	2.39	+0.47	+5.5
Lay 4	640 STC	2.9	1.83	+0.51	+4.5
Lay 5	640 STC/CDL	2.6	2.45	+0.31	+4.0
St. Hilaire burial	114 STC	12.8	0.068	+0.44	+2.0

* Cable manufactured by STC and armored by Simplex Wire and Cable Co.
† Actual temperature during the Green Hill burial operation is thought to have been about 2°C higher than annual mean.

percent per kilofathom, enough to explain +0.37 percent of the cable-shape laying effect in deep water. The difference in slope is equivalent to 1.3 μrad per kilofathom at 30 MHz, enough again to account for 3 μrad of the deep water linear-loss laying effect. This leaves only 0.05 percent cable shape and 1.4 μrad of excess loss angle to account for. Together, they correspond to a misalignment of less than ±0.1 percent across the band.

Any further attempts to explain the remaining laying effect are probably not justified by the accuracy of measurement that was possible. However, the loss-angle component could be accounted for by in-factory aging. At least half the cable used in TAT-6 was extruded and cooled under conditions which could cause an increase in loss angle of 2 μrad after 6 months' storage (see Fig. 9). The remaining cable shape effect could also be accounted for quite easily by an error in the air-gap coefficient. Analysis of factory transmission measurements, including study of any change due to cable turnovers, suggests that the average air-gap at 10°C was less than the figure of 0.002-in. (0.051 mm) adopted for the final design. Assuming there is no air gap present when the cable is on the sea bed, then the change in air gap from ship to sea bed will be less than expected and could reduce the cable-shape laying effect by as much as 0.2 percent.

VIII. CABLE TERMINATIONS

Terminations, frequently called couplings, are used on each end of each cable section to interface the cable to the repeater or equalizer. They must perform the following functions:

(*i*) Transfer tensile loads from the strength member of the cable to the repeater or equalizer end cone.

(*ii*) Provide a signal path between the cable and the coaxial pigtail wire which connects the termination to the repeater or equalizer.

(*iii*) Provide sufficient flexibility to permit handling over 9-ft diameter drums and sheaves on cable ships.

(*iv*) While performing the functions above, be capable of withstanding full sea-bottom pressure and the full system dc operating voltage.

Terminations used on armorless cable differ somewhat in requirements and configuration from those used on armored cable, so it is useful to discuss them separately.

8.1 Terminations for armorless cable

The 8G coupling is used on armorless cable. The specific performance requirements to achieve the functions listed above are:

(*i*) 40,000-lb_f (177.9 kN) tensile capability.

(*ii*) 600 in.-lb_f (67.8 m-N) torque capability at maximum tensile load.

(*iii*) At least 35-dB structural return loss from each end.

(*iv*) Capability to allow at least a 45-degree angle between the cable and the repeater or equalizer axis.

(*v*) 7000-V dc capability at both possible polarities with no breakdown and no noise generation in the transmission band.

(*vi*) Capability to withstand 12,000 pounds per square inch (816 atmospheres) sea pressure.

Figure 16 illustrates the design of the 8G coupling. The center strand strength member of the cable is terminated in epoxy in the termination cone. This in turn is glued to a ceramic ring, and the entire assembly is encapsulated in polyethylene. This molded unit is then placed inside the cone housing which attaches to the gimbal mechanism. The gimbal in turn attaches to the gimbal housing and clamp ring, which threads onto the end cone of the repeater or equalizer. In this way, a tensile load in the cable strand is transferred to the repeater or equalizer housing.

Two half-slots are formed in the polyethylene which encapsulates the strand termination. Two keys, part of the cone housing cap, engage these slots to prevent rotation of the anchor mold assembly and provide the required torque capability.

The molded polyethylene layer between the ceramic of the termination assembly and the cone housing is dimensioned to be thin enough to

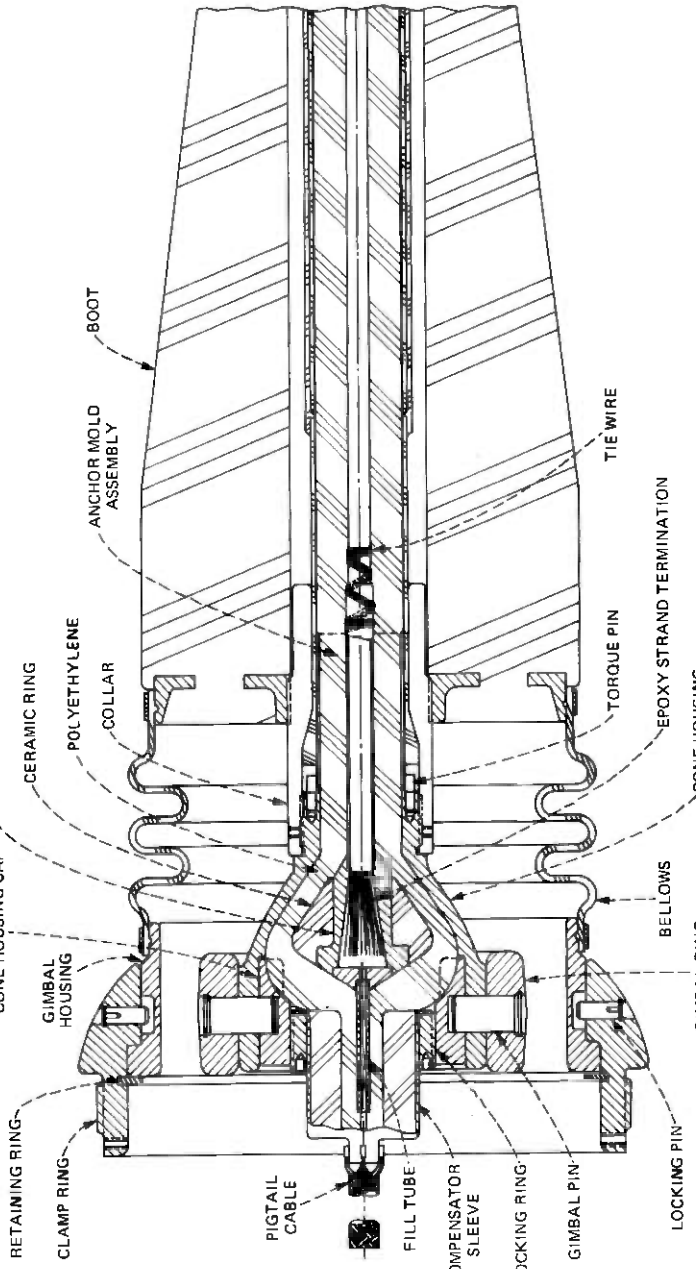


Fig. 16—SG armorless cable coupling.

support the required tensile load and thick enough to provide the high-voltage capability stated above. Within these constraints, it was not possible to achieve the required return loss in the anchor mold assembly as molded. The structure showed excessive capacitance with respect to 50 ohms, so more inductance was needed. To provide this, the large polyethylene sleeve was added at the end of the anchor mold assembly.

The tube on the end of the termination assembly opposite the strand is used for introducing the terminating epoxy into the cone; thus, it must be kept open until the termination is attached to the cable. Since the anchor mold assembly is provided to the cable factory in overmolded condition, the tube must extend outside the molded volume. For this reason, the compensating sleeve could not be made a part of the anchor mold assembly. Instead, it is a separate part which is added after the epoxy termination is complete. The inner conductor of the coaxial pigtail cable is crimped and soldered into the end of the tube and the joint overmolded.

On the opposite end of the anchor mold assembly, the inner conductor of the cable is attached to the conductor tube of the assembly by three turns of three wires to allow for some axial motion under load without loss of continuity. The outer conductor is dimensioned to maintain the 50-ohm cable characteristic impedance up to the termination assembly.

By careful impedance control where possible and by use of the compensating sleeve, the return loss requirement was satisfied.

The gimbal mechanism is so designed that it provides slightly more than 45 degrees of motion between the cable and the repeater or equalizer axis. The bellows is provided to accommodate the gimbal motion while keeping foreign material out of the gimbal mechanism.

The entire termination outside the molded polyethylene is open to sea water, so the sea pressure produces only an additive hydrostatic compression in the various parts. Similarly, the molded assembly is essentially solid so that sea pressure has no detrimental effect.

All the metal parts are made from either copper or copper-beryllium to prevent electrolytic corrosion in sea water.

The large rubber boot serves to limit cable bending radius at the termination and to help guide the whole assembly through the cable machinery on the cable ship.

8.2 Terminations for armored cable

Requirements for the couplings for armored cables differ from those stated above primarily in the tension and torque values. Both the 1-inch (25.4 mm) shielded armored and the 1.7-inch (43.2 mm) single-armored cables have strengths well in excess of 100,000 lb_f (445 kN). After careful

analysis, it was decided that the kind of coupling structure necessary to make the coupling as strong as the cable was unreasonable from both a fabrication and a cost standpoint. Consequently, it was decided to limit the coupling capability to:

(*i*) 100,000 lb_f (445 kN) tension.

(*ii*) 25,000 $lb_f \cdot$in. (2825 N-m) torque.

In addition, in the couplings for armored cable, the tensile load must be carried from the steel armor wires to the copper-beryllium repeater housing. To prevent corrosion in sea water, special attention must be paid to electrically isolating these two materials from each other at some point while maintaining the tension and torque-carrying capability.

Figure 17 shows the 8N coupling which is used with 1.7-in. (43.2 mm) single-armored cable. The 8P coupling, which is used with 1.0-in. (25.4 mm) shielded-armored cable, is very similar to the 8N, differing only in the details in the round nut assembly that must change with cable diameter.

As in previous cable systems, double-armored cables are terminated as single-armored, the outer armor layer being lashed off outside the coupling.

The strength termination is achieved by crimping a ferrule onto the end of each armor wire. The terminated wires are then placed into the slots of the armor ring against which the ferrules bear. The load is transferred by a yoke arrangement through plastic insulators from the steel armor ring to the copper-beryllium armor housing. The plastic insulators prevent electrolytic corrosion between the dissimilar metals.

To carry the required torque, the armor housing is threaded into a copper-beryllium bushing and secured with a key in the threads. The outside of the bushing is hexagonal in shape and mates, through a plastic insulator, with the hexagonal interior profile of the steel armor ring housing. Similarly, the interface surface between the armor ring housing and the armor ring is hexagonal for torque transfer.

Since most of the strength is in the armor wires, a strength termination is not required for the strand. A direct connection is thus made between the cable inner conductor and the coaxial pigtail inner conductor. The joint is overmolded to provide a transition from the cable to the pigtail outer diameter. The round nut assembly provides a connection between the cable outer conductor and the special outer conductor insert designed to accommodate the diameter transition. This transition joint provides the required return loss.

Other features of the armored coupling such as the gimbal, the bellows, and the rubber boot serve the same functions as described in connection with the armorless cable coupling.

In summary, using the designs described above, the 8G, 8N, and 8P couplings met or exceeded all design requirements.

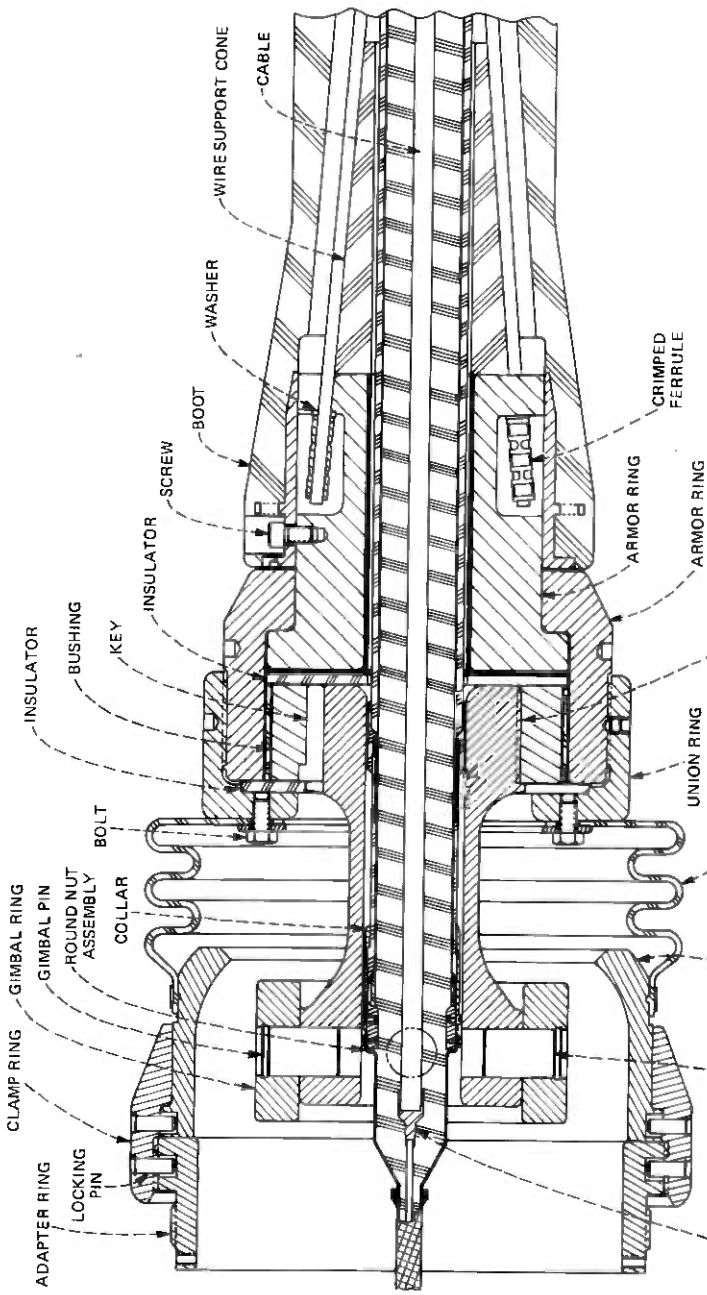


Fig. 17—Typical SG armored cable coupling.

8.3 Pigtail cable

The coaxial pigtail cable mentioned above, which connects the terminated cable to the repeater, has an outer diameter over the insulation of 0.500 in. (12.7 mm) and an inner conductor diameter of 0.146 in. (3.71 mm). These dimensions, used in standard formulas, do not provide a 50-ohm structure. The outer conductor is a copper braid with 0.020-in. (0.51 mm) diameter wires. These wires are much larger than those used in standard braids (typically, 0.007 in. (0.18 mm) diameter), but are required to allow for some loss of material due to corrosion while maintaining continuity. Because these larger wires are used, the outer conductor current is effectively approximately 0.010 in. (0.25 mm) outside the outer surface of the polyethylene insulation. The 0.146-in. (3.71 mm) inner conductor compensates for this effective increase in outer diameter. The braid is covered with heat-shrinkable tubing to hold it in place and to limit water flow for corrosion protection.

When the pigtail is placed in sea water, the conductivity of the water causes a small reduction in the effective coaxial outer diameter, which changes the characteristic impedance of the structure. This change must be accounted for when correlating "wet" and "dry" transmission measurements.

IX. ACKNOWLEDGMENTS

Many individuals within Bell Laboratories and the British Post Office contributed to the development of SG cable. Their efforts are gratefully acknowledged here. In particular, mention must be made of the considerable efforts made by E. F. S. Clarke and E. E. L. Winterborn (both now retired) who were successively responsible for coordinating the development program.

Acknowledgment is made to the Director of Research of the British Post Office for permission to publish this paper.

REFERENCES

1. "SF Submarine Cable System," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 601–825.
2. N. G. McCrum, B. E. Read, and G. Williams, *Anelastic and Dielectric Effects in Polymeric Solids*, New York: Wiley, 1967, Chapter 10.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Terminal Transmission Equipment

By M. BROUANT, C. CHALHOUB, P. DELAGE, D. N. HARPER, H. SOULIER, and R. L. LYNCH

(Manuscript received October 7, 1977)

Terminal transmission equipment has been developed for the SG cable system that, compared to similar equipment for land analog facilities, makes more efficient use of the transmitted bandwidth and provides superior continuity of service. Although economy of development effort dictated the use of existing equipment and conventional design techniques wherever possible, many new designs were required to achieve the desired features. Means for protecting the undersea repeaters from gross signal overload and a carrier generating scheme for hypergroup and wideband-line frequency-translation equipment are examples of new designs that proved particularly challenging. Duplication of higher-level multiplex equipment coupled with automatic changeover switching is the means used to obtain the desired service continuity. In this regard, much care was required to limit amplitude and phase differences between duplicate transmission paths to insure "hitless" maintenance changeover switching. Other features of the terminal include an on-site equalizer design and construction capability at the wideband line and supergroup levels, a separate 3-channel order-wire facility, individual supergroup signal limiters, and the ability to operate with 3-kHz as well as 4-kHz/spaced message channels. Mechanical design follows the current French standard for such equipment. CIT-Alcatel in France carried out the detailed design.

I. INTRODUCTION

The terminal transmission equipment (TTE) for the SG Undersea Cable System provides the vital interface between the inland telephone network and the undersea link. Interconnection to the domestic network is at the basic supergroup level; thus, the TTE consists of *frequency-di-*

vision multiplex and *wideband line equipment* (including a *directional filter* that provides a physical 4-wire to equivalent 4-wire conversion) (Fig. 1). The transmitting multiplex serves to frequency-translate fifty-plus basic supergroups, combining them to form what we call the "baseband." The receiving multiplex performs the inverse operation.

The wideband line equipment provides its customary functions of signal preemphasis and deemphasis, equalization, level adjustment, and the like. Following the conventional equivalent 4-wire undersea cable terminology, the terminal stations at opposite ends of the system are designated A and B. Low-band transmission in the A-to-B direction is at baseband frequencies directly. In the opposite direction, the high band is formed by one further stage of frequency translation of the baseband, which takes place in the wideband line equipment.

The principal system parameters influencing TTE design are summarized below:

(*i*) Nominal frequency band (corresponding to 4000 3-kHz voice channels):
Low band 800 to 13,300 kHz,
High band 16,700 to 29,300 kHz.

(*ii*) Possible extension to:
Low band 500 to 13,900 kHz,
High band 16,100 to 29,500* kHz.

(*iii*) Operation with either 3- or 4-kHz voice channels, or a combination of the two.

(*iv*) Design load per channel: −13 dBm0.

(*v*) Terminal noise contribution (both stations): No more than 250 pW0p (24.5 dBrnc0) in the worst channel, with the system fully loaded and with all transmission levels at their nominal values.

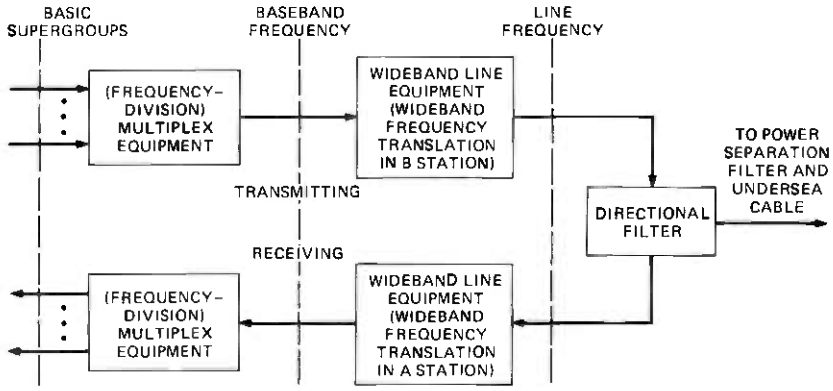


Fig. 1—Summary diagram of the SG terminal transmission equipment.

* Experience gained with the first SG system installation, TAT-6, has led to an upward translation in the extended high band by about 500 kHz (Fig. 4b).

Of course, 3-kHz spaced channel operation is common on long-haul undersea cable routes. It has also been traditional that undersea system design loads be higher than those for similar inland carrier systems for, among other reasons, the potential application to them of speech concentrators such as TASI or CELTIC.

The reason for nominal and extended frequency bands is quite simple. During development, it was not possible for us to be sure of the exact transmission band that would ultimately be achieved. This depends on the degree to which repeater gain can be trimmed to match cable loss at the band edges, and on the degree of success of mop-up equalization in the ocean-block equalizers.[1] We chose to design for the extended bands to encompass the largest bandwidth that optimistically could be achieved, so that the TTE would never restrict channel capacity of the overall system.

Conceptually, we were guided in the design of the TTE for the SG system by two main principles:

(*i*) The available frequency band should be used to the fullest.

(*ii*) The failure of an active device should not cause a loss of service of more than the equivalent of one supergroup.

In addition to the capability of 3-kHz operation, the first principle resulted in a unique multiplex frequency allocation that, in turn, required development of new modulation equipment. The second principle, one of reliability, we achieved largely through automatic protection switching between duplicated equipments.

II. FREQUENCY ALLOCATION

The frequency plan for the SG system was chosen primarily to achieve efficient use of the available transmission band and secondarily to minimize the degree to which standard multiplex equipment and techniques of other undersea and inland systems would have to be modified. For example, the ease with which new carrier frequencies could be produced was a factor in our plans. From the outset, our desire to achieve a high traffic-band efficiency factor (defined as the ratio of the actual frequency spectrum devoted to traffic to the total transmission band) forced us to reject standard CCITT* or Bell System frequency plans developed for inland carrier systems of comparable or greater bandwidth than the SG baseband, e.g., 12- and 60-MHz CCITT international systems and the L4 and L5 systems.[2,3] The traffic-band efficiency factor for the 12- and 60-MHz terrestrial systems is 0.89 and 0.78, respectively, and

* The International Telegraph and Telephone Consultative Committee.

that of the basic Bell System mastergroup itself is only 0.87, whereas that for the SG plan is greater than 0.95.*

2.1 Baseband assignment

The SG baseband comprises six *hypergroups* (HG) numbered 1 through 6 in ascending frequency, as shown in Fig. 2. The term *hypergroup* is British Post Office (BPO) terminology that refers to an arbitrary number of supergroups† assembled in a contiguous frequency band. Hypergroup 1 is direct-formed by supergroup translating equipment (Fig. 3). Positions of the 14 supergroups are those of supergroups 3 through 16 of the standard CCITT 4-MHz system, or equivalently, those of the CCITT basic 15-supergroup assembly.[4–6] Hypergroups 2 to 6 each contain 10 supergroups and together occupy the band 4032 to 16,424 kHz. They are formed in two steps of frequency translation. First, a *basic hypergroup* is formed by supergroup translation (CCITT supergroups 4 through 13), and then by hypergroup translation (Figs. 3 and 2, respectively).

Note that all the carrier frequencies are odd multiples of 124 kHz, a desirable feature from the point of view of carrier generation. Thus, the conventional 8-kHz gap between supergroups within hypergroups, which permits through-supergroup connection, is preserved between hypergroups as well. One exception is the HG 1–HG 2 gap of 4 kHz, an offset

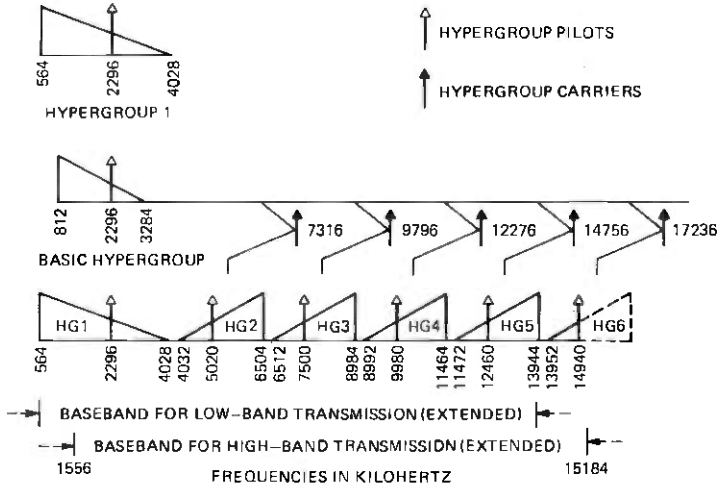


Fig. 2—Frequency allocation—SG basebands.

* At present, the line-to-terminal cost ratio of terrestrial carrier systems relative to their undersea cousins naturally leads to frequency allocations that, while not optimum in traffic-band efficiency, nevertheless achieve an overall efficiency due to less severe terminal filtering needs which facilitate pilot monitoring, automatic gain regulation, and circuit distribution (e.g., blocking and reinserting of large bundles of voice circuits).

† Each supergroup corresponds to 60 4-kHz or 80 3-kHz voice channels.

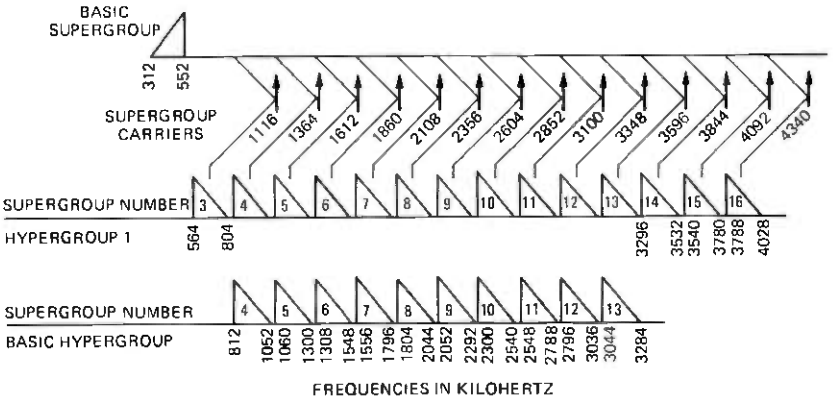


Fig. 3—Frequency allocation—basic hypergroup (and hypergroup 1).

needed to retain the 124-kHz integral multiple characteristic of hypergroup carriers. A secondary advantage of this choice is that all hypergroup carriers fall between voice channels for either 3- or 4-kHz spacing, thereby easing carrier leak requirements.

Several comments are in order concerning the makeup of hypergroups. First, we did not retain supergroup 3 in the basic hypergroup primarily to ease design complexity of, and reduce group-delay distortion caused by, the hypergroup modulator band filter, which must reject the carrier and upper sideband while passing the lower sideband. Next, we included supergroups 14 to 16 in HG 1 (but not in the basic hypergroup) to ease design of the HG 2 modulator band filter, which must reject the basic hypergroup band (because of leakage across the modulator) in addition to carrier and upper sideband. Furthermore, our choice of a 10-supergroup basic hypergroup limits to five the number of new designs of hypergroup translation units.

Although we have spoken of the SG baseband, one can, in fact, distinguish two slightly different spectra as evidenced in Fig. 2. The low-band baseband is simply hypergroups 1 through 5, whereas the high-band baseband includes hypergroups 2 through 5 plus part of hypergroups 1 (supergroups 7 through 16) and 6 (supergroups 9 through 13). The latter allows sufficient separation between the wideband-translation carrier frequency of 31,620 kHz and the upper edge of the high band at line frequency so that design of the associated modulator band filter is easily realizable. Note that this carrier is also an odd multiple of 124 kHz.

In summary, we believe that the SG frequency allocation is a good compromise between the objectives of efficient use of bandwidth and minimum development of new frequency-translating and carrier-generating equipment.

2.2 Order-wire and supervisory-tone band assignment

The order-wire and repeater supervisory-tone band assignments[1] are shown in Figs. 4a and 4b, for the low and high bands, respectively. One can see that both assignments fall within those of the multiplex, so that particular channels in hypergroup 1, supergroup 16, i.e., (1, 16) and (4, 5) in the low band and (1, 16) in the high band, cannot be used for commercial traffic.

2.3 Pilot assignment

A *hypergroup pilot* is inserted at the input of each transmitting hypergroup at a frequency of 2296 kHz, midway between supergroup 9 and 10, to facilitate measurement. (Line-frequency assignment of these pilots is given in Figs. 4a and 4b.) Due to its near-central location within each

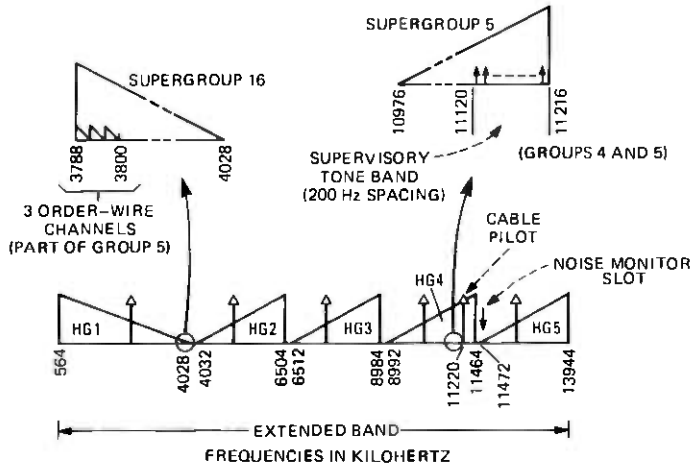


Fig. 4a—Frequency allocation—low band.

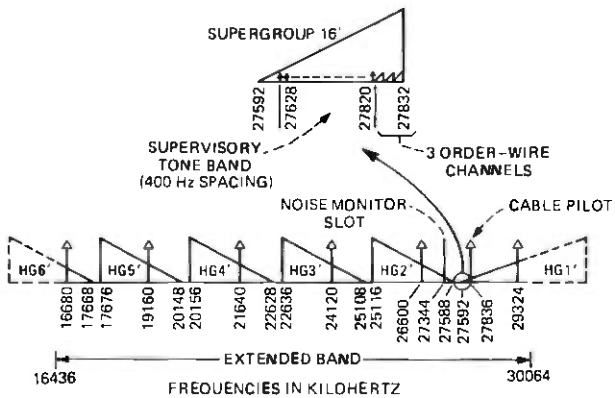


Fig. 4b—Frequency allocation—high band.

hypergroup, it should be representative of the overall transmission level of the hypergroup.

There are no separate assignment of *line* (or *system*) *pilots*, i.e., those which provide information about the behavior of the wideband link (wideband line plus undersea). Instead, the HG 2 and 5 pilots provide this function by being monitored at the input of the transmit and output of the receive wideband line equipment. This is illustrated in Fig. 5.

One additional pilot in each direction of transmission, the *cable pilot*, is inserted and measured as close as practical, from a transmission standpoint, to the undersea system to furnish information about the behavior of only this part of the wideband link. The cable-pilot frequency in the low and high bands was chosen to be near that of the supervisory tones and to fall between supergroups.

The power level of all pilots is −20 dBm0.

III. CONFIGURING THE TTE FOR RELIABLE OPERATION

3.1 Reliability criterion

As mentioned before, the layout of the TTE is such that failure of an active component will not involve loss of more than the equivalent of one supergroup of traffic (80 3-kHz or 60 4-kHz channels). The object of this section is to examine the ramifications of this principle on the configuration of the TTE; in particular, the extent of equipment duplication and automatic changeover switching.

3.2 Features common to transmit and receive

Generally speaking, the supergroup equipment itself (including supergroup signal limiters, modulators, and equalizers, all of which are discussed later) need not be duplicated to meet the above criterion. On the other hand, failure of a supergroup carrier would affect five or more supergroups. Thus, supergroup carrier generators are completely duplicated with automatic protection switching at the outputs for each in-

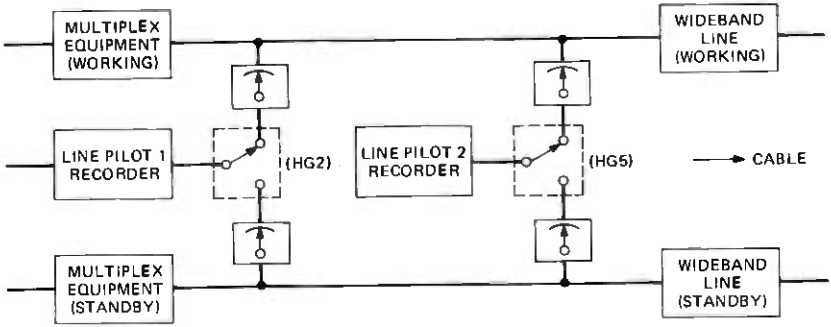


Fig. 5—Line-pilot monitoring (transmit and receive).

dividual carrier. Of course, duplicate sources of the master oscillator frequency of 124 kHz are required.

Referring to Fig. 6, we see that the transmission path between supergroup couplers* and the directional filter is duplicated. In the transmitting direction, this involves hypergroup translation and wideband line equipment. The same is true in the receiving direction but with the addition of hypergroup regulators. The duplicated paths in each direction are called *Path I* and *Path II*. At any one time, the particular paths that connect the directional filter to the supergroup equipment are referred to as *working* and the others as *standby*. If a working path fails, the two are interchanged.

Naturally, hypergroup and wideband translation carrier generators are duplicated, and are referred to as *generator* 1 and *generator* 2.† Connection to the translating equipment is not the same as to the supergroups, because hypergroup and wideband line equipment is itself duplicated and protected by automatic changeover switching. Duplicate generators for each hypergroup carrier can be connected to the transmission paths in the combinations shown in Table I by means of patch links.

The 2296-kHz hypergroup pilot supply is duplicated. Pilot supplies are connected to the transmission paths by means of patch links that allow combinations similar to those in Table I.

The directional filter unit (which includes the receiving path-splitting coupler) is passive but, because a failure here could affect the entire wideband path, we considered it prudent to provide a "built-in" spare which can be quickly interchanged manually for the working unit, albeit not without momentarily interrupting traffic.

3.3 Special transmitting features

Working and standby hypergroup pilots on the transmitting side are monitored at the output of the wideband line equipment just after the changeover switch but prior to the directional filter (Fig. 6). The changeover logic control recognizes pilot *failure* as a change in amplitude from nominal of between 2 and 3 dB. Failure of a working pilot initiates a path changeover provided there is not a concurrent standby pilot failure. Additionally, an alarm inhibit capability is provided on an individual hypergroup pilot basis which, if operated, removes that particular pilot from consideration by the changeover logic.

The cable pilot is introduced throuh a passive coupler after the transmit changeover switch but prior to the directional filter (Fig. 6),

* The couplers themselves are passive and therefore not duplicated.

† Roman numerals I and II are reserved for the duplicated transmission paths protected by automatic changeover switching. All other duplicated equipment is designated 1 and 2.

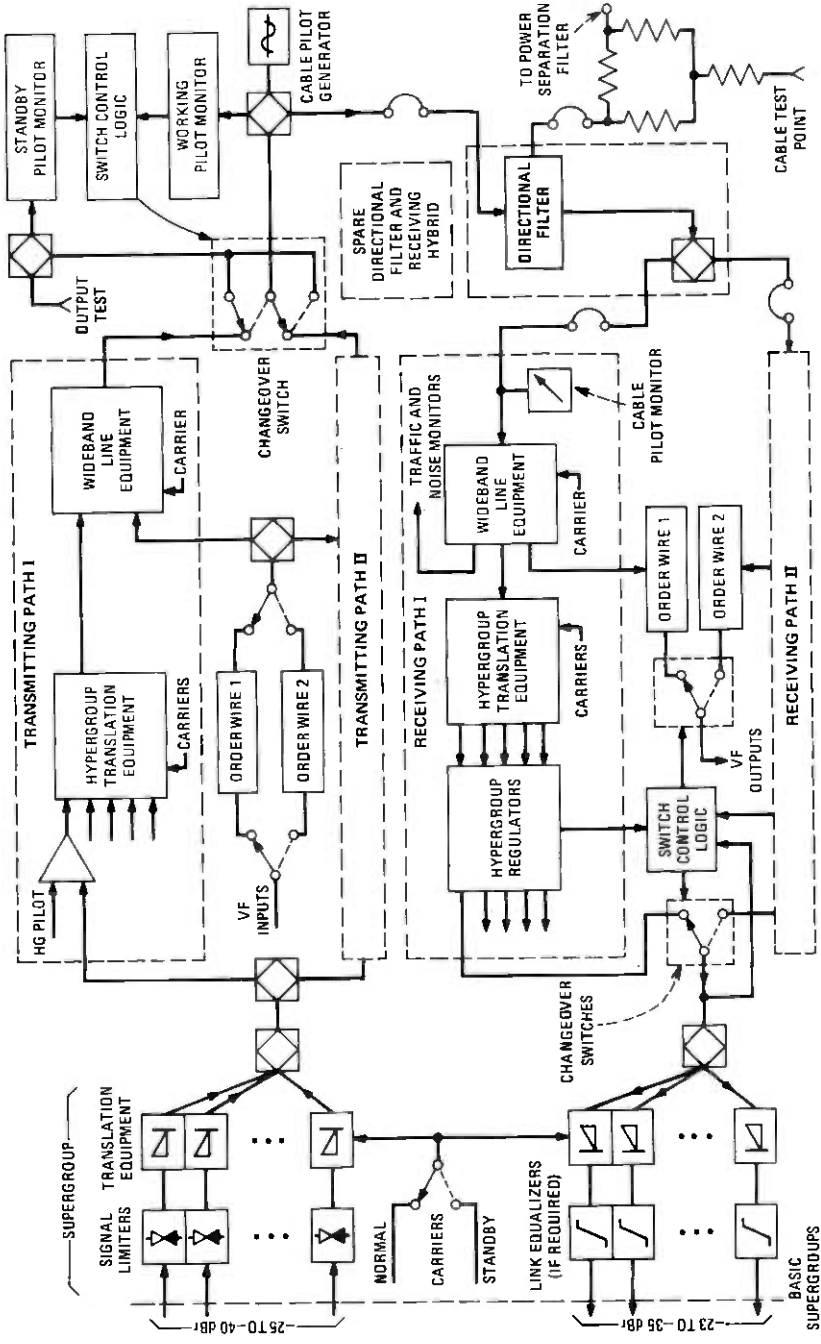


Fig. 6—Overall configuration of the SG terminal transmission equipment.

Table I

Combination No. 1 (normal operation)	Combination No. 2 (source and load paralleled)
Generator 1 → Path I Generator 2 → Path II	Generator 1 → Path I Generator 2 → Path II
Combination No. 3 (interchange of source)	**Combination No. 4 (two loads from single source)**
Generator 1 → Path I Generator 2 → Path II	Generator 1 → Path I Generator 2 → Path II

thereby assuring that its level is independent of all adjustments in the transmitting paths.

3.4 Special receiving features

In the receiving terminal, the cable pilot is monitored just after the directional filter and path splitting coupler so that its level is independent of receiving path adjustments.

Although there is a receive changeover switch for each hypergroup, all are activated simultaneously by the receive path-changeover-logic-and-control unit when it detects a fault affecting even one hypergroup. There are minor advantages to individual hypergroup changeover switch operation, but we felt they are more than counterbalanced by the complexity of maintenance that would result if some working hypergroups were on Path I while others were on Path II.

Inputs to the receiving changeover-logic-and-control unit are from five sources:

(*i*) and (*ii*) Hypergroup pilot level at the output of Paths I and II regulators.

(*iii*) and (*iv*) Gains of Paths I and II regulators.

(*v*) Working hypergroup pilot levels just beyond the changeover switches.

Switching occurs for either of the following conditions:

(*i*) At a changeover switch output, the absolute difference between measured and nominal pilot level becomes greater than 1 dB, whereas at the standby path regulator output the corresponding difference is less than 1 dB.

(*ii*) The difference in gain of corresponding Path I–Path II regulators is more than about 2½ dB, in which case the path whose regulator is providing the smaller correction becomes the working path.

Just as for the transmitting changeover switches, the receiving switches themselves are not duplicated, but pilot monitoring beyond the switches measures their performance as well. Individual pilot-alarm-inhibit capability is also provided.

3.5 Final comment on TTE configuration

The preceding description of TTE duplication and changeover switch operation brings to light the high degree of interdependence of multiplex and line terminal equipment. This is in contrast with most inland systems where multiplex and line terminals constitute separate families of equipment having little in common except impedance and transmission level at their interface. The intimate character of the association is underscored by the lack of separate line pilots in the SG wideband equipment.

IV. WIDEBAND LINE EQUIPMENT

The wideband line equipment (WLE) provides transmission between the multiplex and the terminal *power separation filter*.[7] Figure 7 is a simplified block diagram of the WLE. Aside from the conventional line functions of amplification, preemphasis, deemphasis, equalization, and the like, the SG WLE is designed to:

(*i*) Monitor performance of the undersea link.
(*ii*) Initiate changeover switching, in event of a failure, between duplicate terminal transmission paths.
(*iii*) Protect the undersea repeaters from high-level signal overload.
(*iv*) Provide adjustment capability to compensate for (*a*) that part of the shallow-water cable loss change with seasonal temperature, which is not equalized by temperature-controlled repeaters,[1] and (*b*) that part of transmission aging of the undersea link, which is not compensated for by shore-controlled equalizers.[1]

In some cases, accomplishing these functions provided us with interesting design problems that we had not encountered during development of terminals for earlier undersea systems.[8,9]

4.1 Amplitude and phase equality of duplicated paths

We have designed the changeover switching in such a way that maintenance switching (i.e., that which does not result from a transmission path failure) will not even momentarily disturb transmission more than a negligible amount.* To appreciate how this is accomplished, refer to the schematic in Fig. 8. If the amplitude and phase of the signals on each path are equal, closing of contacts B1 and B2 has no effect. Subsequent opening of contacts A1 and A2 (in either order) will interchange paths and loads without disturbance.

To evaluate the influence of path amplitude and phase differences, consider the following. The ratio in decibels of power delivered to the load in two cases, (*i*) separate paths (P_0) and (*ii*) parallel paths (P_1), is

* So-called hitless switching.

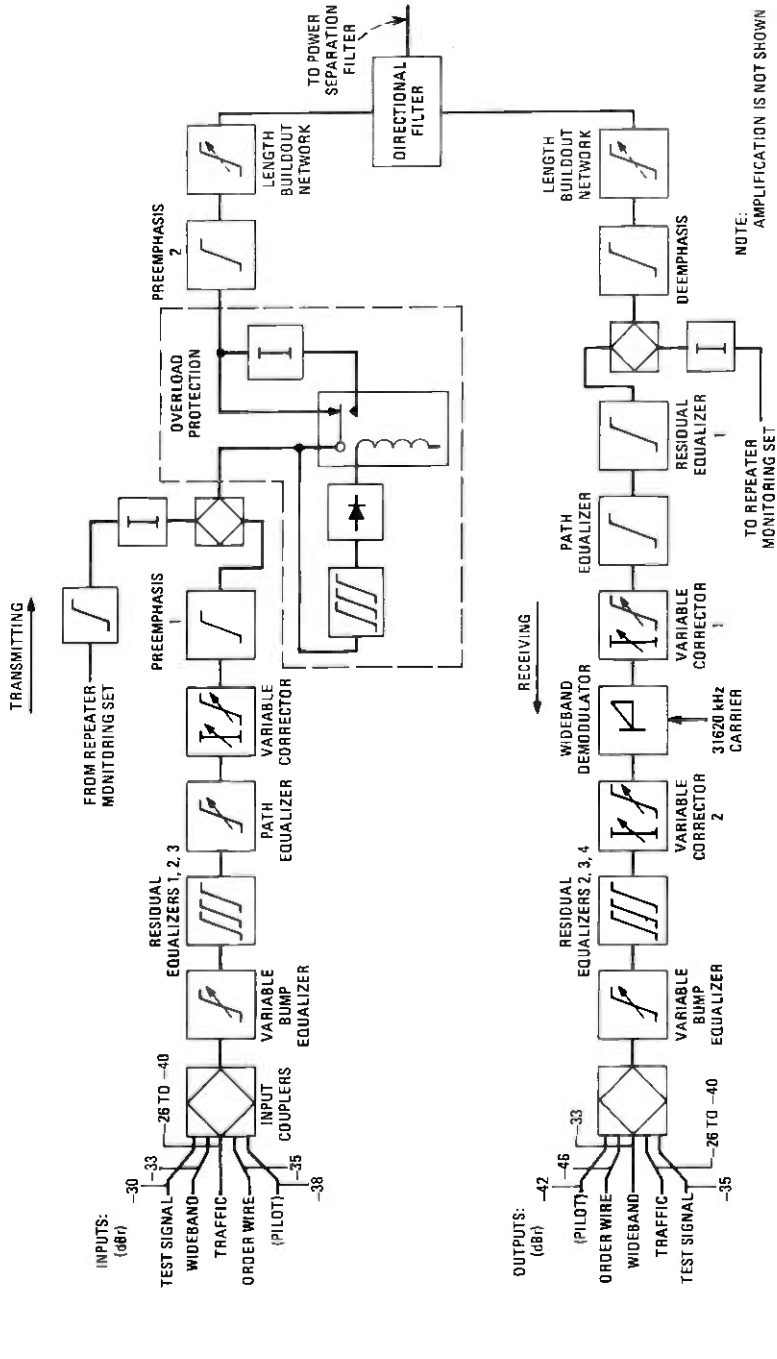


Fig. 7—Wideband line equipment-A station (B station is similar except frequency translation is in the transmitting line and only one preemphasis network is used).

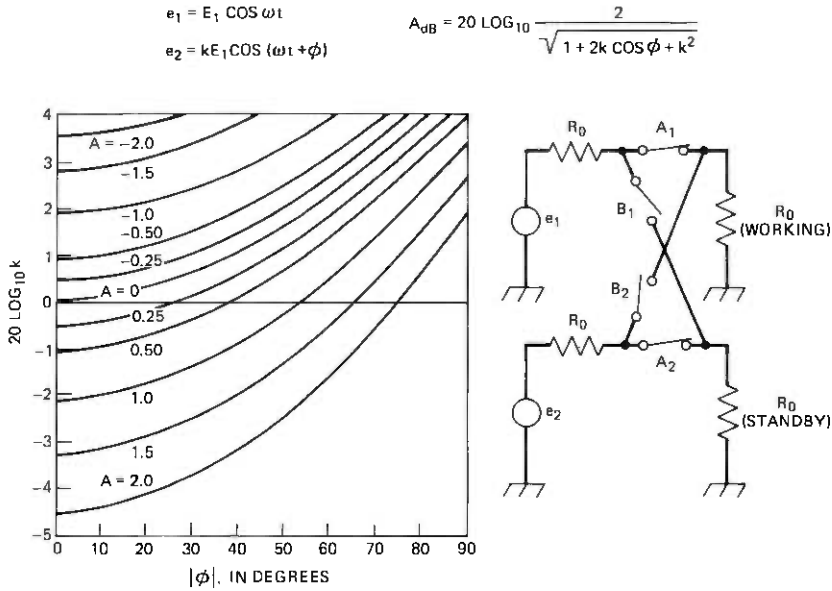


Fig. 8—Effect of path amplitude and phase differences on changeover switching.

$$A = 10 \log_{10} \frac{P_0}{P_1} = 20 \log_{10} \frac{2}{\sqrt{1 + 2k \cos \phi + k^2}},$$

where k and ϕ are defined in Fig. 8. In that figure, contours of constant A are plotted as functions of k and ϕ. One sees that, provided ϕ is not too large, the change in magnitude at paralleling is less than that for complete transfer. For example, if $|20 \log k| < 0.5$ dB and $|\phi| < 30°$, A is always less than about half a decibel. The phase disturbance at paralleling is

$$\tan^{-1} \frac{k \sin \phi}{1 + k \cos \phi},$$

and is always less than for complete transfer.

4.2 Protecting repeaters from signal overload

The peaks of a sufficiently high signal voltage applied to an SG repeater can, in time, cause a change in current gain of the output transistor. We decided, therefore, to limit to a safe value the maximum signal that can be applied to the undersea link from the terminal. This function is accomplished by the *wideband overload protection unit*, shown schematically in Fig. 7. The wideband signal is monitored at the output of the last active circuitry in the transmitting line and shaped with an equalizer before it is amplified and measured by a pseudo-quadratic detector. Should the detected signal exceed a predetermined threshold, a high-loss pad is switched in series with the transmission path. The ideal

frequency weighting in the detection path would be one so that a change in spectral density* of the signal at the input of the overload protection unit has the same effect on the mean power of the signal at the detector as it does at the output transistor in the highest level repeater in the link.

To avoid insertion of the pad for a momentary signal overload, the time constant of the detector is approximately 1 second. On the other hand, once the pad is inserted, it can only be reset manually.

We anticipate that operation of this unit will be extremely rare, because of the action of the supergroup limiters (described in Section 5.1.3), and probably due only to an equipment failure or adjustment error within the TTE itself. (Operation of the overload protection unit in the working transmitting line would, of course, cause changeover to the standby path.)

4.3 Pilot generation and measurement

As explained in Section 2.3, hypergroup and cable pilots serve to monitor operation of the multiplex, line, and undersea links, to initiate automatic protection switching, and to control the gain of receiving hypergroup regulators.

In principle, hypergroup and cable-pilot generators are identical designs, i.e., a quartz oscillator generates a stable frequency that is amplified and clipped to assure level stability. The signal is then filtered to remove harmonics and other spurious energy that could disturb traffic. The generator output is monitored (displayed on a front-face meter) and fed to a test jack so it can be checked with external equipment. An alarm is initiated for changes from nominal of a sufficient amount.

Pilot measuring units are also similar. A narrow-band 2296-kHz quartz filter is located at the input of those units monitoring the basic hypergroup (and hypergroup 1) pilot, and is followed by amplification and detection. On the other hand, for those units monitoring hypergroup pilots at baseband frequencies (except HG 1), this circuitry is preceded by a demodulator whose carrier frequency was chosen to restore the pilot to 2296 kHz. For the cable-pilot measuring units, a demodulator is not used; instead, a narrow-band quartz filter at cable-pilot frequency (11220 and 27836 kHz in the low and high bands, respectively) precedes amplification and detection. For all units, the pilot level is displayed on a front-face meter. A test jack provides access for an independent check of pilot level, and an alarm is initiated for sufficiently large level changes from nominal.

* Peak-to-mean-power ratio remaining constant.

4.4 Traffic and noise monitors

A continuous indication of undersea system noise is furnished by a unit connected near the output of the receive wideband line equipment that measures noise in a frequency slot, which is about the bandwidth of a voice channel, and is located between two supergroups in the neighborhood of the supervisory-tone band. Refer to Figs. 4a and 4b. The measuring apparatus itself consists of a quartz filter centered at 4276 kHz (A terminal*) and at 11,468 kHz (B terminal) followed by amplification and a pseudo-quadratic detector. Noise power is displayed on a meter whose scale is ±7 dB. Preceding this apparatus in the measuring path is a variable attenuator of ±10 dB in 1-dB steps. Fixed levels are such that a zero meter reading can correspond to noise in the range of about 28 to 48 dBrnc0. An alarm is initiated if the measured noise power exceeds the nominal value (zero meter reading) by more than 5 dB.

Also connected near the receive wideband line output is a traffic measuring unit similar to the noise measuring unit except that it is broadband (because it measures the complete traffic band) and is not alarmed. An attenuator adjustment on the front face (0 to 15 dB in 5-dB steps) permits the zero meter reading to correspond to 8, 13, 18, or 23 dBm0. (4200 channels at −13 dBm0/channel corresponds to a broadband power per band of 23.2 dBm0.)

Traffic as well as noise measurements are necessary to monitor an analog transmission system such as SG whose signal-to-noise performance is intermodulation rather than overload limited.

4.5 Alarm and maintenance features

As we have seen, behavior of the transmission link is monitored by means of pilot, noise, and traffic measurements. Pilot generating and measuring apparatus and noise measuring apparatus give visual indication of variations from nominal and initiate alarms when variations exceed specific limits. Upon initiation, alarm lamps on the front face of a unit and at the top of the bay containing that unit are lit, and the alarm is extended to the station alarm equipment which is expected to provide audio as well as visual indications. In this way, a technician is directed to the source of the trouble. A two-position switch on the face plate of the unit, when in the *cutoff* position, inhibits the station alarm as long as the source of the alarm remains. When the trouble is cleared, the alarm is reactivated until the switch is returned to the *normal* position. Finally, an output from pilot, traffic, and noise measuring units can be connected to a chart recorder (12 recorders are provided in a separate bay). This arrangement permits a continuous record to be obtained, when and if it is desired.

* The A station noise monitor is positioned after wideband translation to baseband.

4.6 Principle of primary frequency comparison

From the point of view of carrier generation, an undersea cable link often acts as an interface between national frequency standards. To insure that the frequency offset between terminals is within performance limits (no more than 2 parts in 10^8 for an SG link), it is necessary to compare master oscillator frequencies (primary frequency supplies). Comparison between the two terminals is accomplished in the following manner. Let us designate:

f_g—primary frequency (nominally 124 kHz) (from which all TTE carrier frequencies are derived) at the A terminal.

f_g'—Similarly at the B terminal.

N_1f_g, N_2f_g—Carrier frequency for hypergroups 2 and 5, respectively, at the A terminal ($N_1 = 59, N_2 = 119$).

N_1f_g', N_2f_g'—Similarly for the B terminal.

f_1—Basic hypergroup pilot frequency (nominally 2296 kHz) at the A terminal.

f_1'—Similarly for the B terminal.

f_t—Wide-band translation carrier (nominally 31,620 kHz) at the A terminal.

f_t'—similarly at the B terminal.

The A terminal transmits the HG 2 and 5 pilots at line frequencies of $(N_1f_g - f_1)$ and $(N_2f_g - f_1)$, respectively. At the B terminal, each is demodulated by the corresponding hypergroup carriers, N_1f_g' and N_2f_g' to obtain $f_1 + N_1(f_g' - f_g)$ and $f_1 + N_2(f_g' - f_g)$. The frequency difference between these recovered pilots, $(N_2 - N_1)(f_g' - f_g)$, is displayed directly on an analog meter (pointer movement proportional to the frequency difference) at the B terminal. Additionally, an A terminal reference frequency is made available in the B terminal, which is the difference frequency between received HG 2 and 5 pilots at line frequency, i.e., $(N_2f_g - f_1) - (N_1f_g - f_1) = (N_2 - N_1)f_g$, nominally 7440 kHz.

A similar arrangement provides a frequency-difference display and a B terminal reference signal in the A terminal, even though additional frequency translations by f_t and f_t' are involved.

4.7 Equalizers

We can make the following distinctions with regard to wideband line equalizers:

(*i*) Fixed equalizers meant to achieve the best overall signal-to-noise ratio for an ideal system in the sense of one with no undersea misalignment.

(*ii*) Equalizer networks designed and fabricated on site during commissioning[10] to compensate for actual misalignment measured after completion of installation.

(*iii*) Adjustable equalizers which, during commissioning, can be used

to achieve the best overall signal-to-noise ratio for an actual link with undersea misalignment, and, after commissioning, can be used to compensate for transmission changes resulting, for example, from temperature changes on the continental shelves and, in conjunction with the shore-controlled equalizers,[1] from cable aging.

The terminal equalization plan is discussed in more detail in a companion article.[10]

4.7.1 Fixed equalizers

In the category of fixed equalizers are the *line buildout*, *preemphasis*, and *deemphasis networks*, and the *path equalizers*. The line buildout networks in the transmit and receive paths are fixed for each terminal in that the effective electrical length to the first repeater from the terminal is equal to three-quarters of the loss of a nominal repeater section. In this way, flexibility is maintained in placement of the first repeater, while at the same time reasonably standard terminal equalization is preserved.

Path equalizers compensate for small misalignment unavoidably introduced in the wideband path, by interconnecting cabling and the like, and differences between nominally identical paths.

Given the nominal line equipment, the preemphasis network (transmit) provides the calculated loss shape to optimize overall s/n as mentioned in Section 4.7, item (*i*), above. The deemphasis network (receive) compensates for the preemphasis loss shape as well as for that of the last section of undersea cable (which is not compensated for by the gain of a repeater).

4.7.2 Equalizers fabricated during commissioning

To achieve final alignment of an SG link in a minimum of time after installation, we developed a *residual equalizer unit*. In a single unit, up to five bridged-T[11] networks can be mounted and connected in tandem. The frequency-loss characteristic of each network can be a bump, dip, or slope of wide ranging selectivities. Several such units are used in both transmit and receive paths (Fig. 7). A kit of components (inductors, capacitors, and resistors) covering specific ranges of values, as well as other hardware, is supplied with the WLE. After link misalignment has been measured, bridged-T networks are designed (with the aid of a computer) and constructed on site, and inserted into the wideband line.

4.7.3 Adjustable equalizers

Both broad- and fine-grain equalization capability is provided by the front-panel-adjustable networks which facilitate initial performance optimization and subsequent realignment following undersea loss variations. Broadband adjustment at line frequency is provided in both the transmit and receive paths by a so-called *variable corrector* that contains four types of circuits: slope, curvature, flat, and $\sqrt{f}$. At the frequency of maximum excursion (e.g., the peak of the curvature shape), each is adjustable in 0.3-dB steps over a range of ±3 dB (slope and curvature) or ±6 dB (flat and $\sqrt{f}$). A second type of variable corrector, used only in the receive path, provides additional flat and $\sqrt{f}$ adjustable gain range.

Among other functions, the variable correctors must be able to compensate for certain fault conditions in the undersea link, i.e., those whose effect on transmission is not so severe that service cannot be maintained, albeit at reduced performance, until repair operations commence.

A *variable bump equalizer*[12] permits narrow-band adjustment in the vicinity of a series of discrete baseband frequencies related by geometric progression. Selectivity is such that adjustment at one of these frequencies involves negligible change at adjacent ones. Because of the slightly different basebands, 16 bumps are available in the low band, but only 12 in the high-band direction of transmission.

4.8 Order-wire facilities

The line terminal is equipped with three order-wire circuits, each terminated at voice frequency with a unit permitting two- or four-wire operation and signaling. The frequency-translation circuitry is distinct from the regular multiplex equipment. Each channel undergoes translation to 24 to 28 kHz, as in the French Post Office (PTT) Type 70 12-channel equipment,[13] before the three are contiguously assembled at 60 to 72 kHz, the bottom quarter of the standard group baseband. The final two stages of translation correspond to placement in group 5 of supergroup 16 (see Figs. 4a and 4b).

The order-wire equipment is duplicated; in the transmit direction, either equipment can be connected by a manual switch to the transmitting paths. In the receiving direction, order-wire signals are selected from the working path through an automatic switch controlled by the receive changeover logic. See Fig. 6.

One of the three channels is equipped as a local order wire (i.e., for exclusive use between terminals), with an operator panel containing headset jacks plus ring and call-cutoff pushbuttons.

V. MULTIPLEX EQUIPMENT

The multiplex portion of the TTE is located between supergroup distribution frames and the wideband line equipment (Figs. 1 and 6). It includes supergroup and hypergroup translating equipment (STE and HTE) with the associated carrier generators, hypergroup regulators, supergroup limiters (transmit), and supergroup equalizers (receive). Typically, this circuitry constitutes a link between domestic inland facilities (at basic supergroup) and the wideband line equipment (at SG system baseband).

Design of the multiplex equipment allows it to operate with 3- or 4-kHz spaced channels, with a per-channel load of −13 dBm0 in either case.

The supergroup and hypergroup frequency assignments have already been discussed in Section II. Refer to Figs. 2 and 3.

5.1 Supergroup equipment

5.1.1 Frequency translation

Supergroup translation equipment is derived from standard PTT domestic equipment,[14] which already conforms to CCITT recommendations,[6] but special care was given to reduce carrier leaks. Unfortunately, supergroup carrier frequencies, which are multiples of 4 kHz, do not necessarily fall between channels when operating with 3 kHz spacing as they do with the more conventional 4-kHz spacing. The disturbance level of an ensemble of spurious tones must be less than −65 dBm0/channel, so we felt it prudent to hold the power of a single spurious frequency to no higher than −70 dBm0.

Performance requirements for the supergroup translating equipment are summarized below:

(*i*) Carrier-leak, ≤ −70 dBm0 transmitting, ≤ −40 dBm0 receiving.*

(*ii*) Thermal and intermodulation noise (for +6.0 dBm0/supergroup loading), ≤30 pW0p (15.3 dBrnc0) transmitting or receiving.

(*iii*) Return loss at all ports, ≥25 dB.

(*iv*) Crosstalk ratio (single frequency) ≥85 dB.

5.1.2 Carrier generation

The supergroup carrier generating equipment is of conventional design, a harmonic generator (fed from a 124-kHz master oscillator) driving a parallel string of LC bandpass filters, each selecting the appropriate harmonic for a particular supergroup, followed by a gain-regulated drive amplifier. Suppression of unwanted carrier frequencies at each output is greater than 80 dB. As mentioned in Section 3.2, the supergroup carrier

* Receive carrier leaks fall outside the transmission band.

generating equipment is duplicated, with a changeover switch at the outputs for each carrier. Signal detection permits automatic changeover operation and alarm initiation in case of failure in the working equipment.

5.1.3 Supergroup signal limiters

Application of sufficiently high-powered signals to analog transmission systems will cause excessive intermodulation noise[15] that interferes with traffic and, in a severe case, causes overload which can result in the loss of all traffic. To prevent this situation from occurring on the SG system, we decided to equip each input of the transmitting STE with a signal limiter (Fig. 6). The limiter is transparent (fixed attenuation) to signals in the normal operating range. Should the applied signal exceed preset amplitude limits, however, a controlled amount of attenuation is introduced in a variable loss network in the signal path.

Two monitoring circuits are used to control loss insertion, (*i*) a peak level monitor, and (*ii*) a mean-power level monitor. Refer to Fig. 9. Response of the peak monitor is relatively fast but consistent with maintaining within acceptable limits the generation of unwanted wideband energy. In practice, operation and restoration times of about 1 and 5 ms, respectively, are used. Response time of the mean-power monitor is, on the average, much slower than this, but depends on the magnitude of its overload, i.e., the higher the overload the shorter the response time.

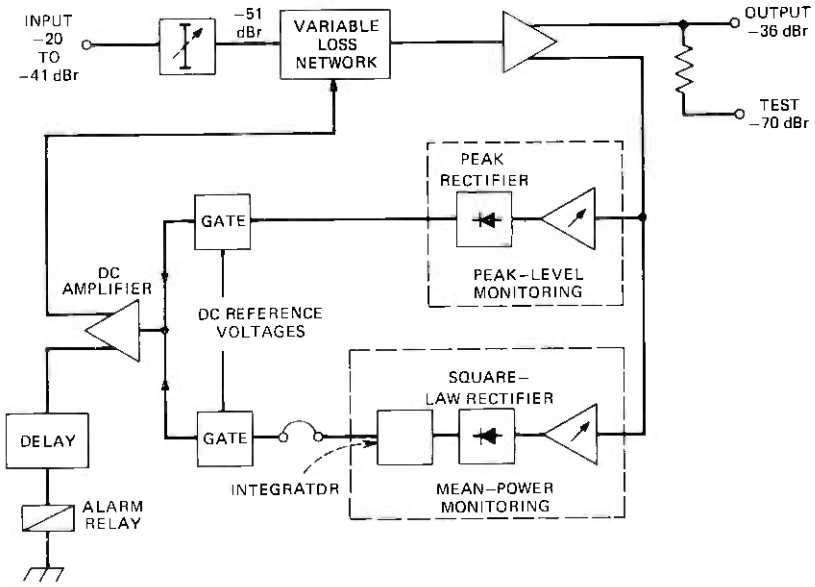


Fig. 9—Supergroup signal limiter.

To allow for differing types of traffic and the actual performance of the undersea link, the operating point of the peak and mean-power limiter circuits are separately adjustable by means of attenuators in each amplifier/detector input. The dc output of each monitor is compared with a reference voltage. When exceeded, the difference voltage is applied (after amplification) to the variable loss network; the larger the difference voltage, the greater the loss increase. A second output from the dc-amplifier energizes an alarm indicative of limiter operation, but a delay network suppresses the alarm if the limiter is activated only momentarily.

An input attenuator to the limiter unit is adjusted at the time of installation for differences among administrations with regard to transmission level at the supergroup input.

The supergroup signal limiter used in the SG terminal is a slight modification of the British Post Office (BPO) limiter[15] as manufactured by GEC, Ltd, and is constructed under SOTELEC 70 equipment practice to conform to the remaining TTE equipment manufactured by CIT-Alcatel of France.

5.1.4 Narrow-band-elimination filters

To assure a clear slot between particular supergroups for subsequent insertion of hypergroup and cable pilots, and to facilitate undersea noise monitoring, narrow-band-elimination filters centered at 308- and 556-kHz are inserted between signal limiter and modulator in certain transmitting supergroups, and between demodulator and equalizer in the corresponding receiving supergroups.

5.1.5 Supergroup link equalizers

Although the primary means of compensating for undersea misalignment in the terminal are the residual equalizers in the WLE, we decided nevertheless to have available for use a supergroup-link equalization capability if needed to satisfy the CCITT recommendation regarding amplitude distortion,[16] viz., no greater than ±1.5 dB relative to a midband frequency which itself must be ±0.1 dB of nominal. When used, an equalizer is placed after the supergroup demodulator as shown in Fig. 6.

A simplified circuit diagram of the equalizer is given in Fig. 10. It is essentially an amplifier whose gain in the 312- to 552-kHz band is determined by damped resonant R-L-C circuits inserted as shunt elements in the feedback network. Up to five such circuits per amplifier can be used. Just as for the residual wideband line equalizer, a specific kit of components is provided for on-site construction, during commissioning, of the feedback and other gain controlling elements.

A sophisticated computer program is used to determine the number

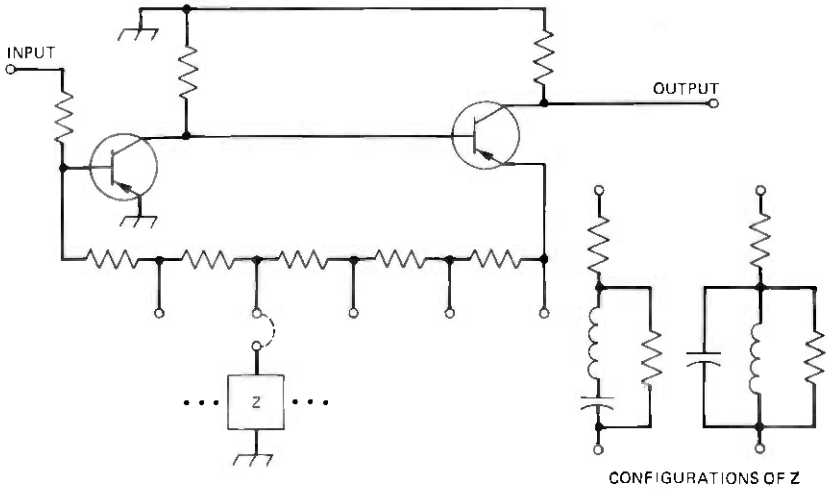


Fig. 10—Supergroup-link equalizer.

and configuration of the feedback circuits and the value of all variable components. The program uses an iterative trial-and-error process which continues until the computed link gain after equalization is within a specified tolerance. If equalization proves particularly difficult, e.g., for those supergroups near the edge of the wideband, a second amplifier can be placed in tandem with the first.

5.2 Hypergroup equipment

5.2.1 Frequency translation

A new design of hypergroup translating equipment is justified for use with undersea systems to achieve high traffic-band efficiency. Moreover, the tight phase and amplitude difference constraints made the new design an interesting one.

Most subassemblies of the HTE are of conventional design. Negative feedback amplifiers are made up of transistor pairs, the modulators use fast diodes in a ring configuration between ferrite core transformers, and the filters are LC type using ferrite inductors and mica capacitors. Solder-adjustable resistive pads permit level alignment. Coupling of the hypergroups is at a low input impedance amplifier on the transmit side and at a low output impedance amplifier on the receive side.

A departure from conventional practice was the addition of a series LC circuit at the input of each carrier amplifier. Tuning the capacitor permits adjustment of carrier phase over a sufficient range to compensate for phase differences between duplicate HTE paths for each hypergroup.

Performance requirements for the hypergroup translating equipment are summarized below:

(*i*) Carrier leak, ≤ −55 dBm0 at transmitting, ≤ −40 dBm0 at receiving.
(*ii*) Thermal and intermodulation noise (for +16.0 dBm0/hypergroup loading), ≤20 pW0p (13.5 dBrnc0) transmitting or receiving.
(*iii*) Return loss at all ports ≥25 dB.
(*iv*) Crosstalk ratio (single frequency) ≥85 dB.

5.2.2 Carrier generation

Design of the hypergroup (and wideband) carrier generating equipment was particularly challenging. The conventional scheme of harmonic generator followed by selective filters was not capable of producing sufficient carrier purity because the ratio of carrier frequency to master oscillator frequency is so large. (In the extreme case of the wideband carrier, the ratio is 255.) Furthermore, it was necessary to maintain the phase difference between corresponding carriers from the duplicated generators to within a few degrees to be consistent with hitless change-over switching between transmission paths as well as hitless manual patching at the duplicated generator outputs, whose combinations are illustrated in Table I.

The arrangement chosen to do the job is shown in Fig. 11 in block form. Notice the harmonic-generator, selective-filter combination is still present, but the desired harmonic of 124 kHz is only used to phase-lock

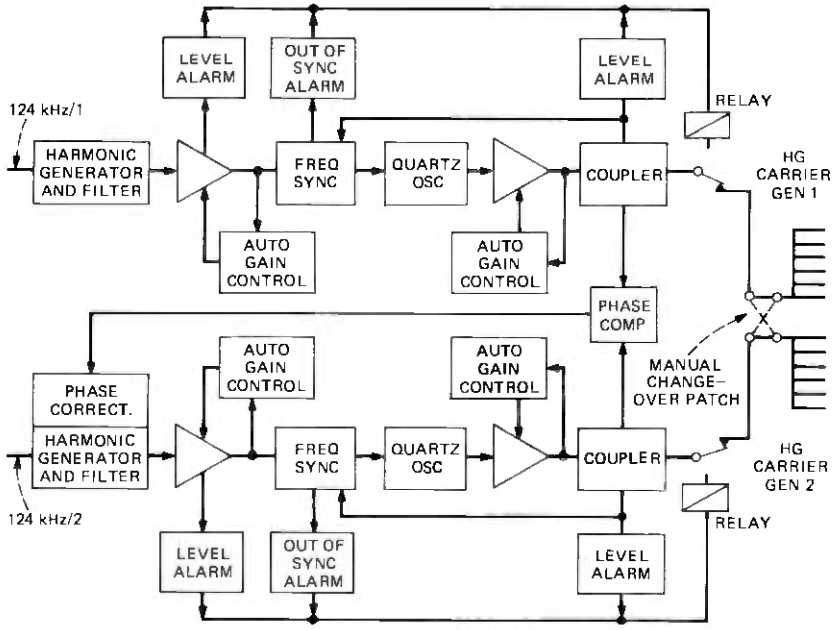


Fig. 11—Hypergroup carrier generation.

a quartz oscillator which is the actual carrier source. The necessary carrier purity is achieved by means of a first-order low-pass filter within the phase-locked loop whose cutoff frequency is only a few Hertz (except during capture). This arrangement produces, in effect, an extremely narrow-band-pass filtering action at the carrier frequency.

A phase comparator monitors the difference between corresponding generators and controls a corrector which automatically compensates phase deviations.

Performance requirements for the hypergroup (and wideband) carrier generating equipment are summarized below.

(*i*) Frequency stability of the master oscillator is preserved.
(*ii*) Output level stability: ±1 dB.
(*iii*) Carrier-to-noise ratio: at least 100 dB in 3-kHz band.
(*iv*) Phase difference between duplicate generators: <5 degrees.

5.2.3 Hypergroup regulators

The hypergroup regulator is a conventional design, i.e., an amplifier whose gain is pilot-controlled. A narrow-band quartz filter in the regulation loop extracts the hypergroup pilot from the traffic band at the amplifier output. This signal is then amplified, rectified, and compared to a dc reference voltage. The resulting difference signal controls the temperature of a thermistor located in the feedback path of the amplifier. The regulator has a compression ratio greater than 10 over an input range of ±2.5 dB about its nominal gain.

Lamps on the front panel of the unit indicate when regulation exceeds ±1 dB and ±3.5 dB, and the latter can be transferred to the station alarm, as is the loss-of-pilot alarm which is initiated if the amount of regulation exceeds 7 dB. Input and output test points are also provided on the front panel.

VI. MECHANICAL ARRANGEMENT

The SG system terminal transmission equipment is produced under a mechanical specification standardized in France under the name SOTELEC 70,[17] whose basic elements are the *frame*, a simple mechanical support for *shelves* which contain a number of plug-in *units*. This total arrangement constitutes a *bay*. Units are provided with a front face plate on which are located labeling, test points, meters and other indicators, and in-service adjustment facilities. The rear face supports a connector that mates with its shelf-mounted partner. Mechanical locking is assured either by the connectors themselves or by a screw on the face plate. All shelves are the same width and depth, but their height can be a multiple of a modular unit of 44.5 mm.

Interconnection of units within a shelf (and occasionally between shelves) is done at the rear. Input and output access to a shelf is through vertical terminal strips, one on each front side of the shelf.

Fig. 12—Typical bay lineup of SG terminal transmission equipment.

A frame supports the shelves, and is in turn held to a base fixed to the floor. Wiring between shelves generally follows along the inside of the vertical frames and connects to the terminal strips.

A typical bay line-up for a terminal is pictured in Fig. 12. Although variations from this bay layout are possible, one must be careful to limit length differences of many intra-bay cabling runs in order to preserve phase equality of duplicate paths. The high degree of functional interdependence of line and multiplex terminal equipment dictates that they be physically close.

VII. CONCLUSIONS

Achieving and maintaining maximum use of the undersea link have been our prime objectives from conception through production of the SG terminal transmission equipment. Accomplishing these ambitious objectives has led in some instances to complex designs whose development has been challenging because of, on the one hand, technological difficulties associated with the relatively high frequencies used (compared to earlier undersea system designs), and on the other the multiplicity and diversity of required functions. In four years it has been necessary to plan for and then design, develop, and produce no fewer than 263 different types of units. Undertaking the work in this interval constituted a wager that could only be won by exerting an all-out effort.

Everyone who has participated in this task confidently believes that the SG terminal transmission equipment will render all the service we have the right to expect from it.

VIII. ACKNOWLEDGMENTS

Many individuals from organizations on both sides of the Atlantic were involved in TTE development from planning through installation as part of the first SG system implementation, TAT-6. Representatives from the Network Planning Department of the British Post Office Telecommunications Headquarters, the National Center for Telecommunications Studies and the Submarine Branch of the French Post Office, the Ocean Cables group of the Long Lines Department, American Telephone and Telegraph Company, and the Undersea Systems Department of Bell Laboratories helped guide development, but the lion's share of the work was carried out most commendably by members of Compagnie Industrielle des Télécommunications (CIT-Alcatel) in France.

REFERENCES

1. C. D. Anderson, W. E. Hower, J. J. Kassig, V. M. Krygowski, R. L. Lynch, G. A. Reinold, and P. A. Yeisley, "SG Undersea Cable System: Repeater and Equalizer Design and Manufacture," B.S.T.J., this issue, pp. 2355–2402.
2. F. H. Blecher, R. C. Boyd, F. J. Hallenbeck, and F. J. Herr, "The L-4 Coaxial Systems," B.S.T.J., *48*, No. 4 (April 1969), pp. 821–839.
3. F. C. Kelcourse and F. J. Herr, "L5 System: Overall Description and System Design," B.S.T.J., *53*, No. 10 (December 1974), pp. 1901–1933.
4. Recommendation G338, "4-MHz Valve Type Systems on Standardized 2.6/9.5 mm Coaxial Cable Pairs," Volume III.1 of CCITT Green Book.
5. Recommendation G343, "4-MHz Systems on Standardized 2.1/4.4 mm Coaxial Cable Pairs," Volume III.1 of CCITT Green Book.
6. Recommendation G233, "Recommendations Concerning Translating Equipments," Volume III.1 of CCITT Green Book.
7. E. T. Calkin, I. Golioto, W. J. Schatz, R. E. Schroeder, D. S. Shull, "SG Undersea Cable System: Undersea System Power," B.S.T.J., this issue, pp. 2497–2522.
8. R. L. Lynch, J. L. Thomas, and C. A. Von Roesgen, "SF System: Shore Terminal Facilities and Fault Location," B.S.T.J., *19*, No. 5 (May–June 1970), pp. 721–748.
9. A. Blanchi, M. Dubois, R. Salvador, and H. Soulier, "La Liaison Sous-Marine Marseille-Alger II," L'Echo des Recherches, July 1973.
10. D. N. Harper, B. O. Larson, and M. Laurette, "SG Undersea Cable System: Commissioning: Final System Alignment and Evaluation," B.S.T.J., this issue, pp. 2547–2564.
11. O. J. Zobel, "Distortion Correction in Electric Circuits with Constant Resistance Recurrent Networks," B.S.T.J., July 1928, pp. 438–534.
12. J. Oswald and C. Chalhoub, "L'Égalisation," Câbles et Transmission Ann., *16*, No. 3 (July 1962).
13. G. Duval, J. Grollemund, and J. C. Duval, "Organes Fonctionnels de Ligne Avec Appel à Fréquence Zéro," Câbles et Transmission Ann., *25*, No. 4 (October 1971).
14. G. Laguerie, M. Louboutin, and M. Knapp-Ziller, "Modulation de Groupe Secondaire, etc.," Câbles et Transmission Ann., *25*, No. 4 (October 1971).
15. N. P. McKay, "Wideband Signal-Limiters," P.O.E.E.J., *62*, Part 4, January, 1970.
16. Recommendation M46, "Bringing International Group, Supergroup, etc., Links into Service," Volume IV.1 of CCITT Green Book.
17. G. Laguerie, M. Lagarde, and J. Menu, "Structure Mécanique du Matérial SOTELEC 70," Câbles et Transmission Ann., *25*, No. 4 (October 1971).

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Undersea System Power

By E. T. CALKIN, I. GOLIOTO, W. J. SCHATZ, R. E. SCHROEDER, and D. S. SHULL

(Manuscript received July 21, 1977)

The power-feed equipment for the SG cable system is of the type that regulates the output current and limits the output voltage. It energizes the serially connected repeaters in the TAT-6 undersea system. This paper describes the power-feed equipment in general terms and, in more detail, several new power techniques. These new techniques include inductive-input inverters in the main power train, pulse-position-modulation in the control and monitoring circuits, precise digital metering of output current and voltage, vacuum switch elements in the transfer switch, and special new components and design concepts in the power separation filter needed to achieve SG-wideband-transmission objectives in the presence of the high, TAT-6, power-feed voltages.

I. INTRODUCTION

The SG power-feed equipment is similar in system configuration to that which powers the SD[1] and SF[2] undersea cable systems previously developed and installed by the Bell System. As examples: (*i*) power-feed redundancy is achieved by connecting two load-sharing, independent, direct-current sources in series, with provision for the automatic assumption of the full load by the survivor should one current source fail, (*ii*) the cable is powered at each end with opposite polarities and a sea-ground power return, and (*iii*) the power supplies at both ends operate

as current sources with their maximum output voltages restricted to be less than the voltage standoff capabilities of the repeaters.

Despite the conceptual similarities among the SD, SF, and SG power-feed systems, the techniques used to realize the system features have changed as the performance requirements have become successively more stringent and new circuit and physical design technologies have become available, or have been invented, to meet the performance requirements. Representative SG requirements are summarized in Table I.

In this paper, we describe the SG power-feed system as applied to TAT-6, emphasizing the following new circuit and physical design features:

(*i*) Inductive-input inverters in the power stages of the main power train.

(*ii*) Pulse-position-modulation in the control and monitoring circuits.

Table I — Representative SG power-feed performance requirements for TAT-6

Output current:	
Nominal	657.00 mA
Set point resolution	±0.10 mA (±0.015%)
Variation within 24 hours	±0.33 mA (±0.05%)
Variation resulting from a ±1000-V load change	±3.30 mA (±0.5%)
Variation resulting from the failure of one of the two converters	±3.30 mA (±0.5%)
Output voltage:	
Nominal voltage	5200 V/station
Maximum voltage	7500 V/station*
Tolerable earth potential	2300 V/station*
Shutdowns:	
Current	+6% of nominal current (circuit duplicated)
Voltage	9250 ± 750 V (static) 10800 V, maximum (dynamic)
Alarms:	
Cable current	±1% (minor), ±3% (major), fixed
Cable voltage	±5% (minor), ±15% (major), adjustable to bracket the cable voltage
Noise:	
Inverter harmonic tones	<0.17 μV in any 3-kHz band from 0.3 MHz to 30 MHz
Impulse (signal transmission path)	Maximum of one pop per 15-min period that exceeds −10 dBmO in any 48-kHz channel (requirement of the entire shore station including the power separation filter)
Power separation filter transmission requirement:	
Return loss	≥20 dB, 0.5 to 30 MHz

* In normal operation, the voltage-limiting-inception level would be set approximately 1000 V above the zero-earth-potential operating level, e.g., at 6200 V in TAT-6 stations. Hence, a readjustment is necessary to take advantage of the total voltage capability of the power supply.

(*iii*) Precise, digital metering of the current and voltage in the high-voltage output circuits.

(*iv*) Vacuum switches in the "hot transfer" switch circuits and vacuum relays in the wideband, signal-transmission path of the shore-station power separation filter.

(*v*) New, high-voltage, wideband, signal-transmission components and a low-pass, skin-effect filter in the power separation filter.

(*vi*) Physical design techniques to control radiated and conducted electromagnetic interference (EMI) within and from the power supply and into the power separation filter.

II. GENERAL DESCRIPTION OF THE POWER-FEED SYSTEM

Figure 1 displays the power-feed and signal transmission connections to the TAT-6 cable. The A terminal (Green Hill, Rhode Island) power supply provides a current of 657 mA at a nominal, positive voltage of 5200 V with a maximum capability of 7500 V. The 657 mA is joined to the signal transmission path within the power separation filter, and together they are routed into the center conductor of the cable. As the current passes through each repeater (as shown in Fig. 1), sufficient voltage is dropped to power its active circuits. At the B terminal (Saint Hilaire de Riez, France), the power-feed arrangements and equipment are identical to those at the A terminal, except that the power supply's output voltage is negative. The maximum end-to-end cable voltage capability of the SG power-feed equipment is 15,000 V. Maximum capability would be exercised only on long cables in the face of adverse earth potentials. Note also that the 7500-V capability of each power supply permits continuance of TAT-6 cable operation with a high-impedance, shunt fault, in the cable or in a repeater, as close as 2900 V to one station (7500 V from the other station) until a repair ship arrives and begins the repair operation.

The power-return-current path is routed from the power supply through the power separation filter and the ocean-ground panel to specially constructed ocean-ground grids.[3]* Potential differences of as much as 2.7 kV have been observed between the ocean-ground grids in the TAT-1 system.[4] The power supplies at each end of the TAT-6 cable are adequately regulated to absorb potentials of similar magnitude while maintaining control of the cable current.

The manually operated transfer switch and adjustable test load, both located in the load transfer bay (Fig. 1), permit the cable to be transferred between power plants without taking the cable out of service. This simple, but important, circuit is also part of the interlock system where

* Although Reference 3 is related to an earlier British system, it provides details of the problems encountered in designing, installing, and operating ocean-ground grids.

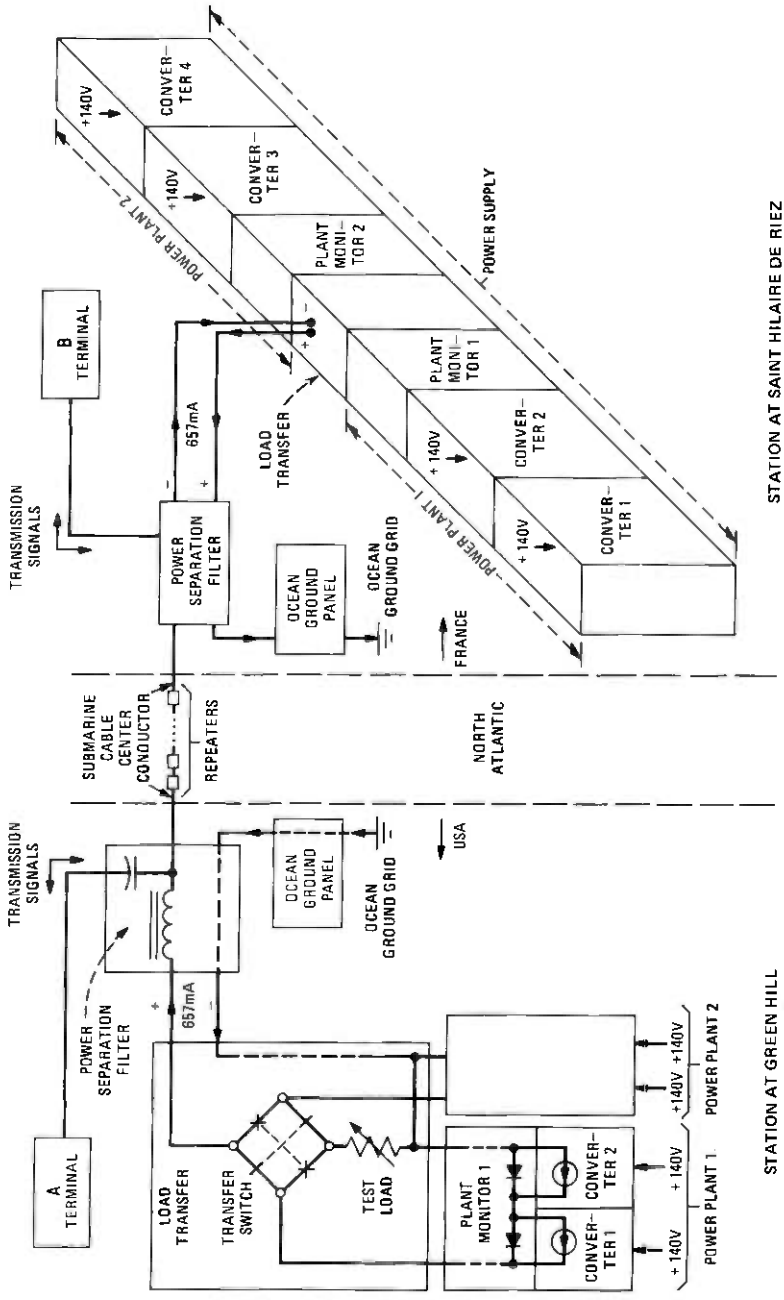


Fig. 1—Power-feed and signal transmission connections to the TAT-6 cable.

it differentiates between energized and unenergized high-voltage areas and permits access to the unenergized parts of the equipment for maintenance and inspection.

Previous Bell System power-feed equipment relies on conservatively designed, mechanically simple, transfer switch mechanisms. Although very rugged mechanically and proven reliable in many existing systems, these transfer switch mechanisms are physically large and cumbersome to operate. Hence, the SG-power-feed equipment makes use of a reliable, smaller, lighter, and simple-to-operate transfer switch.[5] The switch, shown in Fig. 2, has two cams that rotate with the handle of the switch. The cams operate plungers in four vacuum interrupters.

Figure 3 displays an annotated photograph of the seven-bay power supply. The bays are approximately 2.13 m (7 ft) high, approximately 0.71 m (28 in.) deep, and either approximately 0.79 m (31 in.) or approximately 1.14 m (45 in.) wide. The power supply ensemble is approximately 6.86 m (22.5 ft) wide.

Each of the seven bays contains low-voltage and high-voltage spaces. The low-voltage spaces contain alarm, monitoring, and control circuit components that are mounted in plug-in modules. Plug-in modules are

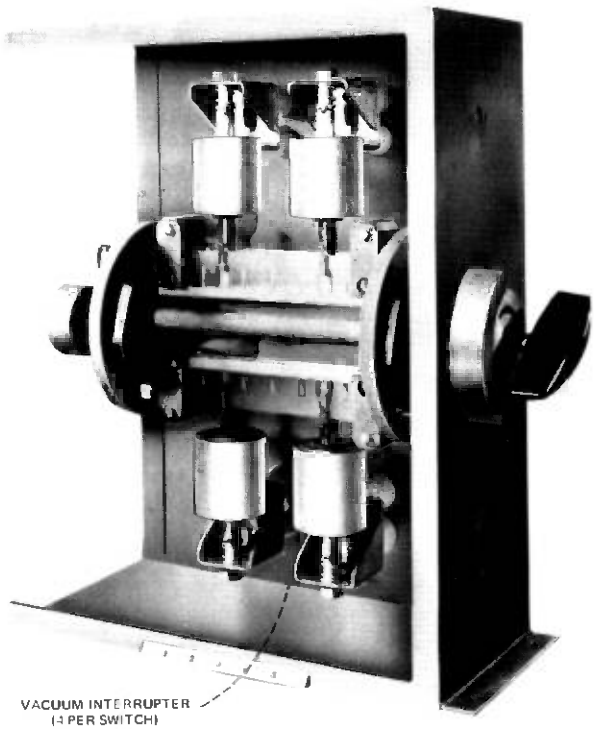


Fig. 2—Transfer switch.

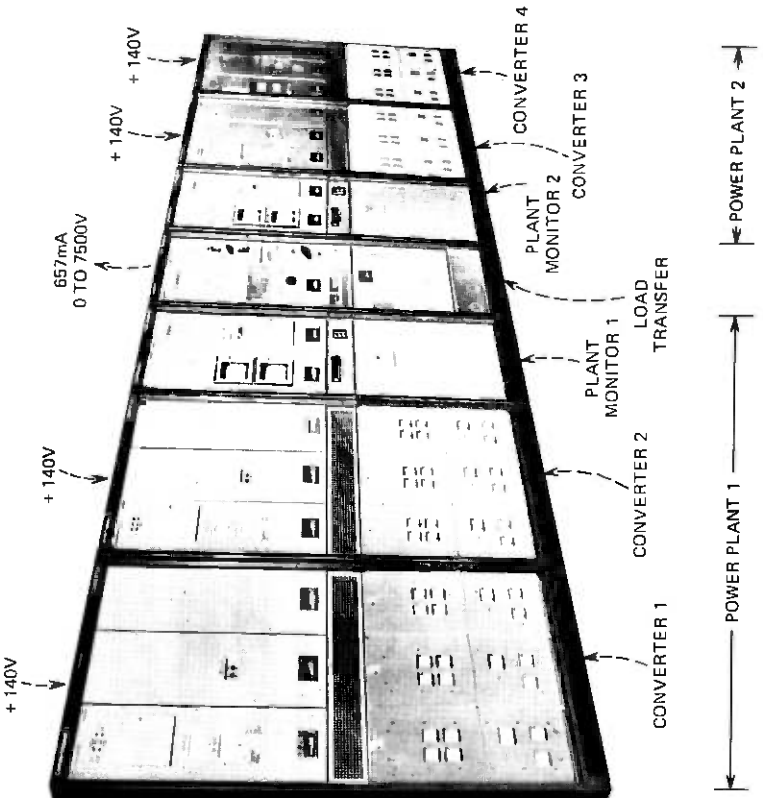


Fig. 3—SG power supply for one end of a long-haul, undersea cable.

located in pull-out slides to provide easy access for both normal operation and maintenance. The high-voltage circuits in the seven bays of the power supply and the high-voltage circuits in the power separation filter are contained in compartments that can be accessed only after appropriate manipulation of the key-controlled, safety-interlock system to assure the disconnection of the input power to the high-voltage space being accessed.

All interbay wiring in the seven-bay power supply is routed through connectors, except for the high-voltage leads, to minimize installation effort and errors. The center conductor and inner shield of a triaxial (shielded coax) cable provide the high-voltage and power-return paths, respectively, between the power supply output and the power-separation-filter bay. The outer shield is useful in reducing EMI conduction from the power supply to the power-separation-filter bay.

The power conversion systems are powered from +140-V battery systems.* In addition, 115-V, 50/60-Hz power is required by the digital meters in the plant monitor and load transfer bays and the test load cooling fans. The 115-V inputs are not shown in Fig. 1.

III. POWER CONVERSION TECHNIQUES

The process of converting power from the +140-V source into a precisely controlled, 657-mA, constant-current level suitable for energizing SG-type cables, at up to 7500 V, is accomplished in the converter bay. The cable-powering requirements are far more demanding than those of previous cables. For example, the 5.0-kW, nominal output capability of each SG converter bay represents a more than eight-fold increase over the capability of the most recent, previous, comparable Bell System converter. The power-train circuits follow the SF system practice of switching at ultrasonic frequencies† to:

(*i*) Minimize the stored energy in filter capacitors.
(*ii*) Avoid generation of high-level, audible noise.
(*iii*) Minimize the physical size of magnetic apparatus.

The increase in power level and the more demanding regulation requirements prompted the use of a pulse-width-control technique instead of the saturable-reactor-type power control used in several previous undersea cable systems. The pulse-width-control technique is used to control switching-regulator transistors in the power path.

The increase to 5.0 kW of converter output power required efficient utilization of the state of the art in power transistor technology. High current (25 A), medium voltage (200 V), and fast switching speed (0.75-μs

* A description of the primary power arrangements inluding the commercial power connections, the rectifier-floated battery systems, and the automatic-or-manual-start, standby engine-alternator systems is beyond the scope of this article.

† The SF power-feed equipment was the first to use static converters switching at ultrasonic frequencies to power long-haul underseas cables.

rise-and-fall times) devices were selected as appropriate to operate with the +140-V input voltage and to minimize the number of required power stages. The +140-V input was selected to minimize current levels. The nominal current drain from the +140-V input for a converter bay is less than 36 A instead of over 100 A if a 48-V input had been used. The lower input current results in:

(*i*) Less I^2R losses, thereby providing greater efficiency in the primary power path.
(*ii*) Lower-strength magnetic fields emanating from the inverter's switching circuits.

The basic dc-to-dc power conversion unit, referred to as a "power stage," operates at 20 kHz (inverter) and 40 kHz (switching-regulator) to develop an output current of up to 690 mA at voltages of up to 2500 V from the +140-V input. Each converter bay contains up to three of these power stages with their outputs connected in series to obtain up to a 7500-V output. During normal operation in a TAT-6 station, six power stages (three in each converter bay) are powering the cable with each power stage operating at less than half its power capability. Figure 4 shows a converter bay with one power stage extended from the bay on its slide and with its EMI-containment covers removed.

A power stage is self-starting and requires only +140-V power, an oscillator-synchronizing signal, and the pulse-width-control signal as inputs for normal operation. Both high-power switching and low-level-control-and-monitoring circuits are incorporated in this plug-in package. Each power stage contains its own base-drive generator rather than receiving base-drive from an external, common generator in the converter bay. This feature assures that the loss of a base drive generator will not cause multiple power train damage. Also, the distribution of switching currents, and the resultant EMI noise radiation within and from the bay, is minimized.

Operation of the power switching transistors to near their maximum current capability requires the use of fast-responding, monitoring-and-protection circuits to prevent damage during load transients. Among the monitoring circuits are those that detect excessive switching currents and sense the high-voltage-output level.

3.1 Description of the power path

The interconnections among the principal elements in the power stage are shown in Fig. 5. The regulator transistor (Q5) switches at 40 kHz to control the power flow from the +140-V input to the inductive-input inverter (Q1, Q2, Q3, and Q4), using a conduction range from 0 to 90 percent. The ratio of Q5 "on" time to the time of a single switching period provides the required average output voltage from the regulator in response to one of the current or voltage control systems in the power plant.

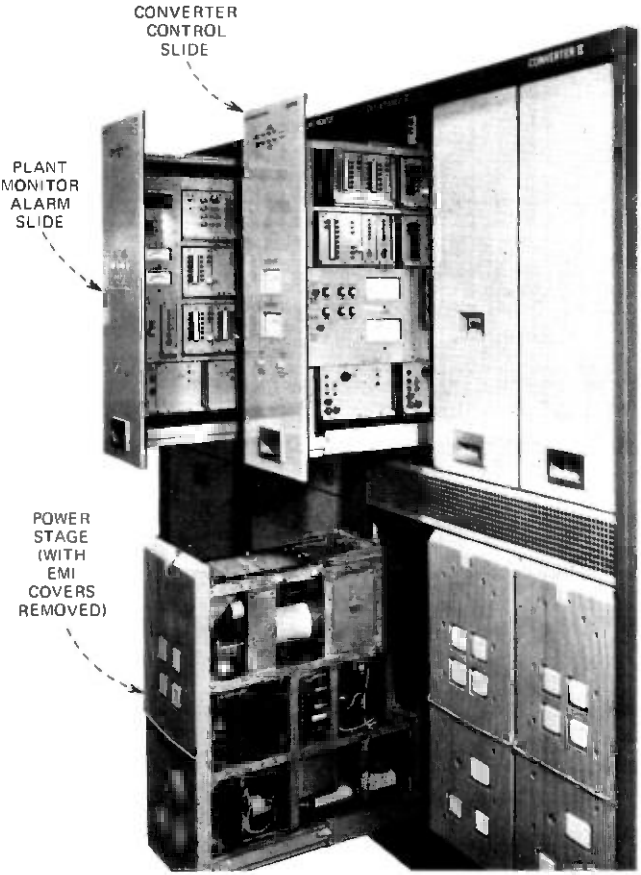


Fig. 4—Power stage extended on its slide from a converter bay.

Identical pulse-width-control signals are applied to all power stage regulators in a converter bay to force each power stage to support an approximately equal share of the load voltage.

The inductor L1, located between the regulator's output and the inverter's input, smoothes the voltage pulses from the regulator to establish a regulated direct-current flow into the inverter. This current flow is inverted into a regulated-amplitude, alternating current by the 20-kHz commutation of the Q1-Q2-Q3-Q4 bridge circuit. The bridge output is applied to a step-up transformer, full-wave rectified, filtered, and applied toward the external load. The inductor L1 is strategically located in the power path to limit the rate of rise of the current in the power semiconductors. The advantages of this special connection between the regulator and inverter in the power path are that it:

(*i*) Permits safe operation of all power semiconductors close to their maximum ratings.

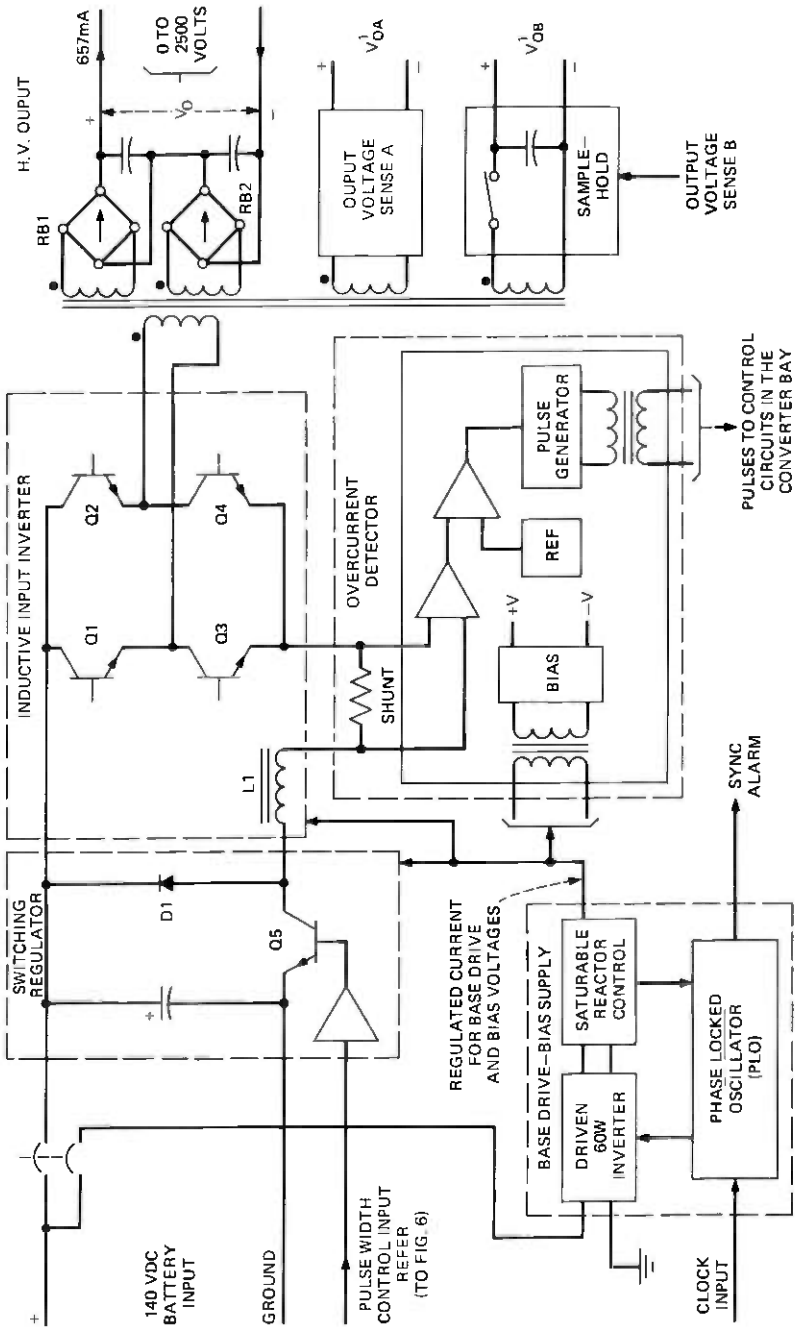


Fig. 5—Principal elements of a power stage circuit.

(*ii*) Controls peak currents in the bridge inverter even during commutation.

(*iii*) Allows easy adaptation of switching-loss-reduction networks which greatly reduce the transistor-peak-power dissipation and improve the efficiency by reducing the average switching loss.

The design and operational details of the basic power circuit topology, described above in general terms, have been documented.[6–10] Hence, the following discussion is concerned primarily with describing the overcurrent protection, the voltage sensing, the base-drive and bias supply, the physical realization, and EMI suppression.

3.2 Overcurrent protection

A fast-responding, overcurrent detector monitors the L1 inductor current that is the input current to the inverter (Fig. 5). The L1 inductor controls the rate of change of current in all semiconductors in the main power path. Thus, only a single overcurrent detector is required in each power stage. The overcurrent detector works in conjunction with the converter bay's control system. This combination limits the peak current in all power switches to be no more than 20 percent above normal under transient conditions, including those occurring in response to the application of a direct short circuit on the output.

The overcurrent detector consists of a shunt, an amplifier to raise the shunt's voltage level by 21 dB, a stable, direct-voltage reference, and a fast-slew-rate comparator that switches a pulse generator to the "on" state if the shunt current exceeds a predetermined threshold. The overcurrent detector has transformer isolation both in the output-pulse signal path and bias-power input path. This is required because the entire overcurrent detector circuit experiences common mode voltage excursions of 200 V at a 1000 V/μs rate during power switching. More details of the overcurrent detector can be found elsewhere.[11]

When the overcurrent detector's pulses are issued from a power stage, the control system in the converter bay halts conduction in the regulators in all power stages in the bay. This blocks further energy flow from the +140-V source to the inverters. Once this control loop becomes active, the cutoff intervals of the regulators are controlled to the proper duration to allow excess stored energy in the power path's magnetic components to be dissipated into the load. As a consequence, the switching currents in the inverter are reduced to normal levels before further regulator conduction is permitted.

3.3 Isolated output voltage sensing

Precise determination of a power stage's output voltage is required by the converter bay's regulation, alarm, and shutdown circuits. Each power stage contains sensing circuits that monitor the output transformer's voltage with sense windings. Direct sensing of the high voltage output is not desirable, since the secondary circuits may operate at several kilovolts above common. (This is a consequence of the series connection of the high-voltage outputs of the power stages.)

Unfortunately, a sense winding does not provide a voltage signal that has an ideal rectangular waveform, but one with large ringing transients and voltage notches at each polarity transition during inverter commutation. These aberrations vary considerably with load. They typically occur in power inverter circuits and cannot be completely avoided.

These disturbances excluded the use of conventional rectifier and passive filter circuits in this accurate, linear, sensing application. Instead, a unique application of a zero-order type sampling circuit is used to synchronously sample the voltage during a brief interval of the positive half-cycle of the sensed waveform midway between the commutation events. The sample switch is then held "off" to ignore the voltage ringing notches associated with commutation and the complete negative half-cycle. The charge stored on the "hold" capacitor maintains a relatively constant voltage across the high impedance load during the nonsampling part of the cycle.

Using this sampling scheme, tracking accuracy between the sense output and high-voltage output is better than ±1 percent of full scale over the 2500-V output range. The sensing circuits are duplicated in each power stage for improved reliability.

3.4 Internal base drive and bias supply

An efficient, 60-W, saturable-reactor-controlled, 20-kHz, quasi-rectangular-wave current source operates from the +140-V input to provide a ±2-A, regulated current to the inverter power transistors for base drive, to the regulator driver as a source for its base drive, and to the bias voltage supply in the overcurrent detector (Fig. 5). This ±2-A source determines the inverter commutation but not the regulator switching, since that is controlled by the converter bay's control circuits through the pulse-width-control input.

A phase-locked oscillator (PLO) operates with the base-drive source in a feedback system to minimize the variable, phase-shift error introduced by the saturable reactor. The PLO synchronizes the inverter's noisy commutation interval to the converter bay's clock with a phase-shift error of less than ±0.2 radian. The commutation transient always occurs during the "off time" of the regulator transistor and therefore does not contribute jitter to either the leading or trailing edge of the conducting period. An alarm is issued if synchronization fails to occur.

This synchronization feature prevents beat frequencies from appearing in the converter bay's output and other anomalous effects. Synchronization of all high-power switching also reduces the possibility of random noise tones exceeding the stringent power-plant-output-noise requirement of less than 0.17 μV for any single frequency in any 3-kHz band from 0.3 MHz to 30 MHz. Note, however, that loss of synchronization would not necessarily result in an increase in EMI.

3.5 Physical realization of and EMI control in the power stage

The power stage is physically realized as a completely shielded plug-in module that has input and output, hand-operated connectors and weighs approximately 56.7 kg (125 pounds). It is also key-interlocked and cannot be removed from its location at the bottom of the converter bay (Fig. 4) unless the high-voltage circuits within the modules are deenergized.

Shielding is utilized to minimize the effects of the radiated EMI noise produced by power-switching semiconductors within the unit. The shielding includes individually shielded compartments within the power stage as well as solid heat sinks on the front and rear of the unit, solid side covers with EMI gasketing, and honeycombed shields on the top and bottom of the unit.

The switching semiconductors, which are the main heat sources, are mounted on a heat sink that serves as the front panel of the unit. This takes advantage of both natural convection and radiation modes of heat transfer. Top and bottom honeycomb shields permit air flow through the power stage for internal cooling.

The transformer that connects the inverter output to the high-voltage rectifier was specially developed for this application. The high-current, low-voltage terminals are located on one side of the transformer, with the low-current, high-voltage terminals on the opposite side. The transformer is mounted on a solid wall separating two shielded compartments with its high-voltage terminals protruding through the wall, thus providing the output rectifier compartment of the power stage with minimum noise contamination from the inverter's switching circuits.

IV. PULSE POSITION MODULATION AND DEMODULATION

A major design problem in power-feed equipment providing regulated currents at high voltage is the implementation of current sensing in the high-voltage side. Very accurate current sensing is necessary for feedback control of the current as well as for alarm, shutdown, and metering functions. Analog magnetic amplifier technology has been successfully employed in previous high-voltage supplies. However, the need in the SG system for high precision (on the order of 0.01 percent) and the desire for less complex and lower cost apparatus made it necessary to develop a new technique for sensing current at high voltages.

The objectives of the new technique are to:

(*i*) Generate a signal that is an accurate linear function of the current being sensed.
(*ii*) Provide dc isolation in coupling the signal from the circuits referenced to the high-voltage side, where the current is sensed, to the signal-processing circuits referenced to the ground side.
(*iii*) Process the received signal to perform various control, alarm, shutdown and metering functions.
(*iv*) Provide operating power to the sensing circuits located in the high-voltage side, without employing an auxiliary power supply having high-voltage isolation.

The pulse-position-modulation (PPM) method[12] is used for current-sensing and information-coding. The PPM circuits convert a sensed analog signal into pulses whose position in time is varied relative to the occurrence of a trigger pulse. Pulse transformers couple the pulse from the PPM circuits, referenced to the high-voltage side of the power plant's output, to processing circuits referenced to the ground side. The pulse transformers are the only link between the PPM and processing circuits. To minimize EMI susceptibility, a low-impedance, balanced transmission line, terminated at each end with pulse transformers, is used to carry the pulse signals between the PPM circuit and the signal processing circuits.

The PPM method provides excellent signal-to-noise characteristics. Superimposed random noise has little effect on the time position of the pulse carrying the modulated signal. The PPM method is therefore well suited for interbay transmission of signals within various wire cables without appreciable reduction of the signal-to-noise ratio.

4.1 Regulation

The basic PPM circuit is used as the control element in multiple feedback loops controlling the output current and voltage of the power plant.

The principal feedback loop regulates the cable current at 657 mA. Figure 6 includes the current control loop. The switching transistor (Q5)* in the regulator section of each power stage is controlled by the state of a set/reset-type flip-flop in the power-stage-control circuit. Each switching transistor is turned on by a start-pulse signal originating at the 40-kHz clock. The start-pulse signal is applied, through isolation transformers, to the set terminal S of the flip-flop, setting it to a state that supplies a base-drive signal from the Q output to the switching transistor in each power stage.

* Same Q5 as in Fig. 5.

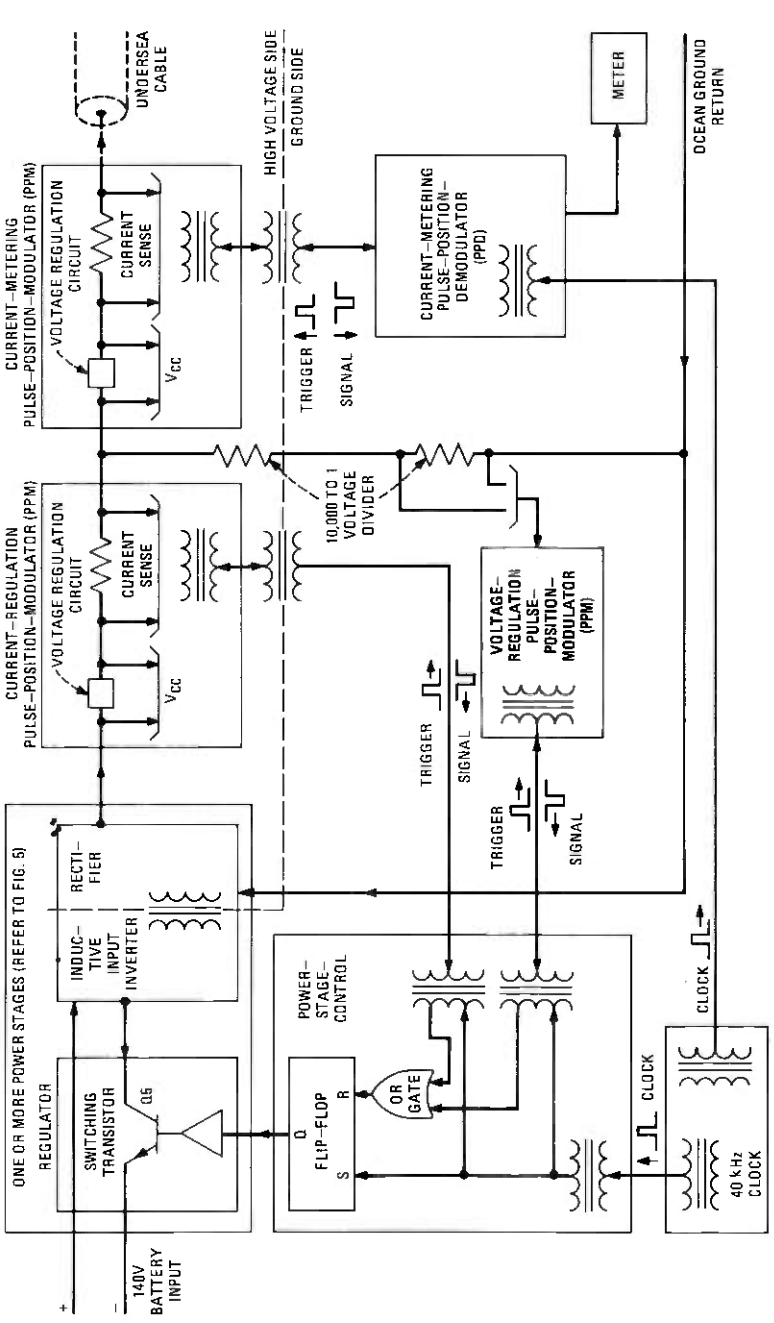


Fig. 6—Simplified circuit showing output-regulation and metering systems using pulse-position modulators.

At the same time, a positive-polarity trigger pulse is also transmitted from the 40-kHz clock to the current-regulation PPM circuit in the high-voltage side. The trigger pulse synchronizes the PPM circuit to the start of the switching-transistor-base-drive pulses. The signal pulse returned from the PPM circuit has a negative polarity. The trigger pulses and signal pulses share a common transmission path. The signal pulse is routed though an OR gate to the reset terminal R of the flip-flop. The flip-flop output is reset by the signal pulse to a state that terminates the base-drive pulse at the output Q. As a consequence, the regulator transistor in each power stage is turned off. The duty cycle (ratio of the time lapse between the trigger and signal pulses to the period of the trigger-pulse train) of the PPM output signal determines the width of the voltage signal to the output of the regulator and, ultimately, the current supplied to the cable.

Power to operate the current-sensing PPM circuits is obtained from the voltage (V_{cc}) developed across the voltage-regulation circuit in the PPM circuits, as shown in Fig. 6. The voltage drop across the voltage-regulation circuit in the PPM circuit is used to furnish 12 V to the PPM circuits. Sixty-mA line current is needed to power the PPM circuits. This permits the PPM circuits to effectively regulate the cable current at any value greater than 60 mA. The voltage-regulation circuit in the PPM circuit bypasses current in excess of 60 mA. The networks required to obtain closed-loop stability are inserted across the error amplifiers (not shown in Fig. 6) located within each PPM.

The current regulation loop has control of the cable current during normal operation. After an initial 5-hour warmup period, the maximum 24-hour current drift is less than ±0.05 percent if the ambient temperature variation stays within ±5°C. This represents a change of ±0.33 mA when the cable current is set to 657 mA.

The voltage-regulation loop is a secondary control loop designed to limit the output voltage to a preset level. Figure 6 illustrates one of the voltage-regulation PPM circuits. The power plant's output voltage is sensed by means of the 10,000-to-1 voltage divider.* A signal pulse is sent from the voltage-regulation PPM circuit through the isolation transformers and the OR gate to the set/reset flip-flop within the power-stage-control circuit. The OR gate, and hence the flip-flow circuit, reacts to the first pulse that arrives at the OR gate during each cycle of operation.

The voltage regulation loop has a combined setability and maximum drift of ±25 V.

* Note the contrast between the direct voltage sensing used in the power plant's output where the circuit common is close to earth potential and the indirect voltage sensing previously described for the converter bay's output where the common of one bay is the high-voltage side of the other.

4.2 Alarms and shutdowns using PPM circuits (not shown in Fig. 6)

The basic PPM circuit is also used in conjunction with pulse-sequence detection and decoder circuits to generate alarm and/or shutdown signals. Alarms of ±1 and ±3 percent are obtained from one PPM circuit sensing the cable current. The PPM circuit receives a trigger pulse from the clock circuit every 25 μs (a 40-kHz rate) and generates a signal pulse in the middle of the time period under normal conditions when the cable current is at its nominal value. The PPM circuit has a negative transfer function. Hence, an increase in the cable current shortens the time delay between the trigger and signal pulses. The signal pulse is sent to four sequence-detection circuits. Each circuit compares the occurrence of the signal pulse to a marker pulse generated by feeding a 2.5-MHz clock frequency* into a decoder circuit that quantizes the period between trigger pulses into 32 discrete time slots, each with a duration of 0.8 μs. Marker pulses are obtained at time slots 4, 12, 20, and 28. The marker pulse occurring at time slot 4 is generated 3.2 μs (4 × 0.8) after the trigger pulse. The marker pulses at time slots 12, 20, and 28 occur at 9.6 μs, 16 μs, and 22.4 μs after the trigger pulse. The PPM circuit has an active range of ±4 percent around the nominal cable current. As a consequence, each time slot is equivalent to 0.25 percent (8 percent ÷ 32) change in the sensed cable current.

No alarms will occur when the cable current is at 657 mA, and the signal pulse appears at time slot 16. When the sensed output current begins to increase, the PPM circuit will send a signal pulse to the sequence-detection circuit earlier than time slot 16. When the signal pulse appears earlier than the marker pulse occupying time slot 12, the sequence-detection circuit will generate a +1 percent current alarm. If the cable current continues to increase until the signal pulse appears before the marker pulse occupying time slot 4, a +3 percent current alarm is generated. Current alarms of −1 and −3 percent are generated when the sensed current decreases enough to cause the signal pulse to appear after the marker pulses occupying time slots 20 and 28, respectively.

Means are provided for in-service checking of the proper functioning of these alarms. Techniques used to check all four alarm circuits in a single operation are described elsewhere.[12]

With slight modifications, a PPM circuit can be used to sense either cable current or cable voltage with an active range suited to detect and send out alarm and/or shutdown signals for any desired percent change in the cable current or voltage.

* A single 2.5-MHz clock is used to provide, by division, the 40-kHz clock signal that synchronizes all clocks in the power supply.

4.3 Metering using PPM and PPD circuits

Figure 6 also includes a pulse-position-demodulation (PPD) circuit used with a PPM circuit to provide a metering function. The PPD circuit converts the time-delayed pulses from the PPM circuit into an analog signal after the pulse signals pass through the isolation transformers into the ground-side circuits.

A synchronizing trigger pulse is sent from the clock through the PPD to the PPM. The resulting PPM signal is a narrow pulse, linearly controlled in time delay by the amplitude of the sensed cable current. The signal pulses are sent to the PPD where they are used to set and reset a flip-flop circuit within the PPD. The output of the flip-flop is a rectangular-wave signal having a width proportional to the time delay of the signal pulse. This rectangular-wave signal is applied to an averaging filter producing a direct current output. The amplitude of the direct current output signal is a linear function of the time delay of the signal pulse. This direct current output can be used for either digital or analog metering.

Expanded-scale meter PPM and PPD circuits are used with digital meters to provide greater accuracy. The circuits have an active range of 16 mA centered around the nominal 657-mA cable current. The overall measurement accuracy at the 657-mA current is ±0.33 mA (±0.05 percent).

A technique for employing a PPD in the feedback loop of the PPM circuit to linearize its transfer function and render the system insensitive to timing errors is described elsewhere.[12]

4.4 Physical design considerations for PPM circuits

The operation of the PPM circuits with narrow pulses and consequentially broad frequency spectrum in a high-electromagnetic-interference and high-electrostatic-voltage environment required that considerable care be exercised in the physical design of both the PPM circuits and the bays where they are mounted.

The components of the PPM circuits are mounted on three-layer printed-wiring boards and are located, with their associated pulse transformers, in the high-voltage areas of the converter and plant monitor bays. The components of the low-voltage PPD circuits are also mounted on three-layer, printed-wiring boards. Several of these boards comprise a low-voltage plug-in module, located in one of the pull-out equipment slides in a converter or plant monitor bay. Access to PPM circuits located in high-voltage areas is possible only when the circuits are deenergized, while the low-voltage PPD circuits located in the ground-side areas are always available to the operator.

The three-layer boards provide the PPM and PPD circuits with two layers of interconnecting paths and a third layer that is used as a ground

plane. The ground plane consists of very wide ground paths to all integrated circuits, thus providing a low-impedance ground that aids in reducing susceptibility to noise pulses.

Figure 7 displays the location of the four PPM circuits (one current-regulation, two alarm and shutdown, and one current-metering) in the high-voltage area of a converter bay. Two PPM circuit boards are mounted on each metal base plate with a perforated metal covering over both boards. This cover and the metal base plate are electrically con-

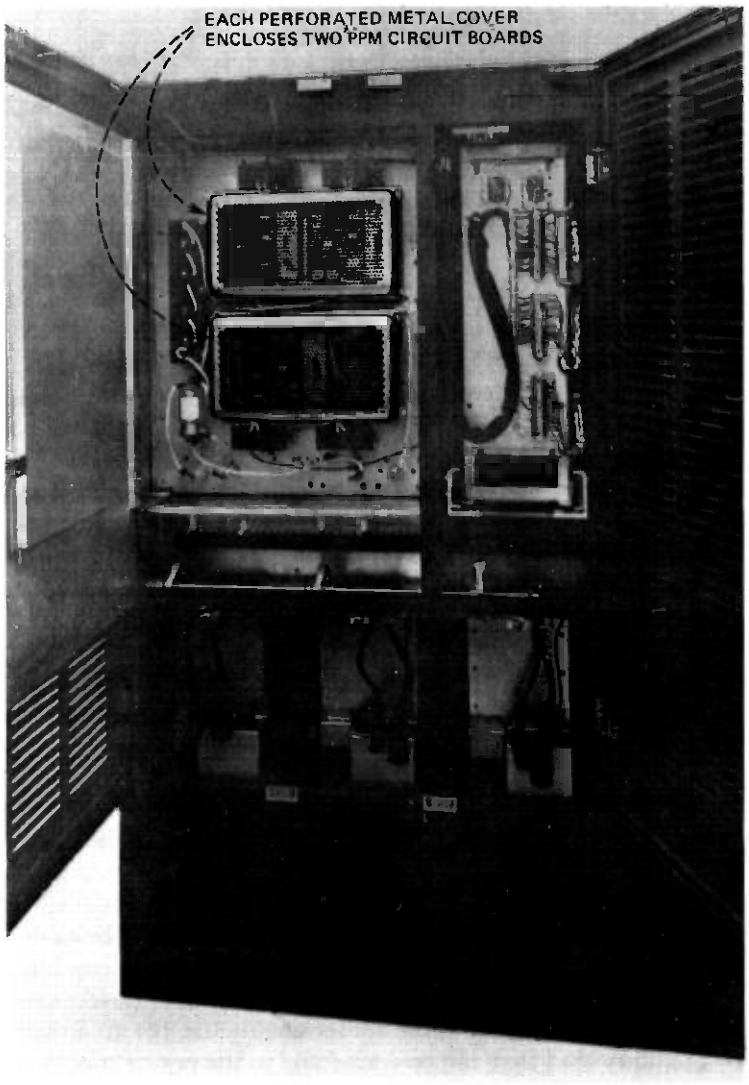


Fig. 7—Pulse-position-modulators shown in the high-voltage area of a converter bay.

nected to the high-voltage circuit to form an equipotential electrostatic field that surrounds the integrated circuits in the modulators. This arrangement prevents any damage to the integrated circuits by eliminating the possibility of large voltage gradients in the surrounding field. The cover and base plate also reduce electrostatic coupling of EMI into the PPM circuits.

The current-regulation PPM in a converter bay is equipped with coarse and fine cable-current-adjust potentiometers, that are electrically at high voltage. These potentiometers are adjusted from the front of the converter bay by means of long, nonconducting shafts.

V. SHORE STATION POWER SEPARATION FILTER

The shore station Power Separation Filter (PSF) bay is so named because it is the bay in which the wideband signal transmission path and the high-voltage dc power path, which share the undersea cable's center conductor, are separated for connection to their respective terminal equipment. Viewed in simplified form, the PSF circuit is a three-port filter, with an all-pass port for connection to the undersea cable, a high-pass port for connection to the wideband line transmission equipment (WLE),[13] and a low-pass port for connection to the power supply.

There are two fundamental but diametrically opposite design requirements for the PSF circuit: (*i*) to pass the wideband signal with some minimum acceptable values of return loss and insertion loss, which requires that components be small and circuit paths short; and (*ii*) to keep high-voltage partial discharge activity ("corona") below some appropriate threshold that requires that components be large and circuit paths widely spaced and long, in order to minimize destructive charge transfer in dielectrics and to satisfy transmission impulse noise objectives. For the SG system, with an upper frequency of 30 MHz and a maximum operating voltage of 7500 Vdc, the resulting dilemma was formidable and required extensive changes in the electrical and mechanical configurations compared to shore station PSFs of earlier undersea cable systems.

In addition to separating the power and signal paths, the PSF bay performs a number of important additional functions in the operation, maintenance, and safety of the cable system. Specifically, it provides: (*i*) proper electrical and mechanical termination of the undersea cable, (*ii*) a 75:50-ohm impedance match between the WLE and the cable, (*iii*) extensive shielding and filtering to prevent EMI from reaching the wideband signal path, (*iv*) facility for quick metallic connection of test equipment to the undersea cable center conductor, (*v*) an adjustable, 5-kW, auxiliary load that can be connected to the power supply in lieu of the undersea cable system, (*vi*) extensive high-voltage protection for

equipment and personnel, and (*vii*) in conjunction with the power supply, a key-interlocked safety system that synchronizes their mode of operation and controls access to hazardous voltage compartments.

A simplified diagram of the PSF circuit is shown in Fig. 8. The broken lines represent copper compartments, which serve the dual purpose of EMI shielding and high-voltage protection. The end of the undersea cable connects to the cable compartment, and the wideband signal path is separated from the high-voltage power path by C1 and L. The signal path proceeds from C1 through the transmission compartment to the WLE port. From L, the high-voltage power path goes through the wall of the inner compartment to switch S, through filter network N, to the power supply port. Switch S is a large, key-interlocked, rotary power switch that controls the mode of operation of the PSF and access to its high-voltage compartments. Auxiliary contacts (not shown) operate vacuum relay K.

The dominant requirement on the design of the signal path is a return loss of ≥20 dB over the entire SG band. This requires that impedance discontinuities be kept to a minimum. These occur primarily in the cable compartment, where high voltage considerations require large clearances, resulting in significant structural variations from the desired signal path impedance. The 20-dB return loss requirement is equivalent to a voltage reflection coefficient $|\rho| \leq 0.1$, since return loss = $20 \log |1/\rho|$. For a cable of characteristic impedance Z_0, terminated in an impedance Z_T,

$$\rho = \frac{Z_T - Z_0}{Z_T + Z_0}.$$

Let the PSF be represented by its equivalent series impedance $Z_S = R_S + jX_S$, and let each signal port be terminated in its characteristic impedance, Z_0 (both ports referred to the same impedance level). Then the reflection coefficient at either port is:

$$\rho = \frac{(Z_S + Z_0) - Z_0}{(Z_S + Z_0) + Z_0} = \frac{Z_S}{Z_S + 2Z_0}.$$

The series resistance is small, and at the higher frequencies lead inductance dominates, so that

$$\rho = \frac{jX_S}{2Z_0 + jX_S},$$

whence for

$$|\rho| \leq 0.1, X_S \leq \sqrt{4/99}\, Z_0 \simeq 0.2\, Z_0.$$

This result provides a means of estimating the maximum allowable lead length through the cable compartment. At the undersea cable port, $X_S = 0.2 \times 50 = 10\ \Omega$, and at the top frequency of 30 MHz, an allowable

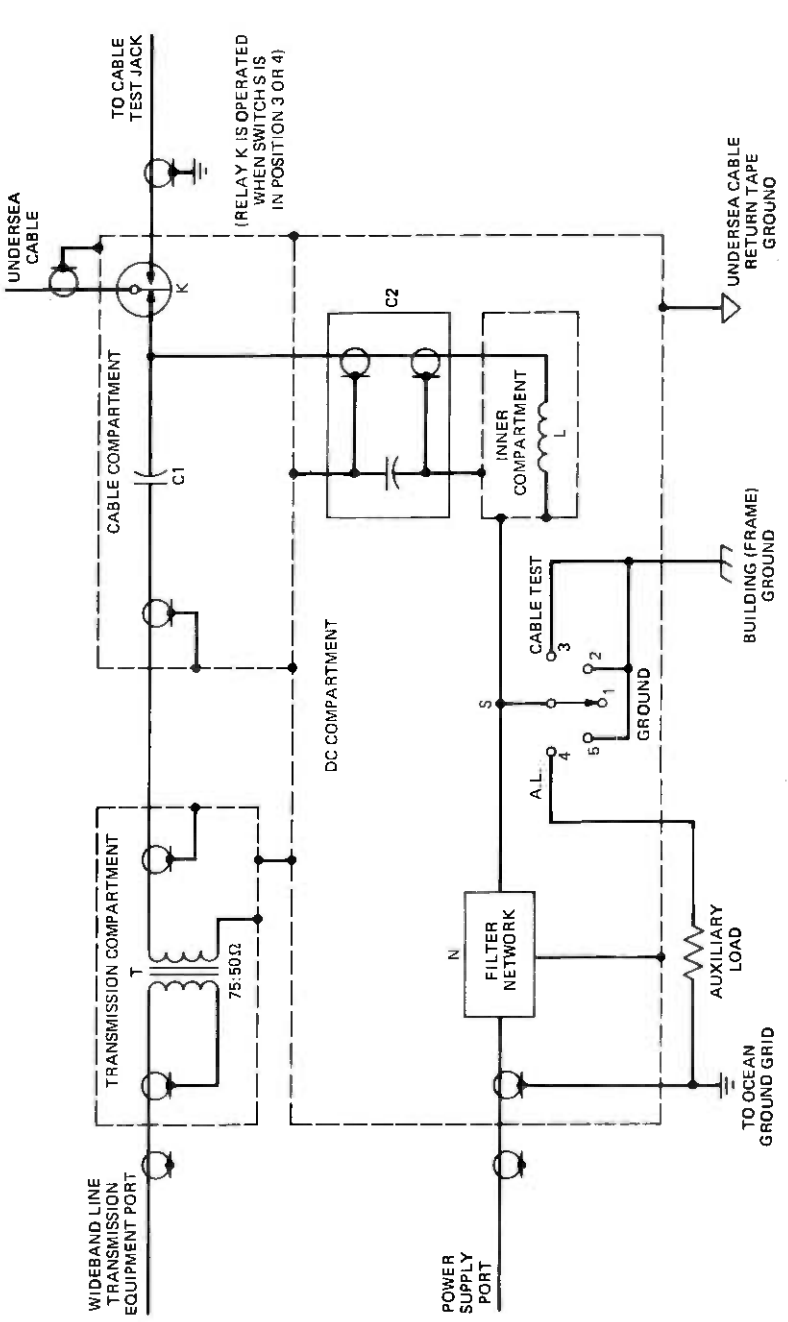


Fig. 8—Simplified circuit of the shore-station power-separation filter.

equivalent series inductance of 53 nH is obtained. Assuming, for illustration, a distributed lead inductance of 3.94 nH/cm (10 nH/inch), the maximum lead length is 13.5 cm (5.31 in.).

To attain the 53-nH requirement, the equivalent series inductance can be reduced by (*i*) shortening the path length, (*ii*) reducing the distributed inductance per inch, or (*iii*) adding shunt compensating capacitance. The path length was minimized by using a small high-voltage vacuum relay (K) to perform the required transmission path switching and by using the small capacitor (type 732A) designed for the SG repeater as the high-voltage blocking capacitor, C1. The distributed inductance of the remaining path was reduced by creating the separate cable compartment for the high-voltage portion of the signal path. The radial dimension of this compartment was made as small as possible (about 5-cm clearance from path to ground plane) consistent with the high-voltage, partial-discharge, activity requirements, thereby reducing the distributed inductance. Shunt compensating capacitance was not added, as this would have required the introduction of additional components into the high-voltage circuits in the cable compartment, increasing the possible sources of impulse noise and reducing reliability.

The actual, achieved path length of open leads through the cable compartment was approximately 20.3 cm (8 in.). The actual achieved return loss was greater than 30 dB over most of the 0.5- to 30-MHz band. Minimum values were 25.3 dB on a laboratory-built exact prototype, and 25.7 dB on the Green Hill PSF.

The principal requirements affecting the design of the high-voltage path are those concerned with high-voltage, partial-discharge activity, which must be kept below some appropriate level to satisfy signal-path impulse-noise objectives and to minimize destructive charge transfer in dielectrics. (The term "partial discharge activity" is preferred over "corona," since it more precisely describes the phenomenon taking place, viz., intermittent partial discharges of the dielectric. Strictly speaking, corona is a continuous discharge accompanied by a glow.) An excellent analysis of high-voltage partial discharge considerations in an SG-type cable system is given by Franke.[14]

Impulse-noise objectives established for SG terminals for the 48-kHz group data bands allow one "pop" (partial discharge pulse) per 15 minutes exceeding a threshold of −10 dBmO. This is equivalent[14] to a peak instantaneous voltage threshold of 100 mV from a 48-kHz bandpass filter placed anywhere in the SG signal band. Since there are a number of possible sources of impulse noise in a terminal other than the PSF, the actual pop rate allocated to the PSF should be significantly less than one each 15 minutes.

Attempting to establish limits to minimize the destructive effects is more difficult. Destructive charge transfer limits for dielectrics are

nebulous and difficult to specify. Franke suggests a maximum value of 10 pC, based on levels permitted in previous undersea cable systems, with a rate corresponding to that of the data transmission interference requirement. For a maximum pulse duration of 50 ns in a 50-ohm system, he calculates a peak voltage threshold of 5 mV.

If the impulse noise and destructive charge transfer requirements are now combined, the resultant partial discharge activity requirement is a threshold of 5 mV, with a rate limitation of under 1 pop per hour.

The creation of a separate cable compartment not only aided in the solution of signal-transmission design problems, but also provided an effective way to isolate the high-voltage and signal paths, and thereby attenuate any noise which might otherwise be introduced via the high-voltage path. High voltage is fed from the dc compartment into the cable compartment via a novel skin-effect, low-pass filter* consisting of L, C2, and the inner compartment. C2 is a feed-through, ground-separating capacitor (type 729A) constructed like a coaxial cable. The feed-through lead is the center conductor, and the outer conductor is composed of two high-voltage, insulated, concentrically wrapped foils, each of which is connected to a ground cap at opposite ends of the cylindrical structure. At dc, the two ground ends of C2 are isolated, but at higher frequencies, they are effectively connected, and C2 is equivalent to a short piece of coaxial cable. One ground of C2 is connected to the cable compartment, while the other ground is connected to the inner compartment. The high-voltage path from the power supply is connected to the outside surface of the inner compartment. Hence, the inner compartment floats at the cable supply voltage, while for ac signals it is effectively an extension of the cable compartment (undersea-cable-return-tape ground).

The inside surface of the inner compartment is connected to L, which bridges the high voltage onto the signal path via the center conductor of C2. The high-voltage path consequently passes through the wall of the inner compartment, and any noise following this path is attenuated by the skin effect of the 0.0794-cm (0.0313-in.) wall. Using the classical formula for current penetration, it can be shown that, for thickness, t (cm), and frequency, f (Hz), at 20°C,

$$\text{Attenuation} = 20 \log_{10} e^{-0.151 t \sqrt{f}},$$

which, for $f = 0.5$ MHz, gives 73.6 dB.

Filter network N provides attenuation for the 20-kHz converter frequency and its harmonics and protects against tones which might be radiated or conducted around the power supply output filters. Together,

* The skin-effect filter concept is from G. H. Deshaies of Bell Laboratories, North Andover, Massachusetts, and was used in the PSFs of the L5 land transmission coaxial system.

N and the skin-effect filter provide a minimum loss of 80 dB from 20 kHz to beyond 60 MHz, as measured from the power supply port to the signal path.

High-voltage performance of the prototype and TAT-6 PSFs was evaluated using a wideband partial discharge detector developed by Franke and Czekaj.[15] With the detector connected to the 50-ohm signal path, the prototype had no pops exceeding 1 mV in 86 hours at +10 kV, and no pops exceeding 2 mV in 56 hours at −10 kV. The Green Hill PSF had no pops exceeding 1.4 mV in 8 hours at +10 kV, and the St. Hilaire PSF had no pops exceeding 1.4 mV in 15 hours at −10 kV. Since normal PSF operating voltages are less than 7.5 kV, these results provide a high degree of confidence that PSF high-voltage performance will not only meet but exceed design objectives.

VI. SUMMARY

The TAT-6, shore-station, power-feed equipment has been described. A modified version has been developed and installed and is in service on Cable Ship *Long Lines*. The shipboard power-feed equipment can power any long-haul, undersea cable in service at the time of the shipboard installation of the power-feed equipment.

VII. ACKNOWLEDGMENTS

The multiple-team concept that produced the SB, SD, and SF power-feed equipment was continued in the SG development. Significant contributions have been made by many individuals; directly by those participating as members of the development teams and indirectly by others. It is impractical to acknowledge all of the team members directly involved.

B. H. Hamilton, as leader of the power-supply-circuit-development team, is primarily responsible for both the system concept and specific techniques used in the power conversion, control, monitoring, protection, and switching circuits. From among the several members of this team, E. T. Calkin, W. J. Schatz, and R. E. Schroeder were selected as contributors to this article.

S. Mottel provided leadership for the power-supply-equipment-development team that is represented among the writers by I. Golioto. The physical design team was responsible for all aspects of the physical design of the power supply's bays, including the development and implementation of physical design techniques related to EMI suppression, maintainability, component selection, and installation methods, and the generation of manufacturing information.

W. G. Ramsey supervised the activities of D. S. Shull and R. E. Curlee, who were responsible for all aspects of the development of the terminal PSF equipment for both cable installation use and permanent system

operation. Mr. Shull did the general planning and circuit development, and Mr. Curlee did the physical design.

Close cooperation among all three teams contributed to the successful development of the power-feed equipment.

REFERENCES

1. J. D. Bishop and S. Mottel, "SD System: Cable Power Facility," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1339–1366.
2. E. T. Calkin and I. Golioto, "SF System: Power Conversion," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 749–765.
3. J. R. Walters and G. B. Fernie, "Submarine-Cable Earth-Electrode Systems," Post Off. Elec. Eng. J., *57* (July 1964), pp. 105–108.
4. G. A. Axe, "The Effects of the Earth's Magnetism on Submarine Cables," Post Off. Elec. Eng. J., *61*, No. 4 (April 1968), p. 41.
5. I. Golioto, "High Voltage Transfer Switch With Cam Controlled Overlap During Transfer," U. S. Patent 4,016,385, applied for October 8, 1975, issued April 5, 1977.
6. E. T. Calkin and B. H. Hamilton, "A Conceptually New Approach for Regulated DC to DC Converters Employing Transistor Switches and Pulsewidth Control," IEEE Trans. Ind. App., *1A-12*, No. 4 (July–August 1976), pp. 369–377.
7. E. T. Calkin and B. H. Hamilton, "Circuit Techniques for Improving the Switching Loci of Transistor Switches in Switching Regulators," IEEE Trans. Ind. App., *1A-12*, No. 4 (July–August 1976), pp. 364–369.
8. E. T. Calkin, B. H. Hamilton, and F. C. La Porta, "Regulated DC to DC Converter with Regulated Current Source Driving a Non-Regulated Inverter," U. S. Patent 3,737,755, applied for March 1972, issued June 1973.
9. E. T. Calkin, B. H. Hamilton, and F. C. La Porta, "Switching Regulator with Network to Reduce Turnon Power Losses in the Switching Transistor," U. S. Patent 3,745,444, applied for March 1972, issued July 1973.
10. E. T. Calkin, B. H. Hamilton, and F. C. La Porta, "Switching Regulator with High Efficiency Turnoff Loss Reduction Network," U. S. Patent 3,736,495, applied for March 1972, issued May 1973.
11. B. H. Hamilton, F. C. La Porta, and R. E. Schroeder, "DC to DC Converter with Regulation Having Accelerated Soft Start Into Active Control Region of Regulation and Fast Response Overcurrent Limiting Features," U. S. Patent 3,879,647, applied for June 7, 1974, issued April 22, 1975.
12. B. H. Hamilton, "Pulse Modulation and Signal Processing Circuits to Perform Precise Control and Monitoring Functions with High-Voltage Isolation," IEEE Trans. Ind. Appl., *1A-12*, No. 4 (July–August 1976), pp. 378–386.
13. M. Brouant, C. Chalhoub, P. Delage, D. N. Harper, H. Soulier, and R. L. Lynch, "SG Undersea Cable System: Terminal Transmission Equipment," B.S.T.J., this issue, pp. 2471–2496.
14. E. A. Franke, "Corona Considerations in Submarine Cable Communications Systems," IEEE Trans. Elect. Insul., *EI-9*, No. 4 (December 1974), pp. 150–154.
15. E. A. Franke and E. Czekaj, "Wide-Band Partial Discharge Detector," IEEE Trans. Elect. Insul., *EI-10*, No. 4 (December 1975), pp. 112–116.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Installation and Maintenance of the Undersea System

By J. E. H. COSIER, A. P. DAVIES, S. W. DAWSON, Jr., R. F. GLEASON, F. E. KIRKLAND, and T. A. McKENZIE

(Manuscript received May 16, 1978)

Many changes were made in Cable Ship Long Lines *to facilitate installation and maintenance of SG cable and repeaters. This paper describes new equipment provided for the ship and modifications made to existing equipment. Both transmission equipment and cable and repeater handling facilities are discussed.*

I. BACKGROUND

A significant part of the development of a new undersea cable system is involved with the equipment required to install and maintain the system. Much of the basic work concerning cable payout and repeater handling was done as part of the earlier SD Submarine Cable System and the design of Cable Ship *Long Lines*.[1] Even so, for the SG system, many specific changes were necessary. This paper describes the most important ones. The areas or conditions that called for revision or new design can be summarized as follows:

(*i*) *Electrical*
  (*a*) Test sets
  (*b*) Power feeding equipment
  (*c*) Transmission equipment
  (*d*) Computer facility
  (*e*) Repeater monitoring set

(*ii*) *Environmental*
  (*a*) Repeater temperature control and measurement
  (*b*) Cable temperature control and measurement

(*iii*) *Physical and Mechanical*
  (*a*) Increased cable tensions and cable size

(*b*) Large number of repeaters and equalizers
(*c*) Burying of cable and repeaters
(*d*) Grapnels.

II. TEST SETS AND PROCEDURES

Installation of a system requires a continuous program of testing beginning when the cable, repeaters, and equalizers are loaded aboard ship and ending with the final splice completing the undersea link. The test equipment, test procedures, and computations are designed to ensure the proper performance of the system, to optimize equalizer settings, and to preserve the acquired data for future use.

2.1 Test set philosophy and design

New test equipment was designed to facilitate the installation of SG submarine cable systems. The design allows the use of identical test equipment on the ship and at shore terminals. The new installation test equipment designs include the cable laying test set, the repeater monitoring set, the SG high frequency line equipment, the shipboard power-feeding equipment, and a new shipboard computer facility.

2.1.1 Cable laying test sets

The cable laying test set (CLTS) was developed for the purpose of making automatic, simultaneous, two-way transmission measurements on SG and SF cable systems during installation and commissioning.[2] As shown in Fig. 1, the cable laying test set has a transmit section and a receive section. The transmit section consists of an oscillator with precision output level control and a digital control unit, while the receive section consists of a selective detector, a digital control unit, and signal monitor unit. Signaling to achieve automatic control between the transmit section and the receive section is accomplished by set command tones. By using a patching arrangement, this signaling scheme allows the transmit and receive sections of the same CLTS to work either directly with each other or with the receive and transmit sections of another CLTS at the other end of a cable system. In the latter case, control is via an order wire channel over the link being installed.

Each section of the CLTS has one manual and five automatic modes of operation. In the manual mode, the transmit and receive digital control units are disabled. The transmit section simply becomes a manually tuned oscillator, and the receive section becomes a manually tuned selective detector with a signal monitoring capability. The signal monitor unit detects the signal power received by the selective detector and provides an audible and visible alarm when the received power level varies by more than ± 1.5 dB.

The five automatic modes of operation are: SF low band, SF high band,

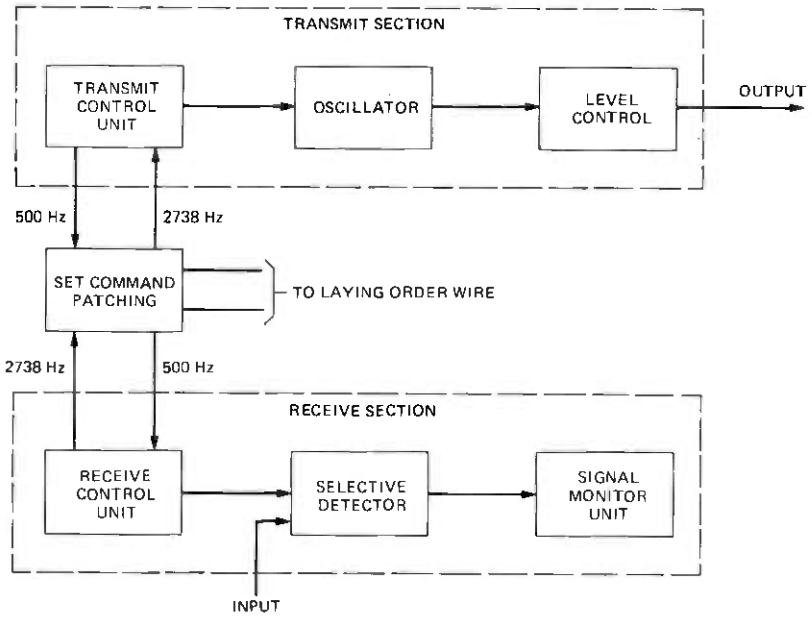


Fig. 1—Cable laying test set.

SG low band, SG high band, and RIPPLE. The RIPPLE mode is the most flexible, and all other modes are special cases of it. In the RIPPLE mode, the transmit and receive operator(s) choose a start frequency, stop frequency, and step frequency interval by setting thumbwheel switches on the front panels of the control units. When the transmit control unit is initialized, it causes the transmit section to begin sending a test tone at the start frequency and also sends a set command tone to the receive section, causing it to measure the received power at the start frequency. When the measurement is complete, the frequency and measured power are recorded via the data translator and teletypewriter (TTY). The receive control unit then sends a set command tone back to the transmit control unit, causing it to step to the next test frequency. To step to the next test frequency, the transmit control unit increments the oscillator frequency in 1-kHz steps at 10,000 steps per second until the next test frequency is reached. The receive control unit increments the measurement frequency in a similar fashion and begins a new cycle. This process continues until the measurement at the stop frequency has been completed and the data have been recorded.

All other automatic modes of operation are similar, except that the start and stop frequencies are fixed, standard values and the step frequency intervals vary in size between the standard frequencies. For these

modes of operation, the start, stop, and step frequency intervals are stored in read-only memory (ROM) in the control units.

The oscillator and selective detector units used in the CLTS are standard Western Electric transmission measuring sets, and thus the basic measurement capability of the CLTS is determined by these units. Both the oscillator and the selective detector are designed to cover the frequency range from 10 kHz to 60 MHz. The level control unit in the transmit section provides an output of 0.00 ± 0.02 dBm (into 75 ohms), which can be attenuated to −99.9 dBm in 0.1-dB increments. The selective detector has a measurement range from 0 to −129.9 dBm. In all automatic modes, measurements can be made over a range from zero to −109 dBm with a readout resolution and repeatability of ±0.01 dB.

All the equipment associated with a CLTS, including power supplies and a blower for cooling, is packaged in one bay. Two complete CLTS bays are installed aboard ship, and two additional CLTS bays are installed (on a temporary basis) at the shore terminal from which cable is being laid, thus providing redundancy at both ends of the cable.

2.1.2 Power feeding equipment

New shipboard power feeding equipment was designed for the installation of SG submarine cables.[3]

2.1.3 Transmission equipment

The SG high-frequency line equipment was developed for the installation of SG cable systems. It is permanently installed aboard the Cable Ship *Long Lines* and temporarily installed at cable stations from which cable is laid. This allows the undersea cable system to be installed independent of the installation of the terminal wideband line equipment. The high-frequency line serves as the interface between all the transmission test facilities and the cable system.

The major function of the transmit high-frequency line is to provide level adjustment capability for transmission test tones and the 12-channel laying orderwire signal to obtain acceptable signal-to-noise ratios for each. At the same time, the broadband power transmitted must be limited to a value such that no repeater is overloaded. These functions are achieved for the conditions of transmitting into an inboard- or outboard-end section of cable (whose length can range from essentially zero to a full repeater section) or into the test lead of an ocean block equalizer.[4]

The receive high-frequency line provides three paths of independently adjustable gain for the broadband signal received from the cable system. First, the gain of the transmission test tone path is adjustable to allow for as much gain as possible while limiting the signal level applied to the

receive cable laying test set to less than 0 dBm. Second, the gain of the orderwire path is adjustable to achieve the proper transmission level at the input to the multiplex equipment. Finally, the gain of the supervisory tone path is adjustable so that the power of the supervisory tones at the input to the repeater monitoring set will lie between −40 and −100 dBm. Again, the implementation of the high-frequency line allows all these requirements to be met for receiving at an inboard or outboard cable end section or at an ocean block equalizer test lead.

Two high-frequency line units, regular and spare, are included in each high-frequency line bay. Each unit consists of a directional filter for separating transmit and receive signals, fixed-gain broadband amplifiers, variable attenuators, and hybrid transformers. Regular and spare SG laying orderwire multiplex equipment is also included in each high-frequency line bay. The multiplex units contain the stages of modulation necessary to translate the existing SF laying orderwire spectrum into the SG bands. In addition, the high-frequency line bay contains a patch panel to facilitate the convenient patching of all signals and test equipment used for transmission testing during the installation.

Because of the many different situations that can arise during laying, flexibility of patching the high-frequency line equipment received considerable attention. For example, the use of broadband amplifiers and attenuators throughout the high-frequency line having essentially flat response over the frequency range from 500 kHz to 60 MHz allows the transmit and receive bands to be determined simply by patching the directional filter. The design of the directional filter requires a large band-to-band rejection to preclude leakage from the relatively high-level transmitted signal from overloading the low-level receive section of the high-frequency line. The SG high-frequency line directional filter has an in-band loss of less than 1 dB in each band and a band-to-band rejection of greater than 65 dB.

2.1.4 Computer facility equipment

The advantages of a shipboard computing facility have been described previously.[5] Figure 2 is a block diagram of the new computing facility aboard Cable Ship *Long Lines*. It consists of a programmable calculator with 15,000-word memory and an internal cassette memory, two additional cassette memory units for program and data storage, a thermal page printer, a plotter, and an optical paper tape reader.

During a cable lay, a detailed record-keeping procedure is employed which consists of entering data and performing computations on a specially prepared set of printed forms. Both the shore terminal and the ship maintain a complete set of all data and computations. The computer software is designed to replace the manual record-keeping procedure

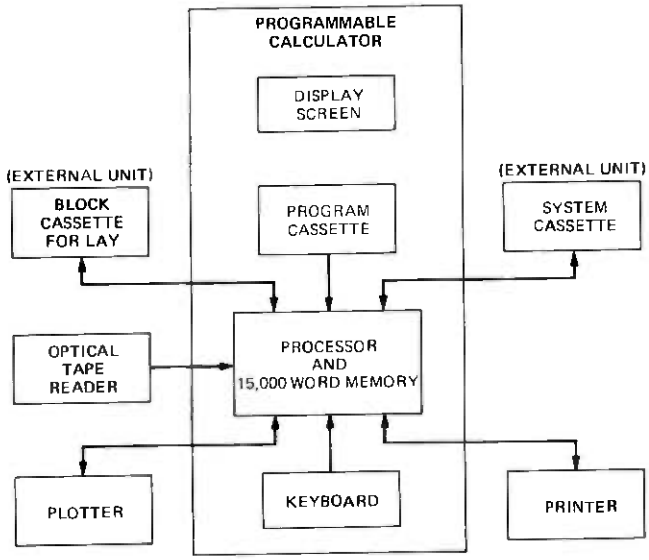


Fig. 2—Shipboard computer facility.

aboard ship. For this reason, it prints out data and computations on pages which replace the forms which would have been filled out manually. Thus, in the event of a computer failure, the manual computations can easily be resumed. In addition, a software package allows operator-controlled manipulation of various data for special calculations during a cable lay.

2.1.5 Repeater monitoring set

The repeater monitoring set is designed to be used on the SG system to measure supervisory tones from repeaters as well as to measure the intermodulation performance of repeaters.[4] Unlike the equipment described above, which is used only during installation and commissioning,[6] the repeater monitoring set is a permanent piece of terminal station equipment which is also used during installation. The supervisory tone measurements allow for data acquisition necessary for system administration and for fault localization. Intermodulation tests, which are normally performed on an out-of-service basis, are helpful in localizing a wide variety of faults.

The repeater monitoring set can be operated in a manual mode or in a semi-automatic mode. In the manual mode, input parameters are set on thumbwheel switches by an operator. In the automatic mode, input parameters are entered from mark-sense cards by a card reader. The repeater monitoring set has a visual output of measured data as well as an output suitable for driving a TTY. The set can also be used to measure supervisory tones on the SF system.

For supervisory tone measurements, the repeater monitoring set uses a phase-locked selective detector technique. The measurement bandwidth is 5 Hz, and all the conversion carriers in the selective detector are phase-locked to a highly precise common reference frequency, allowing accurate frequency measurement of the received tones.

The receiver, whose automatic frequency control loop contains a 50-Hz bandwidth IF filter, has an acquisition range of ±25 Hz for signals above −110 dBm. Once phase lock is achieved, the frequency and power of the received signal are measured automatically over the range of −100 to −40 dBm. Input bandpass filters, used to prevent overload of the detector by the broadband message signal, can be bypassed, enabling the set to be used as a general-purpose selective detector over the 10-kHz to 60-MHz range.

Because the SG repeaters are highly linear, a special technique was required to obtain accurate intermodulation measurements of individual repeaters on an installed system. This particular measurement is made possible by the common-amplifier repeater configuration and by the different distance and hence different round-trip delay to each repeater. Chirp radar and matched-filter techniques are used to obtain the necessary signal-to-noise ratio under the peak power constraints of the repeaters.

A linear FM (chirp) signal 100 kHz wide with a 10-Hz repetition rate is transmitted together with a single frequency tone. In the low band, an 8.6- to 8.7-MHz chirp and a 12-MHz tone are transmitted. Each repeater produces a second-order intermodulation product whose frequency sweeps from 20.6 to 20.7 MHz and is received in the high band. The return from a given repeater is recovered by demodulating the received signals with a swept carrier identical to the transmitted chirp but translated in frequency by an amount corresponding to the round-trip delay to the given repeater. A single IF signal results (in this case, 20 MHz) which is measured by the 5-Hz bandwidth selective detector. Returns from adjacent repeaters are separated by approximately 100-Hz intervals and are rejected by the detector. The effects of delay distortion on the signal are minimized by choosing the transmit and receive frequencies such that the round-trip delay over the 100-kHz chirp band is constant.

Successful implementation of this method requires precise phase control of two synchronized frequency generators.

2.2 Test procedures

2.2.1 Tests during loading

Repeaters, ocean-block equalizers, and shore-controlled equalizers are individually tested during manufacture and then transported to the dock for loading aboard ship. Following loading and prior to splicing,

resistance checks are made to determine that all repeater housings have a high dc resistance to ground and that all equalizer housings are solidly grounded to the ship's hull. The purpose of these pre-checks is to ensure the validity of the outer-conductor dc resistance tests which are used later to determine cable temperature. The cable, repeaters, and equalizers are then spliced together to form an assembled shipload.

2.2.2 Assembled shipload tests

The assembled shipload is subjected to a series of tests to ensure proper operation. The diagram in Fig. 3 shows the equipment and connections used to perform these tests on a block-by-block basis. Individual block testing is necessary because the differences in temperature and pressure between shipboard and ocean bottom environment can result in large temporary end-to-end misalignment which makes a single measurement of an entire shipload impractical. Block-by-block measurements have some useful benefits. Testing procedures and computations can be mechanized in a generally repetitive manner, and single block tests automatically localize faults to a block.

Each end of the assembled shipload is terminated at a power separation filter. The power separation filters allow high voltage to be applied to the center conductor at one end and dc ground at the other. A broadband connection is also provided for test signals. Transmission tests are performed on each of the blocks (including the partial end blocks) at a selected set of standard frequencies in each band.

For discussion of the actual transmission tests, refer again to Fig. 3. All transmission measurements are made by either the cable laying test sets (CLTSs) or the repeater monitoring set (RMS). For these tests, the CLTSs transmit and receive signals at a standard list of stored frequencies. All the test equipment as well as the orderwire is connected to the assembled cable system through the high-frequency line equipment. The CLTSs are used to measure the transmission response of each block. The measurement starts at the first test frequency and automatically sequences through the standard frequency list. The measured data are automatically recorded by means of the data translator and the teletypewriter in both printed page and punched-paper-tape form. The data translator also causes the TTY to print out heading information which identifies the type of measurement made, the date and time of day, the direction of transmission, the test sets used to make the measurement, the run number, and the block number. The CLTSs are also used to measure all the losses and gains associated with the test configuration (e.g., patch cords and power separation filter). The punched paper tapes allow automatic loading of test data into the shipboard test room computer facility.

Computations are performed using the measurement data described

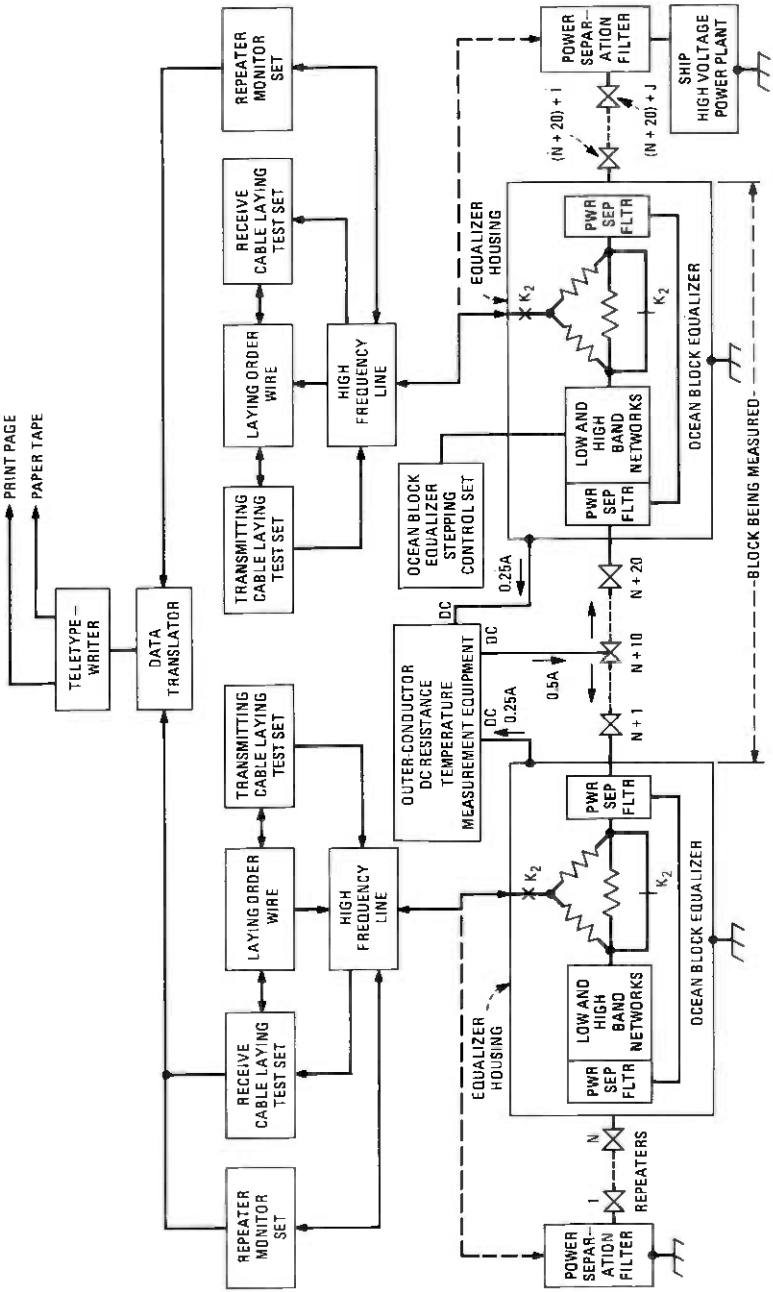


Fig. 3—Assembled shipload transmission test.

above. The primary intent of these calculations is to verify that the assembled shipload is operating properly. To accomplish this, one must be able to predict the expected transmission response. Factory measurements on the actual repeaters, cable sections, and equalizers in each block, which are loaded into the computer memory prior to testing, are used to compute a transmission response to be compared with the shipboard measurements. The computed transmission response must correct for the actual temperature of the repeaters and cable aboard ship, because these are generally not the same as during factory measurements.

2.2.3 Tests during laying

Many of the tests during laying are similar to the assembled shipload tests except that the test path is between ship and shore rather than only within the shipload.

The beginning of a cable lay starts with a cable splice. Before the splice is made, a transmission reconciliation test is made to verify the performance of the previously installed portion of the system. If this reconciliation (comparison with measurements made earlier) is satisfactory, the splice is made. After the splice, another reconciliation transmission test is performed between the shore terminal and the test lead of the most outboard ocean-block equalizer in the shipload. Again, the data are analyzed using the shipboard computer to determine if the measurement is consistent with the previous reconciliation and the assembled shipload measurements.

The primary reasons for transmission tests during laying are to confirm proper operation of the system and to obtain data necessary to choose the optimum ocean-block equalizer settings. Using the test configuration shown in Fig. 4, transmission tests are made at regular intervals. At the end of a standard frequency measurement run, each station has only the measurement data for its receive band. These data are then exchanged via the laying orderwire by using the TTY with a paper tape reader and the data sets. Thus the shore terminal and ship obtain measurement data for both bands.

Between standard frequency runs, while the cable is being payed out, the cable laying test sets are used to send and monitor a single tone in each direction of transmission. The receive cable laying test sets give an automatic alarm when the received power of this tone varies by more than 1.5 dB. Also, the supervisory tone power from each repeater is measured and recorded using the repeater monitoring set, data translator, and TTY. Because adjacent repeaters transmit supervisory tones in opposite directions, the data must be exchanged between the ship and shore terminal. Additional tests performed during laying include (*i*) a transmission test through the next-to-be-laid block to ensure that its

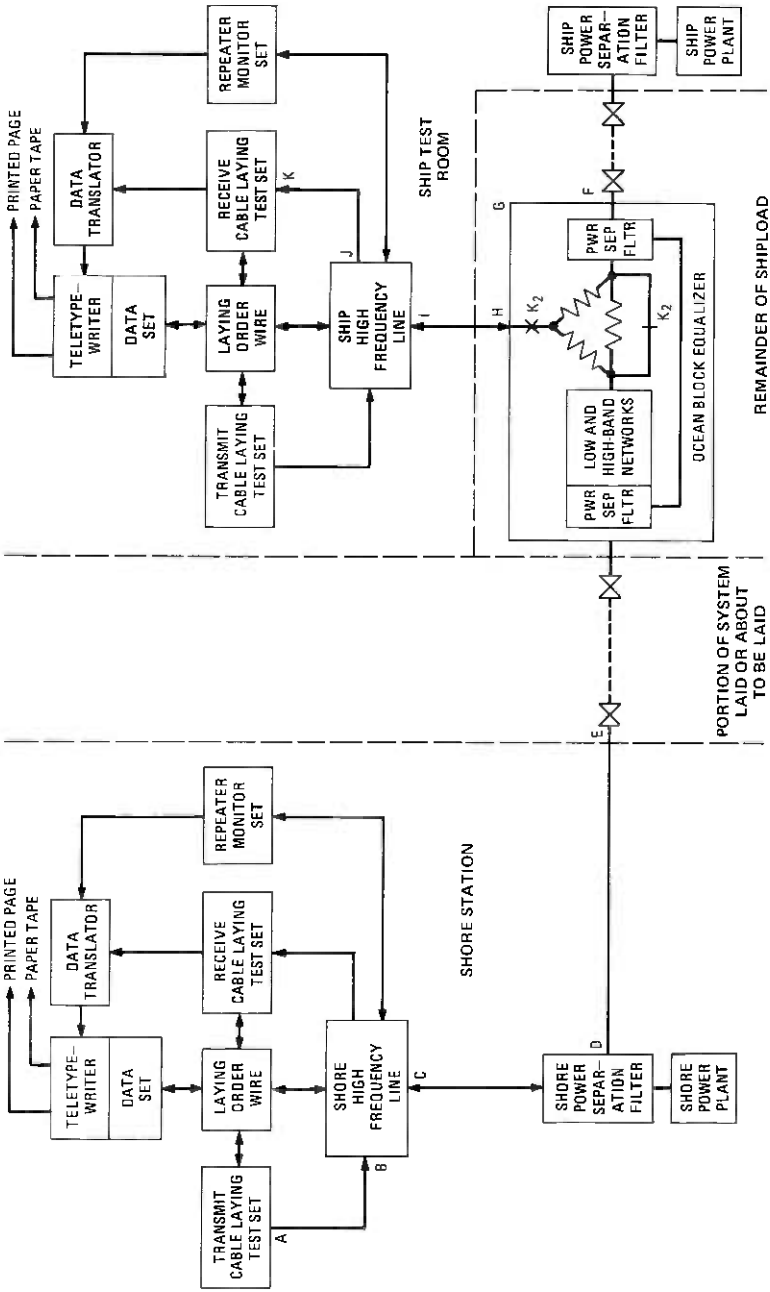


Fig. 4—Transmission tests during laying.

transmission has not changed since the assembled shipload tests, (*ii*) transmission tests at closely spaced frequency intervals, called ripple runs, and (*iii*) monitoring the power plant voltage and current.

The process of choosing the optimum equalizer setting typically starts about three hours before the equalizer is scheduled to go overboard. This interval allows time for the equalizer decision to be made and for the test and stepping leads to be sealed by overmolding. First, a transmission objective is computed using the shipboard computer. The transmission objective is the expected test tone power which would be measured at the input to the receive cable laying test set, assuming that the only misalignment in the system is the planned misalignment.* All the components of this calculation are known either by calculation or by direct measurement with the cable laying test set.

At the time of any standard frequency run, the misalignment of the system consists of the sum of residual misalignment and temporary misalignment. The temporary misalignment is due to the fact that not all the block being laid has stabilized at ocean bottom conditions. Nevertheless, the temporary misalignment must be removed from the measurement data since the equalizer should compensate for only the residual misalignment. This is accomplished as follows. Under steady-state conditions (i.e., constant ship speed, cable payout rate, and shipboard cable and repeater temperature), which are normally closely approximated during laying, temporary misalignment decreases linearly with time (and thus distance) until the equalizer is overboarded. The shipboard computer sorts the standard frequency run data and extrapolates to the expected value corresponding to the time that the equalizer would go overboard. The amount of remaining temporary misalignment due to the cable suspended between the ship and the ocean bottom and due to the thermal inertia of the cable and repeaters is calculated from theoretical considerations and used to correct the extrapolated data.

Direct comparison of the corrected extrapolated data to the transmission objective yields the residual misalignment to be compensated for by the equalizer, which is referred to as block deviation. The computer equalization program then compares the block deviation to the loss shapes available in the equalizer and orders the various setting choices based on a weighted-sum-of-squared-error criterion. The equalizer setting that is selected is typically the one with minimum sum-of-squared-error. The transmission engineers, however, examine the top several choices, taking into account auxiliary information (e.g., anticipated misalignment in the next block), and occasionally choose a different setting.

Once the choice of equalizer setting has been made, four standard

* An example of planned misalignment is gain provided at time of cable installation to precompensate for anticipated increased cable loss with time, i.e., cable aging.

frequency runs are performed. The first two are made immediately before and after the equalizer is switched. A comparison of these two standard frequency runs is made to verify that the correct change in loss has been achieved. Second, the shipboard transmission test lead is moved to the next inboard equalizer, adding the next ocean block to the measurement path. The next two runs are performed before and after the transmission test pad is switched out of the equalizer about to be laid. A comparison of these two standard frequency runs is made to verify that the test pad has, in fact, been switched out. At this point, the equalizer stepping and transmission test leads are sealed, and the equalizer is ready for overboarding. Transmission testing then continues through the next inboard equalizer.

III. ENVIRONMENTAL CONSIDERATIONS AND MEASUREMENTS

3.1 Repeater temperature control

In previous undersea cable systems, no special effort was made to control the temperature of repeaters and equalizers on the cable laying ship. For SG, however, reasonably precise control is required.

3.1.1 Control requirements

Even though repeaters and equalizers are vacuum-dried during manufacture, some water remains, much of it dissolved in polymeric materials. At high storage temperatures a significant amount of water exists as vapor within the free volume of the repeater. If such a repeater were launched into the ocean, where the bottom temperature is typically 2.5°C, there is a possibility that the water would condense on surfaces within the repeater before it could be reabsorbed by the polymers. Condensed water on and between conducting surfaces can cause repeater failure. A careful experimental investigation showed that condensation would not occur if the ambient storage temperature were less than 32°C for unpowered repeaters and 27°C for powered repeaters. The lower of these two values is controlling, since repeaters are powered when they are laid.

If temperature-controlled repeaters[4] are included in a shipload, it is necessary to know their temperature in order to predict what their transmission characteristics should be. To limit positive misalignment in blocks containing temperature-controlled repeaters, the difference between repeater and cable temperature must be controlled. Because the thermal mass of the cable is much greater than that of the repeaters, it is more efficient to control repeater temperature than cable temperature. Thus it is necessary not only to limit the maximum value of repeater temperature but to be able to set it and maintain it with reasonable precision.

Temperature control is also needed for other repeaters, but the re-

quirements are less stringent because temperature changes produce a much smaller change in transmission. Table I summarizes repeater temperature setability, stability, and measurement accuracy objectives.

3.1.2 Temperature-control facilities

To achieve the required setability of repeater temperature, it was decided to use air-conditioning units in each repeater bay and to build tents over each repeater stack. The air-conditioning units are located along the port bulkhead of each repeater bay. The chilled air is piped into a 4-inch-high (102 mm) rectangular telescoping duct located beneath the repeater stack.

When a full load of repeaters is in place, the duct is extended to its full length, spanning the entire stack. As repeaters are payed out during a cable-laying operation, the duct can be shortened ultimately to approximately one-third of its fully extended length. The telescoping feature was provided to keep the incoming air flow concentrated around the remaining repeaters. The return duct for the air flow is on the deck next to the telescoping inlet duct. It contains a series of holes on the top and the side opposite the inlet duct to receive the air. Each hole is fitted with a swivel plate which can be moved to open or close the hole. As repeaters are payed out, the unused holes are sequentially closed.

The repeater stack is covered by an insulating tent with openings on the sides to pass the cable bights. The tent is made in sections so that, as repeaters are payed out, sections can be removed to allow the size of the tent to conform approximately to the size of the stack of remaining repeaters.

Under unusual conditions of very high humidity, condensation can occur on the repeater and equalizer housings in the tent. Drains have been provided to remove this water.

Table I — Shipboard repeater temperature objectives

	Temperature-Controlled Repeaters	Non-Temperature-Controlled Repeaters and Equalizers
Setability	<27°C*	<27°C
Stability in 12 hours	±0.2°C	±2.0°C
Measurement accuracy	±0.2°C	±0.5°C

* The maximum temperature for the temperature-controlled repeaters is determined in part by the cable temperature. The specific requirement depends on a number of factors. In the worst case, it is sufficient to keep the repeater temperature at or below that of the cable.

3.1.3 Repeater temperature measurement

An automatic temperature measurement and recording system has been installed on Cable Ship *Long Lines*. Twenty-one thermocouples can be placed in each of the three repeater stacks. The thermocouples are attached to the periphery of holes in the end cones of the repeaters using clamps which can be removed readily before the repeater is payed out.

The thermocouple identifying number, the indicated temperature, and the time of measurement are printed out on paper tape. The equipment can be programmed to scan and print data from all of the thermocouples at set intervals, e.g., every eight hours. When calibration and correction procedures are used, the temperature measurement accuracy is about 0.05°C.

3.2 Cable temperature control

With previous cable systems, there was no need to control cable temperature. The temperature was measured with thermocouples at 15 points in each tank, giving adequate information for transmission measurements. For SG, cable temperature stability in each tank has to be maintained.

3.2.1 Control requirements

As discussed above, temperature setability requirements have been applied to repeaters rather than to cable. Measurements to determine equalizer setting, however, require that the cable temperature remain stable for the measurement period. Specifically, the stability objective is ±0.2°C over a 12-hour period. Similarly, the measurement accuracy objective is ±0.2°C.

3.2.2 Temperature-control facilities

Measurement of cable temperature during installation of the TRANSPAC-2 SF system showed that the bottom aft portion of the Tank 3 wall was warmer than the rest of the tank. This was traced to the proximity of that area to the ship's main boilers. To improve the uniformity of temperature distribution in the cable, the cable was isolated from the warm area by spacers approximately two inches thick and a sheet of fiberglass to provide a smooth surface adjacent to the cable.

As a further means to provide a uniform cable temperature, the tanks are flooded with water during cable-laying operations. The water level is maintained about two feet below the top of the cable to protect people working in the tank. The water in each tank can be circulated by draining it from drains in the bottom of the tank and pumping it into the top of the tank along the walls. While the tank water temperature does change

slowly in response to sea temperature, the circulating water arrangement provides adequate stability.

3.2.3 Cable temperature measurement

Two means are now provided for measuring cable temperature. The first, similar to that used previously, consists of 12 thermocouples in each tank which are connected to the temperature recording system discussed above for the repeater stacks. The thermocouples are taped to the cable during loading so that they are distributed throughout the tank. Thermocouples have the advantages of simplicity and accuracy but the disadvantage that they provide only point measurements. If the cable temperature is not uniform, it is not possible to determine precisely the average temperature of a cable section or a block from the thermocouple readings.

For transmission measurements, the average temperature of the cable in a block is the most important item of cable temperature data. To provide this information directly, the second cable temperature measurement system was developed. This method consists of measuring the dc resistance of the outer conductor of the cable and inferring from this and factory test data the average temperature.

The repeater and equalizer high-pressure housings are connected in series with the cable outer conductors but contribute negligible resistance. The conductivity of the copper used for the outer conductor of undersea cable is carefully controlled. Since the temperature coefficient of resistivity is directly related to the conductivity,[7] it is thus also carefully controlled. During cable manufacture, the factory determines the dc resistance of the outer conductor of each cable section at 10.0°C. This reference reading and the known temperature coefficient are then sufficient to relate temperature to dc resistance throughout the range of interest. Working the outer conductor during cable handling in the factory and from the factory to the ship increases the resistance by an amount equivalent to approximately 0.25°C temperature change on the average. A correction for this is made in the temperature calculation.

If the ship is stationary, a satisfactory measurement can be made as described above. If the ship is pitching or rolling, however, motion of the coiled cable in the earth's magnetic field induces voltages which can cause errors in the measurement. To overcome this problem, instead of measuring a block directly, the two halves of the block are measured in parallel. In this way, the voltages induced by ship motion in the two halves tend to cancel. As an added precaution, high measuring currents (up to 500 mA) are used to make the voltage drop from measuring current much larger than the motion-induced voltages.

The correlation under carefully controlled conditions between thermocouple and direct current resistance (DCR) measurements has been

quite satisfactory. The high-current DCR method is regarded as the primary cable temperature measurement. The thermocouple results are used for check and backup.

IV. CABLE LAYING FACILITIES—C. S. LONG LINES

References 8 to 11 discuss the construction and operation of the Bell System Cable Ship *Long Lines*, first used for installing SD cable systems. This section discusses the extensive changes made to the ship to permit handling of SG cable and repeaters.

4.1 Reasons for modifications

Cable used in the SG system is larger, heavier, and stronger than the cable used in previous systems. Table II compares the size, strength, and weight of SG armorless cable and SF armorless cable. The larger diameter of SG cable limits a single shipload to approximately 770 nautical miles (1429 km). With 5.1-nmi (9.46 km) repeater spacing, one equalizer for each 30 repeaters and one-half mile of cable between the equalizer and each adjacent repeater, there is an average of 4.8 nmi of cable for each repeater and equalizer. Thus, it is necessary to stow approximately 160 repeater plus equalizer "bodies" on the ship.

4.2 Linear cable engine

The greater weight and resultant greater laying loads for SG cable required extensive modifications to the linear cable engine. In its original configuration, it had one drive power unit (electric motor and hydraulic pump combination), one main brake, and one hydraulic motor each for payout and pickup operation.

4.2.1 New drive motor and drive power unit

To increase the laying tension capability at 8 knots payout speed from 8000 to 16,000 pounds (3629 to 7257 kg), a new hydraulic motor was added and connected, through a gear box, to the upper forward sprocket shaft. The new motor operates only in the payout (hold-back) mode. A second drive power unit was then added to provide sufficient high pressure oil flow to drive both the old and the new motors simulta-

Table II — Comparison of physical properties of SF and SG armorless cable

	SG	SF
Outer diameter (inches/mm)	2.07/52.6	1.75/44.5
Strength (pounds/kg)	37,000/16,783	16,000/7257
Weight per nautical mile, in sea water (pounds/kg)	3,500/1588	2,000/907
Expected laying tension in 3 miles of water (pounds/kg)	10,500/4763	6,000/2722

neously. Isolation and cross-connection features are provided so that the engine can be run with both motors and both power units or, for lower laying loads, in the original one-motor configuration with either power unit. The engine cannot be operated with the new motor only since the old motor cannot readily be disconnected.

4.2.2 Track modification

The shear-limiting feature of the linear cable engine[10] was modified for the SG laying loads. Since the extensional elastic properties of the cable, as well as the weight in water, are determined primarily by the steel center strand as in previous cable designs, the maximum required shear limiter travel remained unchanged at one-half inch. This allowed the modification to be made without major redesign of the roller carriages. The higher loads were accommodated by using higher modulus springs in the roller carriage assemblies and by preloading the springs to approximately twice the load used previously.

New gripper blocks were designed for the larger diameter of SG cable. These blocks will handle SF and SG cable satisfactorily, but not SD cable.

4.2.3 New brake

Increased braking capability is required for the higher SG strength and loads. To provide this, a new hydraulically operated, spring-release brake unit was added on the upper forward sprocket shaft. This provides enough capacity so that, in an emergency, the cable will part before the brakes slip.

4.2.4 Tension displays

The original aft tension displays used a linear scale from 0 to 16,000 pounds (0 to 7257 kg). The system was changed to provide a quasi-logarithmic* scale from 0 to 50,000 pounds (0 to 22,680 kg). The scaling gives 0- to 10,000-pound (0 to 4536 kg) indication on the lower half of a 330-degree display, allowing greater resolution in the most commonly used range.

The forward tension displays were similarly changed. The dual 0- to 16,000- and 0- to 100,000-pound (0 to 45,360 kg) linear displays were replaced by 0- to 100,000-pound quasi-logarithmic displays with 0 to 20,000 pounds (0 to 9072 kg) on the lower half.

* The pointer motion is proportional to $\log (1 + bT)$, with T the tension and b a scaling factor.

4.3 Repeater stowage

In each of bays two and three, repeaters and equalizers are stacked in a rectangular array 16 wide by 4 high, allowing storage of 64 bodies. This is adequate for the 300 nautical miles (557 km) of cable which can be held in each of tanks 2 and 3. Tank 1 can hold 170 nautical miles (315 km). To accommodate this, the repeater stack in bay one is nine repeaters wide by four high. Thus a total of 164 repeaters and equalizers can be carried. For an SF shipload, 116 repeaters and equalizers are required.

4.4 Cable stowage

Figure 5 shows the stowage configuration for cable and repeaters typical of tanks 2 or 3. The arrangement for tank 1 is similar except for the different repeater stack orientation.

The cable is coiled in the tanks in the same way it has been for previous systems. The ends of the cable sections which feed from the tank up to the repeater stack on deck are called up-and-down runners, the up-runner being the end payed out ahead of a repeater, the down-runner the end payed out after the repeater. To prevent crossovers and kinking of the cable, these runners are neatly dressed into slots which lie on the forward wall of the cable tank. They are held in the slots with thick rubber flaps which deform to let the cable come out of the slot during payout.

Because of the larger diameter of SG cable, the slots had to be modified. In tank 1, there are now six up- and six down-runner slots, each deep enough to hold seven runners. This is sufficient to accommodate 41 repeaters, two runners being reserved to lead to the test room or to other cable tanks. In tanks 2 and 3, there are ten up- and ten down-runner slots, each seven runners deep, so 69 repeaters can be accommodated in each of these tanks.

4.5 Grapnels

The SG system uses larger and stronger cables, and closer repeater spacing than earlier systems. These changes led to a need for improved grapnels for repair operations in both shallow and deep water. Three areas of development were undertaken: (*i*) the design of conventional grapnels with increased strength, with dimensions changed for SG cable, (*ii*) the design of a powered cut-and-hold grapnel for deep-sea use, and (*iii*) the design of a simple detrenching grapnel for recovering buried cable.

4.5.1 Conventional grapnels

Stronger versions of three standard grapnels, the Gifford, the Rennie, and the Flatfish, have been designed and manufactured. These grapnels are illustrated in Figs. 6 and 7.

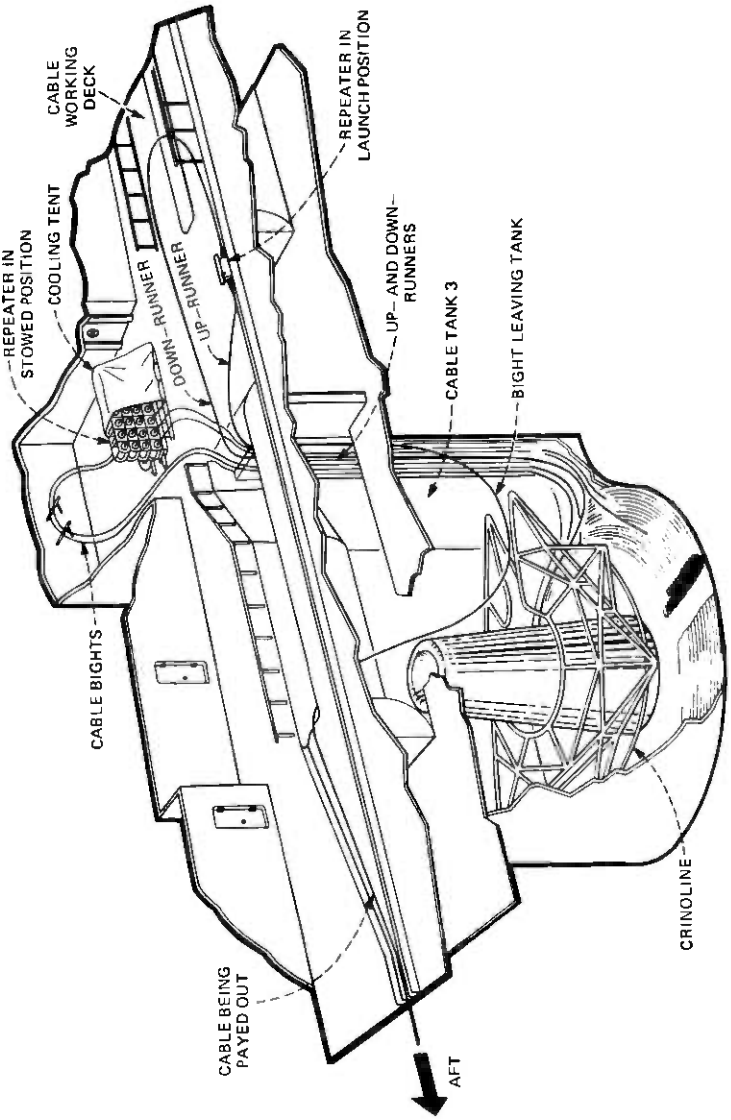


Fig. 5—SG cable and repeater stowage on C. S. *Long Lines*.

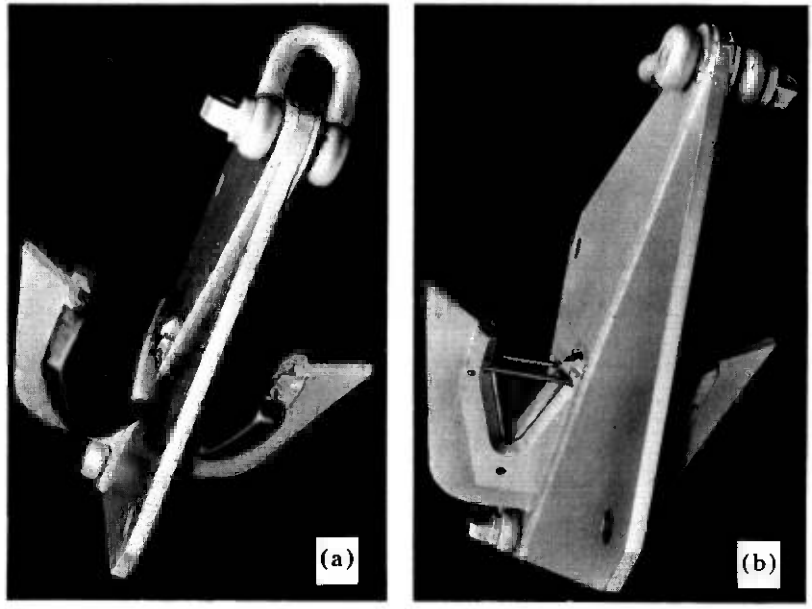


Fig. 6—Flatfish grapnel for SG cable: (a) fitted with "gloves"; (b) fitted with breakaway bars.

The Flatfish grapnel, Fig. 6, has been designed to cut or to hold all types of SG cable. If the grapnel is required for recovering cable, "gloves" are fitted over the blades at their base. For use in the cutting mode, two types of cutting blades have been produced, curved blades for armorless cable and conventional straight blades in a V-formation for armored cable. In addition, the grapnel can be fitted with bars located either side of the open end of the blades. As the cable tension increases, the bars suddenly break, and the cable impacts against the cutting blade, thus increasing the cutting action.

The Rennie grapnel, Fig. 7a, is conventional except for the use of replaceable prongs. The Gifford, Fig. 7b, is also conventional in design except for the inner surface of the "hook"; this has a V-formation, which reduces cable slip through the grapnel.

The new designs were tension-tested to 67,000 pounds (30,390 kg) and proved to be suitable by use during the TAT-6 laying operation.

4.5.2 Cut-and-hold grapnel

When carrying out a deep-sea repair, the cable must be cut before the ends can be brought to the surface. To prevent damage to the repeaters, they should not be dragged along the sea bed during this operation. The cut-and-hold grapnel is designed to cut the cable and also to hold onto

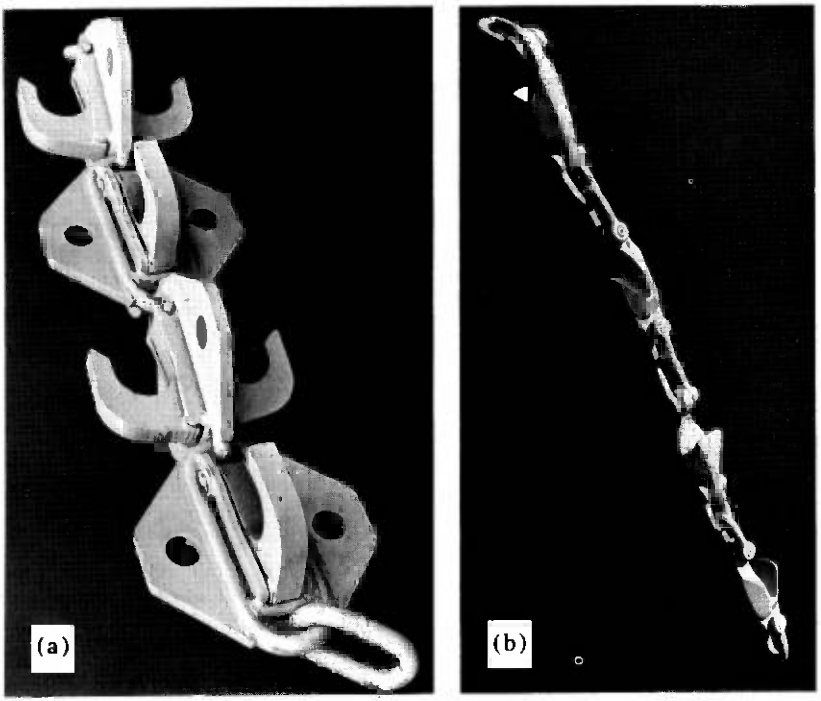


Fig. 7—Rennie and Gifford grapnels.

one end without applying a large tension which would displace the cable from its route.

The grapnel is powered by stored compressed nitrogen, and is hydraulically actuated. When the cable is engaged, a valve is operated which switches oil at 5,000 psi to a large ram with a six-start thread on the periphery of the cylinder. The ramrod is fixed, so the cylinder advances and, as it passes through a fixed nut, it rotates. Protruding horns are fixed to the cylinder and, when one of the horns catches the cable, a bight (loop) is belayed (wound) around the cylinder. At the end of the ram stroke, a second valve is operated which drives two guillotines whose blades can be positioned so that a predetermined side of the cable can be cut away, no matter which side is up when the grapnel lands on the sea bed, while the other cable end is left belayed around the cylinder. Thus one grapnel run cuts the cable and recovers one end.

A sonar surveillance system is built into the grapnel. The rate of transmitted pulses indicates the progress of the operation. The grapnel is suitable for operation to depths of 5000 fathoms.

4.5.3 Detrenching grapnel

An anchor-shaped detrenching grapnel was developed for the recovery

of those sections buried by the cable plow. A prototype grapnel designed for a 36-inch (0.914 m) penetration in a granular sea-bed was successfully tried at sea in June 1976. Towing forces, although manageable, were high, and therefore a reduced penetration of 26 inches (0.660 m) has been chosen for the production grapnels. These should require towing forces of no more than 18,000 pounds (8165 kg).

V. CABLE PLOW

A new Bell System cable plow, designated Sea Plow IV,[12,13] was used to bury cable on both ends of TAT-6. The new plow is capable of burying cable and repeaters to a depth of 24 inches and of working in water depths up to 500 fathoms.

VI. CONCLUSION

This paper summarizes the electrical and mechanical designs and design modifications which were needed to lay, measure, and maintain an SG undersea cable link. The new automated laying test sets and shipboard computer facility were powerful and versatile tools which lifted the pressure of routine chores from test room personnel. The repeater monitoring set proved to be a highly useful and sensitive source of performance information on the undersea system. Modifications of the cable and repeater stowage and handling arrangements served well in the TAT-6 installation. The modified linear cable engine handled the large diameter SG cable smoothly.

The success of these numerous designs, involving several technical disciplines, is due to the expertise, dedication, and teamwork of many contributors, on both sides of the Atlantic. The authors humbly acknowledge their vital contributions.

REFERENCES

1. SD Submarine Cable System, B.S.T.J., *43*, No. 4 (July 1964).
2. D. N. Harper, B. O. Larson, and M. Laurette, "SG Undersea Cable System: Commissioning: Final System Alignment and Evaluation," B.S.T.J., this issue, pp. 2547–2564.
3. E. T. Calkin, I. Golioto, W. J. Schatz, R. E. Schroeder, and D. S. Shull, "SG Undersea Cable System: Undersea System Power," B.S.T.J., this issue, pp. 2497–2522.
4. C. D. Anderson, W. E. Hower, J. J. Kassig, V. M. Krygowski, R. L. Lynch, G. A. Reinold, and P. A. Yeisley, "SG Undersea Cable System: Repeater and Equalizer Design and Manufacture," B.S.T.J., this issue, pp. 2355–2403.
5. W. B. Hirt and D. O. Oldfather, "Transmission Tests, Computations and Equalization During Installation," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 783–798.
6. W. G. Ramsey, Jr., W. B. Hirt, P. P. Theophall, and G. A. Ferguson, "Tuning to Concert Pitch—Installation Techniques," National Telecommunications Conference, 1976, Conference Record.
7. *Copper Wire Tables*, National Bureau of Standards Handbook 100, February 21, 1966.
8. R. D. Ehrbar, "A Cable Laying Facility," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1367–1372.
9. O. D. Grismore, "Cable and Repeater Handling System," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1373–1394.

10. R. W. Gretter, "Cable Payout System," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1395–1434.
11. J. H. Butler, C. J. Altenburg, R. J. McSweeney and L. E. Sutton, "Design and Powering of Cable Ship Long Lines," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1435–1459.
12. G. S. Cobb, D. L. Garren, and T. H. Rose, "Sea Plow IV: Digging-in the Newest Transatlantic Cable," Bell Laboratories Record, *54*, No. 8 (September 1976), pp. 220–224.
13. D. L. Garren, "Two Feet Under Is Ten Times Safer—New Techniques for Plowing Cables and Repeaters," National Telecommunications Conference, 1976, Conference Record.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

SG Undersea Cable System:

Commissioning: Final System Alignment and Evaluation

By D. N. HARPER, B. O. LARSON, and M. LAURETTE

(Manuscript received October 7, 1977)

A comprehensive line-up and test program called commissioning was developed for the SG Undersea Cable System. The program provides instructions for measurement, alignment, and evaluation of various performance parameters of the system. Equalization, optimization of signal-to-noise ratio, and system alignment are the primary objectives of commissioning. Measurement of performance parameters such as differential gain and phase, phase jitter, and impulse noise count distribution with threshold level are included in the test program. This paper describes the commissioning process, interspersing both a general description of procedures and selected test results from the TAT-6 link.

I. INTRODUCTION

Commissioning, a British Post Office term, is the test and adjustment program (or process) performed on a newly installed undersea cable link before it is placed in service. The cable system is measured and aligned, and its performance is optimized. This process starts during installation and continues until the terminal-to-terminal link is ready to be released for service. The bulk of the effort takes place after cable laying[1] is finished; that is, after the final splice, and can take several weeks to complete. This paper describes the commissioning process for the SG system. General descriptions of the test program and illustrative data from the TAT-6 link are provided. Fault localization reference data, although part of commissioning, are not discussed.

Successful completion of commissioning means that the cable and

terminal transmission equipment (supergroup-to-supergroup) are equalized, signal-to-noise ratio is optimized over the transmission bands, and the terminal equipment is adjusted so that service may begin. Accurate commissioning test data are vitally important because these data are used also to:

(*i*) Determine link operating performance.
(*ii*) Compare actual with computed performance.
(*iii*) Provide initial data so that future link performance may be evaluated.
(*iv*) Obtain reference data for maintenance purposes.

Incidentally, the SG system differs from the SF and SD systems[2,3] in that the undersea link transmission can be adjusted from the terminal, by means of shore-controlled equalizers.[4] Five sectors, each approximately 650 nautical miles (nmi) long, were created in the TAT-6 link by using four shore-controlled equalizers. Undersea link transmission is separately adjustable for four of the five sectors using these equalizers (discussed in Section IV); terminal adjustments affect the entire link.

II. THE TRANSMISSION LINK

The system is designed to provide at least 4000 three-kHz channels (from supergroup to supergroup) over ocean distances of up to 4000 nmi. Associated with the undersea single-cable link are two-wire to four-wire terminals located at opposite ends of the link[5] (see Fig. 1). The two terminals are defined "A" and "B" terminals, respectively. At line (cable) frequency, the A terminal transmits in the low band or A-to-B direction, 0.8 to 13.9 MHz, and receives in the high band or B-to-A direction, 16.5 to 30.0 MHz; the B terminal is complementary. A 31.62-MHz conversion frequency is used to translate the high-band baseband, 1.6 to 15.1 MHz, to line frequency at the B terminal and re-establish the baseband at the A terminal. The low-band baseband and line frequencies are identical. These frequency ranges individually and collectively are termed wideband.

Each terminal is composed of separate transmit and receive sections that are used to provide the four-wire supergroup connection to the domestic network external to the terminal. Towards the undersea link, transmit and receive signal paths are combined in a directional filter furnishing the two-wire match with that link. To achieve reliability, major parts of the terminal are duplicated; this feature increased the amount of data-taking and record-keeping needed for commissioning.

For commissioning, the terminal-to-terminal link can be thought of as being made up of three nested links as shown in Fig. 1:

(*i*) Wideband-to-wideband link (wideband link).*
(*ii*) Hypergroup-to-hypergroup link (hypergroup link).†

* The undersea link is part of the wideband link.
† A hypergroup is an ordered frequency-contiguous set of supergroups (see Ref. 5).

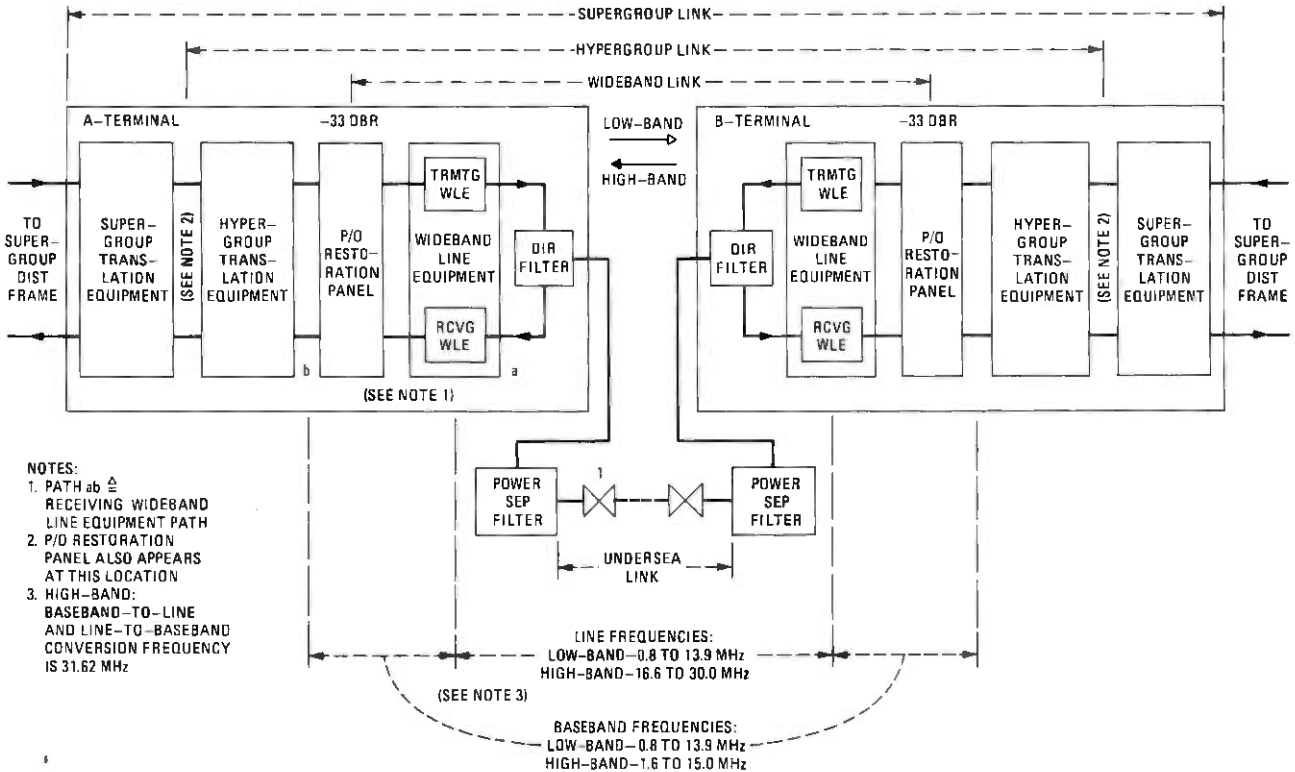


Fig. 1—SG undersea cable system.

(*iii*) Supergroup-to-supergroup link (supergroup link).

Following any necessary adjustments in the undersea link using the shore-controlled equalizers, measurements and adjustments proceed from the wideband to hypergroup to supergroup links.

The wideband/hypergroup interface is the restoration panel.* This interface is a −33 dBr† nominal flat-level point, for both transmit and receive. Transmission measurements between restoration panels in the two terminals define the misalignment of the wideband link which is due primarily to the long undersea section.

III. COMMISSIONING TEST PROGRAM

Development of the commissioning test program is a process that requires data and other information from virtually all areas of cable system development such as repeaters and equalizers,[4] terminals,[5] cable,[6] and power separation filter.[7] Use of specialized test equipment developed for the SG system, such as the cable laying test set, the data translator, and the repeater monitoring set, is integrated into the test program[1] as is the use of a modern table-top computer system.

The commissioning test program is divided into four categories of tests. The four groups of tests are carried out sequentially, each succeeding category building more or less upon the results of its predecessors:

(*i*) Undersea link tests—between cable laying operations.
(*ii*) Shore preparation tests—prior to final splice.
(*iii*) Line-up tests—after final splice.
(*iv*) Reference and special tests—follows link alignment.

3.1 Objectives

Even though the four groups of commissioning tests are done sequentially, it is more illuminating to consider the line-up tests, (*iii*), as being central to commissioning. Most preceding tests, (*i*) and (*ii*), are preparatory to line-up. Most subsequent tests, (*iv*), are designed to provide information that can be used to maintain the lined-up link status or confirm the engineering design of the system. Equalization, optimization of the signal-to-noise ratio (s/n) over each baseband, and terminal adjustments (including pilots) are the major objectives of the line-up procedures.

Transmission data from the line-up tests, however, cannot be referenced in level unless various transmission path gains (or losses) of the terminal, the power separation filter, and the shore section‡ are measured

* The restoration panel provides a manual patching arrangement used to accomplish alternative traffic routing in the event of an undersea transmission failure.
† dBr is equivalent to dB TLP (TLP is transmission level point).
‡ The shore section is the cable between the terminal and the repeater closest to the terminal.

or computed beforehand. For example, the transmission gain of path *ab*, Fig. 1, is measured.

Prior to the shore preparation tests, undersea system tests are made to determine (or estimate) performance parameters such as warm-up time and cable aging.* Figure 2 shows the TAT-6 warm-up characteristic at high-band supervisory tone frequencies (approximately 27.7 MHz) after power turn-up. Most of the gain change occurs over the two continental shelves and takes place within about 10 hours. The change is due largely to internal warm-up of temperature-controlled repeaters.[4] Obviously, it is important that equalization be based on nonchanging data taken after this 10-hour warm-up period.

After the line-up tests are completed, reference and special tests are performed. The objective of the reference tests is to obtain for future comparison initial supervisory tone, pilot, and transmission measurements (made at discrete frequencies between supergroups). Identical measurements can be made in the future without service interruption, thereby providing an excellent way to monitor changes in undersea link transmission.

The objective of the special tests is to obtain by direct measurement such engineering parameters as change in system gain with line current, differential gain and phase, phase jitter, and impulse noise count distribution, to name a few. Figure 3 shows the measured TAT-6 high-band

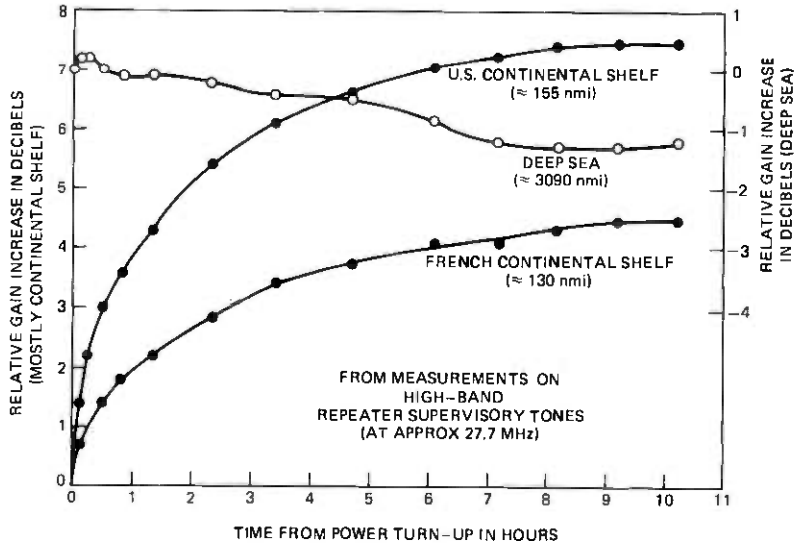


Fig. 2—TAT-6 high-band warm-up characteristic following power turn-up from cold start.

* Warm-up time is the relatively short-term interval required to reach transmission stabilization following power turn-up from a cold start. Cable aging is the long-term cable transmission change with time (Ref. 8).

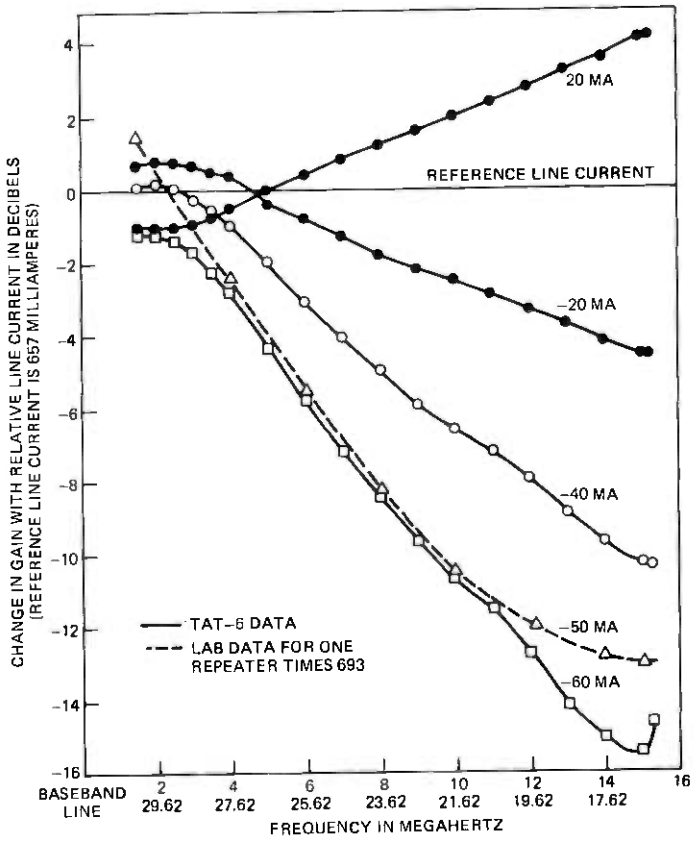


Fig. 3—TAT-6 high-band change in gain with line current.

change in gain as line current is stepped from the nominal 657 milliamperes. Measured gain-change data correlated well with laboratory data on a single repeater scaled by 693 (the number of repeaters in TAT-6). In the event of a shunt fault in the undersea link, there is always the possibility that changes in line current, along with other adjustments, could be used to modify undersea misalignment so that some service could continue until repair was possible. The other engineering parameters listed above are discussed in Section VI.

3.2 Methodology

The lengthy commissioning test program must be grounded in well-defined procedures no more complicated than absolutely necessary, making use of:

(*i*) Standard data forms.
(*ii*) Standard definitions.
(*iii*) Standard list of test frequencies.

Considerable effort was put into developing individual, but standardized, data forms for the various commissioning tests. The forms, in general, represent a method of entering data, and intermediate and final results as a function of frequency. Short equations are provided on the forms, as necessary, to describe how the measured data should be handled to compute an intermediate or final result. These equations can be used as a computer program instruction for keyboard entry into the table-top computer system mentioned previously. A general-purpose program called UTILITY was written to implement this capability. In summary, the computer system served several purposes. It provided:

(*i*) Speed of data manipulation and computation.
(*ii*) Data storage.
(*iii*) Plotting.
(*iv*) Analysis of on-site tests not specifically listed in the test program.
(*v*) Backup, since the computer can be used to compute results based on data from either station.
(*vi*) Supergroup equalizer design (Section 4.4).

Finally, attention was paid to historical and practical ways of defining submarine cable link transmission subpaths to optimize the testing process. Commissioning test frequencies covering the wideband portion of the SG system were chosen to be compatible with the use of the cable laying test set.[1] Defined transmission subpaths and the testing frequencies were integrated into the standardized test forms.

IV. EQUALIZATION

4.1 The equalization process

During installation of the undersea link, residual misalignment is substantially compensated for at the end of laying each ocean block using the associated ocean-block equalizer (OBE).[4] A buildup of misalignment unequalizable in the OBEs invariably occurs, accumulating along the length of the undersea link. This end-to-end residual misalignment is compensated for (to the extent possible) in two ways:

(*i*) By adjustment of undersea link levels on a sector basis using the shore-controlled equalizers.
(*ii*) By terminal equalization adjustments.

4.2 Undersea link adjustment

Four shore-controlled equalizers (SCE),* which divided the undersea link into five sectors, were deployed to compensate for transmission changes due to cable aging[8] over the life of TAT-6 (see Section I). How-

* The shore-controlled equalizers are controllable from the Green Hill Station, with SCE 1 being closest to that station.

ever, sufficient aging (added loss) had occurred over the five-month installation interval to justify a reoptimization of SCE settings following the final splice.

The average transmission level* of each of the four sectors beyond SCE 1 was adjusted to the extent possible to the average level of Sector 1. Since only broadband equalizer shapes are available in the SCEs, the selected settings are a compromise. In the low band only SCE 1 needed readjustment; in the high band, SCE 1, SCE 2, and SCE 3 were reset. Figures 4a and 4b show undersea link levels versus distance at two typical frequencies, 19 and 26 MHz, before and after SCE adjustment.

4.3 Terminal equalization—wideband

4.3.1 The plan

The terminal wideband equalization plan is illustrated in Fig. 5. Equalizer networks are discussed in more detail in Ref. 5. For the ideal situation (no equalization required), *line buildout* (LBO) *networks* are used to build out shore section loss to 3/4 of a cable section (3.83 nmi†) to conform to SG system design. In addition, the transmit combination of *shore section*, LBO *networks*, *pre-emphasis network*, and *path equalizer network* is designed to shape the flat-level signal coming from the inland network to a computed characteristic which provides optimum s/n in the undersea link. The receive part of the terminal at the other end of the undersea link is complementary, restoring the shaped signal to a flat level for transmission to the inland network.

In the real world, however, remaining misalignment, principally in the undersea link, will require terminal equalization. Additional transmit terminal signal shaping will also be needed to optimize s/n in the undersea link, since most levels end up being nonideal. Terminal equalization is carried out in three phases:

(*i*) Wideband equalization, transmit end only (before optimization of wideband s/n).

(*ii*) Wideband equalization, receive end only (after optimization of wideband s/n).

(*iii*) Supergroup equalization, where required.

Transmit-end equalization (before s/n optimization) is accomplished using *residual equalizer networks*, *adjustable corrector networks*, and *adjustable bump equalizer networks*. Normally, most of the transmit-

* The average level of an undersea sector is the average of the level averages of individual blocks in that sector. (The undersea link levels were determined on a broadband basis during cable laying. Subsequently, level can only be measured at supervisory tone frequencies, essentially one frequency in each band. This latter level information is interpolated to other frequencies by fitting an equation of the form $a\sqrt{f} + bf$ to the data.)

† The shore section is between 1/4 and 3/4 of the nominal repeater spacing (5.1 nmi of 1.7-inch cable).

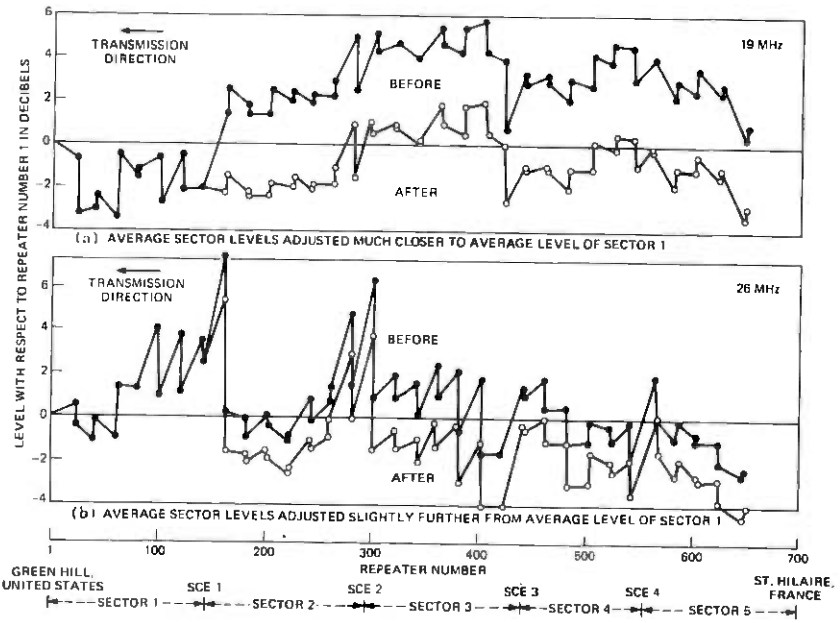


Fig. 4—TAT-6 lengthwise level distribution at 19 and 26 MHz before and after SCE adjustment.

end misalignment (and receive-end misalignment also) is equalized with the residual equalizer networks.

The residual equalizer networks are of the bridged-T type, custom-designed during commissioning using a previously written computer program. For TAT-6, a data link from Green Hill to the Bell Laboratories computer at Holmdel was utilized. Residual networks are built from a stock of high-quality components at the terminal sites and tested there for circuit integrity and transmission performance before being installed in the terminal.

Adjustable corrector equalizer networks are wideband networks of four types, each front-panel-adjustable around a nominal setting:

(*i*) Attenuator.
(*ii*) Square-root-of-frequency.
(*iii*) Slope.
(*iv*) Curvature.

Following s/n optimization (details are discussed in Section V), the remaining misalignment (including level changes introduced by s/n optimization) is equalized in the receive part of the terminal. Networks identical to those mentioned above are used to implement receive-end equalization. Figures 6a and 6b show that the remaining misalignment in the low and high bands, respectively, after terminal equalization is

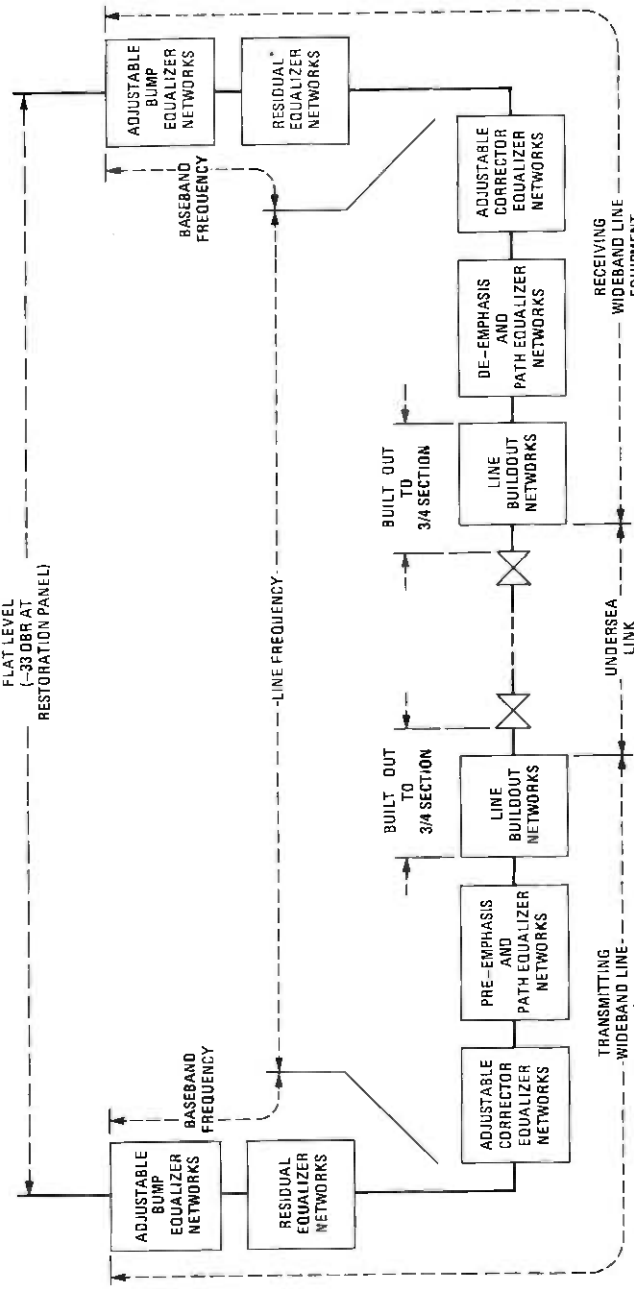


Fig. 5—Terminal wideband equalization plan.

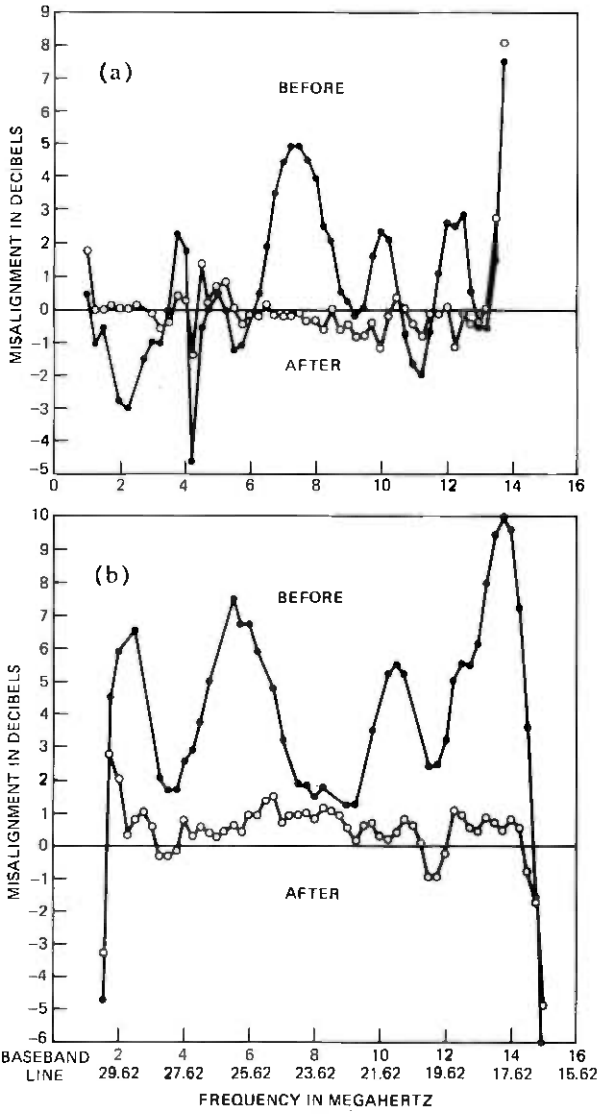


Fig. 6—TAT-6 misalignment before and after terminal equalization. (a) Low band. (b) High band.

for the most part within ±1 dB, except near the band edges. Band-edge misalignment is more conveniently equalized at the supergroup level (discussed in Section 4.4).

4.3.2 Transmit-end equalization

Normally, one-half of the wideband undersea misalignment is equa-

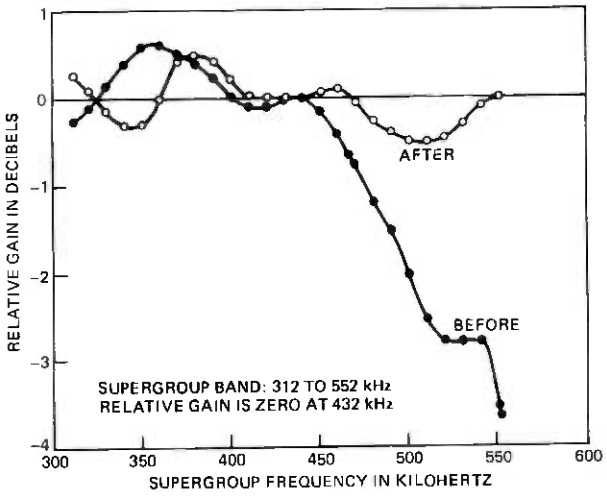


Fig. 7—TAT-6 low-band hypergroup 1, supergroup 4 response before and after supergroup equalization.

lized at the transmit end of the link. Assuming uniform misalignment with distance, this division of equalization should minimize the noise penalty due to misalignment. On TAT-6, however, misalignment is not uniform with distance. Because wideband average level versus frequency had been computed in preparation for the SCE adjustment procedure, it seemed more appropriate to decide the extent of transmit-end equalization by comparing the following three parameters:

(*i*) Average level.
(*ii*) End-to-end misalignment.
(*iii*) Estimated obtainable equalization from residual equalizers.

The final transmit-end equalization decision is obviously a compromise based on computations and experience. This equalization decision is too detailed to be discussed here.

4.4 Terminal equalization—supergroup

Supergroups are equalized individually only if the magnitude of the differences in transmission between the center of the band, 432 kHz, and the rest of the band, 312 to 552 kHz, exceeds 1.5 dB.[9] Supergroup equalization, where required, is carried out after final wideband equalization.

Supergroup equalization is accomplished in the *receive supergroup equalization unit*. For the fifty-some supergroups in each band of TAT-6, at most four band-edge supergroups in each band required equalization, plus two in-band supergroups in the low band. Figure 7 shows how the relative low-band misalignment in hypergroup 1, supergroup 4 (812 to

1052 kHz at line frequency) was equalized at supergroup frequency. An approximate +0.6 to −3.6 dB misalignment range was equalized to within ±0.5 dB.

V. OPTIMIZATION OF WIDEBAND SIGNAL-TO-NOISE RATIO

Optimization of the wideband s/n in the SG system is accomplished by changing one or more of the adjustable corrector or bump equalizer networks in the transmit terminal. This changing reshapes the broadband undersea transmission levels (additional signal shaping). The adjustment process is semi-empirical in that it determines, mostly by trial and error, whether or not a particular equalizer adjustment improves or degrades performance. The test signal load is broadband Gaussian white noise. (Test signal power is varied step-wise to define the maximum s/n at the frequency of test.) The process is actually one of obtaining undersea link levels versus wideband frequency for which the difference between the test signal noise load and the power sum of thermal noise and intermodulation distortion in narrowband test slots (the noise power ratio [NPR] in decibels) is a maximum.

For the TAT-6 link, NPR is measured in several narrow-band test slots in the low and high bands using a white noise test set. The wideband noise load, which is inserted at the transmit −33 dBr flat-level point on the restoration panel, is adjusted to be equivalent to a load of 4000 channels with an average channel power of −13 dBm0. A relatively narrow bandstop filter (part of the test set) is introduced to clear a slot in the transmitted noise load. The noise power which appears at the receive terminal restoration panel is measured at this slot frequency in an even narrower band (using test set filtering), both without and with the bandstop filter. The ratio of these two measurements is the NPR.

The final NPR data from the TAT-6 commissioning is given in Table I, including the equivalent 3-kHz channel noise in dBrnC0. Because of an unforeseen problem which resulted in higher-than-expected third-order intermodulation products, the desired NPR was not attained in

Table I — TAT-6 optimized noise power ratio and equivalent 3-kHz channel noise power

Low Band			High Band		
Freq (MHz)	Noise Pwr Ratio (dB)	3-kHz Ch Noise Pwr (dBrnC0)	Freq (MHz)	Noise Pwr Ratio (dB)	3-kHz Ch Noise Pwr (dBrnC0)
1.248	42.8	31.7	18.142	37.6	36.9
2.438	42.8	31.7	21.494	36.0	38.5
5.340	44.0	30.5	22.948	36.0	38.5
8.672	41.0	33.5	26.280	30.1	44.4
10.126	41.7	32.8	26.970	29.5	45.0
13.478	39.8	34.7	29.182	28.6	45.9

Table II — TAT-6 differential gain in undersea link

Low Band					High Band				
Freq (MHz)	Differential Gain in dB — dB Above Design Load — 2	4	6	8	Freq (MHz)	Differential Gain in dB — dB Above Design Load — 2	4	6	8
1.248	−0.1	−0.2	−0.3	−0.4	18.142	−0.1	−0.2	−0.3	−0.5
2.438	−0.05	−0.2	−0.15	−0.2	20.046	−0.1	−0.3	−0.4	−0.6
3.886	−0.05	−0.15	−0.3	−0.45	21.494	−0.1	−0.2	−0.5	−0.8
5.340	−0.1	−0.2	−0.4	−0.65	22.948	−0.1	−0.3	−0.5	−0.7
8.672	−0.2	−0.4	−0.8	−1.2	26.280	−0.1	−0.4	−0.7	−1.1
10.126	−0.25	−0.65	−1.15	−1.9	26.970	−0.2	−0.5	−0.9	−1.3
11.574	−0.1	−0.2	−0.4	−0.8	29.182	−0.1	−0.3	−0.5	−0.8
13.478	0.0	0.1	0.1	0.1					

Design load is 23 dBm0 per band.

the upper part of the high band.[4,8] In the low band, the attained NPR performance ranged from 3.8 to 8.0 dB better than the objective.

VI. SELECTED PERFORMANCE DATA—TAT-6

6.1 Differential gain—undersea link

Table II gives the differential gain of the undersea link as a function of relative signal load. For this test, a −20 dBm0 tone was inserted in a cleared slot in a broadband white noise test signal load. This test load was varied from 26 dBm0 (the design load) to 8 dB higher, with the difference between the received power of the −20 dBm0 tone at test loading and at design load being differential gain. The data show compression at all frequencies except near the top of the low band, where there is 0.1 dB of expansion. Maximum compression occurs near 10 MHz in the low band and 27 MHz in the high band. Below design load, it was determined that expansion (with respect to design load) was negligible.

6.2 Differential phase—undersea link

Figure 8 shows the differential phase of the undersea link in the high band as a function of signal load relative to high-band design load. A novel method making use of the repeater monitoring set (RMS) was used to measure differential phase. (The RMS is described in Ref. 1.) First, using standard techniques, a composite test signal was sent from the transmit end of the link:

(*i*) High band noise load amplitude modulated at 5 Hz to a depth of approximately 90 percent.

(*ii*) A tone of −20 dBm0 inserted in a cleared slot of the white noise load (the tone was not modulated).

The peak power (crest of modulation) of the modulated load is set at the desired test loading in the high band; the minimum power (valley of modulation) is approximately equivalent to no load. Second, at the re-

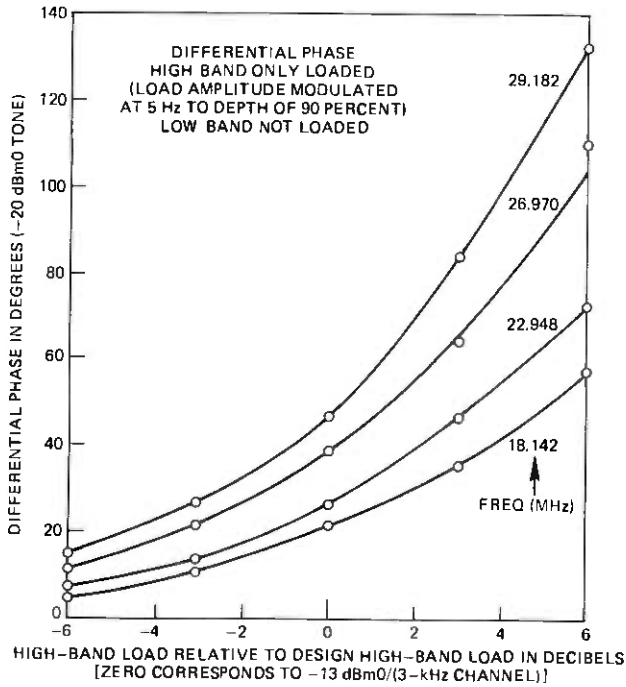


Fig. 8—TAT-6 high-band differential phase in undersea link.

ceive end of the link the composite signal is connected to the RMS. The transmit noise load is then reduced to essentially zero using an attenuator, and the RMS is phase-locked to the tone. Following this operation, the 50-Hz-wide, 100-kHz phase-locked loop of the RMS is opened. The transmit noise load is then reapplied.

The relative positions of the RMS reference 100-kHz signal and the phase-locked-loop 100-kHz error signal are monitored on an oscilloscope. Each of these signals is a square wave, and any movement of the error signal is easily calibrated in degrees. The difference between the position of the error signal at test loading power (crest of modulation) and no load (valley of modulation) is the differential phase of the undersea link. Drifting in the opened phase-locked loop is minimal, as can be observed on the oscilloscope by comparing the 100-kHz reference signal with the error signal at no load.

Maximum differential phase occurs at the top end of the high band, being 47 degrees at high-band design load and 133 degrees at 6 dB above this design load. The 133-degree figure has some relevance because undersea link highband levels will be increased from commissioning levels when the effect of third-order intermodulation products is reduced by terminal cancellation arrangements.[4,8]

Table III — TAT-6 measured phase jitter

Low Band			High Band		
Approximate Line Freq in MHz	Hyper-group—Supergroup	Phase Jitter in Degrees Peak-to-Peak	Approximate Line Freq in MHz	Hyper-group—Supergroup	Phase Jitter in Degrees Peak-to-Peak
1.0	1–4	0.8	16.9	6–11	1.2
4.2	2–13	0.8	19.8	5–12	1.2
7.5	3–9	0.9	22.2	4–12	1.2
10.2	4–9	1.1	24.9	3–13	1.4
13.5	5–5	1.1	26.7	2–10	2.4
			29.4	1–9	3.1

Measurements made in channel 1, group 1 of listed supergroup.

6.3 Phase jitter

Table III gives the measured phase jitter of the TAT-6 link. Phase jitter was measured in channel 1 of group 1 in a sufficient number of supergroups to characterize the low and high bands. Phase jitter is 0.8 degree peak-to-peak at the bottom of the low band and increases with frequency across the low- and high-band spectrum to approximately 3.1 degrees peak-to-peak.*

6.4 Impulse noise

Impulse noise power peaks exceeding approximately 12, 14, 16, and 18 dB above average (continuous) noise power were counted over 15-minute intervals for several supergroups to sample the low and high bands. Measurements were made at the channel level and supergroup level using different test instruments. The impulse noise count data were reduced to the fraction of time that signal peaks exceed peak amplitude threshold. Figure 9 shows these data for the low band at channel level.

Essentially, Fig. 9 confirms that the number of impulse counts measured at different peak amplitude thresholds follows a Gaussian distribution typical of thermal noise. The important observation from Fig. 9 is that, as long as the reduced data curves parallel that of the rmal noise, the source of the impulses is indistinguishable from system noise and not due to non-Gaussian processes such as corona. The fact that the curves do not lie on the thermal noise curve is caused by several factors, the largest being (*iii*), which fortunately causes only a horizontal offset:

(*i*) Error in measurement.
(*ii*) Calibration of test instrument.

* The phase jitter data are believed to be pessimistic because the measurements were influenced by channel noise, particularly at the higher line frequencies.

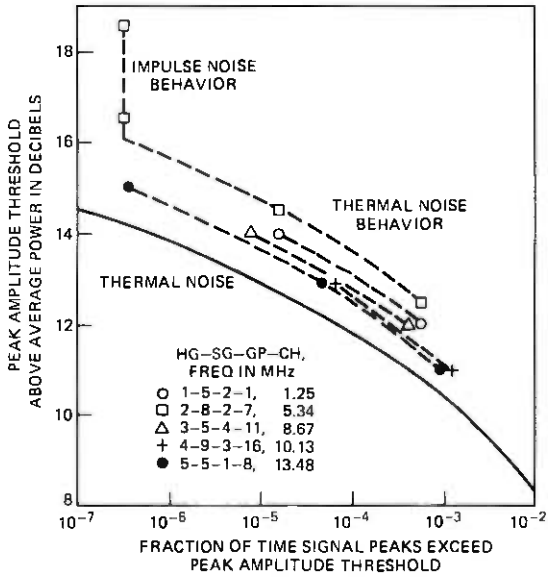


Fig. 9—TAT-6 low-band distribution of impulse noise count data.

(*iii*) Estimate of available number of time slots for which impulses could be counted.

The one indication of an impulse, a count exceeding two different threshold levels, may have been due to inadvertent terminal switching (probably caused by a carrier supply phase problem evident during the time that this test was made).

VII. COMPILATION OF COMMISSIONING DATA

Commissioning data are compiled in three areas, basically:

(*i*) Data books.

(*ii*) Punched paper tape.

(*iii*) Magnetic cassette tape.

The more important data from both stations are compiled at each station; therefore, duplication of data records is achieved. A separate data book is maintained for each of the four categories of tests mentioned in Section III. Each of the data books is sectionalized so that data from specific tests may be easily entered or retrieved.

The commissioning tests are generally run in sequence as directed in the test instructions, each succeeding test relying more or less upon the results of one or more previous tests. As the tests proceed day after day, the books are completed section by section, forming a permanent record of measured and computed data. Supplementing this record are selected

TTY punched tapes of measured data and magnetic cassette tapes. The printed page output from the computer system printer and graphs from the plotter are also entered into the data books where required. The punched tapes and magnetic cassette tapes are especially useful for comparison of data from identical tests separated in time, since these tapes provide an easy way to enter data into the computer.

Regarding residual equalizer design information, a complete record is kept in a separate book at each station:

(*i*) Identifying nomenclature.
(*ii*) Schematic (with element values).
(*iii*) Design and measured transmission.
(*iv*) Notes.

In addition to its record function, this book is used for maintenance purposes

VIII. ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions made by other members of the British Post Office, the Centre National d'Etudes des Télécommunications of France, and Bell Laboratories in carrying out the extensive commissioning test program. Particularly, we want to acknowledge the work of Messrs. P. P. Theophall and W. B. Hirt in adapting many parts of the test program to computer operation.

REFERENCES

1. J. E. H. Cosier, A. P. Davies, S. W. Dawson, Jr., R. F. Gleason, F. E. Kirkland, and T. A. McKenzie, "SG Undersea Cable System: Installation and Maintenance of Undersea System," B.S.T.J., this issue, pp. 2523–2545.
2. "SF Submarine Cable System," B.S.T.J., *49*, No. 5 (May–June 1970), pp. 601–798.
3. "SD Submarine Cable System," B.S.T.J., *43*, No. 4, Part 1 (July 1964), pp. 1145–1479.
4. C. D. Anderson, W. E. Hower, J. J. Kassig, V. M. Krygowski, R. L. Lynch, G. A. Reinold, and P. A. Yeisley, "SG Undersea Cable System: Repeater and Equalizer Design and Manufacture," B.S.T.J., this issue, pp. 2355–2402.
5. M. Brouant, C. Chalhoub, P. Delage, D. N. Harper, R. L. Lynch, and H. Soulier, "SG Undersea Cable System: Terminal Transmission Equipment," B.S.T.J., this issue, pp. 2471–2496.
6. G. E. Morse, S. Ayers, R. F. Gleason, and J. R. Stauffer, "SG Undersea Cable System: Cable and Coupling Design," B.S.T.J., this issue, pp. 2435–2469.
7. E. T. Calkin, I. Golioto, W. J. Schatz, R. F. Schroeder, and D. S. Shull, "SG Undersea Cable System: Undersea System Power," B.S.T.J., this issue, pp. 2497–2522.
8. S. T. Brewer, R. L. Easton, H. Soulier, and S. A. Taylor, "SG Undersea Cable System: Requirements and Performance," B.S.T.J., this issue, pp. 2319–2354.
9. Recommendation M46, "Bringing Internation Group, Supergroup, etc., Links Into Service," Vol. IV.1 of CCITT Green Book.

Contributors to This Issue

Cleo D. Anderson, B.S.E.E., 1960, University of Idaho; M.E.E., 1962, New York University, Bell Laboratories, 1960—Mr. Anderson was involved in systems analysis of the SD and SF undersea cable systems. He then supervised a high-frequency radio group designing facilities for handling overseas and high-seas traffic. Returning to undersea cable work, he was responsible for circuit design of repeaters and the repeater monitoring set for the SG system. Currently, he supervises a group responsible for analyses of undersea cable systems. Member, IEEE, Sigma Tau, Phi Kappa Phi, Eta Kappa Nu.

Sheila Ayers, Birkbeck College, London University; British Post Office, Research Department, 1943–. Ms. Ayers has investigated nonlinear resistors, piezoelectricity, ferroelectricity, magnetostriction, electro-optics, and low-loss dielectrics. Since 1973, she has been head of a group investigating materials problems in submarine cables and measurement techniques associated with these investigations.

S. Theodore Brewer, B.S. (E.E.), 1937, M.S. (E), 1938, Purdue University, Bell Laboratories, 1937—. In his early assignments, Mr. Brewer contributed to the development of broadband coaxial systems and video feedback amplifiers. Later, he explored network control for electronically-controlled switching systems. More recently he was responsible for system and undersea electronics design of the SF Undersea Cable System. Currently, he heads a department which was responsible for system and undersea electronics design for the SG Undersea Cable System. Exploratory work under his direction includes an undersea system of 16,000 channels' capacity. He holds patents on control and feedback systems, switching networks, and repeater circuits. Member, IEEE, Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

Michael Brouant, graduate, Ecole Nationale Supérieure de l'Aéronautique, Paris, 1962; CIT-Alcatel, 1964—. At CIT-Alcatel, Mr. Brouant worked on the design of analog line equipment. Since 1968, he has directed the Undersea Terminal Design Section which has carried out detailed design of terminals for the French S1, S5, and S25 systems, as well as the 60-MHz land system. He also was responsible for detailed design of terminals for the SG undersea cable system. In 1975, he took charge of the lightwave transmission section which is developing applications of optical fibers to telecommunications.

Edwin T. Calkin, B.S. (Eng.), 1961, the Cooper Union School of Engineering; M.S.E.E., 1963, New York University; Bell Laboratories, 1961—. From 1964 to 1975, Mr. Calkin was principally engaged in circuit development for power converters that energize land and undersea cables. Since 1975, he has been responsible for circuits in low- and medium-power converters used in a variety of Bell System applications. Member, IEEE, AES, and Tau Beta Pi.

Christian Chalhoub, graduate, Ecole Supérieure d'Electricité, Paris, 1941. Mr. Chalhoub started his career in the Société d'Etudes pour Liaisons Téléphoniques where he first worked on measurement technology and later on equipment for studio broadcasting. He worked briefly for a firm concerned with medical electronics and then, in 1945, joined CNET (Centre National d'Etudes des Télécommunications), where he was responsible for the laboratory studying high-frequency terminal transmission equipment for carrier systems. In 1956 he joined CIT (Compagnie Industrielle des Télécommunications), a subsidiary of CGE (Compagnie Générale d'Electricité), where he managed the department designing line systems for both land and undersea use. He is currently Deputy Director in charge of Product Policy in the Transmission Division of CIT. Member, SEE (Société des Electriciens, des Electroniciens et des Radioélectriciens).

J. E. H. Cosier has been with the British Post Office Research Department for 41 years, the last 32 on undersea repeater systems. His work has covered the field of housings, high pressure seals, undersea cable jointing methods, terminations and taut wire equipment. He is a co-inventor and patentee of the British design of the linear cable engine, and of the cut-and-hold grapnel. Since 1972, he has been Head of the Mechanical Design and Shipboard Laying Equipment Section of the Research Department.

A. Peter Davies, British Post Office. Mr. Davies started as a telecommunications apprentice on automatic switching in a London telephone district in 1936. He shifted to cable transmission early in World War II, seeing Army service on the undersea cables laid to serve the European front. Later, he was involved with the development of the U.K. coaxial cable network begun about 1948, transferring to undersea cables in 1962, where he has remained ever since. His interests include the practical application of submerged repeater technology in the marine environment. He is presently Head of the BPO section responsible for undersea plant development, installation, and maintenance.

S. W. Dawson, Jr., B.E.E., 1968, University of Virginia; M.S.E.E., 1970, Duke University; Bell Laboratories, 1968—. Mr. Dawson worked on specialized test equipment design for military undersea cable systems before his work on the SG system installation test equipment. He is currently engaged in military hardware systems design. Member, Tau Beta Pit, Eta Kappa Nu.

Paul Delage, graduate, Ecole d'Electricité Industrielle, Paris, 1942. Mr. Delage entered the Société d'Etudes pour Liaisons Téléphoniques et Télégraphiques à Longue Distance in 1942, where he worked on measurement, installation, and splicing of long-distance cables. Following this, he worked in the Alcatel Laboratories, which later became CIT-Alcatel, developing multi-channel analog transmission equipment. This work extended from channel banks to equipment for the 60-MHz coaxial system. He is now Deputy Head of the Analog and Digital Multiplex Laboratories.

Robert L. Easton, B.S. (M.E.), 1953, M.S. (M.E.), 1954, California Institute of Technology; Bell Laboratories, 1954—. Mr. Easton was responsible for system aspects of the earlier SD and SF Submarine Cable Systems. In the development of the SG system, he was responsible for economic studies, signal analysis, and equalization. Currently, he heads a group concerned with undersea system planning and studies associated with a new time-assignment speech interpolation system. Member, Tau Beta Pi.

Robert D. Ehrbar, B.S.E.E., 1937, Johns Hopkins University; Bell Laboratories, 1937—. When he joined Bell Laboratories, Mr. Ehrbar worked on cable carrier systems. During World War II, he helped develop various airborne radar systems and after the war worked on the development of a high-capacity coaxial cable system. In 1955 he became Head, Submarine Cable Systems Department and in 1964 Head, Exchange Transmission Department. In December 1968, he was appointed Director, Exchange Transmission Laboratory and in April 1971, Director, Undersea Cable Laboratory. Member, IEEE.

Arthur E. Ford, British Post Office Research Laboratories, 1938—. Early in his career Mr. Ford was engaged in precision HF techniques, and in the evolution of sweep-frequency scanning and white noise measurement. He left the Research Laboratories in 1962, and joined the Telecommunications Headquarters Division, planning and laying those world-wide undersea cable systems in which the BPO had an interest. He currently heads this division. He was honored by her Majesty Queen

Elizabeth II with the award of Order of the British Empire in 1976 for his services to undersea cables.

W. M. Fox, B.S., 1951, Muhlenberg College; M.S., 1953, St. Lawrence University; The Franklin Institute Laboratories, 1953–1957; Bell Laboratories, 1957—. Mr. Fox has been engaged in transistor development, including devices for the SF and SG undersea systems. Currently, he is working on the development of linear high-frequency power transistors and on low-power, low-noise transistors.

Guy Gerbier, graduate, Ecole Polytechnique of Paris, 1954; Ecole Nationale Supérieure des Télécommunications of Paris, 1957. Upon completing his studies, Mr. Gerbier joined CNET (Centre National d'Etudes des Télécommunications), where he worked on the qualification of components for electronic switching and electronic automatic controls. Next, he headed the Department of Research in Postal Mechanization and later assumed responsibility for the CNET Division of Postal Mechanization. In 1969, he joined the Transmission Lines and Equipment Division of CNET as Deputy Director, and later became Division Director. In this responsibility, he worked on cables and their final realization, long-distance transmission systems, and local networks. Currently, as General Engineer he is responsible for the ground communications network of the French Navy. Member, SEE (Société des Electriciens, des Electroniciens et des Radioélectriciens).

R. F. Gleason, B.S., Engineering Science, 1961, Case Institute; M.S., Engineering Mechanics, 1963, N.Y.U.; Ph.D., 1967, Stanford University. Bell Laboratories, 1961—. At Bell Laboratories, Mr. Gleason has worked on undersea cable system physical design. He has been responsible for stress and vibration analysis of repeaters and equalizers and for design of high pressure feed-through terminals for undersea use. He presently supervises a group with responsibilities including cable and cable termination design, and development of cable handling equipment and methods.

Igor Golioto, M.E. 1961, Stevens Institute of Technology; M.S.M.E., 1963, New York University; Bell Laboratories, 1961–1977. Mr. Golioto was a member of the Power Systems Physical Design Department until his death in April 1977. He was responsible for the physical design of power converters for carrier and microwave systems, for No. 1 ESS, and for the SF, SG, and military undersea cables.

Douglas N. Harper. Mr. Harper joined the British Post Office in 1946. Since 1961, he has been involved in determining intercontinental transmission standards for both cable and satellite facilities and has been responsible for overall planning and preparation of specifications for a number of major submarine cable system projects. These include U.K.-Portugal, MAT 1 and several U.K.-European cables. During the procurement of MAT 1 he was a consultant to Italcable for the detailed planning, installation, and commissioning of the system. More recently, Mr. Harper has been involved in the SG cable system development, particularly in the terminal transmission equipment area, and in commissioning the TAT-6 system. Currently he is Head of the Trans-Atlantic Systems Planning Group, with specific responsibilities as the BPO Co-ordinator for the TAT-7 project.

W. E. Hower, B.S.(E.E.), Newark College of Engineering; M.S. (Management Science), Stevens Institute of Technology; Western Electric, 1961—. Mr. Hower began in Western Electric with assignments in industrial engineering and test set development. He was promoted to Department Chief in 1969 and in December of that year assumed responsibility for undersea cable repeater assembly and test engineering. He is currently responsible for customer-premise power engineering in Kearny.

J. J. Kassig, B.S.E.E., 1952, Massachusetts Institute of Technology; M.S.E.E., 1955, Rutgers University; Bell Laboratories, 1955—. Mr. Kassig has worked on the design of repeaters for the SD, SF, and SG undersea systems. He has also been involved in the areas of semiconductor device characterization, in network optimization techniques using digital computers, and in the application of signal flow graphs to the design of feedback amplifiers. He is currently exploring design of a high-speed regenerator for an undersea optical fiber system.

F. E. Kirkland, B.S.E.E., 1957, Auburn University; Western Electric, 1957–1962; Bell Laboratories, 1962–1976; American Telephone and Telegraph Company, Long Lines Department, 1976—. Mr. Kirkland's early work at Western Electric involved preparation of technical publications. Subsequently, he was engaged in design and development of Naval surface weapons direction equipment, first with Western Electric, then at Bell Laboratories. In 1973, he began design work on new generation undersea cable laying test equipment which was being developed for the SF and SG systems. He participated extensively in the installation

of the HAW-3/TRANSPAC-2 links, and the TAT-6 project. At American Telephone and Telegraph Company, Long Lines, Overseas, he is currently involved with various transmission measurement systems.

Victor M. Krygowski, B.S.M.E., 1942, N.J. Institute of Technology, M.S. (Engineering Management), 1951, Stevens Institute of Technology; Western Electric, 1946—. Mr. Krygowski's areas of responsibility have covered engineering, shop, inspection and testing, quality control, data processing, production service, and plant maintenance. Among the items worked on were PBXs, key equipment, and apparatus and piece parts. He was promoted to Department Chief in Manufacturing Engineering in 1955 and to Assistant Manager in 1959. From 1972 to 1975, Mr. Krygowski was Engineering Assistant Manager in the undersea cable repeater shops in Clark, N.J., where he was responsible for manufacturing and planning for the SDC, SF, and SG undersea repeaters and equalizers. He is presently Assistant Manager, Industrial Engineering and Wage Practices at the Kearny Works.

Bror O. Larson, B.S.E.E., 1958, Pennsylvania State University; M.S.E.E., 1960, New York University; Bell Laboratories, 1958—. Mr. Larson participated in the Telstar satellite development, with responsibility for the command tracker and command transmitter systems. Subsequently, he was involved in the development of single-sideband radio transmitters for the high-seas and point-to-point services maintained by the Long Lines Department of AT&T. He prepared the detailed test specifications and procedures for the installation and the final system alignment and evaluation of the SG Undersea Cable System, and participated in the commissioning of TAT-6. Currently, he is a member of the Undersea Cable Laboratory. Member, IEEE, Eta Kappa Nu, Phi Kappa Phi, Tau Beta Pi.

Michel Laurette, graduate, Ecole Polytechnique de Paris and Ecole Nationale Supérieure des Télécommunications; Centre National D'Etudes Des Télécommunications (CNET), 1967—. Mr. Laurette has participated in the development of French analog transmission systems, including a 60-MHz terrestrial facility and a 25-MHz undersea system. He participated in the SG system project, primarily with regard to the terminal equipment and final system alignment and evaluation of the TAT-6 link. Currently, as Transmission Engineer, he heads the Line and Transmission Equipment Engineering Department at CNET. Mr. Laurette is a member of CCITT, where he is the French representative for Study Commission XV.

Robert L. Lynch, A.S., Kansas City (Missouri) Junior College; B.S.E.E., 1957, Kansas University; M.E.E., 1959, New York University; Bell Laboratories, 1957—. Mr. Lynch's early work included development of cable machinery for the Bell System Cable Ship *Long Lines* and design

of transmission equipment for the shore terminals of the SD and SF undersea cable systems. His first supervisory responsibilities concerned the physical design of the SF system terminal transmission equipment and of the TASI-B system. Subsequently, he was assigned responsibility for the complete SF transmission terminal. More recently, his group has been responsible for a number of facets of SG undersea cable system design including undersea equalization and the testing procedures used during cable laying and commissioning. He provided liaison with the British and French Post Offices on these matters, and, in particular, with CIT-Alcatel during development of the SG terminal transmission equipment. He participated extensively in the installation of the TAT-6 link.

T. A. McKenzie, B.S.E.E., 1955, University of Tennessee; Bell Laboratories, 1955–1978. Initially, Mr. McKenzie worked on military undersea cable systems. In the SG development, he was responsible for the design of specialized test equipment. He retired from Bell Laboratories in 1978. Member, Phi Eta Sigma, Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi.

Geoffrey E. Morse, C. Eng.; British Post Office, Research Department, 1944—. Initially, Mr. Morse was concerned with the design of the British 0.99-in. and 1.47-in. lightweight cables. Since 1956, he has been engaged in undersea cable development. At commencement of the SG project, he became head of the cable group. Member, I.E.R.E.

Paul R. Munk, B.S.E.E., 1951, Purdue University; M.S.E.E., 1952, Purdue University; Bell Laboratories, 1952—. At Bell Laboratories, Mr. Munk designed inductors and transformers for a number of Bell System and military applications, including the Telstar satellite and undersea cable systems. He is currently developing computer software for the Mechanized Loop Testing System. Member, Eta Kappa Nu, Sigma Xi.

G. A. Reinold, B.S.M.E., 1962, Union College; M.S.M.E., 1964, New York University; Bell Laboratories, 1962—. Mr. Reinold was first involved in the development of undersea cable installation and repair techniques. He later worked on methods for installing cables underground with minimal soil disturbance. He then participated in ocean cable design for the military. In 1969, he became supervisor of the undersea apparatus mechanical design group, where he was responsible for the physical design of the SG repeater. He is currently responsible for the physical design of future undersea cable repeaters and for the cable location system on the SCARAB submersible.

Eugene F. Sartori, B.S., 1942, Massachusetts Institute of Technology; M.S., 1961, Stevens Institute of Technology; Bell Laboratories, 1942—. Mr. Sartori was initially concerned with the design of magnetic apparatus for both military and Bell System applications. He was responsible for high-reliability transformer design for the Telstar satellite and for undersea cable systems. He is presently involved in terminal equipment studies and subscriber loop maintenance.

William J. Schatz, B.S. (E.E.), 1965, New Jersey Institute of Technology; Bell Laboratories, 1958—. Since 1965, Mr. Schatz has been involved in circuit development in the Electronic Power Systems Laboratory.

Robert E. Schroeder, B.S.E.E., 1975, New Jersey Institute of Technology; Bell Laboratories, 1967—. Mr. Schroeder has been responsible for circuit development of dc-dc converters and regulators used in the T2 carrier system, the power converters for the SG undersea cable system, and recent exploratory work on high-power dc-dc converter plants.

Daniel S. Shull, B.S.E.E., 1951, University of South Carolina; M.S.E.E., 1960, Carnegie Institute of Technology; Ph.D., 1965, Carnegie Institute of Technology; Bell Laboratories, 1964—. Mr. Shull joined Bell Laboratories with the Power Apparatus Department in Winston-Salem, N.C., where he was involved in the development and design of magnetic devices and power equipment for defense applications. Upon his transfer to the SG Installation Equipment Group in 1971, he assumed responsibility for the development and design of SG-terminal-PSF equipment for both cable installation use and permanent system operation. Since 1974, he has been assigned to the Local Electronic Switching Systems Development Laboratory. He is a senior member of the IEEE, and has served on the Administrative Committee of the IEEE Magnetics Society, and is a past chairman of the technical program committee of the IEEE International Magnetics Conference. Member, Tau Beta Pi, Phi Beta Kappa.

Henri Soulier, graduate, Ecole Polytechnique, Paris, 1958; Ecole Nationale Supérieure des Télécommunications, Paris, 1961. Mr. Soulier began his career at CNET (Centre National d'Etudes des Télécommunications) working on semiconductors. Since joining the Transmission Department of CNET in 1964, his activities have concerned analog transmission: multiplex equipment and particularly system design for coaxial cable facilities, both land and undersea. He was responsible for

planning and specifying the terminal transmission equipment of the SG system. Appointed Chief Engineer in August 1972, he is now Deputy Director of the Cable and Microwave Radio Transmission Division. Member, SEE (Société des Electriciens, des Electroniciens et des Radioélectriciens).

J. R. Stauffer, B.S.E.E., 1962, Purdue University; M.S.E.E., 1964, New York University; Bell Laboratories, 1962—. Mr. Stauffer worked on various aspects of the SF undersea cable repeater, and subsequently supervised a group that developed trunk electronics for military systems. Working with the British Post Office, his group assisted in characterizing the transmission properties of cable use in the SG undersea system. Recently, Mr. Stauffer has been investigating the possibility of applying fiber optics to undersea cable systems. Member, Tau Beta Pi, Eta Kappa Nu.

S. A. Taylor, B.S.C. (Tech.), 1950, Manchester University, England; British Post Office, Research Department, 1950—. Initially, Mr. Taylor worked on the development of wideband feedback amplifiers and the electrical design of submarine repeaters. He is currently head of the Submarine and Inland Cable Systems Division of the Research Department. Member, IEE.

P. A. Yeisley, Jr., B.S. (Physics), 1952, Lafayette College; Bell Laboratories, 1952—. Mr. Yeisley first worked on the development of the L3 coaxial cable system. Since 1954, he has worked on the physical design of the SB, SD, SF, and SG undersea cable systems.

W. H. Yocom, B.A. (Physics), 1940, Oberlin College; S.B. (Electrical Engineering), 1942, Massachusetts Institute of Technology; M.S. (Electrical Engineering), Stevens Institute of Technology; Bell Laboratories, 1942–1956, 1964—. Mr. Yocom had assignments in physical design of radar and test equipment, then studied electron dynamics and did research on traveling wave tubes and optical delay lines. Later he supervised a group developing tantalum thin-film hybrid circuits, and now supervises a group that coordinates the application and design of discrete passive electronic components. Senior Member, IEEE, Sigma Xi.

THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 57 September 1978 Number 7, Part 2

An Approximation for the Variance of the UPCO Offered Load Estimate

By E. J. MESSERLI

(Manuscript received September 26, 1977)

This paper develops a generalization of some available approximations for the variance of the estimate for offered load to a trunk or server group operating in a blocked-calls-cleared mode, using measurements of usage, offered attempts (peg count), and overflow. The analysis takes into account the peakedness of the offered traffic stream, the level of blocking on the group, the duration of the measurement interval, and switch count errors due to sampling usage. The resulting approximation is quite accurate over a wide range of conditions, is easily computable, and clearly displays the role of the basic factors that control the precision of the estimator. The variance approximation is useful in studies of the relationship between traffic measurement errors and the performance of the provisioning and administration processes.

I. INTRODUCTION

The estimation of loads offered to a trunk group or server group operating in a blocked-calls-cleared mode plays an important role in many network-provisioning processes. The preferred measurement combination for developing such load estimates consists of usage, offered attempts (peg count), and overflow attempts (usually referred to in the Bell System as UPCO measurements). This paper develops a generalization of some available approximations for the variance of the UPCO offered load estimate for a single measurement interval. The analysis considers the peakedness of the offered traffic stream, the level of blocking or call congestion for the group, the duration of the measurement interval, and switch count errors due to the sampling of usage at discrete points in time. The resulting approximation is quite accurate

over a wide range of conditions, is easily computable, and clearly displays the role of the basic factors that control the precision of the estimator.

Variance approximations are useful in designing measurements and in studying relationships between traffic measurement errors and the performance of the provisioning and administration processes. For example, the relationship of actual traffic measurement accuracies (which can be further corrupted by wiring, data base, and recording errors) to the quality of the trunk provisioning process was studied in Ref. 1. The variance approximation developed here was useful in quantifying the background accuracy of the process.

This paper is organized as follows. The basic approximation is presented and discussed in Section II. The development of the approximation is given in Section III; supporting analysis of switch count error is developed in the appendix. Concluding remarks are given in Section IV.

II. THE BASIC APPROXIMATION

Figure 1 illustrates UPCO measurements for a measurement interval of length T, with usage scan interval s. The UPCO estimate for the offered load during this measurement interval is given by

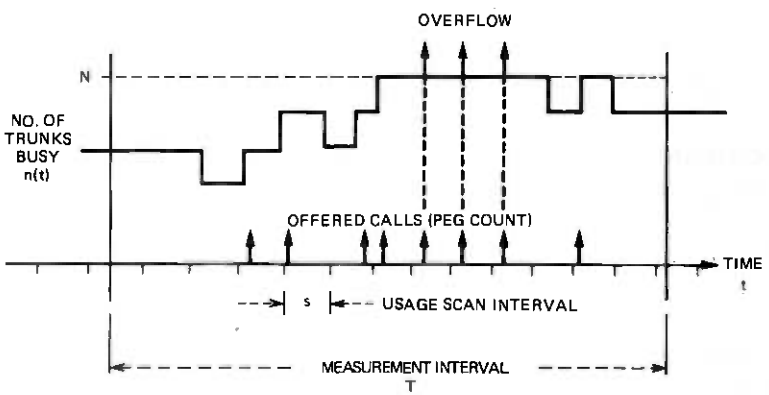


Fig. 1—UPCO measurements.

$$\hat{a} = \frac{\text{average measured usage}}{\text{1-measured blocking}}, \tag{1}$$

where the measured blocking is the ratio of overflow to offered attempts. It is well known that (under reasonable conditions subsequently discussed) this is an unbiased estimate for the true offered load a during this interval.

Early work on analyzing offered load estimators was carried out, among others, by R. I. Wilkinson,[2] who addressed the reliability of holding time estimates. In a 1952 paper,[3] W. S. Hayward, Jr., drawing on some of Wilkinson's analysis, addressed the variance of offered load estimates based on sampled usage. Hayward's model assumed Poisson arrivals, exponential holding times, and no blocking, yielding the result

$$\text{var}(\hat{a}) = \frac{\bar{h}a}{T}(2 + q), \tag{2}$$

where a is the offered load in erlangs, $\bar{h}$ is the average holding time, and T is the length of the measurement interval. The parameter q is given by

$$q = v\frac{1 + e^{-v}}{1 - e^{-v}} - 2, \tag{3}$$

where $v = s/\bar{h}$, and s is the usage scan interval; q determines the variance contribution due to switch count (sampling) error, e.g., $q = 0$ for $s = 0$, the continuous scan case.

In more recent work, Hill and Neal[4] addressed the question of the variance of $\hat{a}$ for peaked traffic,* but did not consider congestion or switch count error. Through the application of an asymptotic result for the variance of the renewals for a peaked traffic stream, they obtained the expression

$$\text{var}(\hat{a}) \cong \frac{2\bar{h}az}{T}, \tag{4}$$

where z is the peakedness factor for the stream.

In this paper, we combine elements of both of these previous analyses

* Peaked traffic refers to overflow traffic, or to streams containing some overflow traffic. The peakedness factor $z(\mu)$ (or z if μ is understood) is the equilibrium variance-to-mean ratio of busy servers when this traffic is offered to an infinitely large group of exponential servers with service rate μ. The peakedness factor is one for Poisson traffic and is larger than one for overflow traffic.

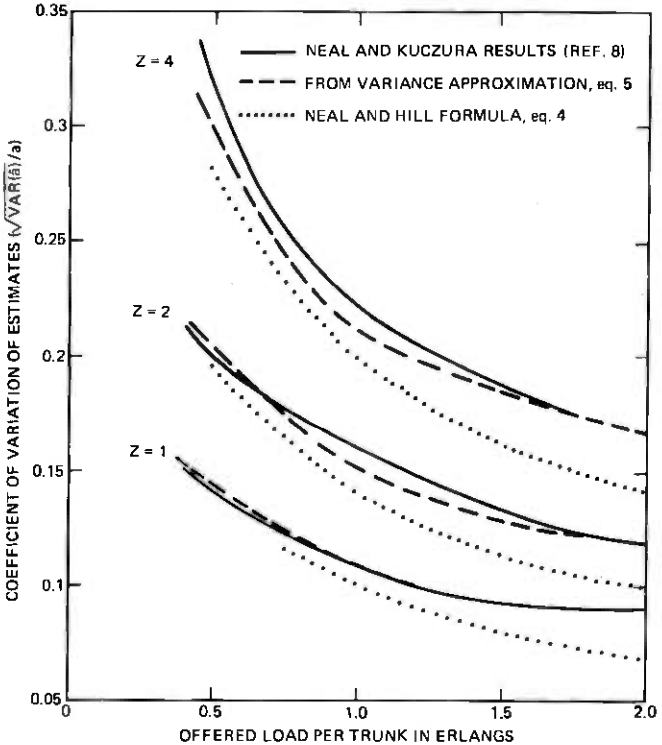


Fig. 2—Comparison of variance approximations for $N = 10$ servers ($\bar{h} = 180$ s, $T = 3600$ s, $s = 100$ s).

and explicitly consider the effect of blocking on the group, to obtain the generalization

$$\text{var}(\hat{a}) \cong \frac{\bar{h}a}{T}\left(2z + \frac{B+q}{1-B}\right), \tag{5}$$

where B is the equilibrium call congestion,* i.e., the fraction of attempts blocked. Thus, congestion basically adds a term to the previous various approximations.

Figures 2 and 3 show comparisons of the variance approximation (5) with the reference approximations obtained via the error theory devel-

* The blocking B is defined in theory as the probability that an arbitrary attempt is blocked. In practice, when the load parameters a,z are given, the blocking or call congestion B is assumed to be defined by the equivalent random method (Ref. 5), so that $B = f(N,a,z)$ where N is the number of trunks in the group. Otherwise, as shown by Holtzman (Ref. 6), the blocking B is not uniquely defined by N,a,z, but may take on a range of values, depending on higher order characteristics of the traffic stream. The actual value of $f(N,a,z)$ may be obtained from traffic tables normally used in administering trunking networks. It may also be estimated by Hayward's approximation, $f(N,a,z) \cong B(N/z,a/z)$ (Ref. 7), thus allowing Erlang $B(.,.)$ tables or formulas to be used.

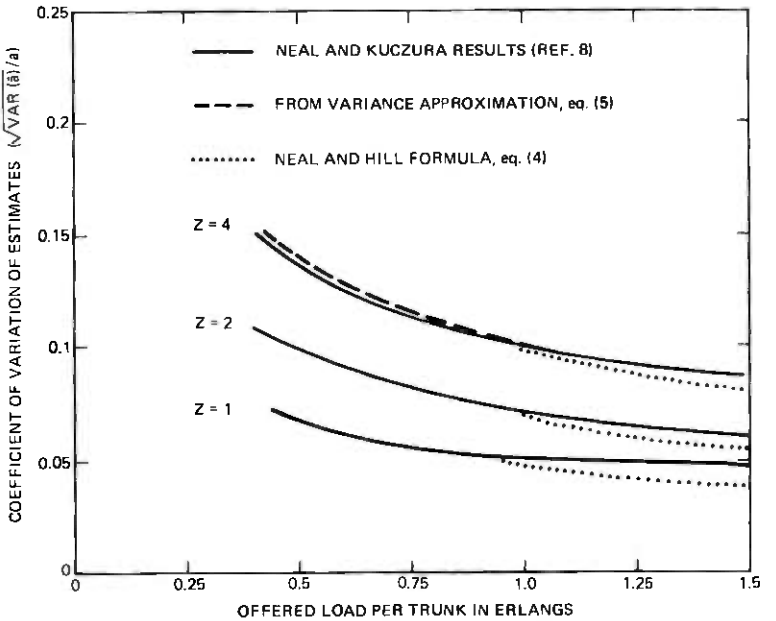


Fig. 3—Comparison of variance approximations for $N = 40$ servers ($\bar{h} = 180$ s, $T = 3600$ s, $s = 100$ s).

oped by Neal and Kuczura.[8*] These results assume that $\bar{h} = 180$ s, $T = 3600$ s (i.e., $\bar{h}/T = 0.05$), and $s = 100$ s. For a wide range of congestion and peakedness conditions, the agreement between eq. (5) and the reference results is very good. Neal and Kuczura also determined by numerical comparisons that switch count error was a small contributor to $\text{var}(\hat{a})$. Since q is small for typical scan-interval-to-holding-time ratios (e.g., $q \cong 0.05$ for $s = 100$ s and $\bar{h} = 180$ s, which are typical scan intervals and holding times for Bell System trunks), this conclusion is also evident from eq. (5).

Figures 2 and 3 also show the behavior of the Neal and Hill result, eq. (4). As the load per trunk increases, it is clear that the contribution of the congestion term in eq. (5) is increasingly important. These higher levels of congestion occur quite commonly on high usage groups, where a substantial fraction of the busy hour loads may be overflowed to an alternate route. As the load is increased to very large values, the coefficient of variation using eq. (4) goes to 0, whereas Figs. 2 and 3 suggest that the coefficient of variation has a positive limit as $a \to \infty$. It can be shown that (for any z) as the attempt rate $\lambda \to \infty$,

$$\lim_{\lambda \to \infty} \text{var}(\hat{a})/a^2 = \bar{h}/TN, \tag{6}$$

* This error theory is applicable to general functions of the e UPCO measurements. The approximation developed for the UPCO offered load estimate is computationally much more complex, as well as less transparent, than eq. (5). The Neal and Kuczura approximation agreed well with simulation results, and hence is a suitable reference for comparing eq. (5).

where N is the number of servers in the group.* Equation (6) has a simple interpretation. The UPCO offered load estimate may be viewed as the product of essentially independent estimators for the attempt rate λ and for the mean holding time $\bar{h}$. As $\lambda \to \infty$, the coefficient of variation for the first estimator goes to 0. Equation (6) represents the squared coefficient of variation for the second estimator, i.e., the positive limit results from having only a finite number of carried attempts from which to estimate mean holding time. For Figs. 2 and 3, the asymptotic limits for the coefficient of variation are 0.071 and 0.035, respectively.

If a is assumed to have a mean a_o and variance $\sigma_a{}^2$, one is often interested in estimating a_o. The results of this section can be applied to obtain $\text{var}(\hat{a}_o)$ for a single measurement period by interpreting them as conditional results, i.e., $\text{var}(\hat{a}|a)$, in the expression

$$\text{var}(\hat{a}_o) = \sigma_a{}^2 + E_a \text{ var}(\hat{a}|a). \tag{7}$$

In many cases, the $\sigma_a{}^2$ term can be a significant contributor. For example, in trunk engineering $\sigma_a{}^2$ may represent a day-to-day variance under an i.i.d. model for busy-hour loads (in this case, a_o is usually estimated from 5 to 20 busy-hour loads) and can be quite large in relation to the other sources of variability.

III. DEVELOPMENT OF THE APPROXIMATION

Consider a full access group of N servers operated in a blocked-calls-cleared mode. The offered traffic process is assumed to be a (nonlattice) renewal process with rate parameter λ, and server holding times are assumed to be exponential with hang-up rate μ. We define the mean and peakedness of the offered load by $a = \lambda/\mu$, $z = \text{var}(n(t))/E(n(t))$, where $n(t)$ is the equilibrium occupancy when the renewal process is offered to an infinitely large group of exponential servers with rate μ. The parameters (a,z) are conventionally used in traffic engineering, and hence it is useful to relate the variance approximation to these parameters.

For a measurement period of length T, let u,p,o denote average measured usage, offered attempts, and overflow attempts, as illustrated by Fig. 1. The average measured usage is defined by $u = 1/m \sum_{j=1}^{m=\lfloor T/s \rfloor} n(js)$ if $s > 0$, and by $u = 1/T \int_0^T n(t)dt$ if $s = 0$, where $n(t)$ is the number of busy servers at time t. It is assumed that equilibrium conditions apply at the beginning of the measurement interval, both for the occupancy on the servers and for the renewal processes corresponding to arrivals and overflows.

* This result is not the same as the limit obtained from eq. (5) as $a \to \infty$, which gives $(1 + q)\bar{h}/TN$. The discrepancy arises because the model for switch count error used in the development of eq. (5) breaks down as $\lambda \to \infty$. For this unrealistic limiting case, the servers are occupied 100 percent of the time, and *no* error is introduced by scanning. The correct result is thus obtained by noting that the carried attempt process approaches a Poisson process with rate $N\mu$ as $\lambda \to \infty$.

The UPCO estimate for the offered load a over the measurement period is

$$\hat{a} = \frac{u}{1 - o/p} = p\frac{u}{c} = \frac{p}{T}\left(\frac{Tu}{c}\right), \tag{8}$$

where $c \triangleq p - o$. Thus, $\hat{a}$ may be viewed as the product of separate estimators for the arrival rate (p/T) and for the average holding time (Tu/c). The approximation for $\text{var}(\hat{a})$ is obtained by introducing an approximate treatment of the scanning error, and then by examining (8) for large T. However, while the *structure* of the approximation is motivated by asymptotic analysis, the *validity* of the approximation is based on its accuracy for realistic values of T.

3.1 Treatment of scanning error

The scanning error for usage affects only the value Tu in (8), which may be expressed as

$$Tu = \sum_{j=1}^{c} \hat{h}_j + r_0 - r_T, \tag{9}$$

where $\hat{h}_j$ is the sampled holding time estimate for the jth call to be accepted by the group, $\hat{h}_j \in \{0,s,2s,\cdots\}$, and r_0, r_T are end effects. In particular, if the jth call to be accepted by the group was hit by k_j scans, then $\hat{h}_j = k_j s$ may be viewed as the sampled holding time estimate for this call. The variable r_0 is the total measurement period usage attributable to calls already in progress at the beginning of the interval, while r_T is the total usage due to accepted calls that would be measured in the subsequent measurement period of length T.

Throughout this analysis, we make the following simplifying assumptions:

(*i*) $\hat{h}_j, j = 1,2,\cdots,c$ are independent random variables.

(*ii*) $\hat{h}_j = h_j + e_j$ where e_j is the scanning error that results when a call with exponential holding time h_j begins at a time which is uniformly distributed between two successive sampling instants.

These simplifying assumptions hold *exactly* for the case $B = 0, s = 0$ (no congestion and continuous scan) and any z, since all calls are carried and the holding times are i.i.d. exponential random variables. They also hold *exactly* for the case $B = 0, s > 0$, and $z = 1$, since for a Poisson process the arrivals in disjoint intervals are independent. Furthermore, given a *fixed* number of arrivals in an interval (in particular, an interval of length s), the arrival times are independent and uniformly distributed within the interval. Thus, the simplifying assumptions—while not always true—can be rigorously justified for some important cases. In general, they can be expected to be reasonable assumptions if the usage on each server in the group does not approach unity, i.e., if congestion is not too severe.

As a result of the simplifying assumptions, the scanning error need only be examined for an isolated call. The analysis for this situation is treated in the appendix, where it is shown that with $e = \hat{h} - h = ks - h$, i.e., the sampled holding time minus the true holding time,

$$E(e) = 0 \tag{10}$$

$$\text{cov}(h,e) = 0 \tag{11}$$

$$\text{var}(e) = \bar{h}^2 \left[v \frac{1 + e^{-v}}{1 - e^{-v}} - 2 \right] \triangleq \bar{h}^2 q, \tag{12}$$

where $v = s/\bar{h}$, $\bar{h} = \mu^{-1}$. For $s = 0$ (continuous scan), $\text{var}(e) = 0$ as expected, and hence these results cover both the continuous or the discrete scan case.

3.2 Asymptotic analysis of variance

Since p corresponds to the arrivals for a renewal process, $x \triangleq p/T$ is asymptotically normal with mean λ and variance of the form $O(1/T)$ (Ref. 9, p. 40). It is established in Ref. 10 that the variance can be approximately expressed in terms of the peakedness z

$$\text{var}(x) \cong (2z - 1)\lambda/T. \tag{13}$$

As noted in Ref. 4, this approximation has been found to be quite good for $a > z - 1$, and $T \geqq 10\bar{h}$. Although the carried calls c do not necessarily correspond to a renewal process (unless $c \equiv p$), c/T is also asymptotically normal with mean $\lambda(1 - B)$, (where $B \triangleq \lim_{T \to \infty}(o/p)$) and variance $O(1/T)$. This follows since if $B > 0$ the overflow process o is a renewal process, and the carried calls between overflows are independent for successive interoverflow periods. The only other asymptotic result needed is the following one, the proof of which is essentially the same as that for the function of sampling moments theorem given on p. 366 of Cramér:[11]

If $g(.,.)$ is a twice continuously differentiable function in some neighborhood of the point $\lambda,\lambda(1 - B)$, then $g(p/T,c/T)$ is asymptotically normal with mean $g(\lambda,\lambda(1 - B))$ and variance $O(1/T)$. It follows that

$$E(g(p/T,c/T)) = g(\lambda,\lambda(1 - B)) + O(1/\sqrt{T}). \tag{14}$$

Now for large T, the end effects r_0,r_T in (9) can be ignored at the outset. In particular, we have $E(Tu) = O(T)$, $\text{var}(Tu) = O(T)$, whereas $E(r_0 - r_T) = o(1)$, $\text{var}(r_0 - r_T) = O(1)$. (In general, ignoring these end effects is valid when $T/\bar{h}$ is reasonably large, e.g., $T/\bar{h} \geqq 10$.) Thus, defining* $y = \Sigma_{j=1}^{c} \hat{h}_j/c$, where the $\hat{h}_j$ satisfy the simplifying assumptions made for handling the scanning error, it follows from (10) to (12) that

* While y can be defined to be 0 for $c = 0$, in order to simplify subsequent notation, we shall assume that $P(c = 0) = 0$. This is reasonable even for the typical values of T that are of interest in practical applications.

$$E(y) = E(\hat{h}) = \bar{h} \tag{15}$$

$$\text{var}(y) = \text{var}(\hat{h})E\left(\frac{1}{c}\right) = \frac{\bar{h}^2(1+q)}{\lambda T(1-B)} + o(1/T), \tag{16}$$

where we have used (14) to evaluate $E(1/c)$.

Turning our attention next to $\hat{a}$, we have

$$\hat{a} = xy \tag{17}$$

$$\text{var}(\hat{a}) = E(x^2y^2) - E^2(xy). \tag{18}$$

In order to simplify this expression, we first note that

$$E(y|c) = \bar{h}$$

and hence

$$E(xy) = E_{p,c}E(xy|p,c) = E_{p,c}(x\bar{h}) = \lambda\bar{h} = E(x)E(y); \tag{19}$$

i.e., x,y are uncorrelated, confirming that $\hat{a}$ is an unbiased estimate of a. By the same conditioning, we also obtain

$$E(x^2y^2) = \bar{h}^2(E(x^2) + (1+q)E(x^2/c)) \tag{20}$$

and since

$$E(x^2)E(y^2) = \bar{h}^2(E(x^2) + (1+q)E(1/c)E(x^2)), \tag{21}$$

$$E(x^2y^2) = E(x^2)E(y^2) + (1+q)\bar{h}^2w, \tag{22}$$

where $w = \text{cov}(x^2,1/c)$. Substituting (19) and (22) into (18) and identifying terms, we have

$$\text{var}(\hat{a}) = E^2(x)\,\text{var}(y) + E^2(y)\,\text{var}(x) + \text{var}(x)\,\text{var}(y) + (1+q)\bar{h}^2w. \tag{23}$$

By direct substitution of the means and variances for x,y

$$\text{var}(\hat{a}) = \frac{a\bar{h}(1+q)}{T(1-B)} + (2z-1)\frac{a\bar{h}}{T} + o(1/T) + (1+q)\bar{h}^2w. \tag{24}$$

It remains to show that $w = o(1/T)$. But $Tw = \text{cov}(x^2,1/(c/T))$ and hence by (14) it follows that $Tw = o(1)$, i.e., $w = o(1/T)$. This completes the analysis; the variance approximation given in eq. (5) corresponds to terms of $O(1/T)$ in (24).

IV. CONCLUSIONS

In this paper, we have developed a simple approximation for the variance of the UPCO offered load estimate commonly used in offered load estimation. This approximation shows clearly the role of source load variation, switch count error, peakedness, congestion, and length of the measurement period. Relative to previous work, the main contribution is the explicit inclusion of congestion. Thus the results are of particular

interest for high congestion situations such as occur in measuring loads on high usage groups.

While the basic approximation is developed here for a single measurement interval, it can be easily applied in analyzing load estimates based on the average load over a number of single measurement intervals.

V. ACKNOWLEDGMENTS

Discussions with S. R. Neal and D. W. Hill and D. L. Jagerman are gratefully acknowledged.

APPENDIX

Analysis of Switch Count Error

In this appendix we analyze, using methods similar to Hayward,[3] the following switch count error model: (*i*) a call with holding time h begins at a time uniformly distributed between two successive sampling instants, (*ii*) the sampling interval is of length s, (*iii*) the holding time is exponentially distributed with rate parameter μ.

For an arbitrary call, the error e between the true holding time h for the call, and the "sampled holding time," is given by $e = ks - h$, where k represents the scan count for the call, $k \epsilon \{0,1,2,\cdots\}$. The scan count for the call is simply the total number of scans that occur during the time the call is in progress.

Since $e \epsilon [-s,s]$, it is convenient to define a normalized error $e' = k - h'$, where $h' = h/s$ is exponentially distributed with rate parameter $\mu' = \mu s = s/\bar{h}$. The density of h' is therefore given by

$$f(t) = \begin{cases} 0 & t < 0 \\ \mu' e^{-\mu' t} & t \geq 0 \end{cases}. \tag{25}$$

Define $x' = x/s$, where x is uniformly distributed in $[0,s]$ and represents the time from a sampling instant to the beginning of a call. Given $x' \epsilon [0,1]$, it is straightforward to show that the conditional probability density of e' at $e' = y$ is

$$g(y|x') = \begin{cases} 0, & y \notin [-(1-x'),x'] \\ \sum_{k=0}^{\infty} f(k-y), & y \epsilon [-(1-x'),x']. \end{cases} \tag{26}$$

The only case for which a negative argument can occur in any term in the preceding sum is for $k = 0, y > 0$. Thus,

$$\sum_{k=0}^{\infty} f(k-y) = \frac{e^{-\mu'}}{1-e^{-\mu'}} \mu' e^{\mu' y} \text{ for } y > 0$$

$$\sum_{k=0}^{\infty} f(k-y) = \frac{1}{1-e^{-\mu'}} \mu' e^{\mu' y} \text{ for } y < 0.$$

Defining $r = e^{-\mu'}$, (26) becomes

$$g(y|x') = \begin{cases} 0 & x' < y \le 1 \\ \dfrac{r}{1-r}\mu' e^{\mu' y} & 0 \le y \le x' \\ \dfrac{1}{1-r}\mu' e^{\mu' y} & -(1-x') \le y < 0 \\ 0 & -1 \le y < -(1-x'). \end{cases} \tag{27}$$

To simplify obtaining of moments for e', we define $G(\alpha) = E(e^{\alpha e'}) = E_{x'}E(e^{\alpha e'}|x')$. Using (27),

$$G(\alpha) = \frac{1}{1-r} E_{x'}\left[r \int_0^{x'} \mu' e^{(\mu'+\alpha)y} dy + \int_{-(1-x')}^{0} \mu' e^{(\mu'+\alpha)y} dy \right]. \tag{28}$$

After integration, one obtains

$$G(\alpha) = \left(\frac{\mu'}{\mu'+\alpha}\right) - \left(\frac{1+r}{1-r}\right)\frac{\mu'}{(\mu'+\alpha)^2} + \frac{(e^{\alpha} + re^{-\alpha})}{1-r}\frac{\mu'}{(\mu'+\alpha)^2}. \tag{29}$$

We have $G(0) = 1$, $G'(0) = 0$, and

$$G''(0) = \frac{1+r}{1-r}\frac{1}{\mu'} - 2\frac{1}{(\mu')^2}; \tag{30}$$

hence,

$$E(e) = 0 \tag{31}$$

$$\operatorname{var}(e) = \bar{h}^2\left(\frac{1+e^{-s/\bar{h}}}{1-e^{-s/h}}\cdot\frac{s}{\bar{h}} - 2\right), \tag{32}$$

which establishes (10) and (12) of the main section.

To establish the covariance between h, e, we note that because of (31), $\operatorname{cov}(h,e) = E(h\,e) = s^2E(h'e')$. But

$$\begin{aligned} E(h'e') &= E_{x'}\left[\int_{-(1-x')}^{x'} \sum_{k=0}^{\infty} y(k-y)f(k-y)dy \right] \\ &= E_{x'}\left[\int_{-(1-x')}^{x'} \sum_{k=0}^{\infty} (-y^2)f(k-y)dy \right] \\ &\quad + E_{x'}\left[\int_{-(1-x')}^{x'} \sum_{k=0}^{\infty} kyf(k-y)dy \right]. \end{aligned} \tag{33}$$

The first term is $-\operatorname{var}(e')$. To evaluate the second term, we note that

$$\sum_{k=0}^{\infty} kyf(k-y) = \sum_{k=0}^{\infty} ky\mu' e^{-\mu'(k-y)} = ye^{\mu' y}\mu' \sum_{k=0}^{\infty} kr^k$$

$$= ye^{\mu' y}\mu' r \sum_{k=1}^{\infty} kr^{k-1} = ye^{\mu' y}\mu' r \frac{d}{dr}\left(\frac{1}{1-r}\right), r = e^{-\mu'}.$$

Therefore

$$E_{x'}\left[\int_{-(1-x')}^{x'} \sum_{k=0}^{\infty} kyf(k-y)dy\right]$$

$$= \frac{r}{(1-r)^2} E_{x'}\left[\int_{-(1-x')}^{x'} y\mu' e^{\mu' y} dy\right]. \quad (34)$$

Thus, we are led to define the function

$$H(\alpha) = E_{x'}\left[\int_{-(1-x')}^{x'} \mu' e^{(\mu+\alpha)y} dy\right].$$

Carrying out the integration yields

$$H(\alpha) = -\frac{2\mu'}{(\mu'+\alpha)^2} + \frac{\mu'}{(\mu'+\alpha)^2}\frac{e^{\alpha} + r^2 e^{-\alpha}}{r}. \quad (35)$$

The expectation in (34) is now evaluated as

$$H'(0) = \frac{1}{\mu'}\frac{(1-r)(1+r)}{r} + \frac{1}{(\mu')^2}\left(4 - 2\frac{1+r^2}{r}\right),$$

giving

$$E_{x'}\left[\int_{-(1-x')}^{x'} \sum_{k=0}^{\infty} kyf(k-y)dy\right] = \frac{1}{\mu'}\left(\frac{1+r}{1-r}\right) - \frac{2}{(\mu')^2} = \text{var}(e').$$

Therefore, $E(h'e') = -\text{var}(e') + \text{var}(e') = 0$, i.e., h' and e' are uncorrelated random variables and

$$\text{var}(\hat{h}) = \text{var}(ks) = \text{var}(h) + \text{var}(e). \quad (36)$$

Remark: Hayward[3] treats switch count error and source load variation separately, assumes independence, and adds the separate variances to obtain an approximate result. He noted that the errors were probably correlated, though weakly, and that (at that time) no method to take this into account was evident (Ref. 3, p. 363). Since $\text{cov}(h,e) = 0$, it follows from this analysis that (for the same model studied by Hayward) the errors are in fact uncorrelated. It was also pointed out by the referee that an alternate proof that $\text{cov}(h,e) = 0$ can be obtained by noting that the scan count k is geometrically distributed for $k \geq 1$. Thus, by directly evaluating $\text{var}(ks)$, one finds that $\text{var}(ks) = \text{var}(h) + \text{var}(e)$, which implies $\text{cov}(h,e) = 0$.

REFERENCES

1. R. L. Franks, H. Heffes, J. M. Holtzman, and S. Horing, "A Model Relating Measurement and Forecast Errors to the Provision of Direct Final Trunk Groups," Proceedings of the 8th ITC (November 1976), pp. 133-1 to 133-7.
2. R. I. Wilkinson, "The Reliability of Holding Time Measurements," B.S.T.J., *20*, No. 4 (October 1941), pp. 365–404.
3. W. S. Hayward, Jr., "The Reliability of Telephone Traffic Load Measurements by Switch Counts," B.S.T.J., *31*, No. 2 (March 1952), pp. 357–377.
4. D. W. Hill and S. R. Neal, "Traffic Capacity of a Probability Engineered Group," B.S.T.J., *55*, No. 7 (September 1976), pp. 831–842.
5. R. I. Wilkinson, "Theories for Toll Traffic Engineering in the U.S.A.," *BSTJ*, *35*, No. 2 (March 1956), pp. 421–514.
6. J. M. Holtzman, "The Accuracy of the Equivalent Random Method with Renewal Inputs," B.S.T.J., *52*, No. 9 (November 1973), pp. 1673–1679.
7. W. S. Hayward, Jr., unpublished work.
8. S. R. Neal and A. Kuczura, "A Theory of Traffic Measurement Errors for Loss Systems with Renewal Input," B.S.T.J., *52*, No. 6 (July–August 1973), pp. 967–990.
9. D. R. Cox, *Renewal Theory*, London: Methuen, and New York: Wiley, 1962.
10. D. L. Jagerman, unpublished work.
11. Harald Cramér, *Mathematical Methods of Statistics*, Princeton: Princeton University Press, 1946.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

Adaptive Equalization of Channel Nonlinearities in QAM Data Transmission Systems

By D. D. FALCONER

(Manuscript received August 23, 1977)

Within the population of voiceband telephone channels, few channel characteristics are as pervasive in their impairment of high-speed data communication as nonlinear distortion, which cannot be removed or equalized in the receiver as easily as can linear distortion. The purpose of this paper is to report on an investigation of a QAM receiver incorporating adaptive equalization of nonlinearities as well as adaptive decision feedback equalization and data-aided carrier recovery for mitigation of linear distortion and phase jitter, respectively. Nonlinearities are equalized by adding to the received in-phase and quadrature signals a weighted sum of nonlinear functionals of the received signal and of modulated previous receiver decisions. The choice of nonlinear terms in the sum is based on a channel model incorporating quadratic and cubic nonlinearities as well as linear dispersive elements. The adjustment of the weighting, or tap, coefficients for the various terms is based on a gradient algorithm, as is the adjustment of the linear tap coefficients and the carrier phase reference. The feasibility of nonlinearity equalization on real voiceband channels was confirmed in a test in which recorded 9600-bps QAM signals, received from a worse-than-average set of 17 voiceband telephone channels, were processed by a computer-simulated version of the proposed receiver (termed the NL receiver). The observed error rates for all channels were lower, in some cases by several orders of magnitude, than those achieved by computer-simulated versions of the linear receiver and of a decision feedback equalization receiver (termed the DFE receiver).

I. INTRODUCTION

The prevalence of nonlinearities and their distorting effect on high-speed data transmission over voiceband telephone channels has long been recognized.[1] The effect of nonlinear distortion on linearly modulated data signals is to introduce nonlinear intersymbol interference and

reduce the margin against noise. For data rates above 4800 bps, nonlinear distortion is the dominant impairment on many voiceband telephone channels. Experimental studies have measured nonlinear distortion and related the observed error rates for specific modulation formats to this and other measured impairments.[2,3] Estimation of performance for data transmission in the presence of nonlinearities can be carried out[4] but gives little insight into the problem of receiver optimization, except for certain simple nonlinear channel models.[5]

Recognizing that nonlinearities in transmission channels generally coexist with linear elements such as filters, one is led to consider a general nonlinear receiver structure, based on a Volterra or Wiener kernel characterization[6] of a general nonlinear system such as that proposed in Refs. 5, 7, and 8, the latter in connection with adaptive echo cancellation. In the present work, we extend this approach by generalizing the structure of a passband decision feedback equalizer, previously studied in connection with linear channel distortion,[9] to process nonlinear as well as linear functionals of the incoming signal and prior decisions.*

The new receiver structure is based on a model of a passband channel with quadratic and cubic nonlinearities, as well as linear filters. We report on the simulation of the new receiver and on comparisons of its performance with two other previously simulated 9600-bps QAM receivers on a worse-than-average set of voiceband telephone channels. The new receiver is referred to as the NL receiver. The other two receivers, designated LE (linear equalization) and DFE (decision feedback equalization), are not designed to compensate for channel nonlinearities. Their performance is compared over the same set of voiceband telephone channels in Ref. 9. The simulated LE receiver is described in Ref. 10.

II. SUMMARY OF THE MAJOR RESULTS

The relative performances of the three simulated receivers on the same set of recorded, received, 9600-bps data signals are briefly summarized as follows: On every channel, the NL receiver yielded a lower error probability than the other two receivers. For 13 out of the 17 channels, the improvement in error rate was equal to or better than about an order of magnitude. Another gauge of the degree of improvement offered by the NL receiver is the fact that it increased the number of channels yielding a better-than-10^{-4} error rate from 8 to 15. On one channel, whose major impairment was second harmonic distortion, the NL re-

* Figure 3a summarizes the structure of the nonlinearity-equalizing receiver.

ceiver's error rate bested that of the DFE and LE receiver by over four orders of magnitude. Figure 5 is a bar graph summarizing the error rate comparisons.

The apparent attractiveness of the NL receiver structure is, however, tempered by its greater complexity. A large number of nonlinear tap coefficients is necessary to account and compensate for the dispersive nonlinear effects typically encountered on voiceband channels. In the simulations summarized above, the LE and DFE receivers each had 32 complex tap coefficients, but the NL receiver was, roughly speaking, comparable in complexity to an LE receiver with 134 complex tap coefficients. Reducing the number of coefficients in the NL receiver lowered its performance margin over the other receivers. Furthermore, the best allocation of a fixed number of tap coefficients varied from one channel to another. These points are explored more fully in later sections.

In spite of the greater complexity of the NL receiver structure, the performance comparison of the three receivers does indicate the importance of alleviating nonlinear distortion for high-speed data transmission.

III. THE CHANNEL MODEL

Obviously, the effect of channel nonlinearities on a passband QAM data signal must be understood before a compensating receiver structure can be suggested. A general representation of a bandlimited QAM signal is as the real part of a complex waveform:

$$x(t) = \mathrm{Re}\left[e^{j2\pi f_c t} \sum_n A(n)F(t - nT) \right], \tag{1}$$

where $j = \sqrt{-1}$, f_c is the carrier frequency, $A(n)$ is a quantized complex number representing the information symbol in the nth symbol interval (for example, in the case of four-level QAM, the real and imaginary parts of $A(n)$ assume one of the four possible values $\pm 1, \pm 3$), T is the reciprocal of the baud, and $F(t)$ is a complex pulse waveform.

In the case of QAM signals, extraction of the information symbols represented by the complex number $A(n)$ requires two receiver outputs, which are derived by appropriate operations on both the received passband signal and on its quadrature version, or Hilbert transform. A phase-splitting filter is used to obtain both in-phase and quadrature versions of a voiceband data signal.

The complex waveform

$$X(t) = e^{j2\pi f_c t} \sum_n A(n)F(t - nT) \tag{2}$$

is assumed analytic;[11] that is, its spectrum is twice the Fourier transform of $x(t)$ for positive frequencies and is zero elsewhere. Furthermore, we assume the spectrum is limited on the high side to frequency $2f_c$. Note that the Fourier transform $\mathcal{F}(f)$ of the complex pulse $F(t)$ is not necessarily symmetric about $f = 0$, but it is assumed to be strictly band-limited to $-f_c < f < f_c$. The Nyquist frequency is $1/2T$ Hz. Figure 1 shows a sketch of $\mathcal{F}(f)$ and of $\mathcal{F}(f - f_c)$, which is the Fourier transform of $e^{j2\pi f_c t}F(t)$.

The notion of analytic signals is a notational convenience. The Hilbert transform, or quadrature version of a signal $u(t)$, is a linear functional of $u(t)$:

$$\check{u}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{u(\tau)}{t-\tau}d\tau.$$

It can be shown that there is a unique analytic signal whose real part is $u(t)$, and that $\check{u}(t)$ is then just the imaginary part of the analytic signal. Conversely, any analytic signal comprises some real signal plus j times its Hilbert transform. Since QAM systems operate on both in-phase and quadrature versions of signals, they are most conveniently represented by means of analytic signals.

The nonlinear receiver structure will be based on the simple nonlinear channel model shown in Fig. 2, using the notation of analytic signals. Filters 1, 2, and 3 are passband with the same bandwidth as the transmitted data signal. The filters may include the receiver's input filter as well as the linear response of the channel. The quadratic and cubic memoryless nonlinearities with attenuated outputs account for second

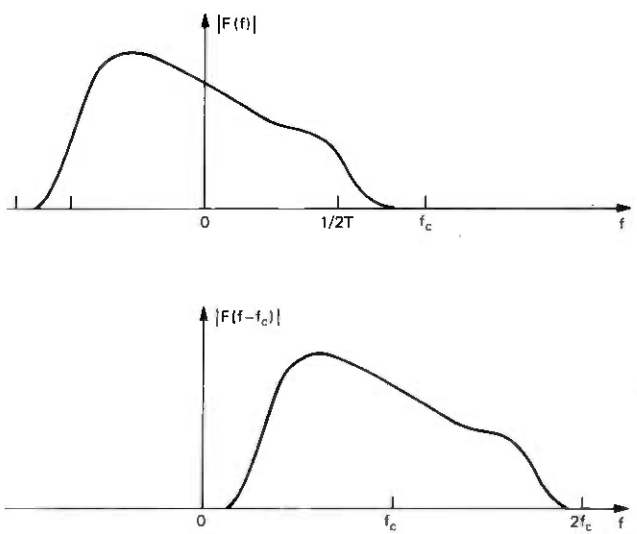


Fig. 1—Fourier transforms of $|F(f)|$ and $|F(f - c)|$.

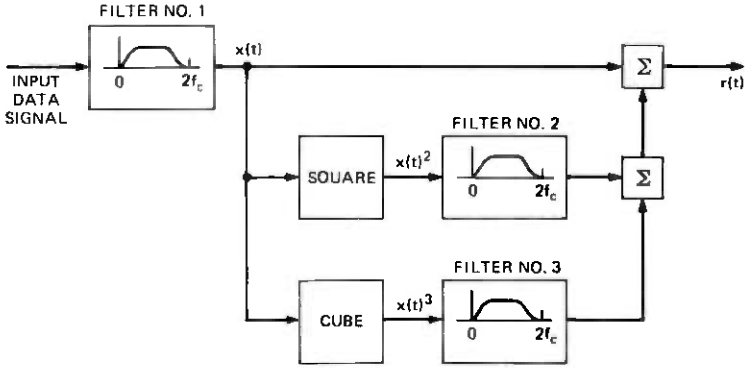


Fig. 2—Model of a nonlinear channel.

and third harmonic distortion, respectively. Additional impairments not shown in Fig. 2 are phase jitter, which implies multiplication of the complex received signal by $e^{j\phi(t)}$, and additive noise.

The result of passing the transmitted waveform through the linear portion of the channel (filter 1) is an analytic waveform in the form of eq. (2). A passband linear[10,12] equalizer (LE) can be used to minimize the mean squared error between its output, sampled at times nT, and a reference $A(n)e^{j(2\pi f_c nT+\hat{\theta}(n))}$, which is the complex information symbol modulated to passband with a receiver phase reference $\hat{\theta}(n)$. In a linear receiver, the passband equalizer output is demodulated [multiplied by $e^{-j(2\pi f_c nT+\hat{\theta}(n))}$] and then quantized to yield a decision $\hat{A}(n)$. A passband equalizer configuration which is theoretically more effective in combatting linear intersymbol interference is the passband DFE, described in Ref. 9.

To motivate a receiver structure which is appropriate for nonlinear distortion as well as linear distortion, we must consider the analytic signals emanating from the quadratic and cubic path elements of Fig. 2.

It is shown in the appendix that the analytic signal output from the model of Fig. 2 is of the form

$$R(t) = U_0(t) + e^{j2\pi f_c t}U_{11}(t) + e^{-j2\pi f_c t}U_{12}(t) + e^{j4\pi f_c t}U_2(t) + e^{j6\pi f_c t}U_3(t), \quad (3a)$$

where

$$U_0(t) = \sum_{n_1,n_2} A(n_1)A(n_2)^* G_0(t - n_1T, t - n_2T) \quad (3b)$$

$$U_{11}(t) = \sum_n A(n)F(t - nT) + \sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3)^* G_{11}(t - n_1T, t - n_2T, t - n_3T) \quad (3c)$$

$$U_{12}(t) = \sum_{n_1,n_2,n_3} A(n_1)^*A(n_2)^*A(n_3)G_{12}(t - n_1T,t - n_2T,t - n_3T) \tag{3d}$$

$$U_2(t) = \sum_{n_1,n_2} A(n_1)A(n_2)G_2(t - n_1T,t - n_2T) \tag{3e}$$

$$U_3(t) = \sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3)G_3(t - n_1T,t - n_2T,t - n_3T), \tag{3f}$$

where asterisks denote complex conjugates.

The various U terms are seen to be linear combinations of products of complex information symbols $A(n)$, $A(n_1)A(n_2)$, $A(n_1)A(n_2)A(n_3)^*$, etc. Each modulates a harmonic of the carrier wave. The term $e^{j2\pi f_c t}U_{11}(t)$ includes the linear response of the channel to the data signal and also a component resulting from cubic distortion. The terms $U_0(t)$ and $U_2(t)$ result from the quadratic nonlinearity and the terms $U_{12}(t)$ and $U_3(t)$ result from the cubic nonlinearity. Additional terms would, of course, result from the assumption of additional nonlinear elements in the model of Fig. 2. The generalization of expression (3) to an infinite power series would be a complex passband version of a Volterra expansion.

IV. THE NONLINEAR RECEIVER STRUCTURE

The receiver structure to be studied here includes the passband QAM decision feedback equalizer discussed in Ref. 9, plus nonlinear processing suggested by the set of eqs. (3). Let $Y(n)$ be the receiver's complex output at time $t = nT$. This output is quantized to form the decision $\hat{A}(n)$, which equals the original transmitted symbol $A(n)$ if no error occurred. Let the demodulator's phase reference at time nT be $\hat{\theta}(n)$. Let $\{W_k^{(1)}\}_{k=-N}^{N}$ and $\{B_k^{(1)}\}_{k=1}^{M}$ be the complex linear forward and feedback tap coefficients respectively, and let $\{R(n)\}$ be the complex receiver input, sampled at times nT. Then

$$Y(n) = e^{-j(2\pi f_c nT+\hat{\theta}(n))} \sum_{k=-N}^{N} W_k^{(1)*} R(n-k) - \sum_{k=1}^{M} B_k^{(1)*} \hat{A}(n-k) + Y_{NL}(n)e^{-j(2\pi f_c nT+\hat{\theta}(n))}, \tag{4a}$$

where $Y_{NL}(n)$ consists of nonlinear functions of $\{R(k)\}$ and $\{\hat{A}(k)\}_{k<n}$.

The linear part of eq. (4a) implies a demodulated linear combination of $2N + 1$ receiver input samples minus a linear combination of M previous decisions.

The nonlinear term $Y_{NL}(n)$ is heuristically suggested by expression (3) in the following way: (i) Assume that at time nT the previous receiver decisions $\hat{A}(k) = A(k)(k < n)$ and that they are available to form the

nonlinear feedback terms. (*ii*) In any terms of expression (3) involving decisions $\hat{A}(k)$ not yet made at time $n(k \geq n)$, replace $\hat{A}(k)e^{j2\pi f_c kT+\hat{\theta}(k)}$ by $R(k)$ to form the forward nonlinear terms. The resulting expression is

$$
\begin{aligned}
Y_{NL}(n) = & \sum_{k_1,k_2} W^{(0)*}_{k_1k_2} R(n-k_1)R(n-k_2)^* \\
& + \sum_{k_1,k_2,k_3} W^{(11)*}_{k_1,k_2,k_3} R(n-k_1)R(n-k_2)R(n-k_3)^* \\
& + \sum_{k_1,k_2,k_3} W^{(12)*}_{k_1,k_2,k_3} R(n-k_1)^*R(n-k_2)^*R(n-k_3) \\
& + \sum_{k_1,k_2} W^{(2)*}_{k_1,k_2} R(n-k_1)R(n-k_2) \\
& + \sum_{k_1,k_2,k_3} W^{(3)*}_{k_1,k_2,k_3} R(n-k_1)R(n-k_2)R(n-k_3) \\
& - e^{j\hat{\theta}(n)} \sum_{\substack{k_1,k_2 \\ \geq 1}} B^{(0)*}_{k_1,k_2} \hat{A}(n-k_1)\hat{A}(n-k_2)^* \\
& - e^{j(2\pi f_c nT+\hat{\theta}(n))} \sum_{\substack{k_1,k_2,k_3 \\ \geq 1}} B^{(11)*}_{k_1,k_2,k_3} \hat{A}(n-k_1)\hat{A}(n-k_2)\hat{A}(n-k_3)^* \\
& - e^{-j(2\pi f_c nT-\hat{\theta}(n))} \sum_{\substack{k_1,k_2,k_3 \\ \geq 1}} B^{(12)*}_{k_1,k_2,k_3} \hat{A}(n-k_1)^*\hat{A}(n-k_2)^*\hat{A}(n-k) \\
& - e^{j(4\pi f_c nT+\hat{\theta}(n))} \sum_{\substack{k_1,k_2 \\ \geq 1}} B^{(2)*}_{k_1,k_2} \hat{A}(n-k_1)\hat{A}(n-k_2) \\
& - e^{j(6\pi f_c nT+\hat{\theta}(n))} \sum_{\substack{k_1,k_2,k_3 \\ \geq 1}} B^{(3)*}_{k_1,k_2,k_3} \\
& \qquad \times \hat{A}(n-k_1)\hat{A}(n-k_2)\hat{A}(n-k_3). \quad \text{(4b)}
\end{aligned}
$$

The formidable-looking expression (4b) is a linear combination of products of receiver inputs and their complex conjugates, minus a linear combination of products of previous decisions and their complex conjugates, modulated by appropriate harmonics of the carrier.

Figures 3a and 3b are block diagrams of the NL receiver. The cross-hatched boxes in Figure 3a show the nonlinear processing that has been added to the basic decision feedback equalization structure described in an earlier paper.[9]

V. ADAPTATION OF RECEIVER PARAMETERS

As in the linear and decision feedback equalization receivers, the parameters $\{W\}$, $\{B\}$ and $\hat{\theta}$ are adjusted in an estimated gradient algorithm to minimize the average value of the squared error magnitude $|E(n)|^2$ defined by

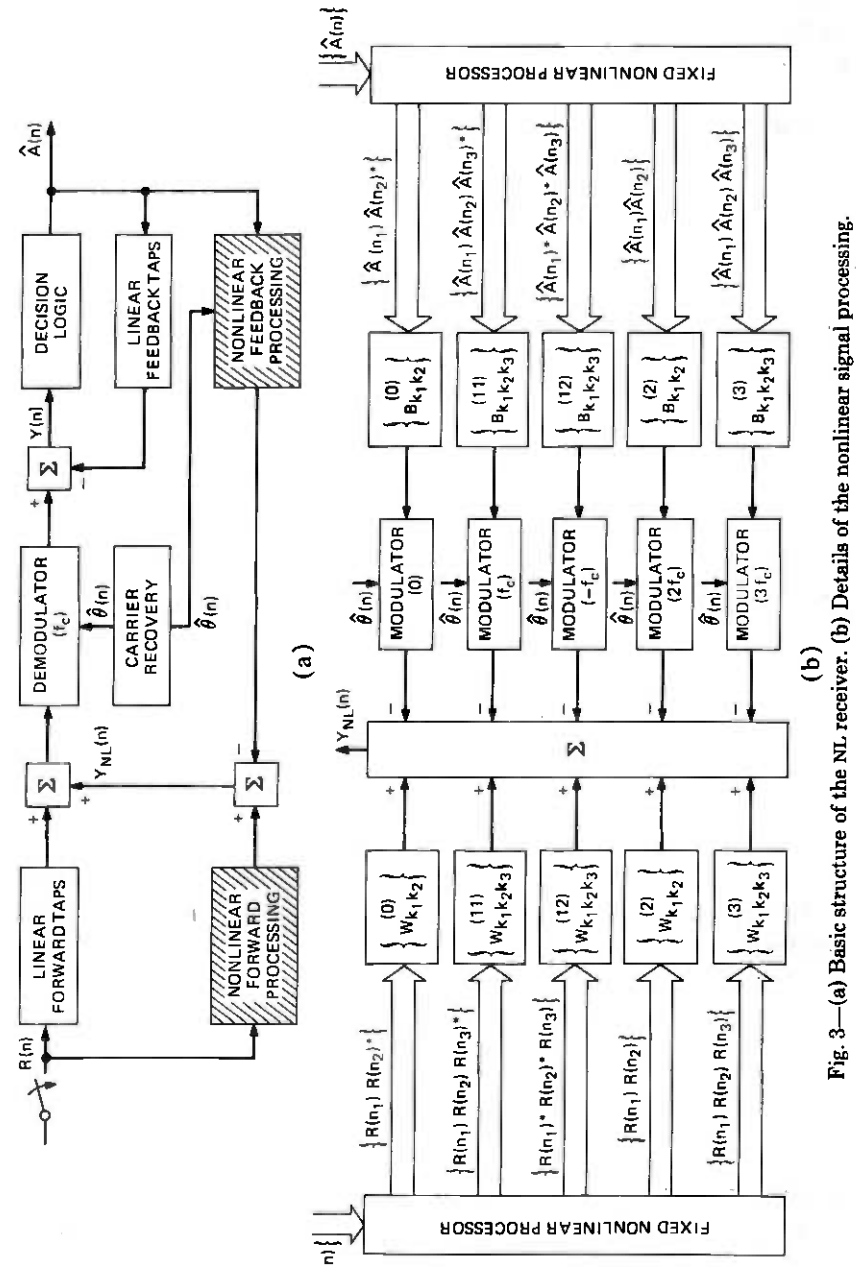


Fig. 3—(a) Basic structure of the NL receiver. (b) Details of the nonlinear signal processing.

$$E(n) \equiv Y(n) - A(n). \tag{5}$$

The error $E(n)$, as in the previous receivers, is a *linear* function of the parameters $\{W\}$ and $\{B\}$; consequently, the expression for $|E(n)|^2$ is convex in these parameters.

In writing the updating equations for the $\{W(n)\}$ and $\{B(n)\}$ coefficients and for $\hat{\theta}(n)$ in the nth symbol interval, it is convenient to use the symbol $\epsilon(n)$ to denote the *observed passband* error after the decision $A(n)$ has been made:

$$\epsilon(n) = [Y(n) - \hat{A}(n)]e^{j(2\pi f_c nT + \hat{\theta}(n))}; \tag{6a}$$

thus, if $\hat{A}(n) = A(n)$, $|E(n)|^2 = |\epsilon(n)|^2$, and the expression for the gradient of $|\epsilon(n)|^2$ with respect to each parameter determines an adjustment algorithm for that parameter. The adjustment equation for $\hat{\theta}(n)$ is as follows:

$$\hat{\theta}(n+1) = \hat{\theta}(n) - \frac{\alpha Im[\epsilon(n)^* Z(n)]}{|\hat{A}(n)|^2}, \tag{6b}$$

where
$$\begin{aligned} Z(n) = \sum_{k=-N}^{N} W_k^{(1)*} R(n-k) \\ + \sum_{k_1,k_2} W_{k_1,k_2}^{(0)*} R(n-k_1)R(n-k_2)^* \\ + \sum_{k_1,k_2,k_3} W_{k_1,k_2,k_3}^{(11)*} R(n-k_1)R(n-k_2)R(n-k_3)^* \\ + \sum_{k_1,k_2,k_3} W_{k_1,k_2,k_3}^{(12)*} R(n-k_1)^* R(n-k_2)^* R(n-k_3) \\ + \sum_{k_1,k_2} W_{k_1,k_2}^{(2)*} R(n-k_1)R(n-k_2) \\ + \sum_{k_1,k_2,k_3} W_{k_1,k_2,k_3}^{(3)*} R(n-k_1)R(n-k_2)R(n-k_3) \end{aligned} \tag{6c}$$

is the sum of all the forward terms comprising $Y(n)$. The adjustment equations for the $\{W\}$ and $\{B\}$ coefficients are as follows:

$$W_{k_1,k_2}^{(0)}(n+1) = W_{k_1,k_2}^{(0)}(n) - \beta_0 \epsilon(n)^* R(n-k_1)R(n-k_2)^* \tag{6d}$$

$$W_k^{(1)}(n+1) = W_k^{(1)}(n) - \beta_1 \epsilon(n)^* R(n-k) \tag{6e}$$

$$W_{k_1,k_2,k_2}^{(11)}(n+1) = W_{k_1,k_2,k_3}^{(11)}(n) - \beta_{11}\epsilon(n)^* R(n-k_1) \cdot R(n-k_2)R(n-k_3)^* \tag{6f}$$

$$W_{k_1,k_2,k_3}^{(12)}(n+1) = W_{k_1,k_2,k_3}^{(12)}(n) - \beta_{12}\epsilon(n)^* R(n-k_1)^* \cdot R(n-k_2)^* R(n-k_3) \tag{6g}$$

$$W_{k_1,k_2}^{(2)}(n+1) = W_{k_1,k_2}^{(2)}(n) - \beta_2 \epsilon(n)^* R(n-k_1)R(n-k_2) \tag{6h}$$

$$W_{k_1,k_2,k_3}^{(3)}(n+1) = W_{k_1,k_2,k_3}^{(3)}(n) - \beta_3 \epsilon(n)^* R(n-k_1) \cdot R(n-k_2)R(n-k_3) \tag{6i}$$

$$B^{(0)}_{k_1,k_2}(n+1) = B^{(0)}_{k_1,k_2}(n) + \gamma_0 \epsilon(n)^* \hat{A}(n-k_1)\hat{A}(n-k_2)^* e^{j\hat{\theta}(n)} \quad (6j)$$

$$B^{(1)}_{k}(n+1) = B^{(1)}_{k}(n) + \gamma_1 \epsilon(n)^* A(n-k) e^{j(2\pi f_c nT + \theta(n))} \quad (6k)$$

$$B^{(11)}_{k_1,k_2,k_3}(n+1) = B^{(11)}_{k_1,k_2,k_3}(n) + \gamma_{11} \epsilon(n)^* \hat{A}(n-k_1)\hat{A}(n-k_2)\hat{A}(n-k_3)^* e^{j(2\pi f_c nT + \hat{\theta}(n))} \quad (6l)$$

$$B^{(12)}_{k_1,k_2,k_3}(n+1) = B^{(12)}_{k_1,k_2,k_3}(n) + \gamma_{12} \epsilon(n)^* \hat{A}(n-k_1)^* \hat{A}(n-k_2)^* \hat{A}(n-k_3) e^{-j(2\pi f_c nT + \hat{\theta}(n))} \quad (6m)$$

$$B^{(2)}_{k_1,k_2}(n+1) = B^{(2)}_{k_1,k_2}(n) + \gamma_2 \epsilon(n)^* \hat{A}(n-k_1) \cdot \hat{A}(n-k_2) e^{j(4\pi f_c nT + \hat{\theta}(n))} \quad (6n)$$

$$B^{(3)}_{k_1,k_2,k_3}(n+1) = B^{(3)}_{k_1,k_2,k_3}(n) + \gamma_3 \epsilon(n)^* \hat{A}(n-k_1)\hat{A}(n-k_2)\hat{A}(n-k_3) e^{j(6\pi f_c nT + \hat{\theta}(n))}. \quad (6o)$$

The set of eqs. (4) through (6) defines the structure of the nonlinear QAM receiver that has been simulated. The α, β, and γ parameters are positive constants, chosen to ensure reasonably fast convergence and stability in the presence of noise. To enable compensation of rapidly varying phase jitter, the phase tracking constant α was set to the relatively large value of 0.4. The other constants chosen were:

$$\beta_1 = \gamma_1 = 0.001, \beta_0 = \beta_2 = \gamma_0 = \gamma_2 = 0.75 \times 10^{-5},$$
$$\beta_{11} = \beta_{12} = \beta_3 = \gamma_{11} = \gamma_{12} = \gamma_3 = 10^{-6}.$$

A judicious choice must be made for the range of coefficient indices k_1, k_2, and k_3 in the nonlinear terms making up $Y_{NL}(n)$, if the total number of $\{W\}$ and $\{B\}$ coefficients is to be reasonable, say on the order of 100. Obviously, the best choice of indices for a fixed number of taps depends on the channel. Trial and error (by no means exhaustive) of various sets of indices used in simulations on several voiceband channels led to the choice of terms shown in Table I. There are 73 "forward" tap coefficients $\{W\}$, of which 22 are linear, and 61 "feedback" tap coefficients $\{B\}$, of which 10 are linear. Note that the nonlinear forward tap indices are confined to the range $-1 \leq k \leq 1$ and the nonlinear feedback tap indices have been confined to the range $1 \leq k \leq 3$.

VI. THE SIMULATIONS

The nonlinear QAM receiver structure described in the previous section was simulated on an IBM 360 computer to process recorded 9600-bps QAM data signals that had been received from 17 voiceband telephone channels. The simulation effort was an extension of that described for linear and decision feedback QAM receivers in Refs. 3 and 9, respectively. The set of recorded QAM signals was the same, permitting the performance of all three receiver types to be compared under identical con-

Table I — Index terms used in voiceband simulations

Terms	Indices k_1	k_2	k_3	Terms	Indices k_1	k_2	k_3
$W^{(0)}_{k_1,k_2}$	−1	−1		$B^{(0)}_{k_1,k_2}$	1	1	
	0	0			2	2	
	1	1			3	3	
	−1	0			2	1	
	0	−1			1	2	
	0	1			3	2	
	1	0			2	3	
	−1	1			3	1	
	1	−1			1	3	
$W^{(1)}_{k_1}$ (Linear) terms −12 to 9 inclusive				$B^{(1)}_{k_1}$ (Linear) terms 1 to 10 inclusive			
$W^{(11)}_{k_1,k_2,k_3}$				$B^{(11)}_{k_1,k_2,k_3}$ and			
and $W^{(12)}_{k_1,k_2,k_3}$	−1	−1	−1	$B^{(12)}_{k_1,k_2,k_3}$	1	1	1
	0	0	0		2	2	2
	1	1	1		3	3	3
	−1	−1	0		1	1	2
	−1	0	−1		1	2	1
	0	0	−1		2	2	1
	0	−1	0		2	1	2
	0	0	1		2	2	3
	0	1	0		2	3	2
	1	1	0		3	3	2
	1	0	1		3	2	3
	−1	0	1		1	2	3
	0	1	−1		2	3	1
	−1	1	0		1	3	2
$W^{(2)}_{k_1,k_2}$	−1	−1		$B^{(2)}_{k_1,k_2}$	1	1	
	−1	0			2	2	
	0	0			3	3	
	1	1			1	2	
	0	1			2	3	
	−1	1			1	3	
$W^{(3)}_{k_1,k_2,k_3}$	−1	−1	−1	$B^{(3)}_{k_1,k_2,k_3}$	1	1	1
	0	0	0		2	2	2
	−1	−1	0		1	1	2
	0	0	−1		2	2	1
	1	1	1		3	3	3
	0	0	1		2	2	3
	1	1	0		3	3	2
	−1	0	1		1	2	3

ditions. The set of 17 channels could be described as "worse than average." Every channel had at least one impairment equal to or worse than the 90-percent point on the nationwide toll connection survey.[2]

The transmitted QAM signals had been generated digitally, with two pseudorandom four-level information symbol streams in quadrature, each repeating after 256 symbols. Each quadrature pair of symbols

therefore conveyed four information bits and the symbol rate was 2400 bauds, making a total bit rate of 9600 bps. The carrier frequency f_c was 1650 Hz, and the double-sideband baseband pulse signal had 12 percent roll-off.

The received signals that were recorded in digital form (12-bit samples, 24-kHz sampling rate) were received from a variety of real and analog-simulated voiceband telephone channels in tandem with an actual 50-km, C2-conditioned, N2-carrier voiceband channel.

As in the simulation of the linear and decision feedback receivers, the adaptive passband signal processors [defined by the set of eqs. (4) and Table I] were preceded by a pair of fixed digital filters that split the incoming signal into in-phase and quadrature components. Each was sampled at time instants $t = \tau + nT$ ($n = 0,1,2,\cdots$). Each simulation was actually of five separate receivers in parallel, with sampling epochs $\tau =$ 0, 0.2T, 0.4T, 0.6T, and 0.8T. The results reported in this paper are in each case for the timing epoch which yielded the best performance. As noted previously in Ref. 9, the decision feedback structure generally produced a relatively small performance spread between the best and the worst timing epochs. The receiver's decisions $\hat{A}(n)$ were formed by quantizing each equalized demodulated output, in-phase or quadrature, into one of the four possible levels $\pm 1, \pm 3$.

Before tabulating the simulation results, we mention some qualitative observations. In the interest of reducing the large numbers of nonlinear coefficients, it would have been desirable that only a few of the observed coefficients be large enough to be significant for all the channels. Unfortunately, this was not the case; no pattern was discernible common to all channels of a significant subset of coefficients; typically, the nonlinear component $Y_{NL}(n)$ in the receiver's output consisted of a large number of small terms, rather than a small number of relatively large terms plus insignificant terms.

Another qualitative observation was that the best values for the adaptation parameters for the nonlinear coefficients were so small that convergence of the nonlinear tap coefficients required at least 2000 symbol intervals, much slower than the convergence rate of the linear coefficients. This is attributed to the high correlation among many of the nonlinear terms. For example, the term $|A_{k_1}|^2 A_{k_2}$ is positively correlated with the linear term A_{k_2}, since $|A_{k_1}|^2$ takes only one of the three possible positive values 2, 10, or 18. Under such circumstances, the $\mathcal{A}$ matrix which describes the correlations among all the terms is expected to have a rather large eigenvalue spread, necessitating small adaptation constants and slow convergence.[13]

During each run, after an initial training period of 2000 symbol intervals to allow the coefficients to converge to nearly stationary values, the simulated receivers switched to a decision-directed mode in which

their decisions $\hat{A}(n)$, right or wrong, were used in the adaptation and decision feedback operations. Since the true transmitted information stream $\{A(n)\}$ was known, the performance was measured by observing the number of decision errors made during 7000 symbol intervals (or 28,000 bits). The empirical probability of the sampled analog error $Y(n) - A(n)$ was also measured, and if no errors were observed during a run, the error probability could be roughly estimated by extrapolating the tail of this distribution, using a computer subroutine by S. B. Weinstein.[14] The tabulated error probability, p_e, is the probability that a four-level symbol is in error; i.e., it is roughly twice the bit error rate. Another tabulated measure of performance was the *output SNR*, defined by

$$\text{output SNR} = \frac{\langle |A(n)|^2 \rangle}{\langle |E(n)|^2 \rangle}$$

where "$\langle \quad \rangle$" denotes the time average.

VII. QUANTITATIVE RESULTS

The simulation results for the NL receiver are tabulated in Table II along with the corresponding results taken from Ref. 9 for the LE and DFE receivers. For each channel, Table II lists the measured impairments and the error probabilities (either observed or extrapolated) for the LE, DFE, and NL receivers. The quantity in parentheses below each error probability is the output SNR in decibels. Error rates below 10^{-5} were extrapolated; in some cases in which the tail of the empirical probability distribution of the quadrature components of $E(n)$ was markedly non-Gaussian, the extrapolation yielded limited accuracy. Figure 4 illustrates the nonlinear compensation for channel 14, which had unusually severe second-harmonic distortion. Figure 4 is plotted on a "probability scale;" i.e., a Gaussian error distribution function would plot as a straight line on it. The distribution function for the linear receiver has distorted tails, indicating the presence of residual nonlinear distortion. However, the curve is nearly straight for the NL receiver, indicating that nonlinear distortion components have been substantially removed.

Comparison of error rates for the three receivers on all the channels is displayed more dramatically by the bar graph of Fig. 5. In all cases, the performance of the NL receiver surpassed that of the other two receivers. (Note that measurable nonlinear distortion was observed on all the channels.) In most cases, the NL receiver afforded a greater improvement in error rate over the DFE receiver than did the DFE receiver over the LE receiver. This is a very significant point. It indicates that if 9600-bps voiceband modems are to be improved by more sophisticated signal processing at the receiver, it is more fruitful to attempt to overcome nonlinear distortion than to concentrate on more sophisticated receiver structures, optimal for linear channel models.

Table II — Experimental comparison of LE, DFE and NL receivers
(facility in tandem with Holmdel-Murray Hill N2-carrier line)

		No.5 none	No. 6 Private N Carrier to White Plains	No. 7 Line Simulator	No. 8 Private T1 Carrier to Newark	No. 9 Line Simulator	No. 10 Line Simulator	No. 11 Line Simulator	No. 12 Line Simulator	No. 13 Line Simulator
Measured Impairments	Slope (dB)	0	2	9	4.4	−2	3	11*	11*	0
	Signal-to-noise ratio (dB)	29.0*	22.5†	31	27*	30	35	24.4*	33	34
	Second harmonic (dB)	33.5	28*	33	35	25†	32.1*	28.6*	33.8	34.4
	Third harmonic (dB)	44	31*	22.5†	40	33	47	36.4	49	30.7*
	Phase jitter (peak-to-peak)	<3°	<3°	<3°	<3°	14°† (120 Hz)	17°† (50 Hz)	<3°	<3°	<3°
Error rates (output SNR)	Linear equalization	1×10^{-8} (28.0 dB)	2×10^{-6} (22.4 dB)	6×10^{-3} (17.9 dB)	1×10^{-5} (20.8 dB)	7×10^{-3} (15.0 dB)	8×10^{-6} (23.9 dB)	1×10^{-4} (18.6 dB)	3×10^{-7} (22.6 dB)	3×10^{-6} (23.8 dB)
	Decision feedback equalization	2×10^{-9} (27.8 dB)	4×10^{-6} (22.8 dB)	3×10^{-3} (19.5 dB)	4×10^{-6} (21.2 dB)	9×10^{-3} (15.0 dB)	5×10^{-7} (23.8 dB)	7×10^{-5} (20.2 dB)	7×10^{-10} (24.6 dB)	3×10^{-8} (23.9 dB)
	Nonlinearity equalization	2×10^{-11} (29.4 dB)	1×10^{-6} (24.9 dB)	1×10^{-5} (23.7 dB)	7×10^{-7} (22.3 dB)	2×10^{-4} (18.2 dB)	3×10^{-8} (24.7 dB)	3×10^{-7} (21.7 dB)	3×10^{-12} (28.6 dB)	1×10^{-12} (27.5 dB)

Table II (cont)

		No. 14 Line Simulator	No. 15 Line Simulator	No. 16 Line Simulator	No. 17 Line Simulator	No. 18 DDD Loopback to Dallas	No. 19 (Private T1 Carrier to Newark	No. 20 Private T1 Carrier to Newark	No. 21 Private T1 Carrier to Newark
Measured Impairments	Slope (dB)	0	0	12†	11.1*	7.8	13†	6	8
	Signal-to-noise ratio (dB)	31	31	23†	29*	29*	28*	24.8*	23.2†
	Second harmonic (dB)	20.6†	27.2*	25.2†	32.2*	36.4	31.8*	24.4†	24.6†
	Third harmonic (dB)	49	32*	30.3†	34.7*	32.2*	37	32.6*	28.6†
	Phase jitter (peak-to-peak)	<3°	<3°	15°† (120 Hz)	10°† (120 Hz)	6°	<3°	<3°	<3°
Error Rates (output SNR)	Linear equalization	2×10^{-4} (19.4 dB)	1×10^{-6} (24.5 dB)	1.7×10^{-2} (14.0 dB)	3×10^{-3} (17.4 dB)	1×10^{-3} (18.5 dB)	2.1×10^{-3} (18.3 dB)	5×10^{-4} (18.4 dB)	1.6×10^{-3} (17.8 dB)
	Decision feedback equalization	8×10^{-5} (19.5 dB)	5×10^{-7} (24.5 dB)	3×10^{-2} (14.1 dB)	2×10^{-3} (18.6 dB)	9×10^{-4} (18.3 dB)	2×10^{-3} (19.9 dB)	3×10^{-4} (18.7 dB)	3×10^{-3} (18.0 dB)
	Nonlinearity equalization	2×10^{-9} (26.8 dB)	1×10^{-10} (28.4 dB)	1×10^{-2} (15.6 dB)	3×10^{-5} (20.8 dB)	5×10^{-5} (20.4 dB)	5×10^{-5} (21.6 dB)	3×10^{-5} (20.6 dB)	7×10^{-5} (19.6 dB)

* Indicates worse than 90-percent point in the nationwide toll connection survey.
† Indicates worse than "worst case" 3002 channel impairment limit.

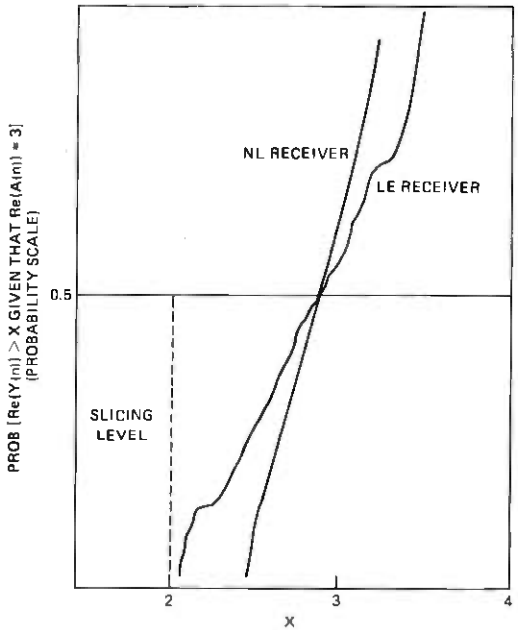


Fig. 4—Comparison of distribution functions of the receiver output $Y(n)$ for the linear and nonlinear receivers (data from channel 14).

Note in Fig. 5 that, for some of the channels, the nonlinearity equalization reduced the error rate by two or three orders of magnitude. However, on other channels, such as 9 and 16* which had most of their impairments in the "severe" category, the error rate was high and the NL receiver afforded very little improvement.

An interesting statistic that can be gleaned from Fig. 5 concerns the ability of the NL receiver to increase the number of channels which yield error rates below a specified maximum. For example, 15 of the 17 channels yield an error rate of better than 10^{-4} with the NL receiver, but only 8 of 17 meet this error rate standard with the LE receiver. For a maximum error rate of 10^{-5}, the number of channels is 10 with the NL receiver and 7 with the LE receiver. For a maximum error rate of 10^{-6}, the numbers of channels are 9 and 3 with the NL and LE receivers, respectively.

The price paid for the better performance of the NL receiver is, of course, its increased complexity, measured by the number of terms comprising $Y_{NL}(n)$ in eq. (4) and its slower convergence. The effect of reducing the number of terms, and therefore the complexity, is treated in the next section.

* Channel 16's impairments, produced by a line simulator, were all "worst case" values.

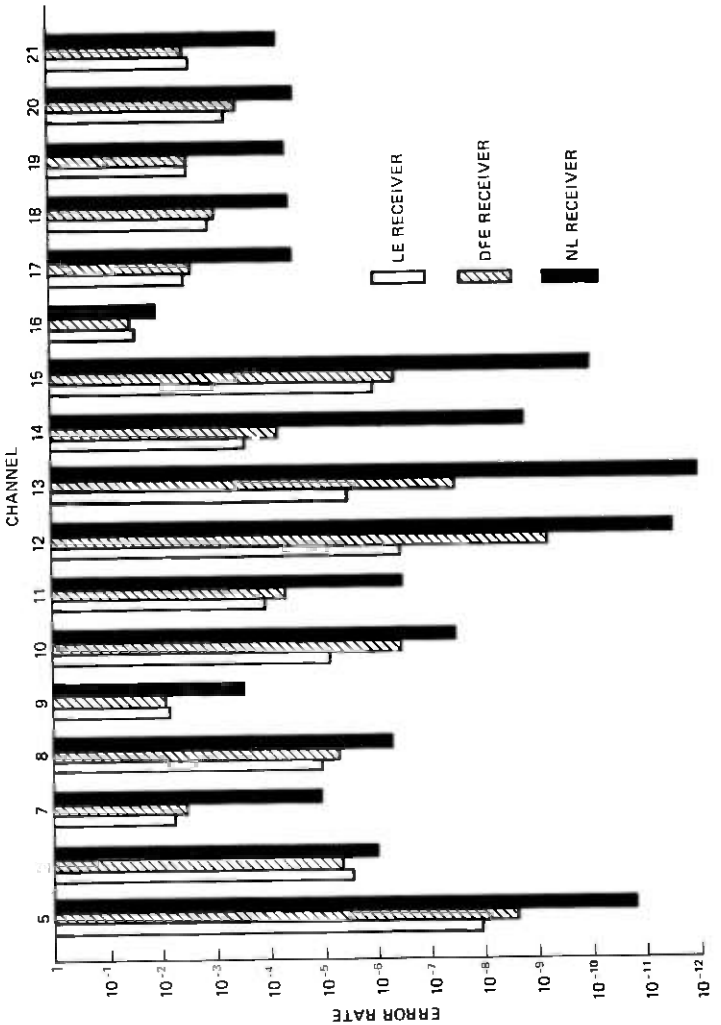


Fig. 5—Comparison of error rates for the three receivers.

VIII. MODIFICATIONS OF THE NONLINEAR RECEIVER STRUCTURE

8.1 Reductions of the number of nonlinear tap coefficients

(*i*) The tap coefficients $\{W^{(12)}_{k1,k2,k3}\}$ and $\{B^{(12)}_{k1,k2,k3}\}$ were set to zero, reducing the total number of nonlinear forward and feedback taps to 37 each. The measured output SNRs for most of the channels were slightly less than those for the full complement of 51 forward and 51 feedback taps, as illustrated in Table III.

(*ii*) A different set of 100 nonlinear terms was created by eliminating all cross-product terms and extending the time span covered by the forward and feedback terms to 10 symbol intervals. Thus, the forward tap coefficients consisted of $\{W^{(0)}_{k,k}\}$, $\{W^{(11)}_{k,k,k}\}$, $\{W^{(12)}_{k,k,k}\}$, $\{W^{(12)}_{k,k}\}$, and $\{W^{(13)}_{k,k,k}\}$, where $-5 \leq k \leq 4$, and the feedback terms consisted of $\{B^{(0)}_{k,k}\}$, $\{B^{(11)}_{k,k,k}\}$, $\{B^{(12)}_{k,k,k}\}$, $\{B^{(2)}_{k,k}\}$, and $\{B^{(3)}_{k,k,k}\}$, where $1 \leq k \leq 10$. Some resulting output SNRs are tabulated in part (*iii*) following.

(*iii*) A smaller set of nonlinear taps was created by taking a subset of 46 of the original set of 102 nonlinear taps. The resulting output SNRs for several channels are shown in Table IV, along with the corresponding set of SNRs from the original NL receiver structure with 102 nonlinear taps and also from the receiver with 100 nonlinear taps, described in item (*ii*), above.

The results of items (*i*), (*ii*), and (*iii*), compared with the original results using the NL receiver with 102 nonlinear tap coefficients indicate that a large number of nonlinear correction terms is necessary to yield substantial performance improvement. Undoubtedly, still better performance would have been attained by using more than 102 nonlinear taps. The results of item (*ii*) also showed that elimination of the cross-product terms degraded performance, even though the remaining nonlinear terms encompassed a longer time span.

Table III — Output SNR (dB) for nonlinear receivers

Channel	102 nonlinear taps	74 nonlinear taps
5	29.4	29.4
6	24.9	23.8
7	23.7	22.7
8	22.3	22.1
9	18.2	17.5
10	24.7	24.5
11	21.7	21.5
12	28.6	27.5

Table IV — Output SNR (dB) for nonlinear receivers

Channel	Original (102 Taps)	(*ii*) 100 Taps	(*iii*) 46 Taps
9	18.2	17.7	17.5
13	27.5	25.8	25.7
14	26.8	23.7	23.0
15	28.4	25.6	26.5

(*iv*) The number of nonlinear taps was also reduced to 46 by eliminating all coefficients $\{W^{(2)}_{k_1,k_2}\}$, $\{B^{(2)}_{k_1,k_2}\}$, $\{W^{(3)}_{k_1,k_2,k_3}\}$, and $\{B^{(3)}_{k_1,k_2,k_3}\}$. The resulting output SNR on channel 14 was only 21.3 dB, as compared to 24.8 dB for 102 nonlinear taps. Thus, it appears that at least the last four sets of coefficients (associated with second and third harmonics of the carrier frequency) are significant and should be retained.

8.2 A variation in the receiver structure tested for channel 20

The forward nonlinear tap coefficients weight various quadratic and cubic products of the sampled received signals. One might speculate that if *linear* distortion were removed from the received samples before their nonlinear processing, the nonlinear distortion remaining in the output might be further reduced. Accordingly, we simulated an NL receiver structure which was the same as that shown in Fig. 4 except that there are no linear feedback taps and the input to the forward nonlinear taps comes from the output of the linear forward taps instead of directly from the phase splitter. Since the adaptive linear forward taps, constituting the passband equilizer, are in tandem with the adaptive nonlinear taps in this structure, the mean squared error is not a convex function of the nonlinear tap coefficients, and hence the question of convergence is more complicated. Nevertheless, this structure was simulated on channel 20. The resulting output SNR was 20.0 dB compared to the 20.6 dB obtained from the original receiver structure. Thus, prior linear equalization did not appear preferable.

IX. CONCLUSIONS

The simulations have demonstrated that nonlinearity-equalizing QAM receivers can provide substantially better performance than can conventional linear or decision feedback equalization receivers over a variety of voiceband telephone channels. This encouraging result may stimulate further research aimed at finding less complicated receiver structures for overcoming channel nonlinearities.

The number of nonlinear terms that can be considered for inclusion in the NL receiver's analog output $Y(n)$ is potentially enormous. For example, the number of different terms $R(k_1)R(k_2)R(k_3)^*$ for all indices k_1, k_2 and k_3 between $-N$ and $+N$ is $(2N+1)^2(N+1)$, which is much more than $(2N+1)$, the corresponding number of linear terms $\{R(k)\}$ in that range of indices. The simulation results indicated that inclusion of a large number of nonlinear terms, including "cross-product" terms for which $k_1 \neq k_2 \neq k_3$, may be necessary. Reductions in the number of terms and a variation of the NL receiver's structure, in which adaptive linear processing preceded nonlinear processing, resulted in worsened performance.

Perhaps the major conclusion to be drawn concerns means for im-

proving the reliability of high-speed data transmission over the population of voiceband telephone channels. The simulations reported in Ref. 9 showed that decision feedback equalization, which is known theoretically to be superior to linear equalization in overcoming severe linear distortion, only moderately bettered the error rate obtained with linear equalization, especially on voiceband channels meeting C2 conditioning standards. However, the results summarized by Fig. 5 indicated that there is more to be gained by mitigating nonlinear distortion than in using more elaborate methods (beyond linear or decision feedback equalization) of mitigating linear distortion.

APPENDIX

In this appendix, we derive the form of the analytic signal that emerges from the summed filtered outputs of the quadratic and cubic nonlinearities. The real and imaginary parts of this analytic signal will then be the in-phase and quadrature components, respectively, of the nonlinearly distorted received QAM signal. The following theorems, proven in Ref. 11, will be required:

Theorem 1: Given real waveforms $u(t)$ and $v(t)$, defined on $-\infty < t < \infty$ with respective Hilbert transforms $\check{u}(t)$ and $\check{v}(t)$, the convolution

$$w(t) = \int_{-\infty}^{\infty} v(\tau)u(t-\tau)d\tau \tag{7}$$

has Hilbert transform

$$\check{w}(t) = \int_{-\infty}^{\infty} \check{v}(\tau)u(t-\tau)d\tau = \int_{-\infty}^{\infty} v(\tau)\check{u}(t-\tau)d\tau. \tag{8}$$

Thus, if $v(t)$ is the input to a filter whose impulse response is $u(t)$, the analytic output signal is

$$\begin{aligned} w(t) + j\check{w}(t) &= \int_{-\infty}^{\infty} (v(\tau) + j\check{v}(\tau))u(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} v(\tau)(u(t-\tau) + j\check{u}(t-\tau))d\tau. \end{aligned} \tag{9}$$

Theorem 2: The analytic signal resulting from the convolution can also be expressed as

$$w(t) + j\check{w}(t) = \frac{1}{2}\int_{-\infty}^{\infty} (v(\tau) + j\check{v}(\tau))(u(t-\tau) + j\check{u}(t-\tau))d\tau. \tag{10}$$

Now we consider an analytic signal of the form

$$X(t) = e^{j2\pi f_c t}\sum_n A(n)F(t-nT), \tag{11}$$

as in expression (2) of the text. The squaring and cubing elements in Fig. 2 operate on $x(t)$, the real part of $X(t)$. The response of the squaring element to $\mathrm{Re}(X(t))$ can be written

$$x(t)^2 = \frac{1}{2}\mathrm{Re}\left[e^{j4\pi f_c t}\sum_{n_1,n_2} A(n_1)A(n_2)F(t-n_1T)F(t-n_2T)\right] + \frac{1}{2}\left[\sum_{n_1,n_2} A(n_1)A(n_2)^*F(t-n_1T)F(t-n_2T)^*\right]. \quad (12)$$

Of the complex expressions in square brackets in (12), the first is complex and analytic, since it is the square of an analytic signal (its spectrum is nonzero only for positive frequencies). Thus, from Theorem 2, the analytic signal that results from passing the first part of expression (12) through a passband filter 2 is of the form

$$U_2(t) = e^{j4\pi f_c t}\sum_{n_1}\sum_{n_2} A(n_1)A(n_2)G_2(t-n_1T,t-n_2T), \quad (13)$$

where $G_2(t-n_1T,t-n_2T)e^{j4\pi f_c t}$ is a complex analytic waveform, whose spectrum has been limited by filter 2 to $0 < f < 2f_c$. The second term in (12) is baseband, real, and not analytic.† However, from Theorem 1, the analytic signal resulting from passing the second term through filter 2 has the form

$$U_0(t) = \sum_{n_1,n_2} A(n_1)A(n_2)^*G_0(t-n_1T,t-n_2T), \quad (14)$$

where $G_0(t-n_1T,t-n_2T)$ is an analytic waveform, whose spectrum is confined to $0 < f < 2f_c$.

The cubic nonlinear terms are handled similarly. The cube of the input signal $\mathrm{Re}(X(t))$ can be written

$$x(t)^3 = \frac{1}{4}\mathrm{Re}\left[e^{j6\pi f_c t}\sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3)F(t-n_1T)\cdot F(t-n_2T)F(t-n_3T)\right] + \frac{3}{8}e^{j2\pi f_c t}\sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3)^*F(t-n_1T)F(t-n_2T)\cdot F(t-n_3T)^* + \frac{3}{8}e^{-j2\pi f_c t}\sum_{n_1,n_2,n_3} A(n_1)^*A(n_2)^*A(n_3)\cdot F(t-n_1T)^*F(t-n_2T)^*F(n-n_3T). \quad (15)$$

The first term in square brackets (15) is analytic, being the cube of an analytic signal. The other two terms in (15) are not analytic, since their

† The ranges of the indices n_1 and n_2 in (12) and (13) are assumed to be the same.

Fourier transforms are not necessarily zero for negative frequencies. The analytic signal resulting from passing $x(t)^3$ through bandpass filter 3 can be written by applying Theorem 2 to the first term of (15) and Theorem 1 to the second and third terms. The resulting analytic signal is the sum of three analytic signals, $U_3(t)$, $U_{11}(t)$, and $U_{12}(t)$, which have the following forms:

$$U_3(t) = e^{j6\pi f_c t} \sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3) \cdot G_3(t - n_1T, t - n_2T, t - n_3T). \quad (16)$$

$$U_{11}(t) = e^{j2\pi f_c t} \sum_{n_1,n_2,n_3} A(n_1)A(n_2)A(n_3)^* \cdot G_{11}(t - n_1T, t - n_2T, t - n_3T). \quad (17)$$

$$U_{12}(t) = e^{-j2\pi f_c t} \sum_{n_1,n_2,n_3} A(n_1)^*A(n_2)^*A(n_3) \cdot G_{12}(t - n_1T, t - n_2T, t - n_3T). \quad (18)$$

The $G(\ )$ signals are complex, and the spectra of the analytic signals $U_3(t)$, $U_{11}(t)$, and $U_{12}(t)$ are all confined to the range $0 < f < 2f_c$ by bandpass filter 3.

REFERENCES

1. R. W. Lucky, "Modulation and Detection for Data Transmission on the Telephone Channel," in *New Directions in Signal Processing in Communication and Control*, ed. by J. K. Skwirzynski, Leyden: Noordhoff, 1975, pp. 321–327.
2. F. P. Duffy and T. W. Thatcher, Jr., "Analog Transmission Performance on the Switched Telecommunications Network," B.S.T.J., *50*, No. 4 (April 1971), pp. 1311–1347.
3. R. R. Anderson and D. D. Falconer, "Modem Evaluation on Real Channels Using Computer Simulation," Proc. National Telecomm. Conf., Dec. 1974, San Diego, pp. 877–883.
4. S. Benedetto, E. Biglieri, and R. Daffara, "Performance of Multilevel Baseband Digital Systems in a Nonlinear Environment," IEEE Trans. Comm., *COM-24*, No. 10 (October 1976), pp. 1166–1175.
5. W. J. Lawless and M. Schwartz, "Binary Signaling over Channels Containing Quadratic Nonlinearities," IEEE Trans. Comm., *COM-22*, No. 2 (March 1974), pp. 288–298.
6. N. Wiener, *Nonlinear Problems in Random Theory*, Cambridge, Mass.: M.I.T. Press and J. Wiley, 1958.
7. T. Arbuckle, "Nonlinear Equalization System Including Self- and Cross-Multiplication of Sampled Signals," U.S. Patent 3,600,681, Aug. 17, 1971.
8. E. J. Thomas, "Some Considerations on the Application of the Volterra Representation of Nonlinear Networks to Adaptive Echo Cancellers," B.S.T.J., *50*, No. 8 (October 1971), pp. 2797–2805.
9. D. D. Falconer, "Application of Passband Decision Feedback Equalization in Two-Dimensional Data Communications Systems," IEEE Trans. Comm., *COM-24*, No. 10 (October 1976), pp. 1159–1166.
10. D. D. Falconer, "Jointly Adaptive Equalization and Carrier Recovery in Two Dimensional Digital Communication Systems," B.S.T.J., *55*, No. 3 (March 1976), pp. 317–334.
11. J. Dungundji, "Envelopes and Pre-Envelopes of Real Waveforms," IRE Trans. on Info. Theory, Sept. 1957, pp. 53–57.
12. R. D. Gitlin, E. Y. Ho, and J. E. Mazo, "Passband Equalization for Differentially Phase Modulated Data Signals," B.S.T.J., *52*, No. 2 (February 1973), pp. 219–238.

13. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission," B.S.T.J., *48*, No. 1 (January 1969), pp. 55–70.
14. S. B. Weinstein, "Theory and Applications of Some Classical and Generalized Asymptotic Distributions of Extreme Values," IEEE Trans. Inform. Theory, *IT-19*, No. 2 (March 1973), pp. 148–154.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

Spectral Sharing in Hybrid Spot and Area Coverage Satellite Systems via Channel Coding Techniques

By A. S. ACAMPORA

(Manuscript received December 1, 1977)

Multiple spot-beam switching satellites employing frequency reuse are considered, and a method for incorporating an area coverage beam to provide service to those regions not covered by the footprint of any spot beam is proposed here. The method consists of employing a convolutional code for the area beam transmission to enable sharing of a common spectral band among the spot and area beams on a noninterfering basis and with no sacrifice in the capacity of the spot beams. A maximum-likelihood algorithm for this purpose is derived, and bounds on the bit error rate performance of all beams are found. Results show that excessive performance degradation arising from cochannel interference is limited to a thin annular ring surrounding each spot beam.

I. INTRODUCTION

Multiple spot beam communication satellites offer the potential for greatly increasing the traffic handling capability relative to wide-area coverage systems, since the allocated spectral band can be reused in the various spot beams.[1,2] A high-level block diagram of the satellite transponders for such a system might appear as shown in Fig. 1. Here, the various service regions are interconnected via an on-board switching matrix operating in the time-division mode, and digital modulation techniques consistent with time-division multiple access (TDMA) are employed.

As previously noted,[3] such a system suffers a serious drawback in that a large blackout region, serviceable by none of the spot beams, is created. The situation is depicted in Fig. 2, which shows the radiation footprints of a hypothetical 11-beam private line system serving the large population regions in the United States. Although most of the traffic load for

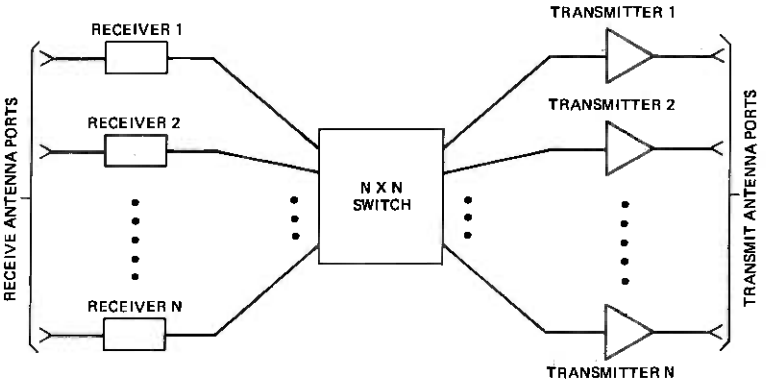


Fig. 1—Satellite transponder.

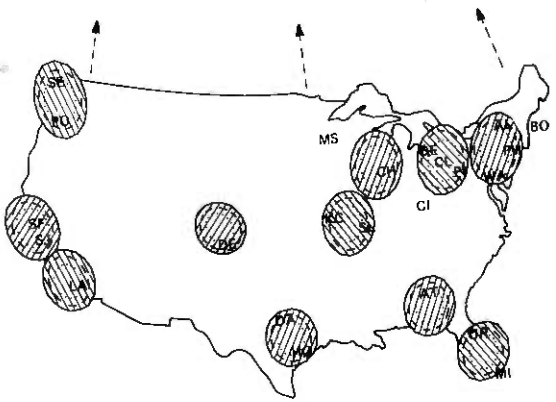


Fig. 2—Footprints of a hypothetical 11-beam system showing −1, −2, and −3 dB contours. Both polarizations are employed.

such an offering would be adequately served by the 11 high-capacity spot beams, it is nonetheless desirable to provide service to the outlying areas.

Among the various techniques proposed in Ref. 3 for coping with this blackout problem, the method of deploying a channel coded area coverage beam, in addition to the various uncoded spot beams, appears most attractive in that the blackout region is reduced to a thin annular ring surrounding each spot beam. This method offers the additional advantage of reducing the required radiated power for the area coverage beam, an important consideration since the gain of the area beam antenna port might be 20 dB lower than that of a spot beam port. In this paper, we review the principles involved in this approach and derive bounds on the resulting bit error rate performance of both the spot and area beams.

In Section II, we discuss the problems associated with sharing a

common spectral band between area coverage and spot coverage satellite beams. Section III is devoted to the derivation of a detection algorithm for such a hybrid system in which convolutional coding is used to alleviate the effects of cochannel interference, and bit error rate bounds are found. In Section IV, these results are applied in a typical communication satellite scenario.

II. PROBLEM DEFINITION

Consider a satellite system consisting of M spot-beam transponders serving M geographically separated, high-traffic demand areas on a noninterfering basis. The allocated spectral band is totally reused in the M spot beams. We wish to deploy an area coverage beam, in addition to the M spot beams, to provide service to the low traffic demand outlying regions serviced by none of the spot beams. The total traffic demand to all outlying regions might be of the same order of magnitude as the demand for one spot beam. Service to the outlying regions must be provided on a noninterfering basis and with no sacrifice to the capacity of the various spot beams. We assume that the spot beams require use of both electromagnetic polarizations to minimize mutual interference among themselves.

Four types of interference are readily identified:

(*i*) *Down-link:* The area coverage radiation is detectable at every spot beam receiving terminal and can thereby interfere with reception of the desired signal at those ground stations.

(*ii*) *Down-link:* The spot beam footprints might typically be useful out to their −3 dB radiation contours. Area-coverage receiving terminals located at the −3 dB through the −20 dB contours of any spot beam thereby suffers interference from that spot beam.

(*iii*) *Up-link:* All up-link transmissions from spot-beam earth terminals are detectable at the antenna port of the area coverage beam and thereby interfere with reception of the area coverage up-link transmission.

(*iv*) *Up-link:* Transmission from an area coverage ground station located between the −3 dB and the −20 dB contour of a spot-beam antenna pattern could interfere with that spot beam's up-link transmission.

Thus, the inclusion of an area beam might make the original spot beams totally unusable. To eliminate these interference problems, one might split the allocated spectral band into two components; one segment would be dedicated to the area coverage transponder and the second segment would be reused among the various spot beams. If this is done, the system designer must choose one of two options:

(*i*) Reduce the throughput of the spot beam transponders by that fraction of the satellite band dedicated to the area coverage beam.

(*ii*) Maintain the original throughput of the spot beam transponders while increasing the effective radiated power on both the up-link and the down-link to overcome the degradation caused by excessive band-limiting.

Option (*i*) results in a sizable decrease in the overall capacity of the satellite. Consider a 10-spot-beam system with each beam occupying the entire spectral band. The normalized throughput of such a system is defined to be 10 units. Suppose that two area coverage beams are added to the system (one using each polarization) and that one-half the band is reserved for the spot beams. Then, under option (*i*), the normalized throughput is reduced to $\frac{1}{2} \times 10 + \frac{1}{2} \times 2 = 6$, and the overall system throughput is reduced by 40 percent. For the same fractional split of the total bandwidth, option (*ii*) could incur a power penalty in excess of 6 dB for a 4ϕ-CPSK (coherent phase shift key) system originally operating at a modest BT (bandwidth-time) product of 1.3. Such a penalty might be acceptable on the up-link, but would typically be unacceptable on the down-link since space platform power is a limited resource.

Thus, to provide service to the outlying area at no sacrifice in either the throughput of the spot beams or in the required spot beam effective isotopic radiated power (e.i.r.p.), one might consider splitting the band, as described above, to eliminate up-link interference. Up-link digits would be regenerated, switched, and reformatted into the appropriate down-link port. A suitable scheme must then be sought to accommodate the down-link.

Channel coding techniques will be investigated as a possibility. The motivation for such an approach is twofold. First, and most important, coding can provide for effective immunity against cochannel interference. Second, we note that because of the difference of about 20 dB between the antenna gains of the area and spot-beam coverage antenna ports, a system would require 20 dB more power for the area coverage port than for a spot-beam port to achieve the same bit error rate performance. Through use of coding, we can effect a considerable reduction in the required power for the global beam.

The scenario envisioned, shown in Fig. 3, would employ uncoded transmission for the spot-beam messages and rate $r = \frac{1}{2}$ convolutionally encoded transmission for the area beam port. The throughput of the area beam port would be one-half that of the spot-beam port, implying that the down-link channel symbol rate for all beams are the same. In addition, since on-board regeneration is employed, all down-link channel symbols can be time-aligned. We will consider 4ϕ-CPSK modulation and explore in detail the situation where the in-phase and quadrature rails of both the area and spot coverage beams are modulated separately and where there is no crossrail coupling. Thus, we can consider baseband performance. The algorithms and other results to be presented are

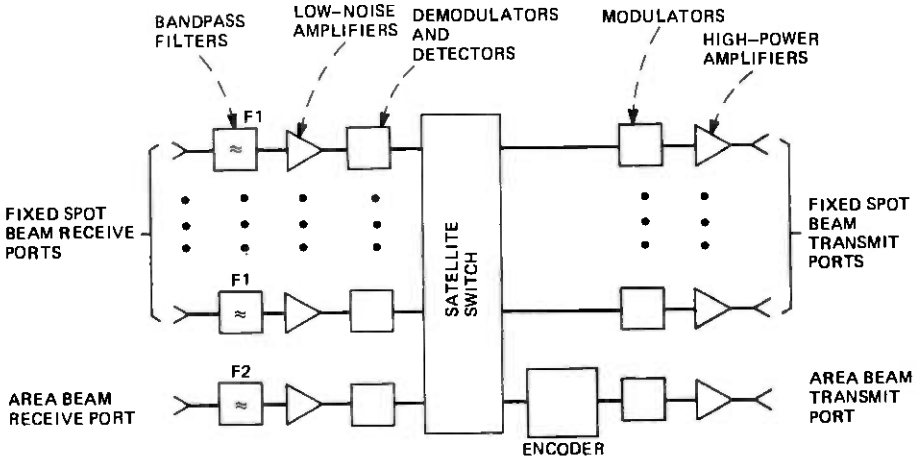


Fig. 3—Regenerative transponder for a hybrid spot-area coverage satellite employing channel coding. The allocated band is split between the spot and area beams on the uplink. The band is totally reused for all the downlinks.

readily generalized to the situation where there is a known, fixed carrier phase shift between the spot and area transmissions.

III. BIT ERROR RATE PERFORMANCE

We now investigate the bit error rate performance of both the uncoded spot beam message and the encoded global beam message in the interference-prone region surrounding one of the spot beams. We need to consider the presence of only one such spot beam since, in the footprint area of that beam, interference from the remaining beams is negligible. At a particular ground station, after coherent demodulation, we observe the following received baseband process:

$$R(t) = \sqrt{E_1} \sum_k b_k h(t - kT) + \sqrt{E_2} \sum_k y_k(\mathbf{a}) h(t - kT) + n(t). \quad (1)$$

In (1) above, b_k is the kth member of the binary data stream $\mathbf{b}$ of the uncoded spot beam message, $\mathbf{a}$ represents the binary data stream for the global beam, $y_k(\mathbf{a})$ is the kth channel symbol of the global beam and is dependent upon $\mathbf{a}$ through the structure of the encoder, $h(t)$ is the impulse response of the channel, $n(t)$ is a Gaussian noise process of spectral power density $N_0/2$, and E_1 and E_2 are, respectively, the received pulse energy of the spot and global beam transmissions. We note that the b_k's are independent and equally likely to be ± 1, and that the y_k's can assume the values ± 1 but are not independent. We assume that intersymbol interference is absent.

A set of sufficient statistics[4] for detecting the $\mathbf{a}$ and $\mathbf{b}$ sequences is formed by the synchronous samples of $R(t)$ taken at the opening of the binary eye. One such sample is:

$$r_k = \sqrt{E_1}\, b_k + \sqrt{E_2}\, y_k + n_k. \tag{2}$$

We assume the various n_k's to be independent. From the samples (2), we form the log-likelihood function or path metric[4]

$$\Lambda(\mathbf{a},\mathbf{b}) = 2\sum r_k[\sqrt{E_1}b_k + \sqrt{E_2}y_k(\mathbf{a})] - \sum[\sqrt{E_1}b_k + \sqrt{E_2}y_k(\mathbf{a})]^2 \tag{3}$$

and decide upon those sequences $\hat{\mathbf{a}},\hat{\mathbf{b}}$ for which (3) is maximized.

The maximum-likelihood algorithm to perform optimum detection is similar to the Viterbi algorithm[5] and is illustrated by the state transition diagram of Fig. 4, drawn for a $K = 3$ convolutional code. The state is defined by the contents of the first two stages of the shift register, and knowledge of the starting state and the next bit entering the encoder uniquely determines the next state and the encoded channel symbols generated. We note that, unlike the ordinary Viterbi algorithm for rate $r = \frac{1}{2}$ codes, each transition between states can occur along four paths, rather than one, because two independent uncoded symbols are also

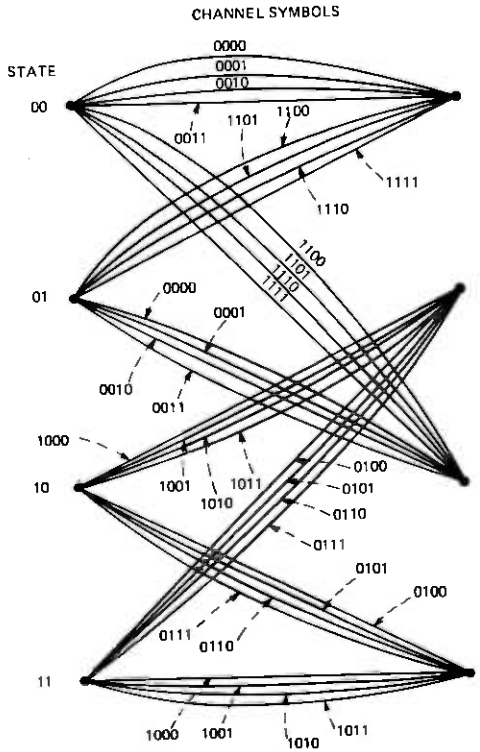


Fig. 4—State diagram for maximum-likelihood detection of interfering coded and uncoded signals. A $K = 3, r = \frac{1}{2}$ convolutional code is assumed. For each transition, the first two channel digits correspond to the coded symbols, and the second two correspond to the uncoded symbols.

generated during each epoch. The first two bits appearing along each branch correspond to the encoded channel symbols for that transition, and the second two digits correspond to one of the four possible two-bit sequences for the uncoded transmissions.

To perform maximum-likelihood detection, we note that, at each state, eight possible branches merge, and the partial path metric of one such merging branch must be the largest. The remaining seven paths then cannot be most likely because any succeeding additions to any one of these seven paths are valid additions to that one path exhibiting the greatest partial metric; succeeding additions, then, cannot cause the overall metric of any of these seven paths to exceed that of the path exhibiting greatest partial metric, and the seven paths having the smaller path metrics can be deleted from further consideration.

Thus, at each point in time, the four most likely paths (one leading to each state) and their associated partial metrics are known. During the next clock cycle, we determine the most likely of eight paths leading into each state by performing, for each of two initial states and for each of four branches for each initial state, the operation

$$\Lambda_n = \Lambda_{n-1} + \sum_{k=0}^{1} [2r_{2n-k} - \sqrt{E_1}b_{2n-k} - \sqrt{E_2}\,y_{2n-k}][\sqrt{E_1}b_{2n-k} + \sqrt{E_2}y_{2n-k}] \quad (4)$$

and saving the path and path metric of the largest for subsequent operations. The values of b_{2n-k} and y_{2n-k}, $k = 0,1$ to be substituted into (4) are determined from the state transition diagram, Fig. 4.

To perform true maximum likelihood detection, the most likely path leading into each state must be stored over the entire past. However, it has been shown that after 4 to 5 constraint lengths have elapsed, the oldest bits in all path memories are the same with a very high probability. Thus, we need to save only the most recent 4K through 5K of data for each state and, once in each epoch, the oldest bits in any one of four path memories can be outputted as detected data. We note that, unlike the ordinary Viterbi algorithm for which each path memory consists of a single rail of data, here we need to store three rails of data for each state. One rail contains the most likely source sequence **a** for the area coverage beam, and the second two contain the first and second source bits emitted each epoch for the uncoded spot beam sequence **b**.

The detector will commit an error for the first time at node n if the partial metric of some path which previously diverged from the correct path and remerges at node n is greater than that of the correct path. Some possible error events are shown in Fig. 5. We now calculate the probability of such an event.

Let **A** correspond to the spot and area coverage information sequence along the correct span, and let $\hat{\mathbf{A}}$ be those along the incorrect span. Then, the path metric difference is given by:

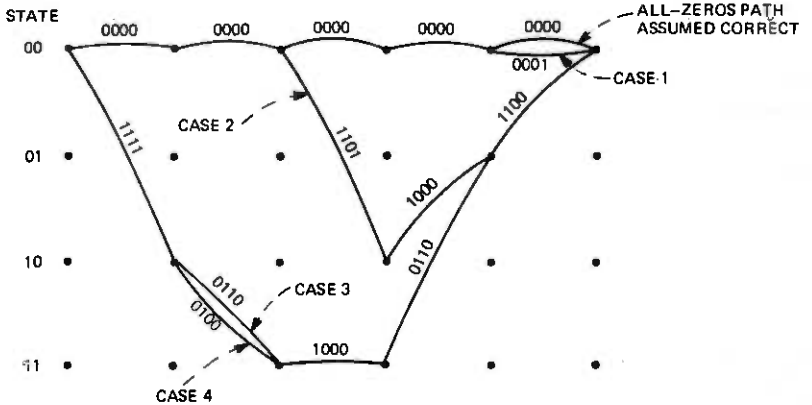


Fig. 5—Select error events for the maximum-likelihood detector.

$$\Lambda(\mathbf{A}) - \Lambda(\hat{\mathbf{A}}) = 2\sum r_k[\sqrt{E_1}(b_k - \hat{b}_k) + \sqrt{E_2}(y_k - \hat{y}_k)] - \sum[(\sqrt{E_1}b_k + \sqrt{E_2}y_k)^2 - (\sqrt{E_1}\hat{b}_k + \sqrt{E_2}\hat{y}_k)^2], \quad (5)$$

where the summation is performed over the unmerged span. Substituting (2) into (5) and recognizing that $b_k^2 = \hat{b}_k^2 = y_k^2 = \hat{y}_k^2 = 1$, we obtain

$$\Lambda(\mathbf{A}) - \Lambda(\hat{\mathbf{A}}) = A + n_{eq}, \quad (6)$$

where

$$A = 4\sum[\sqrt{E_1}\tilde{b}_k + \sqrt{E_2}\tilde{y}_k]^2 \quad (7)$$

and

$$n_{eq} = 4\sum n_k[\sqrt{E_1}\tilde{b}_k + \sqrt{E_2}\tilde{y}_k]. \quad (8)$$

In (7) and (8), we have used the nomenclature

$$\tilde{b}_k = \begin{cases} b_k & \text{if } \hat{b}_k = b_k \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

$$\tilde{y}_k = \begin{cases} y_k & \text{if } \hat{y}_k = y_k \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

The first event error probability P is equal to the probability that $\Lambda(\mathbf{A}) - \Lambda(\hat{\mathbf{A}}) < 0$. From (6) through (8), we conclude that

$$P = Q\left\{\sqrt{\frac{2\sum(\sqrt{E_1}\tilde{b}_k + \sqrt{E_2}\tilde{y}_k)^2}{N_0}}\right\}, \quad (11)$$

where Q is the complimentary error function. From this result, we now derive upper bounds on the bit error rate performance of the coded and uncoded transmission. We do the uncoded first.

Let the unmerged span be L channel digits long. We see from (11) that the first event error probability is dependent upon the correct sequence

along the unmerged span. For each possible error path of length L, we will determine the number of uncoded bit errors experienced along that path, and average (11) over all possible correct sequences. An upper bound on the average bit error rate for the uncoded transmission is then given by the summation over all possible incorrect paths, of the product of the number of bit errors experienced along a particular path and the average probability that the particular path has a metric exceeding that of the correct path.

Let the coded channel bits be different along the correct and incorrect paths in D symbols. Let the number of channel symbols for which an error occurs for both the coded and uncoded bits be denoted by r, and let the number of channel symbols for which an error occurs for the uncoded, but not for the coded, be denoted by s. Since the uncoded transmissions are equally likely to be ± 1, then along any L,D,r,s path, the coded and uncoded symbols may add or subtract, depending on the particular correct path. In $1/2^r$ of the paths, the correct symbols of the coded and uncoded transmissions will algebraically subtract over all r symbols. Similarly, in $\binom{r}{j}/2^r$ of the paths, there will be a subtraction in $(r-j)$ symbols and an addition in j symbols. In s symbols, $\tilde{b}_k^2 = 1$ and $y_k = 0$, while in $(D-r)$ symbols, $\tilde{b}_k = 0$ and $\tilde{y}_k^2 = 1$. There are $r+s$ errors committed in the uncoded transmission. Thus, averaging over all possible correct paths of the same L,D,r,s, we obtain the result that the average probability of error for each path of the same L and D for the uncoded transmission is given by

$$P_b = \frac{1}{2} \sum_{s=0}^{L-D} \sum_{r=0}^{D} \binom{L-D}{s} \binom{D}{r} \frac{r+s}{2^r} \sum_{j=0}^{r} \binom{r}{j} Q(r,s,j,D), \tag{12}$$

where

$$Q(r,s,j,D) \triangleq Q\left\{\sqrt{\frac{2}{N_0}}\left[(r-j)(\sqrt{E_1}-\sqrt{E_2})^2 + j(\sqrt{E_1}+\sqrt{E_2})^2 + (D-r)E_2 + sE_1\right]\right\}. \tag{13}$$

The factor of $1/2$ appearing in front of (12) arises from the fact that two uncoded bits are transmitted per epoch.

Using the inequality that for $x \geq 0, y \geq 0$,

$$Q\{\sqrt{x+y}\} \leq Q\{\sqrt{x}\}e^{-y/2}, \tag{14}$$

we can overbound and simplify (12) and (13) to the following:

$$P_b \leq Q\left\{\sqrt{\frac{2DE_2}{N_0}}\right\} e^{-E_1/N_0}[DXe^{E_1/N_0} + L(1+X) - D] \times [(1+X)^{D-1}(1+e^{-E_1/N_0})^{L-D-1}], \tag{15}$$

where

$$X \triangleq \frac{1 + e^{-4\sqrt{E_1E_2}/N_o}}{2} e^{-(E_1-2\sqrt{E_1E_2})/N_o}. \tag{16}$$

Finally, for a particular convolutional code, we use the code generating function matrix method of Viterbi[5] to identify all L,D paths for that code, and sum the contribution (15) for each such path over all possible paths. To this result must be added the contribution of the trivial case for which no coded errors occur (see Fig. 5, case 1). The contribution of these paths is simply

$$P_b = Q\left\{\sqrt{\frac{2E_1}{N_o}}\right\} + Q\left\{\sqrt{\frac{4E_1}{N_o}}\right\}. \tag{17}$$

The results of this exercise have been applied to the optimum $K = 7$, $r = \frac{1}{2}$ code,[6] and appear in Fig. 6. Plotted there is the uncoded bit error rate bound vs the required energy per information bit-to-noise ratio, e_b/N_o, for various ratios of interference to signal (E_2/E_1). Also plotted is the ideal, interference-free performance. We see here that as E_2/E_1 decreases below 2.5 dB, performance starts to improve. The utility of the maximum likelihood sequence estimation (MLSE) to detect uncoded transmission in the presence of coded area coverage interference is il-

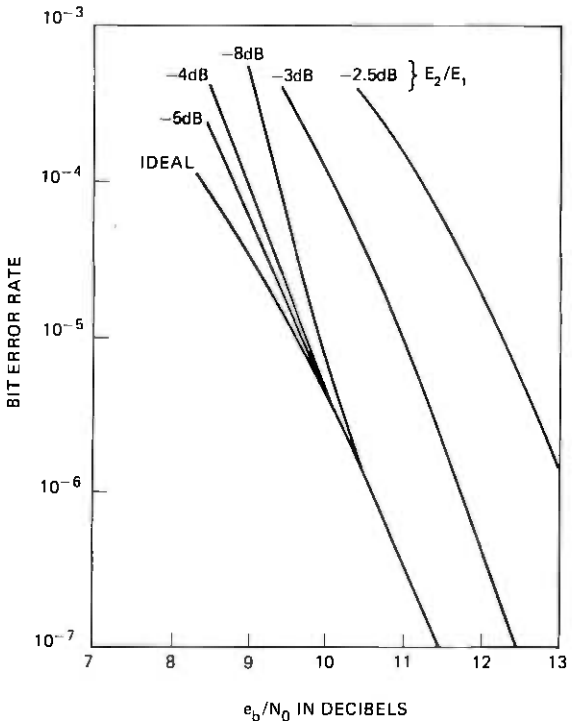


Fig. 6—Bound on uncoded bit error rate performance of maximum-likelihood detector vs e_b/N_o for select values of interference.

lustrated by the following example. Suppose $E_2/E_1 = -3$ dB. Then, if simple bit-by-bit detection is performed, an asymptotic degradation of about 10 dB from ideal would be expected. However, through use of the MLSE, the asymptotic degradation is about 1 dB.

We also see from these curves that, as E_2/E_1 decreases below about -8 dB, there is an apparent degradation in performance. This virtual result is caused by the bounding technique used, and is not experienced in practice. To see how this arises, we note that, as E_2/N_0 becomes small, all paths through the decoding trellis exhibiting a fixed number N of uncoded bit errors become equally likely. The contribution of each such path to the bit error rate bound is, however, summed, indicating a much higher bit error rate than would actually be encountered since only one such incorrect path could actually be selected at any node. For sufficiently small E_2/N_o, in fact, the bound no longer converges. To evaluate performance of the MLSE in the regime where the bound converges poorly, extensive simulation studies were performed and are shown in Fig. 7. These studies show that there is in fact a degradation in performance as E_2/E_1 decreases below -4 dB, but that the worst-case degradation of about 1 dB from ideal occurs for $E_2/E_1 = -10$ dB. As the in-

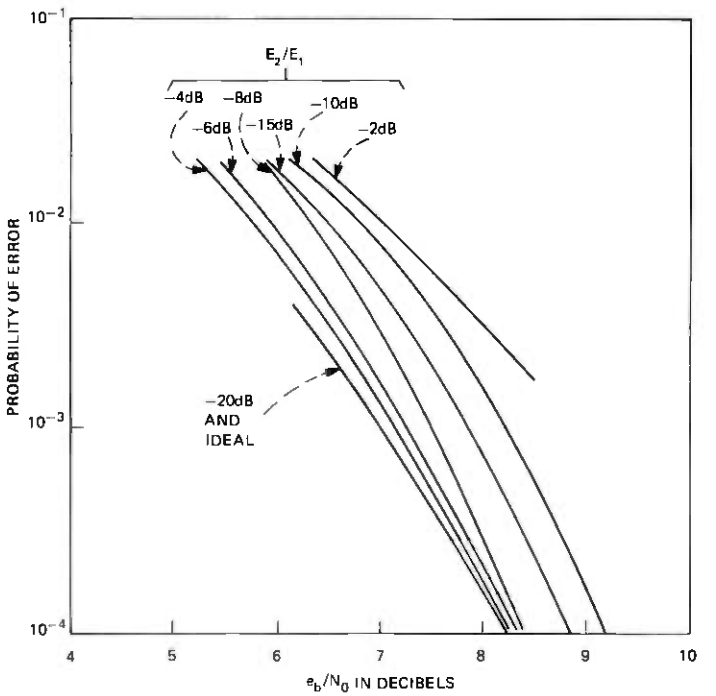


Fig. 7—Uncoded bit error rate performance of maximum-likelihood detector vs e_b/N_o obtained via simulation for select values of interference.

terference becomes smaller, performance begins to approach the ideal, interference-free case as intuitively expected.

We now study the performance of the encoded area beam. Again, let the unmerged span be L channel digits long and let the coded digits be different along the correct and incorrect paths in D channel symbols and N information symbols. Then, averaging overall possible combinations of the uncoded symbols, the average number of area beam bit errors incurred along any L,D,N path is given by:

$$P_b = N \sum_{s=0}^{L-D} \sum_{r=0}^{D} \binom{L-D}{s} \binom{D}{r} \frac{1}{2^r} \sum_{j=0}^{r} \binom{r}{j} Q(r,s,j,D), \tag{18}$$

where $Q(r,s,j,D)$ is given by (13). Invoking inequality (14), we obtain the bound:

$$P_b \leq N e^{-DE_2/N_0} (1 + e^{-e_1/N_0})^{L-D} (1 + X)^D, \tag{19}$$

where X is given by (16). Once again, we use the generating function matrix approach to determine the contribution of each incorrect path.

Results for the optimum $K = 7$ code appear in Fig. 8. Shown there is the bit error rate performance of the encoded area beam message vs e_b/N_o for select values of E_1/E_2, the interference-to-signal ratio. We see that, when E_1 becomes much greater than E_2, performance approaches the ideal, interference-free case since, under these conditions, the MLSE algorithm exploits the large difference between the signal and interference strengths to correctly decode the small signal. For $E_2 > E_1$, the bounding technique again suffers from poor convergence properties, and the results are meaningless. Again, extensive simulation studies were performed and are shown in Fig. 9. We see that, as expected, the ideal interference-free case is approached as E_1/E_2 becomes small.

For all values of E_1/E_2, the MLSE algorithm provides the best attainable performance. However, when E_1/E_2 becomes sufficiently small, the improvement possible via MLSE becomes negligible as shown by the plots of Fig. 10. These data were obtained experimentally and show the bit error rate performance of the ordinary Viterbi algorithm in the presence of a single bit-synchronous cochannel interferer. The ordinary Viterbi algorithm operates as though no interference was present and, unlike the MLSE algorithm, would be useless for $E_1 > E_2$. However, for $E_1 \ll E_2$, performance of the two are about the same, and the slight improvement possible via MLSE is not warranted in view of the additional complexity incurred.

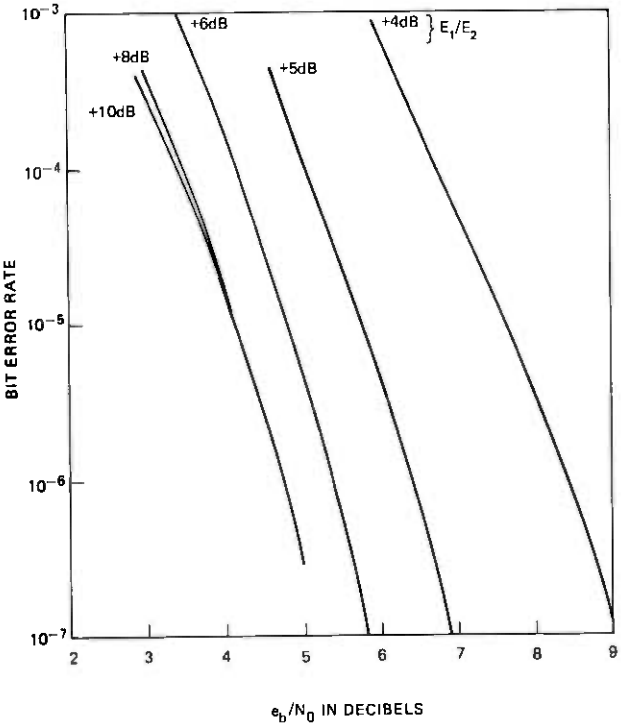


Fig. 8—Bound on coded bit error rate performance of maximum-likelihood detector vs e_b/N_0 for select values of interference.

IV. APPLICATION

We now apply the results of the preceding section to the problem of reducing mutual interference between spot and area coverage beams sharing a common spectral band. Let the spot-beam radiation pattern be Gaussian-shaped and usable to its −3 dB contour. In the absence of cochannel interference, the e.i.r.p. of the coded global beam would be 8 dB lower than that of the spot beam at its −3 dB contour for the same system outage and bit error rate (BER) performance. This 8-dB factor can be broken down into a 3-dB component, since the information rate of the global beam is half that of the spot beam, plus a 5-dB component representing the coding gain of a $K = 7$, $r = \frac{1}{2}$ convolutional code. Suppose we set E_2/E_1 at the 3-dB contour of the spot beam at −8 dB. Then, throughout the spot-beam coverage area, $-11 \text{ dB} \leq E_2/E_1 \leq -8 \text{ dB}$. From Fig. 6, we see that, over this range, the BER performance of the uncoded spot-beam message is degraded by at most 1 dB if MLSE is employed. By contrast, if bit-by-bit detection of the spot-beam message were employed, the degradation would be between 2.9 dB and 4.4 dB.

Let the e.i.r.p. of both the spot and area coverage beams rise by 1 dB.

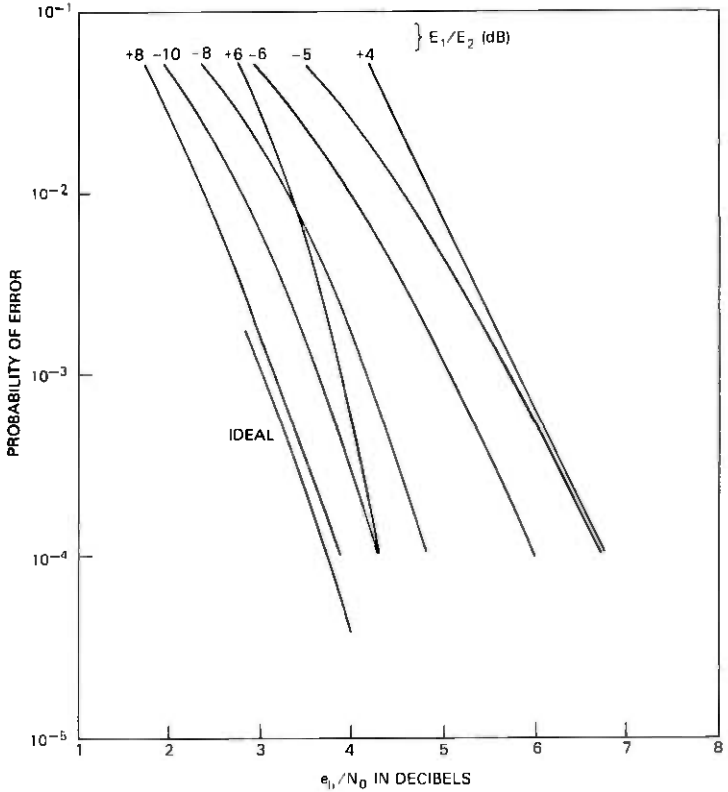


Fig. 9—Coded bit error rate performance of maximum likelihood detector vs e_b/N_o obtained via simulation for select values of interference.

Then, throughout the spot-beam coverage region, the BER performance is at least as good as that obtained in the absence of interference with 1 dB less power. From Fig. 8, we see that, beyond the −3 dB contour of the spot beam, we can communicate via the area coverage beam in conjunction with MLSE with at most 1-dB degradation from the ideal interference-free situation provided $E_1/E_2 > 5.5$ dB. Finally, from Fig. 10, we see that we can use the area beam with the ordinary Viterbi algorithm provided $E_1/E_2 < -12$ dB. From these observations, we can construct the plot of Fig. 11, which shows the one-dimensional radiation patterns of a spot beam and the area beam and the usable regions for the spot and area coverage beams in the vicinity surrounding a spot beam. Implicit in this illustration is the fact that the e.i.r.p. of both the spot and area beams is increased by 1 dB to provide the same grade of service as possible with 1 dB less power in the absence of cochannel interference. We see that communication via the spot beam, in conjunction with MLSE, is employed out to the −3 dB contour of the spot beam. From $\theta = \theta_{3dB}$ out to $\theta = 1.4\theta_{3dB}$, we can communicate via the area beam, even though

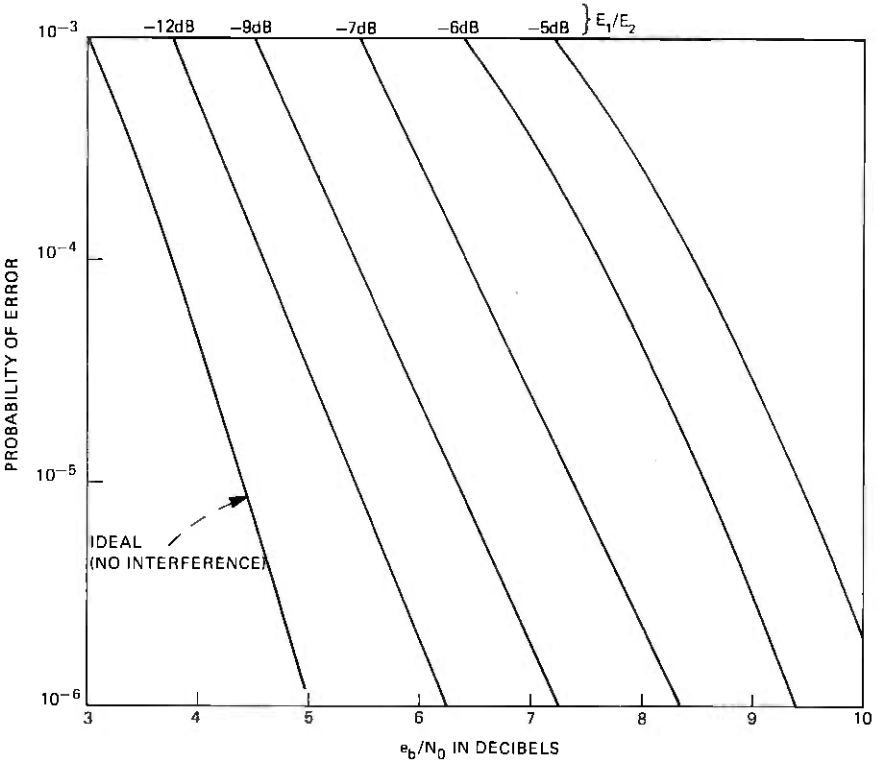


Fig. 10—Measured bit error rate performance of a $K = 7$, $r = \frac{1}{2}$ convolutional code vs e_b/N_o for select values of interference. The interference is bit-synchronous with the encoded channel bits, and the ordinary Viterbi algorithm with soft (3-bit) quantization is employed.

the interference is stronger than the desired signal. Between $\theta = 1.4\theta_{3dB}$ and $\theta = 2.75\theta_{3dB}$, the performance degradation of the area beam exceeds the allotted 1 dB, and the desired grade of service cannot be provided. This region, then, is blacked out. Finally, for $\theta > 2.75\theta_{3dB}$, communication via the global beam is again possible.

Thus, through utilization of an area coverage beam in conjunction with channel coding and MLSE, the blackout region of a multiple spot-beam communication satellite is reduced from the entire region not serviced by any spot beam to a thin annular ring surrounding each spot beam. There is no sacrifice in the capacity of the spot beams, and the power penalty is 1 dB for all beams.

Let us now consider a specific example. We assume the existence of 10 spot beams, half of which employ one polarization and half the orthogonal polarization. In the absence of interference, each spot beam transponder uses a 3-watt final power amplifier, and the difference between the spot and area beam antenna gains is 20 dB. Suppose we deploy

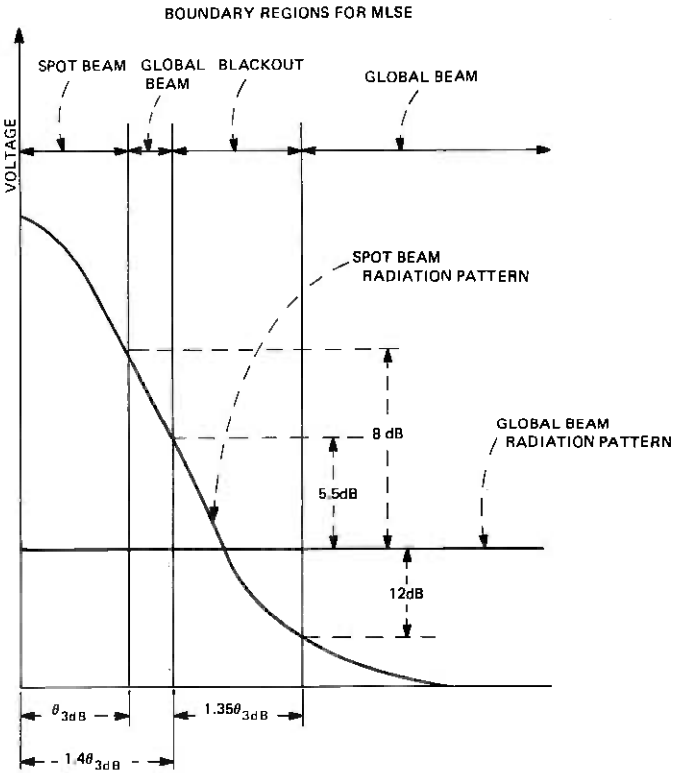


Fig. 11—One-dimensional plot showing the usable regions attainable via MSLE for a hybrid spot-area beam satellite system employing a $K = 7$, $r = 1/2$ code for the area beam transmission. The spot beam antenna pattern is Gaussian shaped.

a single-area beam transponder employing one of the two polarizations; the capacity of this beam is one-half that of a spot beam, and a $K = 7$, $r = \frac{1}{2}$ code is employed. In the absence of interference, the RF power required of the area beam would be $20 - 8 = 12$ dB higher than any spot beam. The total required RF power for the hybrid system outlined above is then

$$P = 1.25 \times [3 \times 10 + 47.5] = 97 \text{ watts.} \tag{20}$$

By contrast, if we employ the band-splitting technique described in Section II, we would need 6 dB more power for each spot beam, and the power required for the area beam would be 17 dB higher than that required for the spot beam in the absence of interference, since coding is not employed. The total power, then, would be

$$P = 4 \times 3 \times 10 + 50.1 \times 3 = 270 \text{ watts.} \tag{21}$$

Considering a 30-percent efficiency for the final TWT, the total dc power

required via coding is 323 watts, while that needed for the alternative band-splitting approach is 900 watts.

Since, through use of coding, the dc power required for an area beam is only 158 watts, we might consider deploying a second area beam using the orthogonal polarization. Then, not only do we double the capacity into the outlying region, but we also eliminate the blackout region, since each spot beam is used in only one polarization. Area coverage communication to the blackout region of one polarization can thereby be provided in the second polarization. The dc power required for this approach is 442 watts.

V. MAXIMUM-LIKELIHOOD ALGORITHM WITH FIXED PHASE SHIFT

In Section III, we derived a maximum-likelihood algorithm which allows joint area and spot-beam coverage sharing a common spectral band whenever there is no carrier phase shift difference between the area and spot beam transmissions. We now derive the proper algorithm for use when there is a fixed phase shift difference, θ. During the kth clock cycle, the spot beam source emits two bits, $b_{1,k}$ and $b_{2,k}$, and the area beam source emits a single bit a_k and two encoded channel bits $y_{1,k}(\mathbf{a})$ and $y_{2,k}(\mathbf{a})$. The data $b_{1,k}$ and $b_{2,k}$ are modulated onto a carrier via 4ϕ-PSK, as are $y_{1,k}$ and $y_{2,k}$. Thus, we transmit:

$$R(t) = \sqrt{E_1}b_{1,k}\cos(\omega t + \theta) + \sqrt{E_1}b_{2,k}\sin(\omega t + \theta) + \sqrt{E_2}y_{1,k}\cos\omega t + \sqrt{E_2}y_{2,k}\sin\omega t. \quad (22)$$

The receiver locks onto the phase of the encoded area beam and, during the kth clock cycle, the receiver observes, after coherent demodulation, the two test statistics:

$$r_{1,k} = \sqrt{E_2}y_{1,k} + \sqrt{E_1}b_{1,k}\cos\theta + \sqrt{E_1}b_{2,k}\sin\theta + n_{1,k} \quad (23)$$

$$r_{2,k} = \sqrt{E_2}y_{2,k} - \sqrt{E_1}b_{1,k}\sin\theta + \sqrt{E_1}b_{2,k}\cos\theta + n_{2,k}. \quad (24)$$

The path metric now takes the form:

$$\Lambda(\mathbf{a},\mathbf{b}) = \sum_k [r_{1,k}(\sqrt{E_2}y_{1,k} + \sqrt{E_1}b_{1,k}\cos\theta + \sqrt{E_1}b_{2,k}\sin\theta) + r_{2,k}(\sqrt{E_2}y_{2,k} - \sqrt{E_1}b_{1,k}\sin\theta + \sqrt{E_1}b_{2,k}\cos\theta)] - \sqrt{E_1E_2}[y_{1,k}(b_{1,k}\cos\theta + b_{2,k}\sin\theta) - y_{2,k}(b_{1,k}\sin\theta - b_{2,k}\cos\theta)]. \quad (25)$$

As before, we define the state of the encoder by the contents of the first $K-1$ stages of its shift register, and each state can be accessed via eight paths. Along each path, we compute the partial metric:

$$\Lambda_k(\mathbf{a},\mathbf{b}) = \Lambda_{k-1}(\mathbf{a},\mathbf{b}) + \sqrt{E_2}r_{1,k}y_{1,k} + \sqrt{E_2}r_{2,k}y_{2,k} + \sqrt{E_1}(r_{1,k} - \sqrt{E_2}\,y_{1,k})(b_{1,k}\cos\theta + b_{2,k}\sin\theta) - \sqrt{E_1}(r_{2,k} - \sqrt{E_2}\,y_{2,k})(b_{1,k}\sin\theta - b_{2,k}\cos\theta), \quad (26)$$

and save the path and metric of the larger. Thus, with a fixed known phase shift, maximum-likelihood decoding is also possible.

VI. CONCLUSIONS

In multiple spot-beam communication satellite systems, it is often desirable to provide service to remote areas not covered by any spot beam. This additional service should neither diminish the capacity of the spot beams nor cause a severe downlink power penalty. We considered deployment of an area beam transponder, in addition to the fixed spot beams, and saw that satisfaction of the above requirements implies considerable downlink cochannel interference at all ground stations located in the vicinity of any spot beam. The use of binary convolutional codes for the area beam transmission was shown to greatly curtail the performance degradation resulting from this cochannel interference and also reduce the prime power requirements of the area beam transponder.

A maximum-likelihood algorithm was derived to optimally detect either the uncoded spot beam transmission or the coded area beam transmission, and performance of this algorithm was evaluated. Use of this algorithm was shown to provide for reliable spot-beam communication in the presence of cochannel interference. It is also possible to reliably communicate via the global beam in the presence of a much stronger spot-beam interference. These results were then applied to a scenario in which interference was reduced on the uplink via the simple technique of band-splitting between the area and spot beams. Such a technique is unsuitable for the downlink because of the power penalty incurred. On board, the uplink bits are regenerated and switched into the appropriate downlink beam, and a $K = 7$, $r = \frac{1}{2}$ code is employed for the downlink area beam. Results show that the degradation from cochannel interference is contained to be less than 1 dB over the entire service area except for a thin annular ring surrounding each spot beam. Traffic originating within or destined for these blackout rings might be backhauled to the nearest serviceable region, or else a second area beam, employing the dual polarization, might be deployed such that, for any given spot beam, the blackout region is contained to only one polarization. Since the spot beams use both polarizations to minimize interference among themselves, the MLSE algorithm must still be used at all spot-beam ground stations to provide spot-beam service with minimal performance degradation.

The satellite prime power demands to satisfy RF radiated power requirements were evaluated and shown to be within the capability of the Thor-Delta class. Thus, the use of spot and area coverage beams, sharing a common spectral band, in conjunction with channel coding techniques, appears to be an acceptable method for providing universal service via high-capacity digital switching satellites of the future.

VII. ACKNOWLEDGMENTS

The author wishes to thank his colleagues, D. O. Reudink and Y. S. Yeh, for their stimulating discussions and contributions, and also Mrs. Phyllis Arnold who wrote the programs for the simulation studies.

REFERENCES

1. L. C. Tillotson, "A Model of a Domestic Satellite Communication System." B.S.T.J., *47*, No. 10 (December 1968), pp. 2111–2137.
2. R. Cooperman and W. G. Schmidt, "Satellite Switched SDMA and TDMA Systems for Wideband Multi-Beam Satellite," ICC Conference Record, 1973.
3. A. S. Acampora, D. O. Reudink, and Y. S. Yeh, "Spectral Re-Use in 12 GHz Satellite Communication Systems," ICC Conference Record, 1977.
4. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Part I, New York; Wiley, 1968.
5. A. J. Viterbi, "Convolutional Codes and Their Performance in Communication Systems," IEEE Trans. Comm. Tech., *COM-19* (October 1971), pp. 751–772.
6. K. S. Gilhousen, "Coding Systems Study for High Data Rate Telemetry Links," Linkabit Corporation, San Diego, California, January 1971.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 9, September 1978
Printed in U.S.A.

Reliability of a Microprocessor-Based Protection Switching System

By G. S. FANG

(Manuscript received July 7, 1977)

High-capacity transmission systems usually include one or more hot spares for protection. When a regular transmission channel fails, its signal is rapidly transferred to the spare channel under the control of protection switching circuits so that there is little signal degradation or interruption. This paper studies the reliability of a microprocessor-based terminal protection switching system. Some new and interesting behavior patterns for transmission systems with automatic protection switching are revealed. Also, some new memory self-checking algorithms are presented which increase the capability of microprocessor system fault recognition.

I. INTRODUCTION

In high-capacity transmission systems, any failure may affect a large number of message circuits. Such systems usually include one or more hot spares to increase system reliability. When a regular transmission channel fails, its signal is rapidly transferred to the spare channel under the control of protection switching circuits so that there is little signal degradation or interruption. This paper studies the reliability of a microprocessor-based terminal protection switching system (TPSS). The specific transmission facility under consideration is the L5E coaxial cable analog system, which is an expanded version of the L5 system.[1] The L5E multiplex equipment, or multimastergroup translators (MMGT), carry up to eight mastergroups, or 4800 telephone circuits. The TPSS will automatically switch into service a protection MMGT in the event of a failure of any one of up to 20 MMGTs.

Reliability theory has been studied by numerous authors,[2,3] and almost every Bell System transmission facility with automatic protection switching has been the subject of at least one reliability study.[4,5] The present analysis was undertaken for several reasons. First, many simplifying assumptions were made in the previous studies. Not all the

effects of the reliability of the switch, the protection switching control circuit, and the monitor circuit failures were taken into account. Second, in most cases, exponentially distributed restoration time has been assumed. This means that the probability of restoration at any instant after a failure is assumed to be independent of how much time has already been spent on restoring the failure. This assumption is rarely true in high-capacity transmission systems. Third, only steady-state analyses were made. A system with hidden failures will not reach its steady state in its lifetime. Fourth, a microprocessor-based protection switching control circuit has not been studied in such detail before. Finally, past experiences have shown that maintenance-induced service outages contribute to a very big share of the total outage time. This study also tries to take these outages into consideration.

With the MMGT system as an example, the present study attempts to analyze the same reliability problem in more detail and with less restrictive assumptions. Section II describes the protection switching arrangement. Section III explains the specific approaches used in this paper. Section IV presents the results graphically to emphasize the various reliability trends. Section V summarizes the conclusions obtained. Appendix A investigates some new microprocessor self-checking algorithms and Appendix B presents the derivations.

II. MMGT PROTECTION SWITCHING SYSTEM DESCRIPTION

Figure 1 is a simplified MMGT-system block diagram which illustrates the $1 \times n$ protection switching arrangement. There is one protection channel in each direction of transmission. Under the command of the microprocessor, each protection channel protects up to n regular channels, where n is equal to 20 in the TPSS. The same processor is used to control the switching actions of both directions of transmission. The switches are all solid-state devices, and their normal states are indicated in the figure. The crucial output switches are dual-powered. Parts of the output switch are designated the through switch and the substitute switch for later reference.

When there is no alarm from the various regular pilot detectors, the processor exercises the input switches for each channel sequentially to detect possible protection failures. In the event of a failure of one of the regular channels, the corresponding pilot detector sends an alarm to the processor. If the protection channel is available, the processor will first switch the input signal through the input switches to feed the protection channel. Whether the protection detector indicates a good signal or not, the processor will complete the 1×2 output switch. The regular detector is now monitoring the signal supplied by the protection channel via the output substitute switch. If the regular detector still alarms after the protection switch, the switching action will be reversed. The 1×2 output

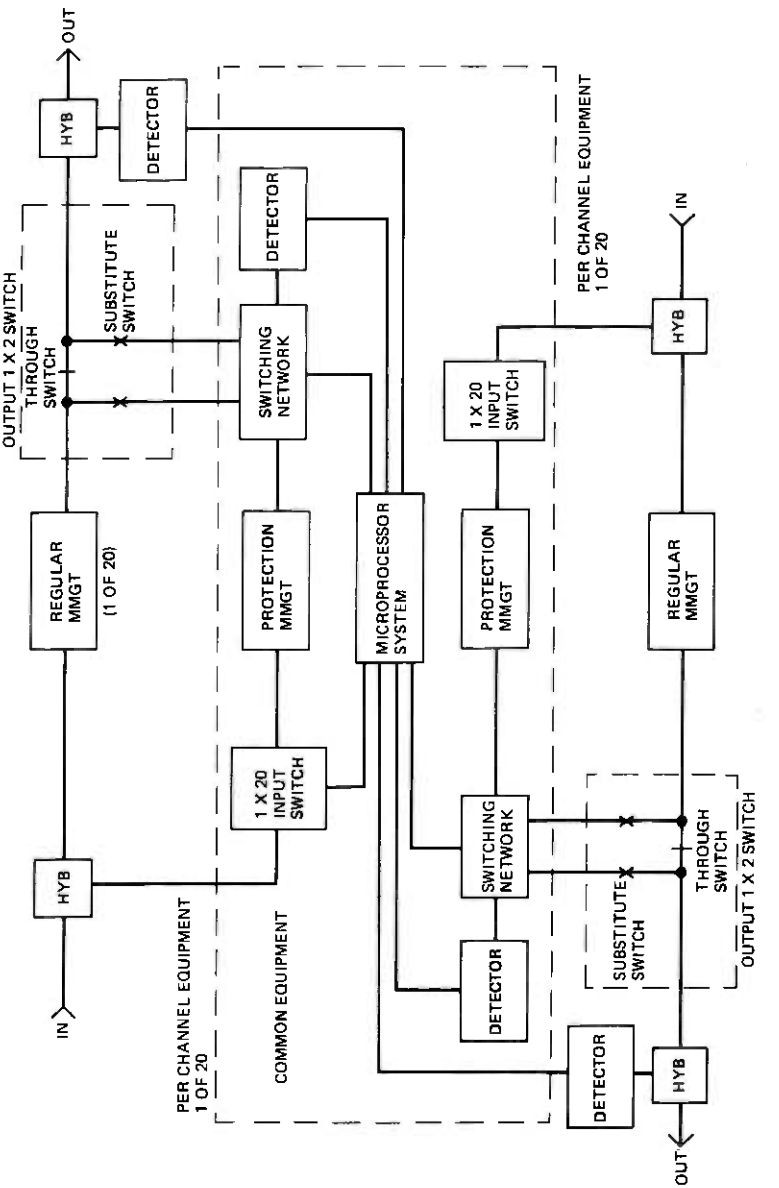


Fig. 1—TPSS block diagram.

switch will be deactivated and the input switch released. If the regular detector stops alarming after the output switching, a successful protection switch has been made, and the protection detector is monitoring the failed regular channel. When the failed channel is repaired, the protection detector will see a good signal, and the switches will return to their normal states. The protection channel is then free to service another regular channel failure.

Service outages can occur in many ways. In addition to multiple transmission failures, they can also be generated by the failures of the detectors, the switches, or the microprocessor system. The various failure modes are taken into account in later derivations.

III. APPROACHES

Two reliability measures of interest in transmission systems are used in this study. The first measure is the probability of service outage due to equipment failures. This probability translates directly to the system outage time per year and is the most commonly used figure of merit in determining transmission system reliability. The second measure is the probability of having maintenance activities going on. This measure will be abbreviated as the probability of activity. It is believed to be closely related to the probability of having maintenance-induced outages. This probability of activity is greater than the probability of having alarms because there are failures that cannot be detected locally. For instance, if the pilot detector for a failed regular channel is stuck to the state of no alarm, the failure can only be detected by downstream offices. Thus there may be maintenance activities in an office but no alarm. The probability of activity is less than the probability of having failures because there are undetectable failures such as the breakdown of an output substitute switch. A reliable system should have a small probability of outage and a small probability of activity.

Two additional criteria are used to measure the effectiveness of the overall protection plan. The improvement factor (IF) is defined as the ratio of the probability of outage without protection switching to that with protection switching. The activity factor (AF) is defined as the ratio of the probability of activity with protection switching to that without protection switching. These definitions agree with the common notion that an effective protection plan should provide more improvement and less activity. Thus, a better protection system has a bigger IF and a smaller AF. The activity factor is always greater than one.

The probabilities discussed above are derived under the assumptions that the various failures are statistically independent and the failure rates are constant. These are very simple assumptions considering the complexity of the problem. The assumption of statistical independence is made to avoid estimating conditional failures, although there is

probably dependency between the through switch and the substitute switch. The constant failure rate implies exponentially distributed failures, i.e., any working item is as good as new. This is a reasonable assumption for solid-state devices after the initial "burn-in" period. Notice that no distributional assumption is made on the restoration time. Based only on the failure rates and the restoration times of the components of the system, the various probabilities are derived from the basic definitions of conditional probability. Not only does this approach require little mathematical background, but the result is more general and more accurate than the usual method of Markoff chain or birth-and-death stochastic processes,[2,3] which assume that both failure and restoration times are exponentially distributed.

IV. DETAILED RESULTS

Table I introduces the notations and gives the estimated failure rates in FITS (number of failures per component per 10^9 hours), restoration times in hours, and the availabilities of the various components. The restoration time is the sum of the detection time and the equipment replacement time. The mean value of the replacement time t is assumed to be 1 hour. Some failure rates are expressed in terms of other failure rates to show their relative dependence. This is necessary in later parameter sensitivity studies. The failure of a substitute switch can only be detected when its use is called for. Thus, its detection time is the mean time between transmission failures of its corresponding channel, i.e., $1/(\lambda_r + \lambda_t + \lambda_0)$. The same is true for the detection time of a regular detector, except that the assumed probability that a failed detector gives a no-alarm indication is 1/4. In both cases, the equipment replacement time is ignored since it is small compared with the detection time.

The detection times of the hidden CPU (central processing unit) and EROM (erasable read-only memory) failures should also be similarly calculated. However, the failure of the regular channels to be exercised sequentially should provide local craftspeople with the indication that something is wrong. Therefore, the detection times are assumed to be 24 hours. The availability[3] of an item is the probability that the item is working. It is a function of time with an initial value of one and with a steady-state value equal to the mean time to failure divided by the sum of the mean time to failure and the mean restoration time. If a component has a short failure detection time, the transient portion in its availability value vanishes quickly, and the steady-state theoretical availability approximates the actual availability very well. For example, the steady-state availability of the regular channel is $p_r = 1/1.000001$. The reliability function of the regular channel is $e^{-10^{-6}t}$. It takes only 1 hour for the reliability function to reach its steady-state availability value.

Table I — Estimated failure rates

	FITS	Mean Restoration Time (hr)	Availability
Regular channel	$\lambda_r = 1000$	$\mu_r = t$	$p_r = \frac{1}{1 + \lambda_r \mu_r}$
Through switch	$\lambda_t = 150$	$\mu_t = t$	$p_t = \frac{1}{1 + \lambda_t \mu_t}$
Output switch	$\lambda_0 = \frac{1}{3}\lambda_t$	$\mu_0 = t$	$p_0 = \frac{1}{1 + \lambda_0 \mu_0}$
Substitute switch	$\lambda_s = \frac{2}{3}\lambda_t$	$\mu_s = \frac{1}{\lambda_r + \lambda_t + \lambda_0}$	$p_s = \frac{1}{1 + \lambda_s \mu_s} + \frac{\lambda_s}{(\lambda_s + \mu_s^{-1})^2 T}\left[1 - e^{-(\lambda_s + \mu_s^{-1})T}\right]$
Regular detector	$\lambda_d = 300$	$\mu_d = \frac{1}{4}\frac{1}{\lambda_r + \lambda_t + \lambda_0}$	$p_d = \frac{1}{1 + \lambda_d \mu_d} + \frac{\lambda_d}{(\lambda_d + \mu_d^{-1})^2 T}\left[1 - e^{-(\lambda_d + \mu_d^{-1})T}\right]$
Protection detector	$\lambda_D = 300$	$\mu_D = t$	$p_D = \frac{1}{1 + \lambda_D \mu_D}$
Protection channel	$\lambda_p = \lambda_r + 4\lambda_s + 100$	$\mu_p = t$	$p_p = \frac{1}{1 + \lambda_p \mu_p}$
CPU	$\lambda_c = 500$	$\mu_c = 24 + t$	$p_c = \frac{1}{1 + \lambda_c \mu_c}$
EROM	$\lambda_e = 300$	$\mu_e = \mu_c$	$p_e = \left(\frac{1}{1 + \lambda_e \mu_e}\right)^4$
RAM	$\lambda_a = 400$	$\mu_a = t$	$p_a = \left(\frac{1}{1 + \lambda_a \mu_a}\right)^2$

These arguments do not hold for failures requiring long detection times. For instance, the mean time to failure and the mean restoration time of a substitute switch are in the order of hundreds of years, while the life span of the equipment is expected to be only 40 years. To obtain an appropriate availability in such cases, one would observe that the restoration time of the substitute switch is exponentially distributed. This is due to the fact that the replacement time is ignored and the detection time depends on the transmission failures which are exponentially distributed. Thus the availability function can be derived explicitly as

$$A_s(t) = \frac{1}{1 + \lambda_s \mu_s} + \frac{\lambda_s}{\lambda_s + \mu_s^{-1}} e^{-(\lambda_s + \mu_s^{-1})t}.$$

The availability p_s given in Table I is the $A_s(t)$ averaged over the life span T of the equipment. The availability of the detector p_d is obtained similarly. The availability expressions of the EROM and the RAM reflect the use of 4 EROMs and 2 RAMs in the TPSS.

To gain insight and to study the sensitivity of the derived probabilities to the estimated failure rates and restoration times, the various estimated parameters are varied one at a time to show the system reliability trends. The results are presented graphically in the figures. In each figure, the solid line corresponds to the ordinate at the left and the dotted line to that at the right.

Figures 2 through 7 present the variations of the outage and the activity probabilities as functions of the regular channel, the detector, the switch, the CPU, the EROM, and the RAM failure rates, respectively. Most of the curves are almost linear because, for the small failure rates of interests, they are still in their linear regions. As far as the probability of outage is concerned, undetectable failures are the most damaging. The hidden detector and the substitute switch failures contribute to the bigger slopes in Figs. 3 and 4. Increasing the microprocessor system failures adds very little to the outage probability, as can be seen from Figs. 5 to 7. The probability that has the fastest increase is the switch failure rates because there are so many switches in the system. Figure 8 indicates that service outage can increase substantially if the replacement time for failed equipment is long. Figure 9 shows the effect of varying the detection time of the hidden microprocessor failure. Neither the outage nor the activity probability is sensitive to the detection time. Figure 10 shows the effect of varying the number of regular channels equipped. The discrete points in the figure are connected to show the almost linear trends. When the system is fully loaded, i.e., $n = 20$, there are about 2 minutes of service outage each year due to equipment failures and there is about half an hour of maintenance activities. It should be emphasized that the curves present the right trends

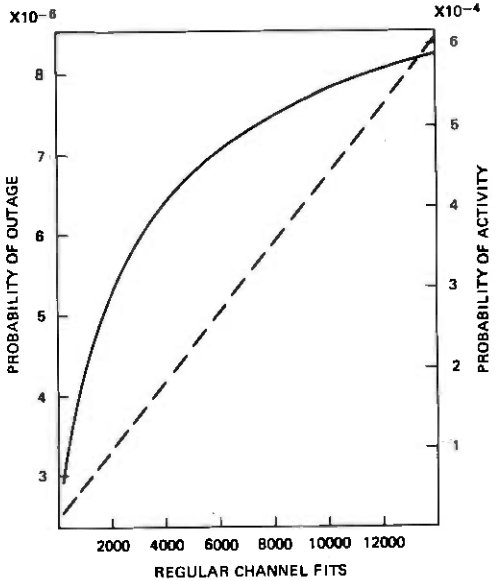


Fig. 2—Probabilities of outage and activity as functions of regular channel failure rate.

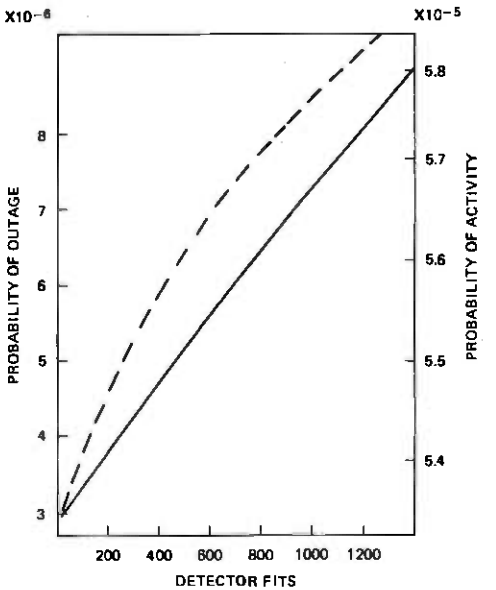


Fig. 3—Probabilities of outage and activity as functions of detector failure rate.

rather than numerical accuracy. From Fig. 2, if the failure rate of the regular channel is increased by ten times, there will be 4 minutes of outage and 4 hours of activity each year. Figure 10 shows the two

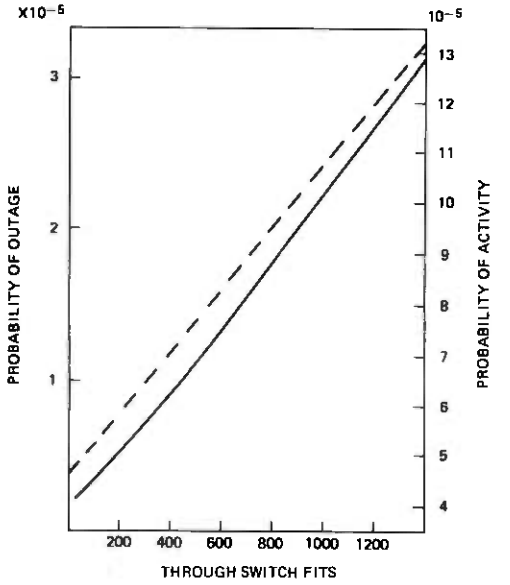


Fig. 4—Probabilities of outage and activity as functions of through switch failure rate.

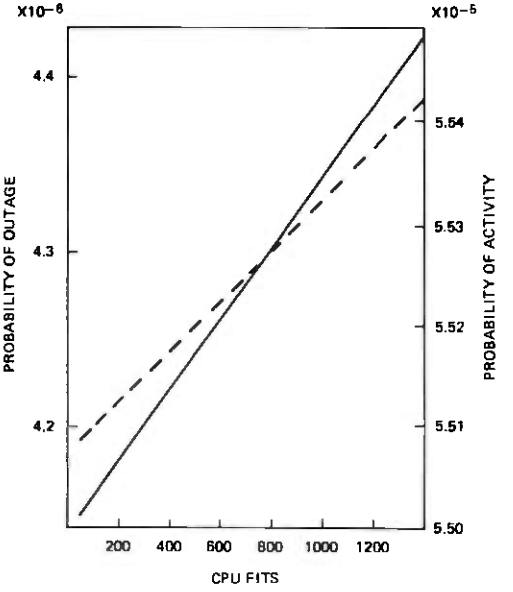


Fig. 5—Probabilities of outage and activity as functions of CPU failure rate.

probabilities as functions of the number of regular channels. The discrete points are connected to indicate trends. For terminal circuits which usually have small failure rates, there is scarcely any need for a second

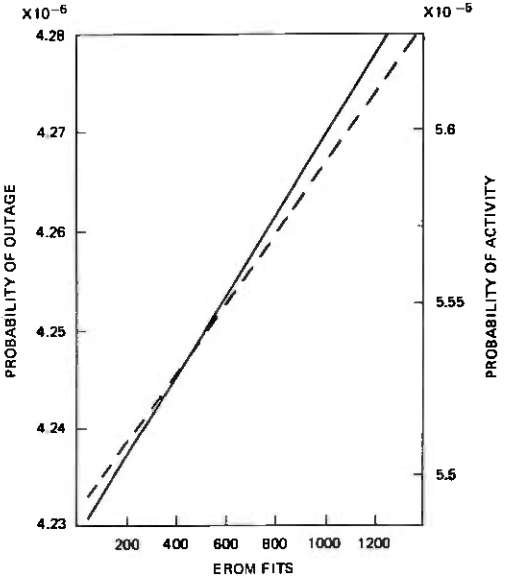


Fig. 6—Probabilities of outage and activity as functions of EROM failure rate.

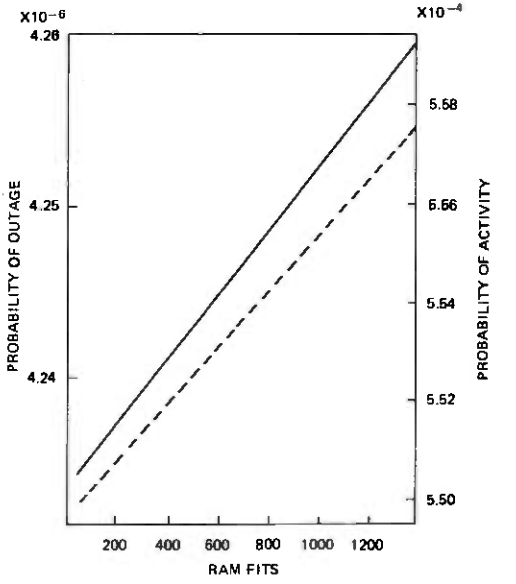


Fig. 7—Probabilities of outage and activity as functions of RAM failure rate.

protection channel even when the number of regular channels is large.

A system without protection switching has only the regular channels and their corresponding detectors to indicate alarms. The switches and

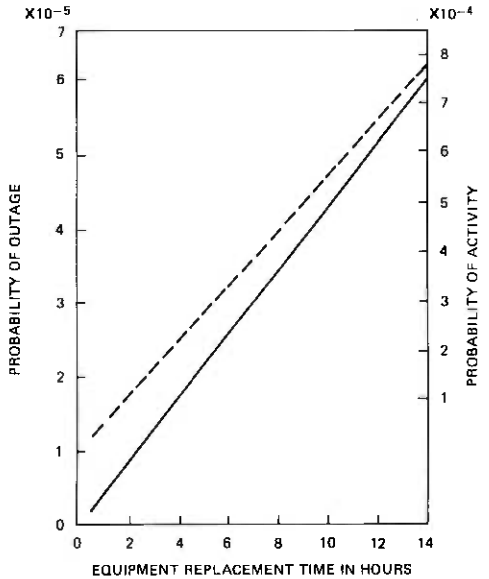


Fig. 8—Probabilities of outage and activity as functions of equipment replacement time.

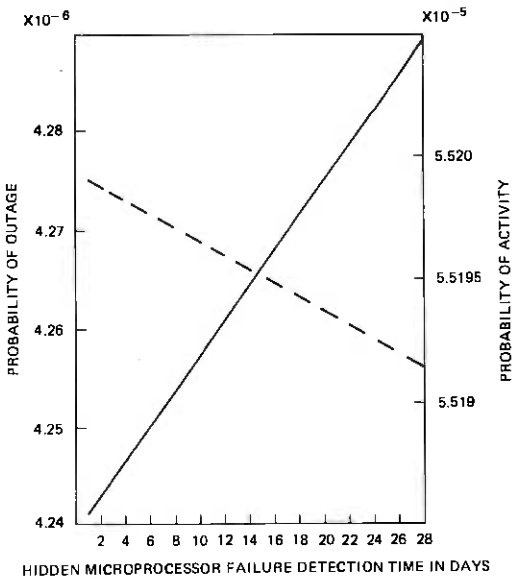


Fig. 9—Probabilities of outage and activity as functions of hidden microprocessor failure detection time.

the microprocessor devices are not required. Thus there is definitely less activity in the maintenance offices. Figure 11 shows the trend that, for small regular channel failure rates, the IF can be less than unity, i.e.,

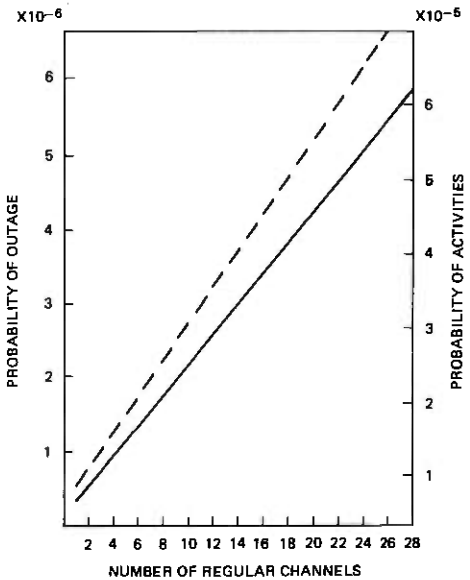


Fig. 10—Probabilities of outage and activity as functions of number of regular channels.

having protection switching actually causes more service outage. This is true when the failure rate of the regular channel is small compared with those of the protection switching circuits. Furthermore, protection switching generates many more activities at low regular channel failure rates. Figure 12 amplifies this fact by examining the 1 × 1 configuration. The IF is so small and the AF is so big that implementation of a 1 × 1 protection plan is questionable at low failure rates. Figure 13 gives the variations of the two factors with detector failure rates. Since detector failures have little effect on the outage probability of an unprotected system, the IF decreases with increasing detector failure. The interesting shape of the AF curve is due to the relatively rapid increase in the probability of activity for an unprotected system when the detector failure rates are small. This behavior is unique to the variation of the detector failure rate because an unprotected system is equipped only with the transmission channels and the detectors.

Figure 14 again indicates the important role played by the output switch. If its failure rate is high enough, the IF can reduce to less than unity. With a perfect switch, the outage of a protected system can be hundreds of times less than that of an unprotected system. The curves showing the two factors as functions of the CPU, the EROM, and the RAM failure rates are not given here. These curves can be simply deduced from Figs. 5 to 7 because the various probabilities of an unprotected system are independent of microprocessor failures. Similarly, the factors involving hidden microprocessor failure restoration time can be obtained

from Fig. 9. Figure 15 shows that both the IF and the AF are not very sensitive to how long it takes to replace failed equipment. Figure 16 varies the number of regular channels. It indicates that more than 10 regular

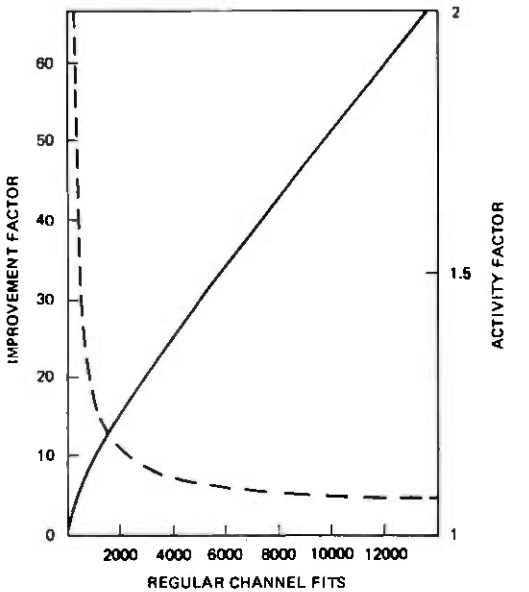


Fig. 11—Improvement and activity factors as functions of regular channel failure rates.

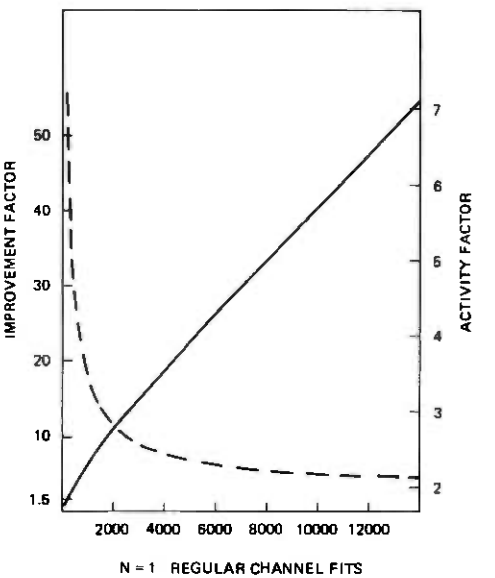


Fig. 12—Improvement and activity factors as functions of regular channel failure rates.

channels should be used to take advantage of the protection switching arrangement.

Figure 17 exhibits an interesting behavior of general protection

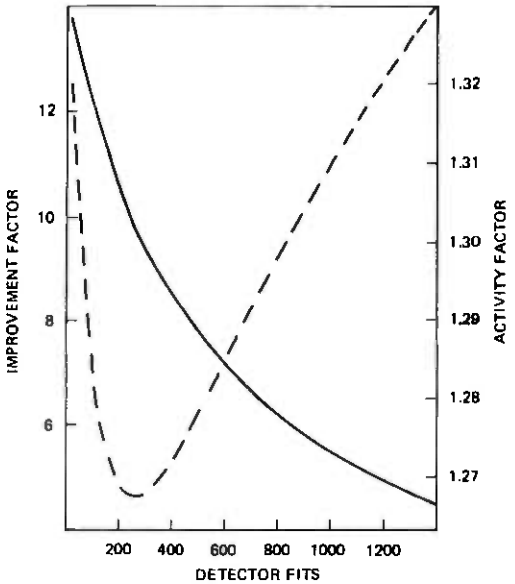


Fig. 13—Improvement and activity factors as functions of detector failure rates.

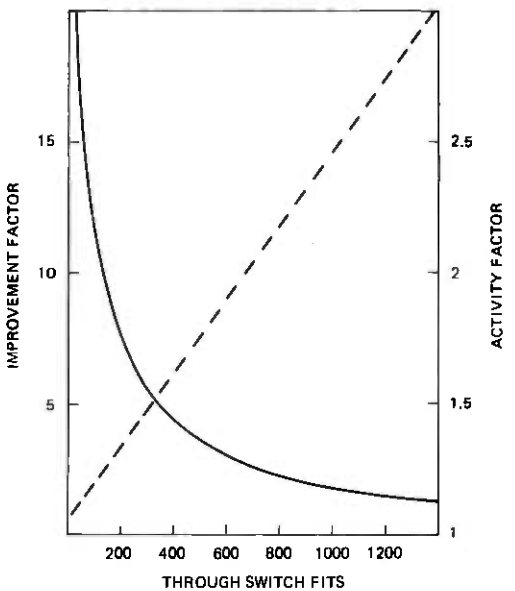


Fig. 14—Improvement and activity factors as functions of through switch failure rates.

switching systems. As the failure rate of the regular channel increases, the IF increases from less than one to a maximum and then starts to decrease. When the failure rate becomes very large, the outage probability is close to 1 with or without protection switching. Thus the IF approaches 1 eventually. The maximum IF shown in the figure occurs at around 150,000 FITS. Although it is unlikely for a terminal multiplexer to possess so high a failure rate, a line transmission system with many cascading repeaters may very well have a failure rate of this order. Therefore, whenever a line protection switching system is planned, the reliability should be studied to determine the length of the protection span so that the IF does not fall in its decreasing region. Of course, the outage probability should also be taken into account to meet any prescribed service objectives.

V. CONCLUSIONS

The reliability of the microprocessor-based TPSS has been studied in detail using conditional probability. Consideration of the four criteria; i.e., the probability of outage, the probability of activity, the improvement factor, and the activity factor, should provide an adequate description of the effectiveness of the overall protection plan. Several conclusions can be drawn from the analysis. First, terminal circuits usually have low failure rates so that one protection channel is adequate for the protection of many regular channels without having excessive

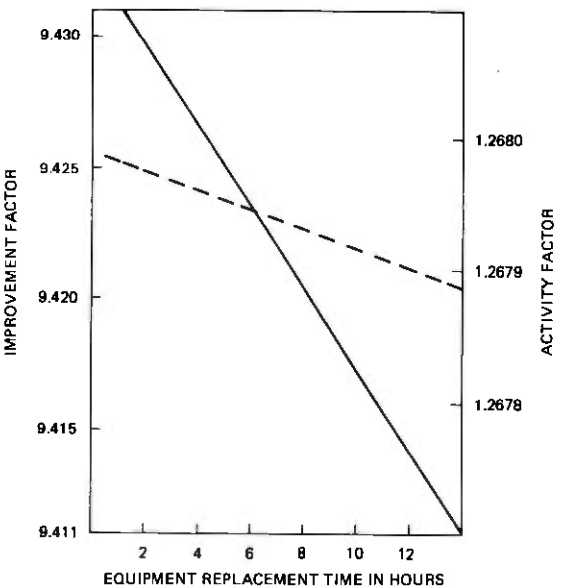


Fig. 15—Improvement and activity factors as functions of equipment replacement time.

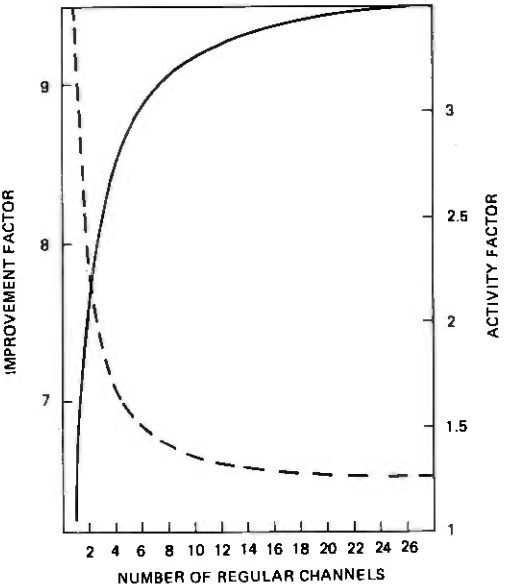


Fig. 16—Improvement and activity factors as functions of number of regular channels.

probability of service outage. Second, undetectable failures are usually the prime causes for increased outage probability and decreased improvement factor. If preventive maintenance is ever to be carried out, the hidden failures should be the principal targets. Third, the microcomputer is reliable as a protection switching controller. Although microprocessor system failures can cause false switching all by themselves, they contribute only a very small amount of the total outage if adequate self-checking is implemented. Reliability could be further improved by providing hardware interlock logic to guard against an insane microprocessor. For example, logic circuit can be provided in the TPSS to prevent the operation of an output switch whenever its input switch is inactive. Fourth, all the figures indicate that, around the various estimated failure rates of interest, the outage probabilities increase almost linearly with the failure rates. Thus there is no "preferred" range of failure rates that any equipment should be designed to. Only the sensitivities of the outage probabilities to the various estimates are different. Fifth, for any TPSS, the implementation of a 1×1 protection plan should be studied carefully. Even if there is improvement in the outage probability due to equipment failure, the increased activity will generate more maintenance-induced outages, not to mention increased costs.

The above comments do not apply in line protection switching systems, which have much higher regular channel failure rates because of the cascaded repeaters. Finally, Fig. 17 suggests one more consideration in determining the length of a line protection switching span. The failure

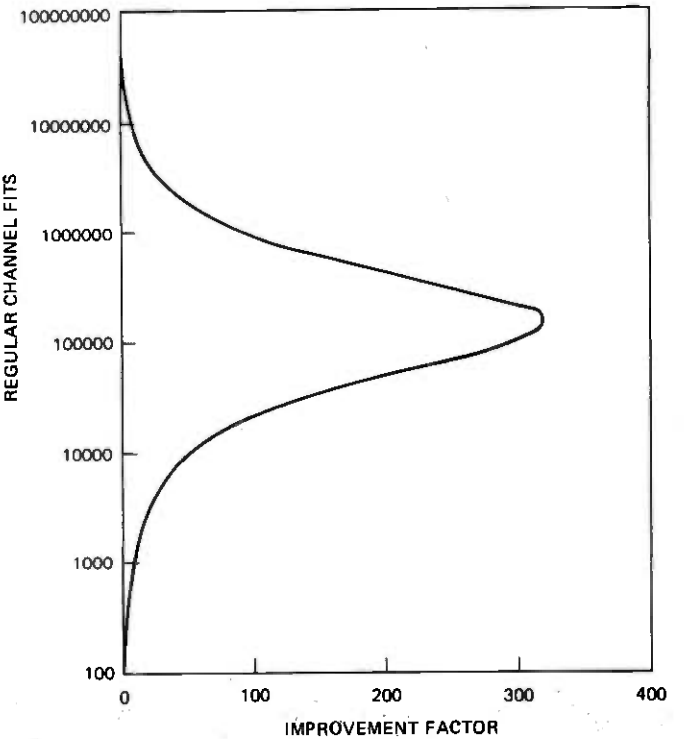


Fig. 17—Regular channel failure rates as functions of improvement factor.

rate of the line should preferably not fall into the decreasing region of its improvement factor. The last two points are obvious and interesting protection switching behavior patterns which seem not to have been explictly pointed out before.

APPENDIX A

This appendix discusses microprocessor self-test algorithms whose purpose is to generate alarms as early as possible to initiate maintenance actions. The test should be exhaustive but should not require too much additional program memory. An 8-bit microprocessor is used in the TPSS application.

When the power is turned on, the microprocessor immediately performs a thorough RAM check. Static RAMs are used, so there is no pattern sensitivity problem. The checking algorithm is to write the least-significant 8-address bits of each RAM byte into that specific RAM location. After all RAM locations are loaded, the contents of each byte are compared with its least-significant 8-bit address. After a byte is checked, its contents are complemented and checked again. The complemented contents will remain in those bytes already checked. This algorithm is

able to detect any bit, any data pin, and any combination of address pins stuck to zero or one. It can also discover data and address lines shorted together. Thus most RAM failures can be detected.

The ROMs are checked immediately following the RAM check. Two consecutive bytes in each ROM are reserved for self-test. One byte is used for parity check and the other for short-circuits in address and data lines. The microprocessor reads out every byte in the ROM and performs a cumulative odd parity check through an exclusive-OR operation on each bit. It will be seen first that, as far as independent ROM bit failures are concerned, it is adequate to use only one byte to check the parity of all ROMs no matter how many ROMs are used in the system. Let ℓ be the number of ROM bytes (excluding the reserved checking byte) used in the system and ϵ be the probability of a ROM bit failure. The probability of having parity violations is $1 - (1 - p)^8$, where p is[6]

$$\left[\frac{1 - (1 - 2\epsilon)^{\ell}}{2} \times (1 - \epsilon) + \frac{1 + (1 - 2\epsilon)^{\ell}}{2} \times \epsilon\right].$$

The probability of having bit errors is simply $1 - (1 - \epsilon)^{(\ell+1)\times 8}$. For $\ell\epsilon \ll 1$, both probabilities can be approximated by $8 \times (\ell + 1) \times \epsilon$. Thus the single byte parity check is adequate when $\ell\epsilon \ll 1$. It can be seen below that this condition is always valid in practice. Since the experimental failure rate of the 1K-byte EROM is 300 FITS, the failure rate of each bit cannot be more than $300/(8 \times 1024) \approx 0.037$ FIT. If a ROM failure can be discovered in 24 hours, then $\epsilon < 10^{-9}$. The number ℓ is limited by the microprocessor addressing capability which is 64K. Therefore, $\ell\epsilon \ll 1$. The reason that one parity byte is used in each ROM is to detect address and data lines stuck to one or zero. Since the ROM has a capacity equal to a power of 2, a stuck output looks like an even number of ones or zeros and violates the odd parity. A stuck address will cause half the bytes to be read twice and again violate the odd parity.

The contents of the bits of the other byte used for self-test are alternating ones and zeros. When this byte is read, short-circuits in data lines are detected. If this byte is located at an address whose 10 least-significant address bits are alternating ones and zeros, reading this byte will most likely detect short-circuits among these address lines. The probability is very small that within the same ROM another byte which also contains alternating ones and zeros is read because of shorted address lines. To detect some of the short-circuits in the remaining six most significant address lines, complemented numbers are stored in these checking bytes according to their address parities. Each ROM can select one of two hexidecimal numbers, AA or 55, to store at one of two addresses. For the first ROM with 0000 starting address, the two addresses are 0155 and 02AA.

The two consecutive checking bytes must be preceded by a jump or

branch instruction to bypass them in normal program execution. It is obvious that, if a single parity checking byte is located at an address with alternating ones and zeros, it alone can detect all ROM failures mentioned above except shorted data lines. It is sometimes possible to make use of the opcode and the operand of the jump or branch instruction to check the shorted data lines. If any failure occurs in the first ROM where the checking program is stored, the failure cannot always be detected. Duplicating the first ROM may be a possible solution.

After the two memory tests, a few instructions are exercised to test the CPU. Then the microprocessor starts executing the main program. Under normal circumstances, the program never comes back to the above RAM, ROM, and CPU tests. Different checks are performed in the main program. To avoid delaying the program execution, only distributed checks on the memory system are made. For example, in going through a program loop, only one RAM byte is tested and only one ROM exclusive OR is taken. However, the ROM check uses the same algorithm discussed above. The RAM check uses alternating ones and zeros which detect only shorted data lines and stuck bits because the exhaustive RAM check discussed before will destroy the temporary data stored, in addition to requiring long execution time. After each cycle of the nonexhaustive RAM check, an additional test[7] is made. Zeros are stored in the first RAM byte. Ones are stored only in RAM bytes with addresses 2^i, $i = 1,2,\cdots$. Every time all ones are loaded into an address, the contents of the first all-zero byte are also checked. The check is also distributed so as not to delay normal program execution. Most remaining RAM failures can be discovered by this additional test.

The effectiveness of the two RAM checking algorithms discussed above is similar. The first one used when turning on the power requires fewer steps and is faster. The second one does not destroy any temporary data because every check involves at most two RAM bytes (the first byte and the 2^ith byte) whose contents can be temporarily stored into CPU registers.

No CPU check is performed in the main program. A restarting sanity timer is employed to detect CPU failures. Under normal operation, the program retriggers the timer at durations shorter than the length of the timer. If the timer times out, an alarm is generated and the microprocessor system will go through its power on restart cycle again. The restarting sanity timer detects complete CPU failures. It can sometimes catch other CPU failures (for example, program counter skipping). It also reduces the damages that are caused by power transients because it restarts the system. RAM failures sometimes cause the timer to time out. ROM failures have similar effects but are more difficult to be self-detected. Output failures can only be detected by reading back the output bits immediately after each output operation.

APPENDIX B

This appendix derives the probabilities of outage and activity with and without protection switching. Figure 1 shows the configuration for a $1 \times n$ protection switching system in each direction of transmission. The microprocessor is responsible for the switching actions of $2n$ regular channels. The unprotected system has only the regular transmission channels plus pilot detectors for alarm.

The events of interests in deriving the outage probabilities are

S: service outage without protection switching.
S_P: service outage with protection switching.
G_1: all regular channels are good.
G_2: both protection channels are good.
G_3: all regular detectors are good.
G_4: all through switches are good.
G_5: all substitute switches are good.
G_6: the microprocessor system is good.
G_7: all output switches are good.

The events G_i's are assumed to be statistically independent. Their probabilities are given by

$$P\{G_1\} = p_r^{2n}$$
$$P\{G_2\} = p_p^2$$
$$P\{G_3\} = p_d^{2n}$$
$$P\{G_4\} = p_t^{2n}$$
$$P\{G_5\} = p_s^{2n}$$
$$P\{G_6\} = p_m = p_c p_e p_a$$
$$P\{G_7\} = p_0^{2n},$$

where the notations are defined in Table I. The symbol q with appropriate subscripts is defined to be $1 - p$ with the same subscript. Let $\overline{G}_i$ be the complement of G_i and g be the joint events of the G_i's with subscripts denoting the complemented events. For instance,

$$g_0 = G_1 G_2 G_3 G_4 G_5 G_6 G_7$$

and

$$g_{35} = G_1 G_2 \overline{G}_3 G_4 \overline{G}_5 G_6 G_7.$$

If these events represent all the possible failure modes of the system, then

$$P\{S_P\} = P\{S_P g_0\} + P\{S_P g_1\} + \cdots + P\{S_P g_6\} + P\{S_P g_7\} + P\{S_P g_{12}\} + \cdots + P\{S_P g_{234567}\} + P\{S_P g_{1234567}\}. \quad (1)$$

There are a total of 2^7 terms in (1). Half the terms involve the event $\overline{G}_7$, which generates service outage regardless of the other events. Therefore,

$$P\{S_P\} = 1 - p_0^{2n} + P\{S_P g_0\} + \cdots + P\{S_P g_6\} + P\{S_P g_{12}\} + \cdots + P\{S_P g_{23456}\} + P\{S_P g_{123456}\}. \quad (2)$$

The 2^6 unknown terms in (2) are to be evaluated. Since the derivations of each term are very similar, only the details in obtaining the more involved $P\{S_P g_{1345}\}$ and $P\{S_P g_{26}\}$ will be given. From the definition of conditional probability,

$$\begin{aligned} P\{S_P/g_{1345}\} &= P\{S_P/g_{1345}, \text{ three or more channel failures}\} \\ &\quad \cdot P\{\text{three or more channel failures}/g_{1345}\} \\ &\quad + P\{S_P/g_{1345}, \text{ two channel failures}\}P\{\text{two channel failures}/g_{1345}\} \\ &\quad + P\{S_P/g_{1345}, \text{ one channel failure}\}P\{\text{one channel failure}/g_{1345}\}. \quad (3) \end{aligned}$$

It is obvious that two protection channels cannot protect three failures; hence

$$P\{S_P/g_{1345}, \text{ three or more channel failures}\} = 1.$$

The joint event of three or more regular channel failures and $\overline{G}_1 G_2 \overline{G}_3 \overline{G}_4 \overline{G}_5 G_6 G_7$ has the conditional probability

$$P\{\text{three or more channel failures}/g_{1345}\} = \frac{[1 - p_r^{2n} - 2np_r^{2n-1}q_r - n(2n-1)p_r^{2(n-1)}q_r^2] \times p_p^2(1 - p_d^{2n})(1 - p_t^{2n})(1 - p_s^{2n})p_m p_0^{2n}}{P\{g_{1345}\}}. \quad (4)$$

The second term in (3) will be evaluated next. The various events will be abbreviated by their initials after their full names are introduced; e.g., tcf represents two channel failures.

$$\begin{aligned} P\{S_P/g_{1345}, \text{ tcf}\} &= P\{S_P/g_{1345}, \text{tcf, both failures in the same direction of transmission}\} \cdot P\{\text{both failures in the same direction of transmission}/g_{1345}, \text{tcf}\} \\ &\quad + P\{S_P/g_{1345}, \text{tcf, one failure in each direction}\} \cdot P\{\text{one failure in each direction}/g_{1345}, \text{tcf}\} \\ &= 1 \cdot \{n(n-1)p_r^{2(n-1)}q_r^2 p_p^2(1 - p_d^{2n})(1 - p_t^{2n})(1 - p_s^{2n})p_m p_0^{2n}\}/ \\ &\quad P\{g_{1345}, \text{tcf}\} + P\{S_P/g_{1345}, \text{tcf, one failure in each direction}\} \cdot P\{\text{ofied}/g_{1345}, \text{tcf}\}. \quad (5) \end{aligned}$$

Equation (5) follows because one protection channel cannot protect two failures in the same direction of transmission. The second term of (5) gives

$$\begin{aligned} P\{S_P/g_{1345}, \text{tcf, ofied}\} &= P\{S_P/g_{1345}, \text{tcf, ofied, two associated detectors are not both good}\} \cdot P\{\text{two associated} \end{aligned}$$

$$\text{detectors are not both good}/g_{1345},\text{tcf,ofied}\}$$
$$+ P\{S_P/g_{1345},\text{tcf,ofied, two associated detectors good}\}$$
$$\cdot P\{\text{two associated detectors good}/g_{1345},\text{tcf, ofied}\}$$
$$= 1 \cdot [n^2 p_r^{2n-2} q_r^2 p_p^2 (1 - p_d^2)(1 - p_t^{2n})(1 - p_s^{2n}) p_m p_0^{2n}] /$$
$$P\{g_{1345},\text{tcf,ofied}\} + P\{S_P/g_{1345},\text{tcf,ofied,tadg}\} \cdot P\{\text{tadg}/g_{1345},\text{tcf,ofied}\}. \quad (6)$$

$$P\{S_P/g_{1345},\text{tcf,ofied,tadg}\} = P\{S_P/g_{1345},\text{tcf,ofied,tadg,}$$
$$\text{both associated substitute switches good}\} \cdot P\{\text{both}$$
$$\text{associated substitute switches good}/g_{1345},\text{tcf,ofied,tadg}\}$$
$$+ 1 \cdot [n^2 p_r^{2n-2} q_r^2 p_p^2 p_d^2 (1 - p_d^{2n-2})(1 - p_t^{2n})(1 - p_s^2) p_m p_0^{2n}] /$$
$$P\{g_{1345},\text{tcf,ofied,tadg}\}. \quad (7)$$

$$P\{S_P/g_{1345},\text{tcf,ofied,tadg,bassg}\} = P\{S_P/g_{1345},\text{tcf,ofied,}$$
$$\text{tadg,bassg, both associated through switches good}\}$$
$$\cdot P\{\text{both associated through switches good}/g_{1345},\text{tcf,ofied,}$$
$$\text{tadg,bassg}\} + P\{S_P/g_{1345},\text{tcf,ofied,tadg,bassg, not both}$$
$$\text{through switches good}\} \cdot P\{\text{not both through switches}$$
$$\text{good}/g_{1345},\text{ttcf,ofied,tadg,bassg}\}$$
$$= 1 \cdot [n^2 p_r^{2n-2} q_r^2 p_p^2 p_d^2 (1 - p_d^{2n-2}) p_t^2 (1 - p_t^{2n-2}) p_s^2 (1 - p_s^{2n-2}) p_m p_0^{2n}] /$$
$$P\{g_{1345},\text{tcf,oefied,tadg,bassg}\} + P\{S_P/g_{1345},\text{tcf,ofied,tadg,}$$
$$\text{bassg, nbtsg}\} \cdot P\{\text{nbtsg}/g_{1345},\text{tcf,ofied,tadg,bassg}\}. \quad (8)$$

For the first term in (8), it is known that not all through switches are good because of $\overline{G}_4$. The outage probability is one because if the two failed channels have good through switches, the rest of the through switches must have failure. Finally,

$$P\{S_P/g_{1345},\text{tcf,ofied,tadg,bassg,nbtsg}\} = P\{S_P/g_{1345},\text{tcf,}$$
$$\text{ofied,tadg,bassg,nbtsg, no other through switch failure}\}$$
$$\cdot P\{\text{no other switch failure}/g_{1345},\text{tcf,ofied,tadg,}$$
$$\text{bassg,nbtsg}\} + P\{S_P/g_{1345},\text{tcf,ofied,tadg,bassg,nbtsg, other}$$
$$\text{through switch failure}\} \cdot P\{\text{other through switch}$$
$$\text{failure}/g_{1345},\text{tcf,ofied,tadg,bassg,nbtsg}\}$$
$$= 0 + [n^2 p_r^{2n-2} q_r^2 p_p^2 p_d^2 (1 - p_d^{2n-2})(1 - p_t^2)(1 - p_t^{2n-2})$$
$$\cdot p_s^2 (1 - p_s^{2n-2}) p_m p_0^{2n}] / P\{g_{1345},\text{tcf,ofied,tadg,bassg,nbtsg}\}. \quad (9)$$

In (9), the first conditional outage probability is zero because all the failures are protected by the two protection channels. The above derivations illustrate one of the basic approaches. Each event and its complement are assumed until the conditional probability of outage is either one or zero.

The third term in (3) is similarly derived.

$$P\{S_P/g_{1345},\text{ocf}\} = P\{S_P/g_{1345},\text{ocf, associated detector bad}\}$$

$$\cdot P\{\text{associated detector bad}/g_{1345},\text{ocf}\} + P\{S_P/g_{1345},\text{ocf, associated detector good}\} \cdot P\{\text{associated detector good}/g_{1345},\text{ocf}\}$$
$$= 1 \cdot [2np_r^{2n-1}q_r p_p^2 q_d(1-p_t^{2n})(1-p_s^{2n})p_m p_0^{2n}]/P\{g_{1345},\text{ocf}\} + P\{S_P/g_{1345},\text{ocf,adg}\}P\{\text{adg}/g_{1345},\text{ocf}\}. \quad (10)$$

$$P\{S_P/g_{1345},\text{ocf,adg}\} = P\{S_P/g_{1345},\text{ocf,adg, associated substitute switch good}\} \cdot P\{\text{associated substitute switch good}/g_{1345},\text{ocf,adg}\} + 1 \cdot [2np_r^{2n-1}q_r p_p^2 p_d(1-p_d^{2n-1}) \times (1-p_t^{2n})q_s p_m p_0^{2n}]/P\{g_{1345},\text{ocf,adg}\}. \quad (11)$$

$$P\{S_P/g_{1345},\text{ocf,adg,assg}\} = P\{S_P/g_{1345},\text{ocf,adg,assg, one other through switch bad}\} \cdot P\{\text{one other through switch bad}/g_{1345},\text{ocf,adg,assg}\} + 1 \cdot \{2np_r^{2n-1}q_r p_p^2 p_d(1-p_d^{2n-1})[1-p_t^{2n-1}-(2n-1)p_t^{2n-2}q_t] \cdot p_s(1-p_s^{2n-1})p_m p_0^{2n}\}/P\{g_{1345},\text{ocf,adg,assg}\}. \quad (12)$$

Equation (12) indicates that the status of the through switch associated with the failed regular channel has no effect on the outage probability.

$$P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb}\} = P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb, bad through switch in other direction of transmission}\} \cdot P\{\text{bad through switch in other direction}/g_{1345},\text{ocf,adg,assg,ootsb}\} + 1 \cdot [2np_r^{2n-1}q_r p_p^2 p_d(1-p_d^{2n-1})p_t^n(n-1)p_t^{n-2}q_t p_s \times (1-p_s^{2n-1})p_m p_0^{2n}]/P\{g_{1345},\text{ocf,adg,assg,ootsb}\}. \quad (13)$$

$$P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb,btsiod}\} = P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb,btsiod, bad switch has good detector}\} \cdot P\{\text{bad switch has good detector}/g_{1345},\text{ocf,adg,assg,ootsb,btsiod}\} + 1 \cdot [2np_r^{2n-1}q_r p_p^2 p_d q_d n p_t^{2n-2}q_t p_s(1-p_s^{2n-1})p_m p_0^{2n}]/P\{g_{1345},\text{ocf,adg,assg,ootsb,btsiod}\}. \quad (14)$$

$$P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb,btsiod,bshgd}\} = P\{S_P/g_{1345},\text{ocf,adg,assg,ootsb,btsiod,bshgd, corresponding substitute switch bad}\} \cdot P\{\text{corresponding substitute switch bad}/g_{1345},\text{ocf,adg,assg,ootsb,btsiod,bshgd}\} + 0 \cdot P\{\text{corresponding substitute switch good}/g_{1345},\text{ocf, adg,assg,ootsb,btsiod,bshgd}\}$$
$$= 1 \cdot [2np_r^{2n-1}q_r p_p^2 p_d^2(1-p_d^{2n-2})np_t^{2n-2}q_t p_s q_s p_m p_0^{2n}]/P\{g_{1345},\text{ocf,adg,assg,ootsb,btsiod,bshgd}\}. \quad (15)$$

From (3) through (15),

$$P\{S_P g_{1345}\} = p_p^2 p_m p_0^{2n}\{(x+x_3)(1-p_d^{2n})(1-p_t^{2n})(1-p_s^{2n})$$

$$+ x_1[q_d(1-p_t^{2n})(1-p_s^{2n}) + p_d(1-p_d^{2n-1})(1-p_t^{2n})q_s$$
$$+ p_d(1-p_d^{2n-1})\cdot[1-p_t^{2n-1}-(2n-1)p_t^{2n-2}q_t]p_s(1-p_s^{2n-1})$$
$$+ p_d(1-p_d^{2n-1})\cdot(n-1)p_t^{2n-2}q_tp_s(1-p_s^{2n-1})$$
$$+ p_dq_dnp_t^{2n-2}q_tp_s(1-p_s^{2n-1}) + p_d^2(1-p_d^{2n-2})np_t^{2n-2}q_tp_sq_s]$$
$$+ x_4[(1-p_d^2)(1-p_t^{2n})(1-p_s^{2n}) + p_d^2(1-p_d^{2n-2})(1-p_t^{2n})(1-p_s^2)$$
$$+ p_d^2(1-p_d^{2n-2})p_t^2(1-p_t^{2n-2})p_s^2(1-p_s^{2n-2}) + p_d^2(1-p_d^{2n-2})$$
$$\cdot(1-p_t^2)(1-p_t^{2n-2})p_s^2(1-p_s^{2n-2})]], \quad (16)$$

where

$$x_1 = 2np_r^{2n-1}q_r$$
$$x_2 = 1 - p_r^{2n} - 2np_r^{2n-1}q_r$$
$$x_3 = 1 - p_r^{2n} - 2np_r^{2n-1}q_r - n(2n-1)p_r^{2n-2}q_r^2$$
$$x_4 = n^2p_r^{2n-2}q_r^2$$
$$x = n(n-1)p_r^{2n-2}q_r^2.$$

To evaluate $P\{S_Pg_{26}\}$, the events

H_1: CPU is good
H_2: ROMs are good
H_3: RAMs are good

will be considered separately. Let h represent joint events similar to those for g, for example, $h_2 = H_1\overline{H}_2H_3$. As before,

$$P\{S_P/g_{26}\} = P\{S_P/g_{26}, \text{both protection channels bad}\}P\{\text{both protection channels bad}/g_{26}\} + P\{S_P/g_{26}, \text{one protection channel bad}\}P\{\text{one protection channel bad}/g_{26}\}P\{S_P/g_{26},\text{bpcb}\}$$
$$= P\{S_P/g_{26},\text{bpcb},h_1\}P\{h_1/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_2\} \times P\{h_2/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_3\}P\{h_3/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_{12}\}P\{h_{12}/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_{13}\}$$
$$\times P\{h_{13}/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_{23}\}P\{h_{23}/g_{26},\text{bpcb}\} + P\{S_P/g_{26},\text{bpcb},h_{123}\}P\{h_{123}/g_{26},\text{bpcb}\}. \quad (17)$$

The microprocessor operation is so complicated that simplifying assumptions have to be made before (17) can be further evaluated. There are two kinds of CPU failures. The first kind is a partial failure which may not be detected by the self-checking method discussed in Appendix A. For instances, program counter skipping and one CPU transistor failure within the CPU may not always be detectable. This partial failure may generate false switching and result in service outage. The second kind is a complete failure, and the CPU operation stops altogether. No false switching will be made in this case, and the sanity timer will detect the

failure immediately. It is assumed that partial failures accounts for 20 percent of the total CPU failures.

When the CPU is partially failed, it executes the contents of the ROMs insanely. Every "instruction" has a finite probability of generating a false switching. The TPSS software contains approximately 4000 bytes of which 100 can be I/O instructions. Out of the $2n + 5$ hardware addresses, $2n$ have outputs controlling the switches. If a correct parity bit and an appropriate output switch control bit are stored in the accumulator, an I/O instruction will operate the output switch. If the protection channels are bad, the operation of the output switch will generate service outage regardless of the status of the input switch. Thus the probability p_1 that any instruction will cause an outage is approximately

$$p_1 = \frac{100}{4000} \cdot \frac{1}{4} \cdot \frac{2n}{2n+5}.$$

When the protection channels are working, the same probability is now

$$p_2 = \frac{100}{4000} \cdot \frac{1}{8} \cdot \frac{2n}{2n+5}$$

because the input switch should be inactive for the false output switching to generate service outage. It is to be noted that false switching can also occur randomly if the 8-bit "instruction," the 16-bit "address," the parity bit, and the switch control bit happen to match the real instruction and address. This probability is of the order $2n/2^{26}$ and is negligible compared with p_1 and p_2. On the average, each instruction takes about 4 microseconds. Thus before restoration, about

$$n_1 = \frac{\mu_c \times 60 \times 60 \times 10^6}{4}$$

"instructions" are executed. The probability p_3 that an outage will occur is

$$\begin{aligned} p_3 &= p_1 + q_1 p_1 + \cdots + q_1^{n_1-1} p_1 \\ &= p_1 \frac{1 - q_1^{n_1}}{1 - q_1} \\ &= 1 - q_1^{n_1}. \end{aligned}$$

When the protection channels are good, the corresponding probability is

$$p_4 = 1 - q_2^{n_1}.$$

After a false switching, it is possible that insane CPU may deactivate the switch and restore service. It may also operate other output switches to generate additional service outages. These two conditional proba-

bilities are small. If they are ignored, the outage probability assuming partial CPU failure and bad protection channels is then $p_3 t/\mu_c$. If only one of the two protection channels is bad, let

$$p_5 = \frac{100}{4000} \cdot \frac{1}{8} \cdot \frac{n}{2n+5} + \frac{100}{4000} \cdot \frac{1}{4} \cdot \frac{n}{2n+5}.$$

The outage probability is $p_6 t/\mu_c$ where

$$p_6 = 1 - q_5^{n1}.$$

When a memory failure occurs, the program counter jumps to an arbitrary location. The initial effect is somewhat like that of a partially failed CPU. Experiments indicate that outage is unlikely to occur if it has not occurred during the initial period. Since 25 out of the 4000 bytes are used to activate the output switches in normal program operation, a jump to these bytes will cause a false switching. Therefore, the false switching probability is

$$p_7 = \frac{25}{4000} + p_1$$

or

$$p_8 = \frac{25}{4000} + p_2,$$

depending on whether the protection channels are bad or good. If only one of the two protection channels is bad, the probability is

$$p_9 = \frac{25}{4000} + p_5.$$

It will be assumed that all RAM failures can be detected. Most of the RAM bytes are used for stack. The effects of the ROM and the RAM failures are assumed to be identical, but their restoration times are different because not all ROM failures are self-detectable. When the CPU fails, memory failures are assumed to have no effect on the system. This makes the evaluation of the fourth, the fifth, and the last terms in (17) unnecessary once the first term is evaluated. It is further assumed that when there are both ROM and RAM failures, the trouble can be detected immediately. Given the previous assumption, then

$$\begin{aligned} P\{S_P/g_{26},\text{bpcb},h_1\} &= P\{S_P/g_{26},\text{bpcb},h_1,\text{ complete failure}\}P\{\text{complete failure}/g_{26},\text{bpcb},h_1\} \\ &\quad + P\{S_P/g_{26},\text{bpcb},h_1,\text{ partial failure}\}P\{\text{partial failure}/g_{26},\text{bpcb},h_1\} \\ &= 0 + \frac{p_3 t}{\mu_c} \frac{p_{10} q_p^2 0.2 q_c p_e p_a}{P\{g_{26},\text{bpcb},h_1\}}, \end{aligned}$$

where

$$p_{10} = (p_r p_d p_t p_s p_0)^{2n}. \tag{18}$$

$$P\{S_P/g_{26},\text{bpcb},h_2\} = \frac{p_7 t}{\mu_e}\frac{p_{10}q_p^2 p_c q_e p_a}{P\{g_{26},\text{bpcb},h_2\}}$$

$$P\{S_P/g_{26},\text{bpcb},h_3\} = p_7\frac{p_{10}q_p^2 p_c p_e q_a}{P\{g_{26},\text{bpcb},h_3\}}$$

$$P\{S_P/g_{26},\text{bpcb},h_{23}\} = p_7\frac{p_{10}q_p^2 p_c q_e q_a}{P\{g_{26},\text{bpcb},h_{23}\}}.$$

Hence

$$P\{S_P,g_{26},\text{bpcb}\} = p_{10}\left\{\frac{p_3 t}{\mu_c}0.2\,q_c + \frac{p_7 t}{\mu_e}p_c q_e p_a + p_7 p_c q_a\right\}. \quad (19)$$

The expression $P\{S_P,g_{26},\text{opcb}\}$ can be similarly evaluated. Finally,

$$P\{S_P g_{26}\} = p_{10}\left\{2p_p q_p\left[\frac{p_6 t}{\mu_c}0.2q_c + \frac{p_9 t}{\mu_e}p_c q_e p_a + p_9 p_c q_a\right] + q_p^2\left[\frac{p_3 t}{\mu_c}0.2q_c + \frac{p_7 t}{\mu_e}p_c q_e p_a + p_7 p_c q_a\right]\right\}. \quad (20)$$

After deriving (16) and (20), the remaining terms in (2) are easy to obtain. They will not be given here. Thus the outage probability with protection switching $P\{S_p\}$ is obtained from (2). It should be emphasized that, because there are hidden failures, multiple equipment failures cannot be neglected in evaluating the various terms in (2). In fact, the term that contributes the most to the outage probability is $P\{S_p g_{135}\}$, which involves both of the undetectable failures (detector and substitute switch).

Since the detectors used to generate alarms do not affect signal transmission, the outage probability without protection switching is simply

$$P\{S\} = 1 - p_r^{2n}. \quad (21)$$

The improvement factor is

$$\text{IF} = \frac{P\{S\}}{P\{S_P\}}. \quad (22)$$

Next, the probabilities of activity with and without protection switching will be considered. The additional events of interest are

A: activity without protection switching
A_P: activity with protection switching
G_5: protection detectors are good.

G_5 is redefined because protection detector failures generates maintenance activities, but the hidden substitute switch failures are assumed to cause no activity. To calculate the probability of activity with protection switching, notice that whenever $\overline{G}_1$, $\overline{G}_4$, and $\overline{G}_7$ occur, there will definitely be maintenance activity. Furthermore, the events $\overline{G}_2$ and $\overline{G}_5$ are detectable when G_6 is true. Therefore

$$P\{A_P\} = 1 - (p_r p_t p_0)^{2n} + (p_r p_t p_0)^{2n} p_m (1 - p_p^2 p_D^2) + P\{A_P g_0\} + P\{A_P g_3\} + P\{A_P g_6\} + P\{A_P g_{26}\} + P\{A_P g_{36}\} + P\{A_P g_{56}\} + P\{A_P g_{236}\} + P\{A_P g_{256}\} + P\{A_P g_{356}\} + P\{A_P g_{2356}\}. \quad (23)$$

In (23), $P\{A_P g_0\}$ is always zero. The last seven terms are negligible compared with $P\{A_P g_3\}$ and $P\{A_P g_6\}$. It is assumed that 10 percent of the CPU and the ROM failures will not generate alarm. The derivation of $P\{A_P g_6\}$ is similar to that of (17). For example,

$$P\{A_P/g_6 h_1\} = P\{A_P/g_6, h_1, \text{undetectable failure}\} P\{\text{undetectable failure}/g_6 h_1\} + P\{A_P/g_6, h_1, \text{detectable failure}\} P\{\text{detectable failure}/g_6, h_1\}$$

$$= 0 + \frac{t}{\mu_c} \cdot (p_r p_d p_t p_0)^{2n} (p_p p_D)^2 \cdot 0.9 \cdot q_c p_e p_a / P\{g_6 h_1\}.$$

Thus,

$$P\{A_P g_6\} = (p_r p_d p_t p_0)^{2n} (p_p p_D)^2 \left[0.9 \frac{t}{\mu_c} q_c + 0.9 \frac{t}{\mu_e} p_c q_e p_a + p_c q_a \right]. \quad (24)$$

If it is assumed that, when a detector fails, the probability that it is stuck to an ON state is 0.25, then

$$P\{A_P/g_3\} = P\{A_P/g_3, \text{one detector bad}\} P\{\text{one detector bad}/g_3\} + \cdots + P\{A_P/g_3, 2n \text{ detectors bad}\} P\{2n \text{ detectors bad}\}. \quad (25)$$

The ith term in (25) is

$$P\{A_P/g_3, \text{idb}\} = P\{A_P/g_3, \text{idb, all bad detectors on}\} P\{\text{all bad detectors on}/g_3, \text{idb}\} + P\{A_P/g_3, \text{idb, some bad detectors off}\} \cdot P\{\text{some bad detectors off}/g_3, \text{idb}\}$$

$$= 0 + \frac{t}{\mu_d} \cdot \frac{(p_r p_t p_0)^{2n} (p_p p_D)^2 p_m \binom{2n}{i} p_d^{2n-i} q_d^i (1 - 0.25^i)}{P\{g_3, \text{idb}\}}.$$

Therefore,

$$P\{A_P g_3\} = \frac{t}{\mu_d} p_m (p_p p_D)^2 (p_r p_t p_0)^{2n} \sum_{i=1}^{2n} p_d^{2n-i} q_d^i (1 - 0.25^i). \quad (26)$$

Equations (23) through (26) yield the probability of activity with protection switching $P\{A_P\}$. The probability of activity without protection switching $P\{A\}$ is simply

$$P\{A\} = 1 - p_r^{2n} + \frac{t}{\mu_b} p_r^{2n} \sum_{i=1}^{2n} \binom{2n}{i} p_b^{2n-i} q_b^i (1 - 0.25^i),$$

where

$$p_b = \frac{1}{1 + \lambda_d \mu_b}$$

and

$$\mu_b = \frac{1}{4\lambda_r}$$

is the detector restoration time without protection switching. The activity factor is given by

$$\text{AF} = \frac{P\{A_P\}}{P\{A\}}. \quad (27)$$

REFERENCES

1. "L5 Coaxial-Carrier Transmission System," B.S.T.J., *53*, No. 10 (December 1974), pp. 1897–2268.
2. J. A. Buzacott, "Markov Approach to Finding Failure Times of Repairable Systems," IEEE Trans. on Reliability, *R-19* (November 1970), pp. 128–134.
3. B. V. Gnedenko, Y. K. Belyayev, and A. D. Solovyev, *Mathematical Methods of Reliability Theory* (Russian orig. and English transl. edited by R. E. Barlow), New York: Academic Press, 1969.
4. I. Welber, H. W. Evans, and G. A. Pullis, "Protection of Service in the TD-2 Radio Relay System by Automatic Channel Switching," B.S.T.J., *34*, No. 3 (May 1955), pp. 473–510.
5. W. Y.-S. Chen, "Estimated Outage in Long-Haul Radio Delay Systems With Protection Switching," B.S.T.J., *50*, No. 4 (April 1971), pp. 1455–1485.
6. G. S. Fang, "Alarm Statistics of the Violation Monitor and Remover," B.S.T.J., *55*, No. 8 (October 1976), pp 1197–1217.
7. B. A. Zimmer, "Test Techniques for Circuit Boards Containing Large Memories and Microprocessor," IEEE Computer Society Conf. Proceedings, Semiconductor Test Symposium, Oct. 19 to 21, 1976, Cherry Hill, N. J.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

Offset Multireflector Antennas with Perfect Pattern Symmetry and Polarization Discrimination

By C. DRAGONE

(Manuscript received September 20, 1977)

Conditions are derived that are useful for designing reflector antennas with excellent cross-polarization discrimination. These conditions ensure circular symmetry and absence of cross-polarization everywhere in the far field of an antenna, provided a suitable feed such as a corrugated horn is employed. The spherical wave radiated by the fundamental mode of such a feed has circular symmetry around the axis, and it is everywhere free of cross-polarization. An arbitrary sequence of N confocal reflectors (hyperboloids, ellipsoids, paraboloids) is combined with such a feed. It is shown that it is always possible to ensure circular symmetry (and absence of cross-polarization) in the antenna far field by properly choosing the feed axis orientation. If the final reflector is a paraboloid, a simple geometrical procedure can be used. It is also shown that the asymmetry caused by an arbitrary number of reflections can always be eliminated by properly introducing an additional reflection. An application to the problem of producing a horizontal beam using a vertical feed is discussed. Two arrangements are described that may be useful for radio relay systems.

Use of orthogonal polarizations is often required in radio systems to double transmission capacity. Antennas providing good discrimination between the two polarizations are then needed. The main purpose of this paper is to derive and discuss certain conditions that ensure excellent discrimination. When two or more reflectors and a suitable feed are arranged in accordance with these conditions, the antenna far field has, in all directions, the same polarization of the feed excitation. Furthermore, its pattern has circular symmetry. The above conditions also minimize astigmatism, and for this reason they are also useful* in the design of multibeam antennas (with several feeds).

* This is the subject of an article being prepared.

I. INTRODUCTION

A suitable feed for the antennas considered here is realized by properly corrugating the walls of a circular horn.[1–4] The spherical wave radiated by the horn then has circular symmetry and, by placing the feed at the focus of a paraboloid, an antenna with circular symmetry in the far field is obtained, provided the paraboloid is centered around the feed axis. Furthermore, the polarization of the plane wave reflected by the paraboloid then coincides with that of the feed excitation.

However, in the centered configuration the reflected wave is in part blocked by the horn.* To avoid this, the horn axis can be offset as in Fig. 1, but unfortunately this causes asymmetry in the pattern after reflection, resulting in an undesired cross-polarized component.[5,6] The same behavior occurs if, instead of a paraboloid, an arbitrary reflector system with a single axis of revolution is used. In Fig. 1, the asymmetry of the reflected wave increases with the angle of incidence α of the ray corresponding to the horn axis. This particular ray will be called *principal ray*.

Although a single offset reflection always causes some asymmetry, it

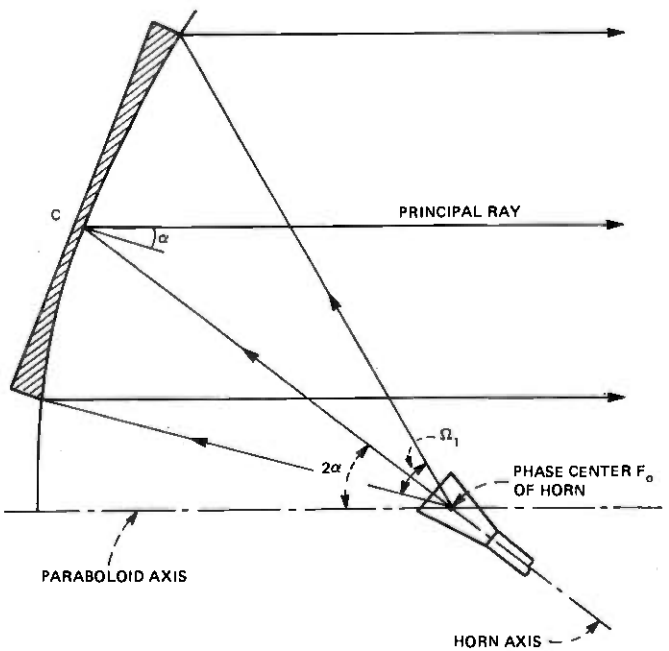


Fig. 1—The spherical wave radiated from F_0 by a corrugated feed is transformed by an offset paraboloid into a plane wave.

* This blockage impairs gain, side-lobes level, return loss, and cross-polarization discrimination.

is possible to combine two reflections with nonzero angles of incidence so as to ensure perfect symmetry after the two reflections.[7–10] In this paper we generalize and extend the results of Refs. 7 to 9 in several respects. First, the analysis here is not restricted to only two reflections, nor does it assume the final reflector is necessarily a paraboloid. Second, very simple conditions that guarantee symmetry after the final reflection are obtained. These conditions are shown to be direct consequences of a general principle of equivalence (see the appendix). Third, a general solution is given to the problem* of restoring the symmetry of a wave whose initial symmetry has been distorted by an arbitrary number of reflectors.

In Section III, two arrangements with excellent performance in cross-polarization are described. Both arrangements produce a horizontal beam using a vertical feed and may therefore be useful for microwave radio systems.

The following analysis is based on geometrical optics. Furthermore, the far field for the antennas of Figs. 12 and 13 is not derived in Section III, but it is important to note that the principle of equivalence of the following section allows the aperture field distribution for both antennas to be derived replacing the reflectors with a single paraboloid, centered around the feed axis. The aperture field distribution and far field of such a paraboloid are well known.[1–5] As pointed out at the beginning of this introduction, the entire aperture will be polarized in one direction if the feed is linearly polarized. The far field is thus free of cross-polarization, neglecting secondary effects such as edge diffraction.

II. THE EQUIVALENT REFLECTOR AND THE ORIENTATION OF ITS AXIS

Suppose a spherical wave from F_0, initially with symmetrical pattern, is successively reflected N times, using paraboloids, hyperboloids, and ellipsoids as shown in Fig. 2 for $N = 3$. The reflectors are properly arranged so that a spherical wave is produced after each reflection. Thus, if F_n is the focal point after the nth reflection, the nth reflector Σ_n transforms a spherical wave centered at F_{n-1} into a spherical wave centered at F_n. Note that some of the points $F_0, F_1, \cdots, F_N$ may be at ∞, in which case the corresponding spherical waves become plane waves. In Fig. 2, F_3 is at ∞, and therefore the last reflector is a paraboloid.

It is shown in the appendix that *such a sequence of confocal reflectors is always equivalent to a single reflector* which will be either an ellipsoid, a hyperboloid, or a paraboloid. This equivalent reflector produces, after a single reflection, the same reflected wave† as the given sequence of

* An interesting formulation of this problem is given in Ref. 10.

† Thus, if one considers the field distribution over a wavefront reflected by the equivalent reflector, it will coincide with the field distribution over the corresponding wavefront produced by the given sequence of reflectors.

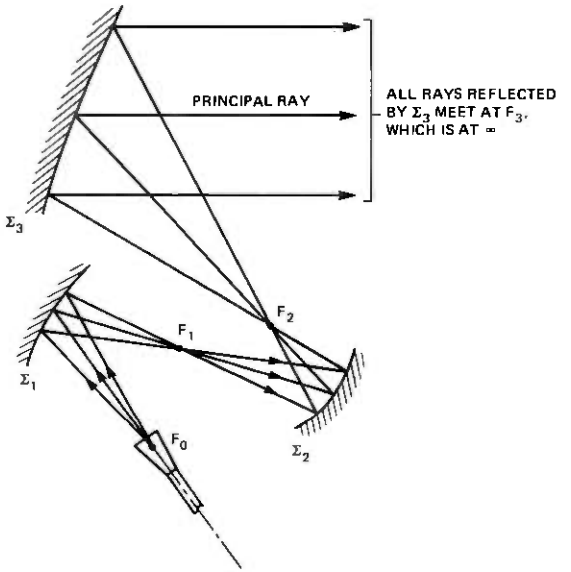


Fig. 2—The spherical wave from F_0 is transformed into a plane wave by three confocal reflectors. The nth reflector transforms the spherical wave from F_{n-1} into a spherical wave converging towards F_n.

reflectors. Thus, for the purpose of determining the properties of the reflected wave, one may replace the N reflectors with the equivalent reflector. This reflector has an axis of symmetry, which passes through F_0, and will be called the *equivalent axis*. It is clear that in order that the symmetry of the incident beam be preserved, *the principal ray must coincide with the equivalent axis*.*

2.1 The central rays, their closed path, and the equivalent axis

Consider first $N = 1$. Suppose the reflector Σ_1 and one of its foci, F_0, are given, but the exact location of the axis of Σ_1 is not known and must be found. Then one may proceed as follows. Let a ray from F_0 be reflected twice by Σ_1, as shown in Fig. 3, and let $\vec{s}$ and $\vec{s}''$ be the initial and final directions of the ray. Then, from Fig. 3,

$$\vec{s} = \vec{s}'' \tag{1}$$

only when the ray coincides with the axis. Thus, the axis can be found by searching for a ray that satisfies this condition. Note from Fig. 3 there are two such rays, with opposite directions.

Next consider $N > 1$. Since a confocal sequence of reflectors $\Sigma_1, \cdots, \Sigma_N$ is equivalent to a single reflector Σ_e, the above procedure is appli-

* Since one can travel along the equivalent axis in two opposite directions, two opposite orientations can be chosen for the principal ray.

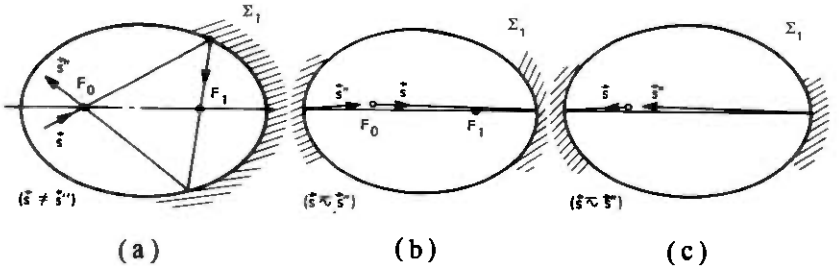


Fig. 3—The axis of Σ_1 is determined by varying $\vec{s}$ until $\vec{s} = \vec{s}''$.

cable also to this case. Thus, to determine the axis of Σ_e (equivalent axis), one must consider a ray from F_0, with initial directions $\vec{s}$. This ray must be reflected *twice* by Σ_e, and $\vec{s}$ must then be chosen so that $\vec{s}'' = \vec{s}$. Notice that the two reflections by Σ_e imply a total of $2N$ reflections in the original configuration. The first N reflections take place in the order $\Sigma_1, \cdots, \Sigma_N$, while the last N have the reverse order $\Sigma_N, \cdots, \Sigma_1$. The final ray passes again through F_0, with the same direction as the original ray. In Fig. 4a, $\vec{s} \neq \vec{s}''$. In Fig. 4b, on the other hand, condition (1) is satisfied, and therefore the ray through F_0 gives the correct orientation of the equivalent axis (and the principal ray for which symmetry is preserved).

Notice that if, after the above $2N$ reflections, the ray in Fig. 4a is reflected $2N$ more times it will not follow the same path of the first $2N$ reflections. In Fig. 4b, on the other hand, the path of the first $2N$ reflections is closed. This closed path, which determines the equivalent axis, will be called the *central path*. The two rays that proceed along the central path in opposite senses will be called the *central rays*.

We show next that condition (1) leads to a straightforward geometrical procedure for determining the equivalent axis when Σ_N is a paraboloid.

*2.2 The equivalent axis when the last reflector Σ_N is a concave paraboloid**

It is now shown that, when the last ellipsoid in Fig. 4a is replaced by a concave paraboloid, the final ray direction $\vec{s}''$ becomes independent of the initial direction $\vec{s}'$. This constant value of $\vec{s}''$ then gives the direction of the equivalent axis, which can thus be found straightforwardly.

Notice the path of Fig. 4a involves two successive reflections by the last ellipsoid Σ_N. Let ψ be the angle between the axis of Σ_N and the ray produced after the second reflection (see Fig. 5). The parameters of the ellipsoid Σ_N are now gradually modified, keeping the vertex V and the focus F_{N-1} fixed, increasing the distance between F_N and F_{N-1} until

* The following considerations apply also when Σ_N is a convex paraboloid, but this case is of little practical interest and will therefore be ignored.

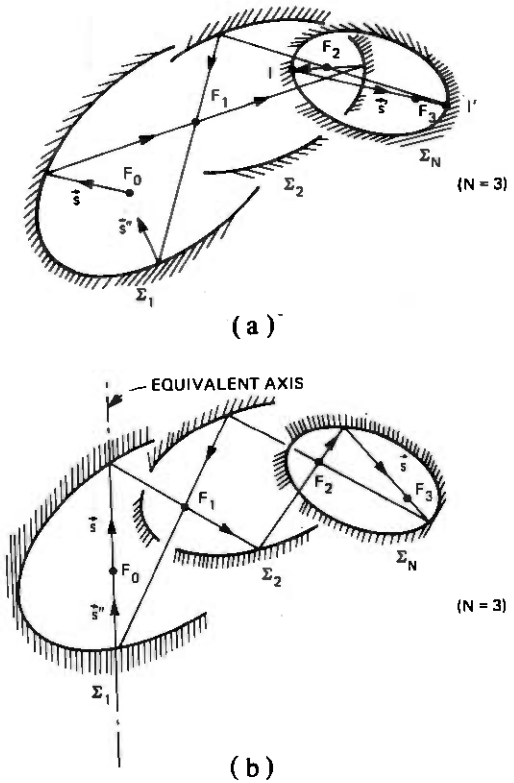


Fig. 4—(a) $2N$ successive reflections. (b) The central path. The equivalent axis through F_0 is obtained by varying in (a) the initial direction $\vec{s}$ until $\vec{s} = \vec{s}''$ as shown in (b).

$F_N \to \infty$. The ellipsoid then becomes a paraboloid with focus F_{N-1} and from the figure $\psi = 0$, which shows that

> If a ray from the focus F_{N-1} of a paraboloid is reflected twice by the paraboloid, so that the second reflection occurs at ∞, the final ray coincides with the paraboloid axis and it has the direction going from F_{N-1} towards the vertex V of the paraboloid. (2)

This implies that, when in Fig. 4 the last ellipsoid Σ_N is replaced by a paraboloid, the direction of $\vec{s}''$ becomes independent of $\vec{s}$, and it can be determined by tracing the ray $F_{N-1}V$ as shown in Fig. 6. The direction $\vec{s}''$ so obtained gives the equivalent axis, as one may verify considering a ray with *initial* direction given by the above value of $\vec{s}''$. One can see from Fig. 6 the path of this ray closes, after $2N$ reflections. Thus,

To obtain the equivalent axis of a sequence of $N - 1$ re-

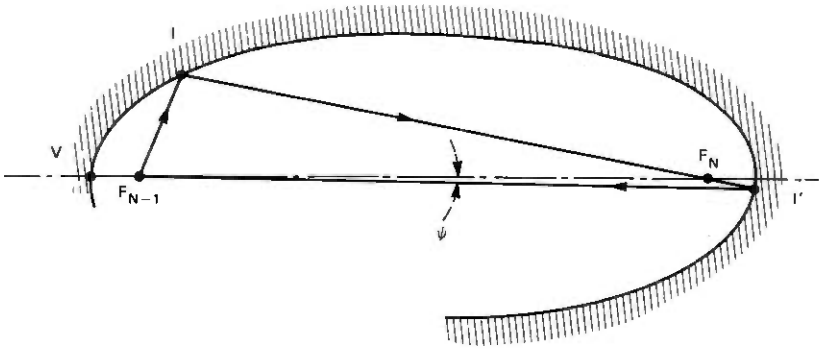


Fig. 5—As the distance of F_N from the other focus F_{N-1} is increased, keeping V and F_{N-1} fixed, the ellipsoid approaches a paraboloid with vertex V and focus F_{N-1}; for the ray reflected at I' one has $\psi \to 0$.

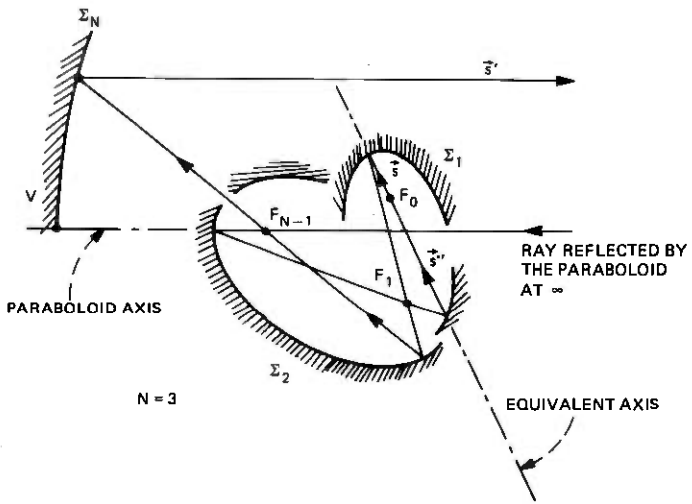


Fig. 6—By tracing from ∞ the path of the ray defined by the paraboloid axis one obtains after $N-1$ reflections the equivalent axis through F_0. If a symmetrical feed is placed at F_0, centered around the equivalent axis, a symmetrical pattern will be reflected by the paraboloid.

flectors $\Sigma_1, \Sigma_2, \cdots, \Sigma_{N-1}$ followed by a paraboloid Σ_N with focus F_{N-1} and vertex V, simply reflect $N-1$ times the ray $F_{N-1}V$ by $\Sigma_{N-1}, \Sigma_{N-2}, \cdots, \Sigma_1$. The final ray through F_0 is the equivalent axis and, therefore, the principal ray along which symmetry is preserved. (3)

As an example, consider $N = 2$, and assume the first reflector is not

a paraboloid.* Then four different arrangements are obtained depending on whether the first reflector is an ellipsoid or an hyperboloid, and is convex or concave. In each case (see Figs. 7 and 8), the equivalent axis† is determined by the intersection I' of the paraboloid axis with the first reflector. The equivalent axis is the line F_0I'. Note the axis of the paraboloid intercepts the first reflector Σ_1 in two points, but only one, I', is acceptable.‡ The acceptable point is the point of reflection of the ray F_1V. Since only one side of the surface Σ_1 is reflecting, only one of the above two points can be considered a point of reflection for the above ray.

From Figs. 7 and 8, since in all cases the equivalent axis and the paraboloid axis meet on Σ_1, the angles 2α and 2β giving their inclinations from the axis of Σ_1 are related,

$$\tan \alpha = m \tan \beta, \tag{4}$$

where m is the axial magnification of Σ_1 given by the distances of the reflector vertex V_0 from the two focal points F_0 and F_1,

$$m = \frac{|F_0V_0|}{|F_1V_0|}. \tag{5}$$

Note that if e is the eccentricity of the reflector, in Figs. 7 and 8,

$$m = \frac{e+1}{e-1}, \frac{e-1}{e+1}, \frac{e+1}{1-e}, \frac{1-e}{1+e}, \tag{6}$$

respectively. In Fig. 7 one has $e > 1$, whereas in Fig. 8, $0 < e < 1$.

In the two cases of Figs. 7a and 8a, eq. (4) is equivalent to eq. (1) of Ref. 9. In the other two cases, on the other hand, eq. (1) of Ref. 9 is not applicable [to obtain a correct relation, one has to replace α with β in eq. (1)].

Another useful relation, derived in the following section, is

$$\tan i = \frac{M}{1-M} \tan p. \tag{7}$$

It relates the angles of incidence i and p of the central ray on the two

* The case where Σ_1 is a paraboloid is treated in Section 2.6.

† That is, the beam orientation for which symmetry is preserved.

‡ Notice for the purpose of deriving the equivalent axis that the entire surfaces of the various ellipsoids, hyperboloids, and paraboloids must be considered to be reflecting. Thus, both branches of an hyperboloid must be considered. Of course, an actual antenna will use only certain sections of the various surfaces.

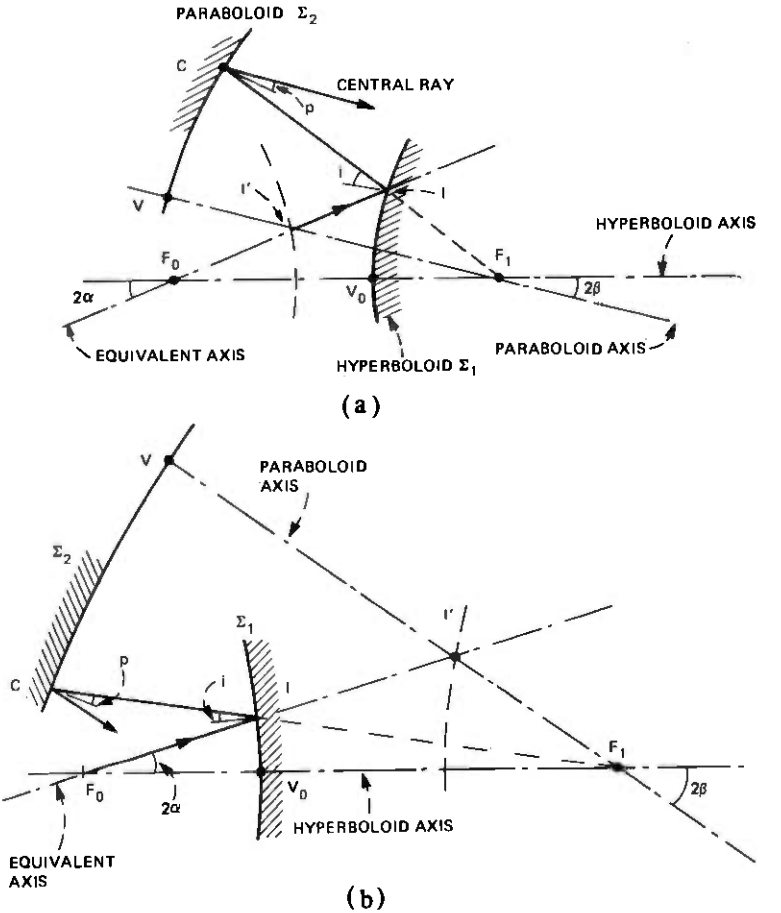


Fig. 7—How to determine the central path and the equivalent axis of a paraboloid combined in (a) with a convex hyperboloid and in (b) with a concave hyperboloid.

reflectors (see Figs. 7a and 8) to the magnification M, defined as

$$M = \pm\frac{|F_0 I|}{|I F_1|}, \tag{8}$$

I being the point of incidence of the central ray on the first reflector. In eq. (8) one has to take the positive sign when F_0 and F_1 are on opposite sides of the tangent plane at I, as in Fig. 8; otherwise, as in Fig. 8, $M <$ 0. The angles of incidence must be taken with opposite sign in Figs. 7a and 8, where the two reflections have opposite senses; in Fig. 7b, on the other hand, i and p have the same sign.

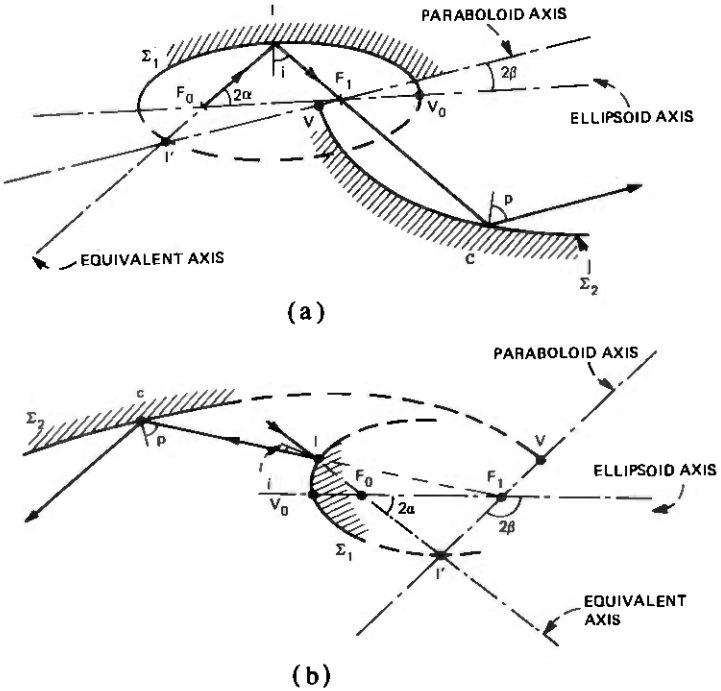


Fig. 8—How to determine the central path and the equivalent axis of a paraboloid combined in (a) with a concave ellipsoid and in (b), with a convex ellipsoid.

The magnification* M determines the ratio between the angular width Ω_0 of the beam incident as I and the width Ω_1 of the reflected beam. More precisely,† for small Ω_0,

$$M = \frac{\Omega_1}{\Omega_0}. \tag{9}$$

If M is specified, eq. (7) gives the angles of incidence i and p that result in a symmetrical beam after two reflections.

A very general relation, which reduces to eq. (7) in the particular case where Σ_N is a paraboloid, is derived in Section 2.4.

---

* Another important significance of M is that the paraxial focal length f_e, for any of the arrangements of Figs. 7 and 8, in the vicinity of the central ray, is $f_e = Mf_p$, where f_p is the paraboloid focal length $f_p = CF_1$; f_e has the significance that a small lateral displacement δs of a feed initially placed at F_0 will cause an angular displacement $\delta\theta = \delta s/f_e$ of the beam reflected by the paraboloid.

† Thus, if a beam of small angular width Ω_0 is transformed by a sequence of N reflectors with magnifications $M_1, \cdots, M_N$, the final beam has angular width

$$\Omega_1 = M_t \Omega_0,$$

where $M_t = M_1 M_2 \cdots M_N$.

2.3 Relations governing the reflections of a central ray by the first or the last reflector

The restriction that Σ_N must be a paraboloid is now removed. The closed path of the central ray in Fig. 4 involves two successive reflections by Σ_1. Consider these two reflections and assume for the moment Σ_1 is a concave ellipsoid as shown in Fig. 9a. The central ray in Fig. 9a first passes through F_1 with direction $\vec{a}$, it is successively reflected at I' and I, and it then passes again through F_1 with direction $\vec{c}$.

Let $2i$ and $2i'$ be the angles of the two reflections and M and M' the corresponding magnifications,

$$M = -\frac{\ell_1}{\ell_2}, M' = -\frac{\ell_1'}{\ell_2'}. \tag{10}$$

ℓ_1, ℓ_2, etc. being defined in Fig. 9a. Then, if $2\gamma = 2i + 2i'$ is the total angle of reflection (given by the angle between the final and initial rays $\vec{c}$ and $\vec{a}$) it is shown in Section A.3 of the appendix that

$$\tan i = \frac{M}{M-1} \tan \gamma \tag{11}$$

and

$$\tan i' = \frac{1}{1-M'} \tan \gamma. \tag{12}$$

Thus, if the parameters (M,i, or M',i') of either reflection are given, the total angle of reflection for a central ray can be calculated. Note that eqs. (11) and (12) apply also to the two consecutive reflections of the central ray by the last reflector Σ_N.

In Fig. 9a, the reflector Σ_1 is a concave ellipsoid, but eqs. (11) and (12) are valid also if Σ_1 is an hyperboloid or is concave, as shown in Figs. 9b, c, and d. Note in cases 9c and 9d the central ray is first reflected at I', then passes through the point at ∞ and is then reflected again at I. Figs. 7a,b and 8a,b correspond to Figs. 9b, 9c, 9a, and 9d, respectively.

2.4 How to arrange two reflectors

Consider Fig. 10a showing a principal ray from F_0 reflected by two reflectors Σ_1 and Σ_2. We wish to show that, in order that this ray be a central ray, i.e., that symmetry be preserved after these two reflections, their parameters M, M', i, and i' must satisfy the condition

$$\tan i = M \frac{1-M'}{1-M} \tan i'. \tag{13}$$

Consider the ray reflected by Σ_1. Let this ray be reflected *twice* by Σ_2, and then again *twice* by Σ_1, as in Fig. 10b. If 2γ denotes the total angle of the first two reflections by Σ_2 and $2\gamma'$ the angle of the other two re-

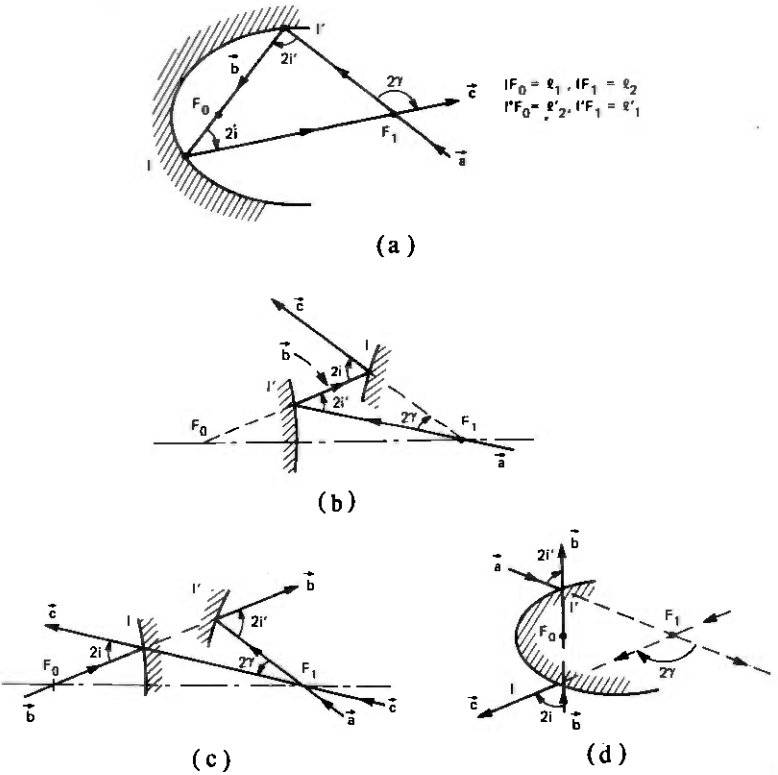


Fig. 9—Two successive reflections. (a) By concave ellipsoid. (b) By convex hyperboloid. (c) By concave hyperboloid. (d) By convex ellipsoid.

flections, one must have

$$2\gamma + 2\gamma' = 2\pi, \tag{14}$$

if the path of the ray is to close (which is necessary for it to be a central ray) after the four consecutive reflections. Now $\tan\gamma$ is given by eq. (11), and $\tan\gamma'$ by eq. (12) with γ replaced by γ'. Thus, by requiring condition (14), one obtains condition (13). In the particular case where the second reflector is a paraboloid,

$$M' = 0$$

and eq. (13) give Eq. 7 (with $i' = p$).

2.5 Restoration of beam symmetry after an arbitrary number of reflections

Suppose an arbitrary sequence of $N-1$ reflections $\Sigma_1, \cdots \Sigma_{N-1}$ have distorted the initial symmetry of a spherical wave originating from F_0. We wish to restore symmetry by introducing an additional reflector Σ_N. Let the principal ray through F_0 be reflected $N-1$ times by the given reflectors as shown in Fig. 11a for $N = 3$. The reflector Σ_N must be chosen

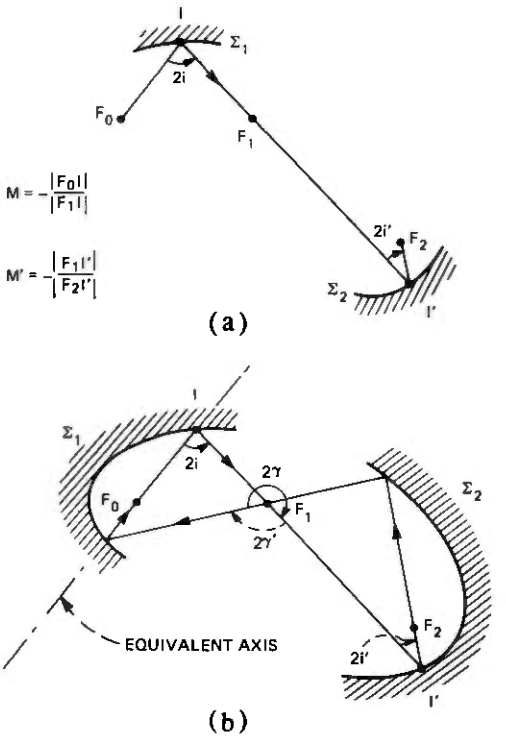


Fig. 10—Central path and equivalent axis of a combination of two ellipsoids.

so that this ray is one of the two *central rays* of the sequence $\Sigma_1, \cdots, \Sigma_N$. This means the path of the ray must *close* after $2N$ successive reflections. Now a part of this path, the section determined by the reflections of Σ_1, $\Sigma_2, \cdots, \Sigma_{N-1}$, is fixed in advance. Therefore let this part of the central ray be determined first. It starts at F_{N-1} and, after $2(N-1)$ reflections, it ends again at F_{N-1} with direction $\vec{a}$ as shown in Fig. 11a. Since its final direction $\vec{a}$ is given, its initial direction $\vec{c}$ can be found by tracing the ray backwards. Once $\vec{c}$ is known, the condition that Σ_N must satisfy is simply eq. (12), with γ given by the angle between $\vec{c}$ and $\vec{a}$, shown in Fig. 11.

2.6 How to determine the first reflector if the remaining ones are given

The above argument applies also to the problem where the first reflector, rather than the last, is to be found and the remaining reflectors are given. The only difference in this case is that one must use eq. (11), instead of eq. (12), as shown by the following example. To consider a situation of practical interest, suppose the last reflector Σ_N is a paraboloid as shown in Fig. 11b. Assume that all the reflectors except the first one are given and that Σ_1 must be chosen so that the central ray passes

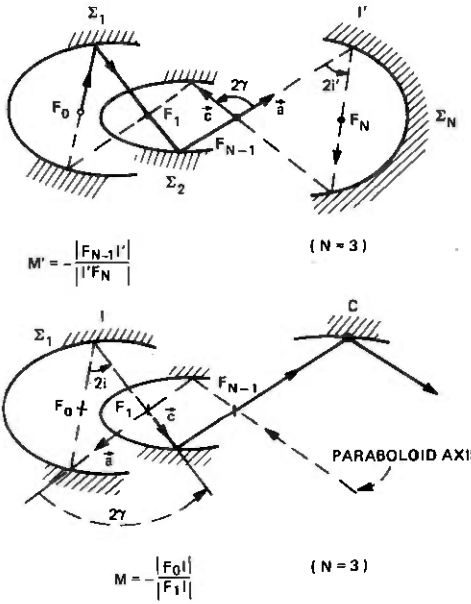


Fig. 11—(a) How to determine the last reflection if the first $N-1$ are given. (b) How to determine the first reflection if the last $N-1$ are given.

through the center C of the paraboloid aperture. Then, as in the previous problem, one notices that a part of the desired central path is fixed in advance. This part starts as F_1 with direction $\vec{c}$ and, after $2(N-1)$ reflections, it ends at F_1 with direction $\vec{a}$ as shown in Fig. 11b. Once $\vec{a}$ is found (by ray tracing), the condition that Σ_1 must satisfy is given by eq. (11), with γ given by the angle shown in the figure between $\vec{c}$ and $\vec{a}$.

2.7 The first and the last reflector are paraboloids

Consider first $N = 2$, in which case eq. (13) with $M = M' = \infty$ demands that the angles of incidence on the two paraboloids be identical, except for a difference in sign. For this to happen, the axes of the two paraboloids must coincide, in which case one can show that the two angles of incidence coincide *for any* choice of the principal ray. These remarks apply also to $N > 2$, since the last $N-1$ reflectors can always be replaced by an equivalent paraboloid. Thus,

> In order that symmetry be preserved, when both Σ_1 and Σ_N are paraboloids, the axis of Σ_1 must coincide with the equivalent axis of $\Sigma_2, \cdots, \Sigma_N$, in which case symmetry is preserved by any choice* of the principal ray. (15)

* A little thought shows that there is another case where the central ray is undetermined: namely, when the equivalent reflector is a flat plate.

III. AN APPLICATION

The most important example of an offset arrangement is perhaps the horn reflector,[11] an antenna consisting of a horn combined with a paraboloid. The excellent properties of this antenna (negligible return loss, very low level of the far sidelobes, etc.) are well known. However, the angle of incidence on the paraboloid is 45 degrees, and this causes in the far field a cross-polarized component of about −20 dB in certain directions.[11] The 45-degree angle of incidence is required to produce a beam orthogonal to the feed axis, which is an important requirement* for radio relay systems. In this section it is shown, with two examples given in Figs. 12 and 13, how this requirement can be fulfilled using two or more reflectors satisfying condition (7). In both Figs. 12 and 13, the feed is of the type described in Refs. 1 to 4, and therefore the antenna beam is essentially free of cross-polarization *everywhere* (see the last remark in the introduction).

Figure 12 shows two large reflectors, a paraboloid and an hyperboloid, arranged to satisfy simultaneously condition (7) and the requirement $i + p = 90°$, without aperture blockage. This arrangement is of the type shown in Fig. 8b of Ref. 7. In Fig. 13, three reflectors, a large paraboloid Σ_3, and two small hyperboloids Σ_2 and Σ_1 are used. This arrangement is more compact, and it requires less total reflecting area, than the one of Fig. 12. It is thus particularly attractive when the antenna aperture is large, i.e., the far-field beamwidth is small. The angle of incidence i and the magnification M of the first reflector Σ_1 satisfy condition (7), with p given by the angle shown in Fig. 12. To understand the significance of p, replace the last two reflectors Σ_2 and Σ_3 by their equivalent paraboloid. According to (3), the axis of this paraboloid is obtained from the axis of Σ_3 by reflecting it once, onto Σ_2, as shown in Fig. 13. Then $2p$ is the angle the central ray makes with this equivalent axis. Note that p is equal to the angle of incidence on this equivalent paraboloid (not shown in Fig. 13). This angle of incidence must satisfy eq. (7). One can verify from the figure that

$$\tan p = \frac{\tan \alpha + m_2 \tan \beta}{1 - m_2 \tan \alpha \tan \beta}, \tag{16}$$

α and β being the angles (see Fig. 7a) of the central ray and the axis of Σ_3 with respect to the axis of Σ_2, and

$$m_2 = \frac{|V_2F_1|}{|V_2F_2|} = \frac{e_2 + 1}{e_2 - 1}, \tag{17}$$

e_2 being the eccentricity of the hyperboloid Σ_2. Also,

$$2i = 90° + 2\beta - 2\alpha, \tag{18}$$

* Of course, this is not the only requirement that must be satisfied. Other requirements will be discussed in an article being prepared.

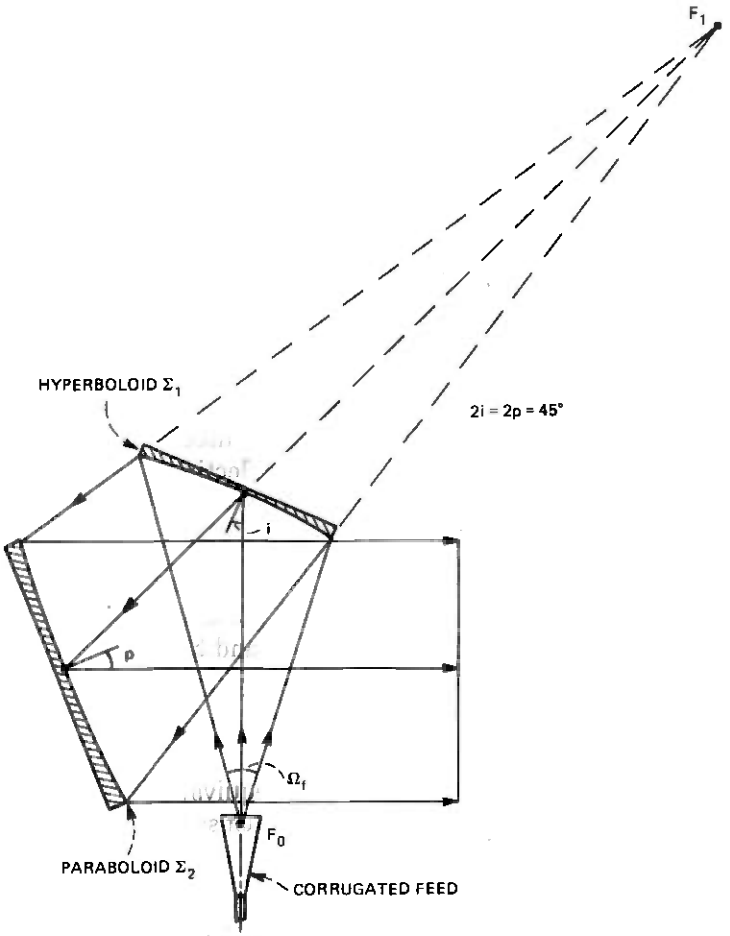


Fig. 12—A vertical feed and two reflectors with $i + p = 45$ degrees producing a horizontal beam without symmetry distortion.

and from eq. (7), solving for M,

$$M = \frac{\tan i}{\tan i + \tan p}. \tag{19}$$

Using eqs. (16) to (19), one can express M directly in terms of α, β, m_2.

An important property of Figs. 12 and 13 is that there is no aperture blockage even for relatively large values (as large as 30 degrees) of the angular width Ω_f of the beam radiated by the feed. Another important property, to be discussed in a future article, is that, if the feed is slightly displaced so as to cause a small angular displacement of the antenna beam, the resulting aberrations are very small. This is a consequence of

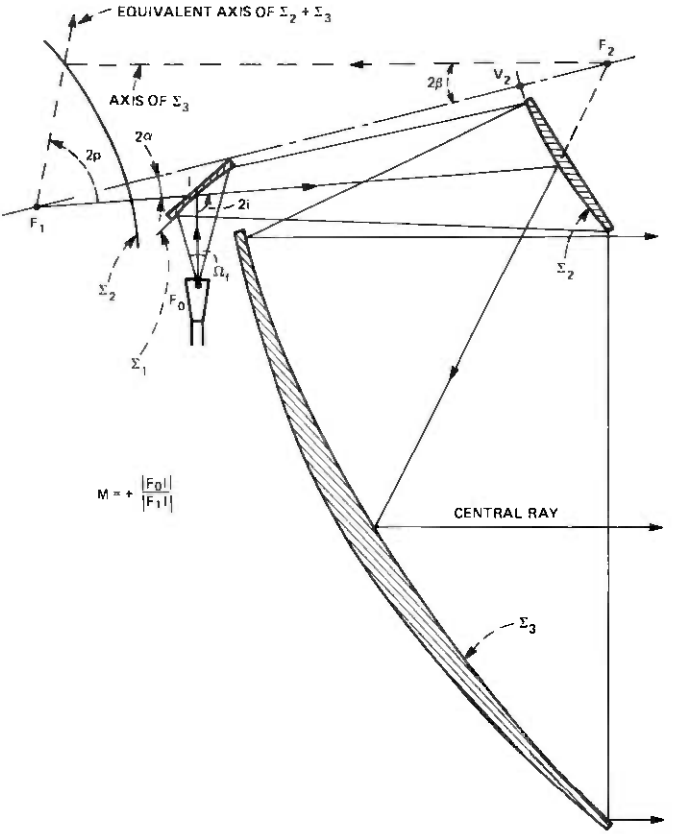


Fig. 13—A vertical feed and three reflectors producing a horizontal beam without symmetry distortion.

condition (7), and it implies that several beams can be produced efficiently by placing several feeds in the focal plane.

IV. CONCLUSIONS

The transformation of a symmetrical beam by an arbitrary arrangement of N confocal reflectors has been studied. It has been shown that it is always possible to choose the principal ray (i.e., the axis of the input beam) so that symmetry is preserved by the transformation. This is a consequence of the principle of equivalence shown in the appendix, according to which an arrangement of several reflectors can always be replaced by a single reflector producing the same transformation. Thus, in order that symmetry be preserved, the principal ray must coincide with the axis of symmetry of this equivalent reflector, i.e., the equivalent axis. A property of the equivalent axis is that the path of a ray having initially its direction becomes closed after $2N$ successive reflections. Because of this property, the equivalent axis can be found by a

straightforward geometrical procedure if the last reflector is a paraboloid. A simple relation [eq. (11) or (12)] has been given for determining the angle of incidence and the magnification of the first or last reflector so as to guarantee symmetry. In Section III, the problem of modifying the horn reflector to eliminate the asymmetry and cross-polarization due to the paraboloid has been discussed. Two solutions have been described.

APPENDIX

General Properties of a Sequence of N Confocal Reflectors

The results of this paper are consequences of the principle of equivalence stated at the beginning of Section II. This principle is now derived.

As pointed out in the introduction, the reflectors we consider are ellipsoids, hyperboloids, or paraboloids; let $F_0, F_1, \cdots, F_N$ be $N+1$ arbitrary points, let a point source be placed at F_0, and let a sequence of N reflectors $\Sigma_1, \cdots, \Sigma_N$ be used to successively transform the spherical wave from F_0 into spherical waves through $F_0, F_1, \cdots, F_N$. The nth reflector, Σ_n, with its focal points of F_{n-1} and F_n then transforms the spherical wave incident from F_{n-1} into a spherical wave through F_n.

Draw two spheres S and S' centered at F_0 and F_N. For each point P of S, there is, on S', a corresponding point determined by the ray through P. This mapping has the following properties.

A circle on S′ corresponds to each circle on S. In fact, it is well known[12,13] that a circular cone of rays from F_{n-1} is transformed by the nth reflector into a circular cone of rays through F_n.

The mapping is conformal,* and therefore two orthogonal curves of S are transformed into two orthogonal curves of S'.

Another property is that, if the point source at F_0 is linearly polarized and the lines of the electric field $\vec{E}$ on S are given, then the corresponding lines defined on S' by the above mapping give correctly the lines of $\vec{E}$ on S'. This result is true in general[14] for arbitrary reflectors, not necessarily paraboloids, hyperboloids, or ellipsoids. It allows the polarization of S' to be determined straightforwardly once the relationship between corresponding rays through F_N and F_0 is known.

A.1 The central rays

Draw a line through F_0, to cut the sphere S at two antipodal points. We show that it is always possible to choose the line orientation so that the corresponding points of S' are also antipodal points.

* This property is valid in general for an arbitrary wavefront S which is transformed by an arbitrary number of reflections (by arbitrary reflectors, not necessarily of the type considered here) into a wavefront S'. The mapping determined between S and S' by the rays orthogonal to S (and S') is a conformal mapping.

Let L_1, L_2 and M_1, M_2 be antipodal points of S (see Fig. 14; the sphere S is not shown). Let L'_1, L'_2 and M'_1, M'_2 be their corresponding points on S'. Through L'_1, L'_2 M'_1, M'_2 draw two great circles. The two circles will intersect in two antipodal points $0'_1$ and $0'_2$, as shown in Fig. 14. We show that the corresponding points are also antipodal points. In fact, 0_1 and 0_2 are the points of intersection of the two circles of S that correspond to the two circles of S'. Since the circles of S contain the antipodal points L_1, L_2 and M_1, M_2, they are great circles and therefore their intersections 0_1 and 0_2 are antipodal points. Q.E.D.

An important significance of the points 0_1, 0_2, $0'_1$, and $0'_2$ is the following. Let a ray from F_0 be reflected by the sequence of N reflectors *twice*, first in the order $\Sigma_1, \Sigma_2, \cdots, \Sigma_N$ and then in the reverse order $\Sigma_N, \Sigma_{N-1}, \cdots, \Sigma_1$. After these $2N$ reflections, the ray will pass again through F_0, but its direction will in general differ from the direction given initially, and therefore the ray will not in general follow the same path if reflected $2N$ more times. However, a little thought shows that, since the three points $0_1, F_0, 0_2$ are collinear and so also are $0'_1 F_N 0'_2$, the path of a ray from 0_1 (or from 0_2) will become closed after $2N$ reflections. The same observation applies to the ray from 0_2, which will follow, in the opposite direction, the same path of the ray from 0_1.

The path of the rays from 0_1 and 0_2 will be called the *central path* and the two rays *central rays*. This definition is consistent with the one given in Section II. As we shall see, there is in general only one central path, except when both Σ_1 and Σ_N are paraboloids (see Section 2.6) or when the equivalent reflector is a flat plate [$m_e = 1$ in eq. (21)].

The axial ray $F_0 0_1$ is now chosen as reference axis. Let a particular plane through this ray be chosen as reference plane. Consider a particular ray from F_0, and let θ be its angle with respect to the axis and ϕ the angle its plane makes with the reference plane. After N reflections, both the ray in question and the axial ray pass through F_N. Let θ' be the angle between the two rays at F_N, let ϕ' be the angle their plane makes with an arbitrary reference plane (chosen through the axial ray). We wish to

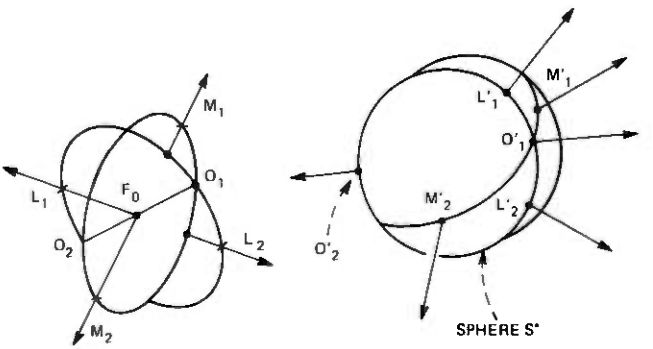


Fig. 14—How to determine the central rays.

show that

$$\phi' = \pm\phi + \phi_0 \tag{20}$$

and

$$\tan\frac{\theta'}{2} = m_e \tan\frac{\theta}{2}, \tag{21}$$

where m_e is a constant determined by the N reflectors and ϕ_0 is determined by the orientation of the two reference planes which will be chosen so that

$$\phi_0 = 0. \tag{22}$$

A.2 Derivation of eqs. (20) and (21)

First consider on S a great circle, through the two axial points 0_1 and 0_2, given by

$$\phi = a, \tag{23}$$

where a is a constant. Since the corresponding circle on S' must pass through $0_1'$ and $0_2'$, it is a great circle, given by

$$\phi' = a', \tag{24}$$

where a' is a constant. This shows that ϕ' depends only on ϕ, not on θ. We now recall that the mapping of S' must be conformal and therefore the angle between two circles through $0_1'$ must equal the angle between the corresponding circles of S. This implies eq. (20).

Next we derive eq. (21). Since the sign in front of ϕ in eq. (20) depends on the definition of ϕ', and can therefore be chosen arbitrarily, we choose for the following derivation

$$\phi' = \phi.$$

Since a circle θ = constant is orthogonal to a circle ϕ = constant, the corresponding circles on S' must be orthogonal. This implies θ' is a function of θ only. To determine this function, consider on S three points of coordinates:

$$(\theta,\phi), \quad (\theta + d\theta,\phi), \quad (\theta,\phi + d\phi).$$

Let

$$(\theta',\phi), \quad (\theta' + d\theta',\phi)(\theta',\phi + d\phi)$$

be the corresponding coordinates on S. Let $d\ell_1$ and $d\ell_2$ denote on S the distances of the first point from the other two. Then

$$d\ell_1 = rd\theta, \quad d\ell_2 = r\sin\theta d\phi, \tag{25}$$

r being the radius of the sphere S. Similarly, for the corresponding distances on S',

$$d\ell_1' = r'd\theta', \quad d\ell_2' = r'\sin\theta' d\phi. \tag{26}$$

Since the mapping is conformal, one must have

$$\frac{d\ell_1'}{d\ell_2'} = \frac{d\ell_1}{d\ell_2},$$

which gives the condition

$$\frac{d\theta}{\sin\theta} = \frac{d\theta'}{\sin\theta'}. \tag{27}$$

Integrating this gives eq. (21), where m_e is a constant of integration.

When $N = 1$, eqs. (20) and (21) are nothing new. In fact, then the reflector system reduces to a single reflector whose eccentricity determines the parameter m_e. When $N > 1$, eqs. (20) and (21) show the N reflectors are equivalent to a single reflector with eccentricity specified* by m_e.

A.3 Derivation of eqs. (11) and (12)

Consider the ellipsoid shown in Fig. 15. Then

$$\tan\alpha \tan\alpha' = 1 \tag{28}$$

and

$$\tan\alpha' \tan\psi' = \tan\alpha \tan\psi = m,$$

where

$$m = \frac{|F_0V_0|}{|F_1V_0|}. \tag{29}$$

Therefore, taking into account that $\gamma = 90° - \psi - \psi'$,

$$\tan\gamma = \frac{1 - \tan^2\psi \tan^2\alpha}{\tan\psi(1 + \tan^2\alpha)} \tag{30}$$

Also, $i = 90° - \alpha - \psi$, and therefore

$$\tan i = \frac{1 - \tan\alpha \tan\psi}{\tan\alpha + \tan\psi}. \tag{31}$$

Now the magnification M of I is defined as

$$M = -\frac{|IF_0|}{|IF_1|}, \tag{32}$$

and from Fig. 15 is related to the angles ψ and α,

$$M = -\frac{\sin 2\psi}{\sin 2\alpha}$$

$$= -\frac{\tan\psi}{\tan\alpha}\,\frac{1 + \tan^2\alpha}{1 + \tan^2\psi}, \tag{33}$$

---

* The value of m_e can be calculated using the formula

$$m_e = \pm M_1 \cdot M_2 \cdots M_N,$$

where $M_1, \cdots, M_N$ are the magnifications calculated for the N reflections of the central ray chosen as reference axis. The sign of m_e depends on the sign convention for ϕ in eq. (20).

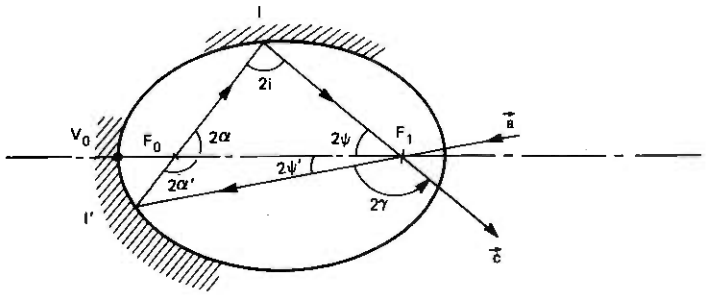


Fig. 15—Two successive reflections by a concave ellipsoid.

which gives

$$\frac{M}{M-1} = \frac{\tan\psi(1+\tan^2\alpha)}{(\tan\psi+\tan\alpha)(1+\tan\psi\tan\alpha)}. \tag{34}$$

From eqs. (30), (31), and (34), one obtains eq. (11). The derivation of eq. (12) is entirely analogous. The case where the reflector is convex, or is a hyperboloid, can be treated in the same way.

REFERENCES

1. V. H. Rumsey, "Horn Antennas with Uniform Power Patterns Around Their Axes," IEEE Trans., *AP-14* (1966), pp. 656–658.
2. H. C. Minnet and B. M. Thomas, "A Method of Synthesizing Radiation Patterns with Axial Symmetry," IEEE Trans., *AP-14* (1966), pp. 654–656.
3. P. J. B. Clarricoats and P. K. Saha, "Propagation and Radiation Behaviour of Corrugated Feeds, Part I—Corrugated Waveguide Feed," Proc. IEE, *118*, No. 9 (September 1971), pp. 1167–1176.
4. C. Dragone, "Characteristics of a Broadband Microwave Corrugated Feed: A Comparison Between Theory and Experiment," B.S.T.J., *56*, No. 6 (July–August 1977), pp. 869–888.
5. T. S. Chu and R. H. Turrin, "Depolarization Properties of Off-set Reflector Resonators," IEEE Trans., *AP-21* (May 1973), pp. 334–345.
6. M. J. Gans, "Cross-Polarization in Reflector-Type Beam Waveguides and Antennas," B.S.T.J., *55*, No. 3 (March 1976), pp. 289–316.
7. Hirokaau Tanaka and Motoc Mizusawa, "Elimination of Cross-Polarization in Offset Dual-Reflector Antennas," Electronics and Communications in Japan, *58-B*, No. 12 (1975).
8. R. Graham, "The Polarization Characteristics of Offset Cassegrain Aerials," IEEE International Conference on Radar—Present and Future, No. 105 (October 1973). Also United States Patent No. 3,792,480.
9. Y. Mizugutch, M. Akagawa, and H. Yokoi, "Offset Dual Reflector Antenna," Digest of 1976 AP-S International Symposium on Antennas and Propagation, October 1976.
10. K. N. Coyne, unpublished work.
11. A. B. Crawford, D. C. Hogg, and L. E. Hunt, "A Horn-Reflector Antenna for Space Communication," B.S.T.J., *40*, No. 4 (July–August 1961), pp. 1005–1116.
12. M. Mizasawa and T. Kitsuregawa, "A Beam-Waveguide Feed Having A Symmetric Beam for Cassegrain Antennas," IEEE Trans. Ant. Propag., *AP-21*, No. 6 (November 1973), pp. 884–886.
13. M. Mizusawa and T. Katoigi, "A Property of the Series of Mirrors on Quadratic Surface of Revolution," Trans. IEEE Japan, *53-B* (November 1970), pp. 707–708.
14. C. Dragone, "New Grids for Improved Polarization Diplexing of Microwaves in Reflector Antennas," IEEE Trans. Ant. Propag., *AP-26*, No. 3 (May 1978), pp. 459–463.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 9, September 1978
Printed in U.S.A.

Radiation Patterns from Parallel, Optical Waveguide Directional Couplers—Parameter Measurements

By V. RAMASWAMY and R. D. STANDLEY

(Manuscript received November 22, 1976)

A new method for measuring the parameters of optical, parallel, waveguide directional couplers is presented. Basically, we observe the changes in radiation pattern obtained by placing a high, refractive, index coupling prism on the coupled guides as a function of position along the coupler. For a coupler, in a Z cut, Ti-diffused $LiNbO_3$ substrate with 3-μm guides and 3-μm separation, the transfer length is about 1.8 mm at 7266 Å.

I. INTRODUCTION

Parallel coupled waveguides are the basic building block for a number of integrated optical devices; these include switches,[1–6] modulators, and channel dropping or adding filters.[7] The techniques used to measure coupling parameters are often visual in nature. The simplest approach is to observe the energy exchanges between the parallel guides from the surface scattering of these guides viewed through a microscope. However, this is not always feasible; e.g., operation at longer wavelengths away from the visible, with low-loss surface scattering guides, and in cases where the energy at the surface is rather low, as it happens with Ti diffused guides in $LiNbO_3$. In such cases, the technique developed by Ostrowsky et al.[8] is quite useful. They observed the fluorescence from RhB-doped polyurethane film over the strip guides pumped by an argon laser.

In this paper, we present a method found useful in measuring the parameters of such couplers. Basically, the method consists of observation of the interaction length dependence of the coupling via radiation pattern measurements;[9] the radiation patterns are obtained by moving an output coupling prism along the coupled waveguide region.

II. THEORY

2.1 Synchronous couplers

Figure 1 depicts two coupled parallel waveguides where a is the guide width, c is the guide spacing, and L is the length over which the guides are coupled; i.e., the interaction length. We consider the ideal case which assumes that the guides are identical in width and thickness so that perfect synchronism of the unperturbed propagation constants exists; for this case, the normalized field amplitudes in the two guides as a function of length z can be shown to be[10]

$$\begin{aligned} R &= \cos \kappa z \\ S &= j \sin \kappa z, \end{aligned} \tag{1}$$

where R is the field amplitude in the initially excited guide, S is that of the auxiliary guide, and κ is the coupling strength per unit length. We are interested in determining the coupling strength κ per unit length for a coupler of known physical parameters; knowledge of κ permits the selection of L for a coupler of desired overall coupling strength. If, at some point z along the parallel coupled region, we place a prism whose refractive index is higher than that of the waveguides, then power will be radiated from the two waveguides. Thus in the far field we observe a radiation pattern due to the interference of the fields from the coupled waveguides over the coupling length of the prism coupler. If we keep the prism coupling length small compared to $1/\kappa$, say, less than a millimeter, then the far-field radiation pattern would truly be representative of the pattern from two slits separated by a distance of d having relative amplitudes given by eq. (1).

If we assume constant transverse field amplitudes, as seen from Fig. 2, the expression for the radiation pattern is

$$|E|^2 = \frac{\sin^2 u}{u^2}\left(1 + \sin 2Z \sin 2u \frac{d}{a}\right), \tag{2}$$

where

$$Z = \kappa z$$

$$u = \frac{\pi a}{\lambda} \sin \theta$$

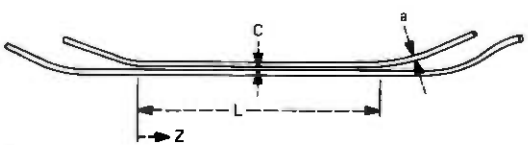
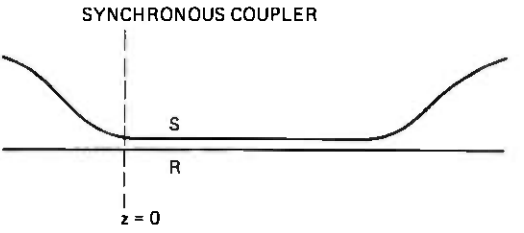


Fig. 1—Parallel waveguide directional coupler where a is guide width, c is the guide spacing, and L is the interaction length.



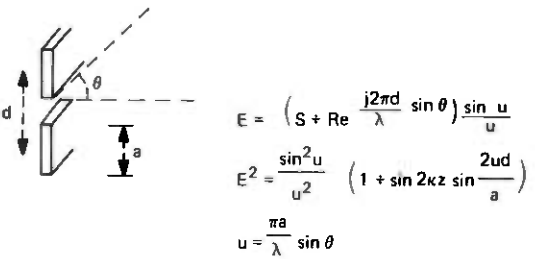


Fig. 2—Radiation field amplitudes under far-field conditions due to a pair of sources of width a, separated by a distance d. θ is measured in the plane perpendicular to the plane containing the waveguides.

$$d = a + c$$

θ is radiation angle.

Figure 3 shows computed plots of $|E|^2$ as a function of u for the case $d/a = 2$ with Z as the parameter. Except when all of the energy is in one guide, e.g., at $Z = o$, the radiation pattern is asymmetrical about $\theta = o$. This is true even for the case when $Z = \pi/4$, when the field amplitudes in both guides are equal, and differ by a phase shift of 90 degrees. When Z is increased from $\pi/4$ to $\pi/2$ in specific increments, the patterns remain the same as Z is varied from $\pi/4$ to 0, for the same shape, i.e., for example, identical patterns are observed for the cases when $Z = \pi/16$ and $7\pi/16$, $\pi/8$ and $3\pi/8$, $3\pi/16$ and $5\pi/16$, etc. At $Z = Zo = \pi/2$, complete energy transfer occurs. When Z is varied from $\pi/2$ to $3\pi/4$ and back to π, the graphs shown in Fig. 3 can be used with change in sign of abscissa. The whole series of patterns repeat themselves in this manner with increasing Z.

2.2 Asynchronous couplers

If the waveguides differ in width, thickness, or refractive index, their propagation constants will differ. This could occur as a result of errors in the fabrication process. For such asynchronous couplers, complete power transfer from one guide to the other is not possible. If we define

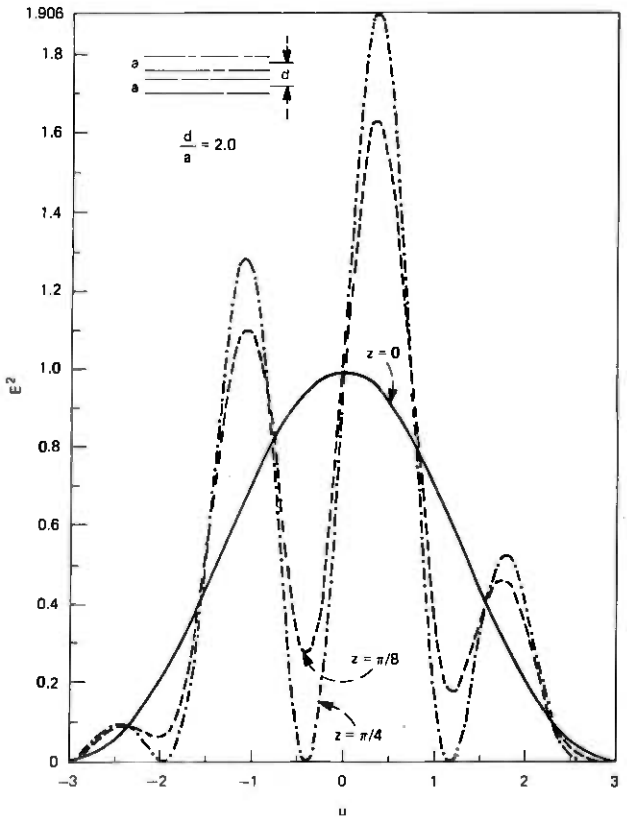


Fig. 3—Computer plots of energy distribution as a function of normalized radiation angle u for the case $d/a = 2$, with $Z = \kappa z$ as the parameter.

the difference in unperturbed propagation constants in the two guides as $\Delta\beta$, then the normalized field amplitudes[6] as a function of z become

$$R' = \cos\alpha - j\frac{\gamma}{\sqrt{\gamma^2+1}}\sin\alpha$$

$$S' = j\frac{\sin\alpha}{\sqrt{\gamma^2+1}}, \tag{3}$$

where

$$\gamma = \Delta\beta/2\kappa$$

$$\alpha = \sqrt{\gamma^2+1}\,\kappa z.$$

Here, again, R' is the field amplitude in the initially excited guide and S' is that of the auxiliary guide. With these field amplitudes, the radia-

tion pattern is given by

$$|E'|^2 = \frac{\sin^2 u}{u^2}\left(1 + \frac{\sin 2(\gamma^2+1)^{1/2}Z}{(\gamma^2+1)^{1/2}}\sin\left(2u\frac{d}{a}\right) - \frac{\gamma}{(\gamma^2+1)}\left[1-\cos 2(\gamma^2+1)^{1/2}Z\right]\cos\left(2u\frac{d}{a}\right)\right). \quad (4)$$

The power in the coupled guide is obtained by squaring eq. (3) and is given by

$$|S'|^2 = \frac{\sin^2[(\gamma^2+1)^{1/2}Z]}{(\gamma^2+1)} \quad (5)$$

and

$$|R'|^2 = 1 - |S'|^2.$$

We find the maximum value for the coupled power to be $(\gamma^2+1)^{-1}$ at $Z = (m\pi/2)(\gamma^2+1)^{-1/2}$. Plots of $|E'|$ show the expected result that the information content in the radiation patterns decreases rapidly with increasing asynchronism. However, useful information is obtained by recognizing the transfer period as indicated by all the power being present in the input guide.

III. COUPLER FABRICATION AND MEASUREMENT TECHNIQUE

The procedures used in the fabrication of the experimental couplers are described. Z-cut lithium niobate substrates were coated with polymethyl-methacrylate (PMMA) electron resist approximately 0.5 micron in thickness. A thin layer of aluminum (100 Å) is evaporated onto the

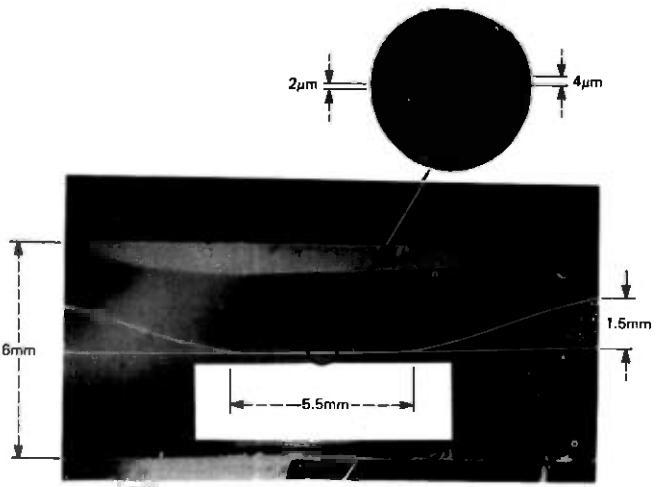


Fig. 4—Guide tracks defined in PMMA after electron beam exposure and development.

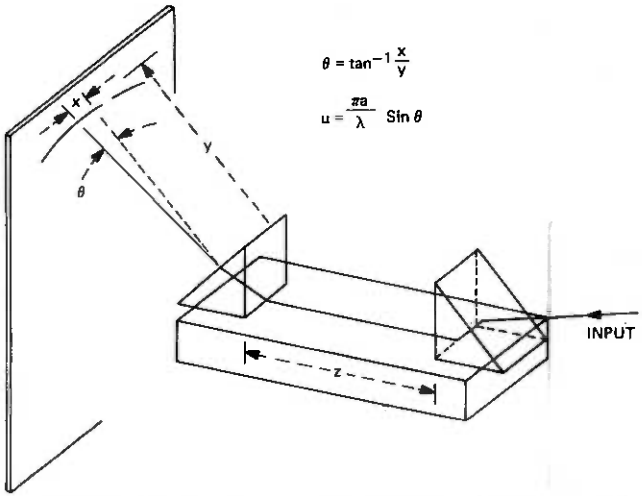


Fig. 5—Illustration of the setup to measure the coupler radiation pattern.

PMMA to eliminate charging problems. The coated substrate is then mounted onto a scanning electron microscope (SEM) stub using a conducting silver paste. Using the appropriate scan generator, the first guide of the coupler is exposed. The scan generator output amplitude is then attenuated and the writing beam moved by electronic adjustment of the fine shift coil current; the auxiliary guide is then exposed. For exposure, a specimen current of 10^{-9} A is typically used with an exposure time of about 25 s to obtain 3-μm wide guides 15 mm in length. The sample is then removed from the SEM. A brief soak in dilute NaOH removes the aluminum layer. The PMMA is then developed for about 30 s in a 3-to-1

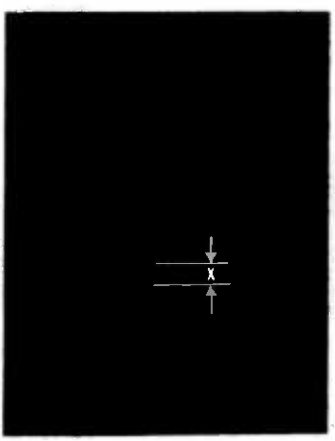


Fig. 6—A typical radiation pattern—in this case, the energy is very close to the position where all the energy is one of the guides.

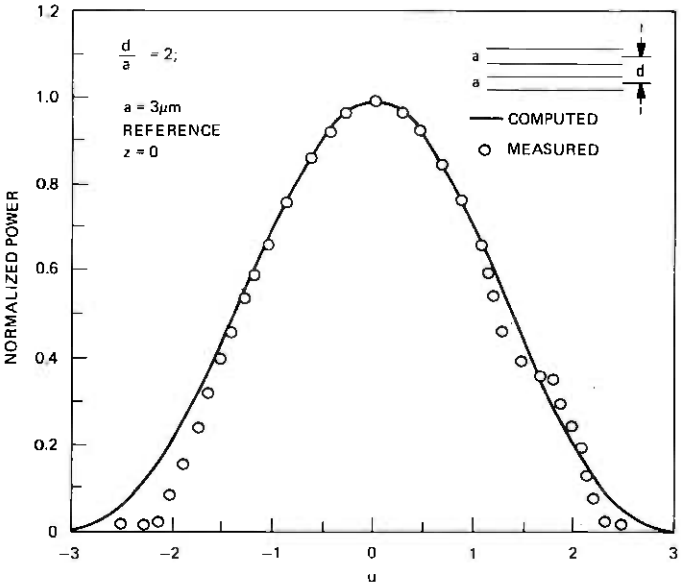


Fig. 7—Radiated power output as a function of normalized radiation angle u where all the energy is in one of the guides, $z = 0$. $d/a = 2$ for this coupler.

mixture of isopropyl alcohol and methylethyl ketone. The guide tracks are now defined in the PMMA (Fig. 4). The sample is blown dry with dry nitrogen and mounted in a sputtering system for deposition of a Ti layer usually about 300 Å thick. The PMMA and excess Ti are next removed by soaking the sample in acetone. At this point, we have a sample with Ti where we want the waveguides. The sample is next placed in an oven and brought to 1000°C in an argon ambient. Following the 1000°C soaking for about three hours, the furnace is turned off and the ambient changed to oxygen. The resulting guides exhibit single TE mode operation.

The experimental set-up used to measure the coupler radiation pattern is shown in Fig. 5. The lasers employed were He-Ne operating at 6328°A and a nile-blue dye laser covering the wavelength 6900°A to 7500°A. The latter source was pumped by the 6471°A line of a krypton laser. The prisms were made of rutile. The input prism was quite flat, allowing strong coupling, whereas the base of the output prism had a curvature in it to ensure the coupling region to be much less than that of a millimeter. Although the amount of energy coupled out is rather small, the resulting radiation pattern is primarily due to the energy of the guide at the output prism location and does not include the effects of long coupling lengths. As the output prism was moved along the guides, the radiation pattern was scanned using an iris. Figure 6 is a photograph of a typical nearly synchronous coupler radiation pattern. The pattern in Fig. 7 resulted from a coupler operating at 7266°A con-

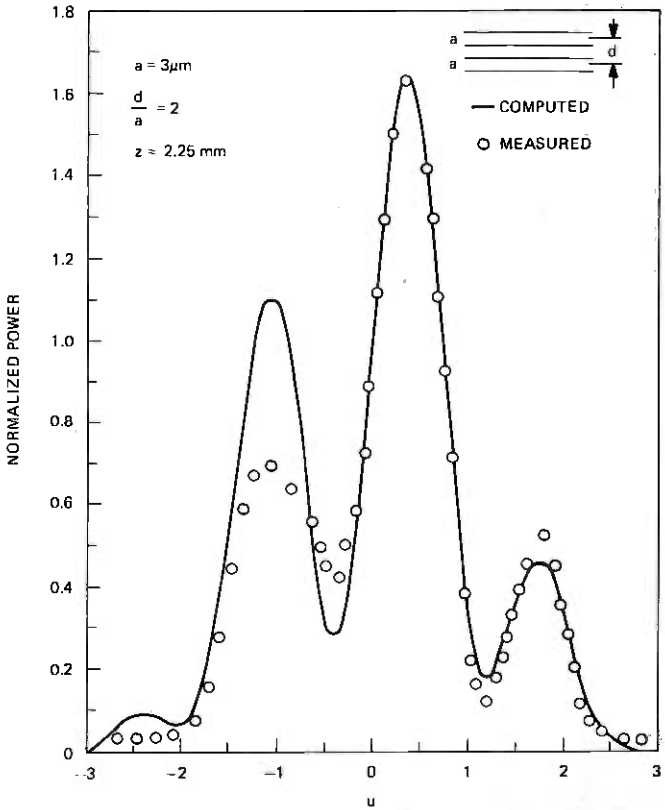


Fig. 8—Radiation pattern of the same coupler shown in Fig. 7, but at a different $z \simeq 5\pi/8$, past the location of complete energy transfer.

sisting of 3-μm guides. The measurement was made at a position $Z = 0$ along the coupler. Where all power was essentially in one of the guides, measured distribution agrees well with the theory. By moving the output prism to a place where all the energy is in the other guide, the transfer length can be measured. However, if the prism is not placed exactly at this location, one can infer this information by noting the nature of the asymmetry and measuring the radiation pattern. For example, Fig. 8 shows the output radiation pattern for the same coupler, but at a different longitudinal position $z = 2.25$ mm. In this case, the power in the two guides is nearly equal, resulting in sidelobe development in the observed radiation pattern. From Fig. 3 for $d/a = 2.0$, the separation Δu between minima is 1.6. At 7266°A, for $a = 3$ μm, this translates into a separation $\Delta x = 2.24$ cm between minima at a distance $y = 18$ cm from the output prism coupling position. This compares favorably with the measured value of 2.25 cm. By a series of observations on this coupler, we can infer an interaction length for full power transfer $L_o = \pi/2\kappa$; the

best fit for curve in Fig. 8 occurs at $Z = \pi z/2L_o = 5\pi/8$, from which the transfer length L_o is inferred to be 1.8 mm for this coupler. The separation of the minima agrees very well, although the peaks do not. Considering that we analyze uniform distribution of energy in the waveguides, the agreement is rather good.

IV. CONCLUSION

We have described a method for measuring the coupling strength of synchronous optical waveguide directional couplers by observing the length dependence of the radiated signal. As indicated earlier, the technique is useful, with care in implementation, as a laboratory tool.

V. ACKNOWLEDGMENTS

The authors are grateful to M. D. Divino and F. A. Braun for their assistance, and to Mrs. D. Vitello for programming the computer plots presented in this paper.

REFERENCES

1. H. F. Taylor, "Optical Switching and Modulation in Parallel Dielectric Waveguides," J. Appl. Phys., *44*, no. 7 (1973), p. 3257.
2. F. Zernike, "Integrated Optical Switch," WA5 Topical meeting on Integrated Optics, New Orleans, 1974.
3. M. Papachon et al, "Electrically Switched Optical Directional Couplers: Cobra," Appl.Phys. Lett., *27* (Sept. 1975), p. 289.
4. J. C. Campbell et al, "GaAs Electrooptic Directional Coupler Switch," Appl. Phys. Lett., *27* (August 1975), p. 202.
5. H. Kogelnik and R. V. Schmidt, "Switched Directional Couplers with Alternating $\Delta\beta$," IEEE J. Quantum Electronics (July 1975).
6. V. Ramaswamy and R. D. Standley, "A Phased, Optical, Coupler Pair Switch," B.S.T.J., *55*, No. 6 (July-August 1976), p. 767.
7. V. Ramaswamy and R. D. Standley, patent pending.
8. D. B. Ostrowsky et al., Appl. Opt., *13* (March 1974), p. 636.
9. R. D. Standley and V. Ramaswamy, "A New Method for Measuring Parallel Waveguide Directional Coupler Parameters," MD3 Topical Meeting on Integrated Optics, Salt Lake City, January 1976.
10. S. E. Miller, "Coupled Wave Theory and Waveguide Applications," B.S.T.J., *33*, No. 3 (May 1954), p. 661.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

Speech Signal Power in the Switched Message Network

By W. C. AHERN, F. P. DUFFY, and J. A. MAHER

(Manuscript received December 15, 1977)

Speech signal power at the main distributing frame in class 5 switching offices is characterized in terms of equivalent peak level (EPL) and average conversational signal power measures. The results indicate that there is little dependence of speech signal power on call destination or originating class of service. Small differences between various subpopulations are explained for the most part by loop characteristics. The switched telecommunications network is essentially transparent to customers in the sense that talker signal power has not been found to be sensitive to factors which affect the transmission path between class 5 central offices.

Present-day speech volumes for toll calls, which average −21.6 VU (volume units), are substantially lower than those found in a survey conducted in 1960,[1] which averaged −16.3 VU, and the ranges of volumes within all call destination categories are substantially smaller than the 1960 ranges. Several substantial changes have been introduced into the telephone plant since 1960 which tend to increase the uniformity of service in the network from the viewpoint of speech volumes. These include a decrease in the proportion of toll grade battery, loss plan improvements, replacement of the 300-type telephone set with the 500-type set, and an increase in direct trunking between class 5 offices.

I. INTRODUCTION

The characterization of speech signal power on Bell System message circuits is an essential step in the determination of signal power loading and crosstalk objectives. Knowledge of speech signal characteristics is also important to designers of a wide variety of telecommunications equipment.

Speech levels at the class 5 office were last characterized in the 1960 Speech Volume Survey[1] in terms of volume units (VU). In the years since

the last survey, there have been substantial changes in the Bell System network. For example, the proportion of toll grade battery has been substantially reduced, the 300-type telephone set has been almost completely phased out, direct distance dialing is now virtually universal, and new loop and trunk design methods have been introduced. Also, in the intervening years, research in speech signal measurement has led to a new measure of speech level known as the equivalent peak level (EPL).[2] This, together with advanced digital data acquisition technology, has facilitated the measurement of speech signal power with greater precision than was possible in 1960.

This paper presents the results of a speech signal power survey made in 1975–1976. The measurements were made at 36 class-5-office main distributing frames (MDFs), which constitute a statistical sample of acceptable precision from all the MDFs within the Bell System. The class 5 (local or end) office MDF was selected as the measurement interface because it has access to all customer loops and all classes of local and toll traffic; dialed address information is readily available; only the customer's loop and station equipment is interposed between the customer and the point of measurement; and the customer's loop current may be measured. A three-stage statistical sampling scheme was employed, which resulted in measurements of near-end and far-end talker power on more than 10,000 calls originating from approximately 2500 loops. Average conversational signal power (averaged over the entire observation interval) and EPL were the measures used for talker signal characterization. Loop dc current, class of service, switch type, and call destination were also recorded.

Survey results are presented in Section II, the methodology is presented in Section III, and comparisons of the present survey results with prior survey results are given in Section IV.

II. SURVEY RESULTS

Table I summarizes the findings of this survey. The results indicate that there is little dependence of speech signal power on call destination or originating class of service. In the sections that follow, it is shown that the small differences between various subpopulations are explained for

Table I—Summary of speech signal powers

Subclass	Near-End Mean Equivalent Peak Level (dBm)
Residence	−11.0
Business	−10.4
Local	−10.8
Toll	−10.1
Combined	−10.7

the most part by loop characteristics, and there is little if any variation in speech signal power that may be attributable to psychological factors such as call distance, perception of received volume, etc. The indication from the data is that the switched telecommunications network is essentially transparent to customers in the sense that talker signal power has not been found to be sensitive to call distance, local or toll call classification, or other factors that affect the transmission path from class 5 to class 5 central office.

2.1 General

In this survey, speech signal power measurements were made on customer loops at class 5 switching office main distributing frames (MDFs) during actual telephone conversations. The parties originating calls on sampled loops are referred to as the "near-end" speakers in the following discussion; the called parties are referred to as the "far-end" speakers. The far-end speakers were more distant than the near-end speakers from the MDFs at which the measurements were made, except for some intrabuilding calls.

The survey results characterize near- and far-end speech signal powers in terms of the equivalent peak level (EPL) and average conversational signal power measures, which are discussed in Section 3.3.3. The differences are also characterized between near- and far-end signal powers and between the EPL and average power measures. In addition, the influences of loop current, originating class of subscriber service, call destination, call distance, originating switch, and demographic features upon speech signal powers are investigated.

2.2 Speech signal powers at main distributing frames

The distributions of speech signal power at main distributing frames can be approximated by normal distributions. Histograms and cumulative distribution functions (CDFs) are given for the EPL and average power measures of speech signal power for the near- and far-end speakers in Figs. 1 through 4. The "bell" shapes of the histograms and the straight line shapes of the CDFs, which are plotted on normal probability grids, attest to the normality of these distributions. Because of this, the distributions are completely defined by the means and standard deviations listed in the first four lines of Table II.

While the near- and far-end signals encounter similar populations of station set and subscriber loop losses, the far-end signals also encounter end-office-to-end-office transmission losses. As a result of these additional losses, which will be referred to as the "apparent network loss" during the remainder of this paper, the average far-end signal power is generally lower than the average near-end signal power. The apparent network loss is a function of call destination, i.e., the greater the call

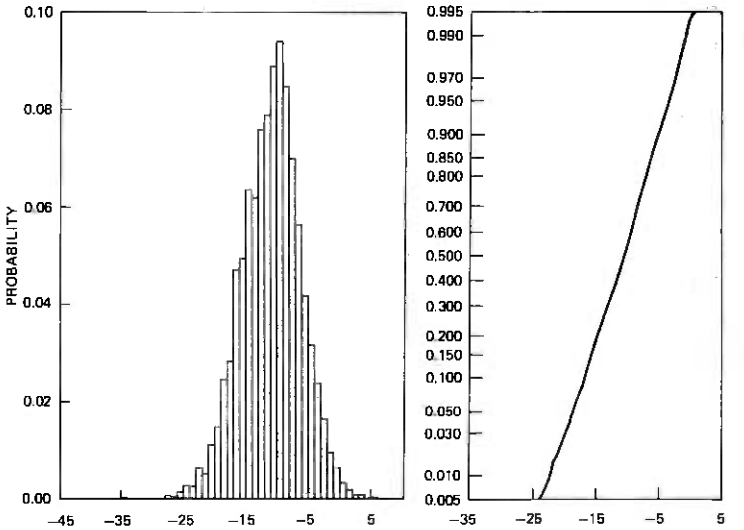


Fig. 1—Near-end equivalent peak level (dBm) distribution.

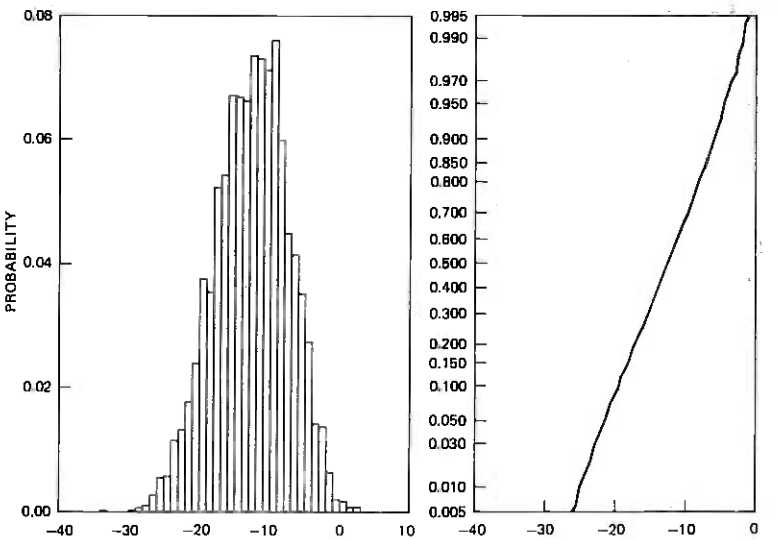


Fig. 2—Far-end equivalent peak level (dBm) distribution.

distance between end offices the more the signals are attenuated. This source of variation explains the greater variability among the far-end signal powers. These near-end, far-end differences exist for both EPL and average power; however, a comparison of the near- and far-end EPL results gives a difference of 2.1 dB, while a similar comparison for the average power measures gives a difference of 2.9 dB. In the following

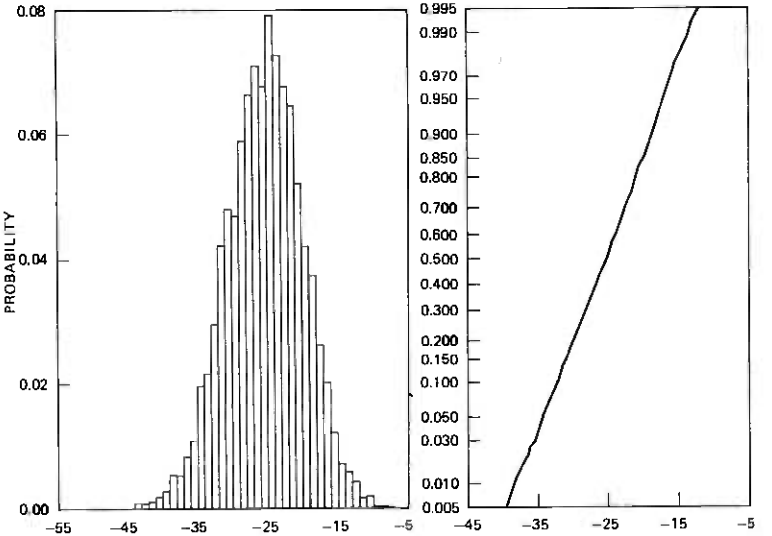


Fig. 3—Near-end average signal power (dBm) distribution.

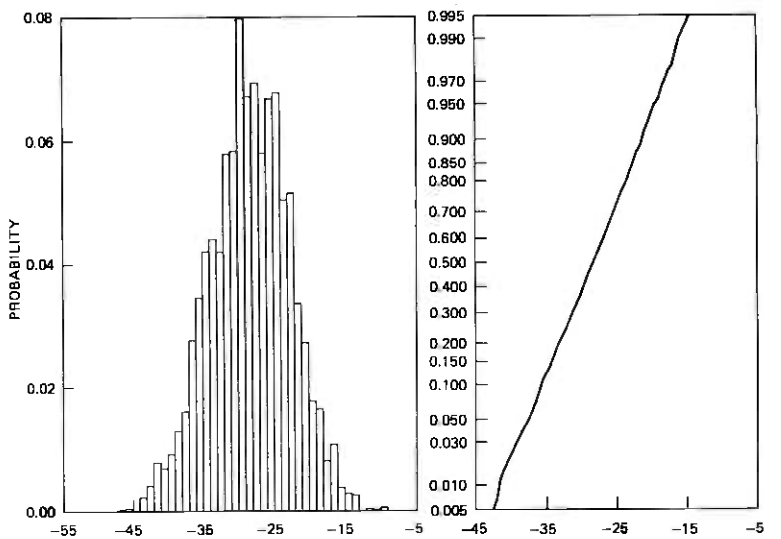


Fig. 4—Far-end average signal power (dBm) distribution.

paragraph, this apparent anomaly is shown to be caused by a difference in the speech activity of near- and far-end speakers.

The EPL, which is derived from the speech samples exceeding a threshold, is a measure of the speaker's peak signal power, and therefore is unaffected by silent periods in the conversation. The average signal power for conversational speech, however, includes intervals of speech

and silence alike. Therefore, the average power measure is lower than the corresponding EPL. This is illustrated by the results in Table II, which show that the average difference between EPL and average power is 14.6 dB for the near-end measures and 15.6 dB for the far-end measures. Such differences represent activity factors in the sense that they are logarithmically related to the amount of silence during a conversation.[3] They indicate that calling parties (near-end) tend to speak more than called parties (far-end) during telephone conversations. Due to these different speech activity characteristics, the apparent network loss result based upon average power is overestimated by about 1 dB. This finding explains the apparent anomaly noted above, and suggests that EPL is more appropriate than average power for estimating apparent network loss.

Comparisons of near-end EPL and average power with the far-end measurements are provided in the scatter diagrams in Figs. 5 and 6. The correlation coefficients are 0.31 and 0.57 for the EPL and average power comparisons, respectively. While the relationships are statistically significant, the modest positive correlations indicate that the signal power of one speaker is not strongly influenced by the signal power of the other.

Average signal power is strongly related to EPL. The results of the linear regressions of the near- and far-end EPLs on the corresponding average powers are given in Figs. 7 and 8, respectively. The near-end regression shows that average power = $-14.27 + 1.04$ EPL, and the far-end regression shows that average power = $-15.40 +$ EPL. The values of R^2, the square of the correlation, on the figures indicate that approximately 85 percent of the variation in average signal power is accounted for by the regression fits.

Signal power at the MDF is dependent upon loop loss and the telephone set electroacoustic efficiency. While these parameters were not measured, the near-end loop current, which was measured, has been found to relate to the overall loop and telephone set loss.[4] The histogram

Table II—Systemwide speech signal power results

Transmission Characteristic	Mean	90% C.I.	Std. Dev.	Sample
Near-end EPL (dBm)	−10.7	±0.5	4.6	10251
Far-end EPL (dBm)	−12.7	±0.4	5.2	8976
Near-end average power (dBm)	−25.3	±0.5	5.3	10251
Far-end average power (dBm)	−28.3	±0.4	5.6	8976
Near minus far-end EPL (dB)	2.1	±0.4	5.9	8478
Near minus far-end average power (dB)	2.9	±0.4	6.7	8478
Near-end EPL minus average power (dB)	14.6	±0.1	2.1	10251
Far-end EPL minus average power (dB)	15.6	±0.1	2.1	8976
Near-end loop current (mA)	42.2	±1.9	12.8	10749

90% C.I. = 90-percent confidence interval for the mean estimate.
Std. Dev. = Standard deviation of the signal power or loop current population.
Sample = Total sample size in calls used to calculate estimates.

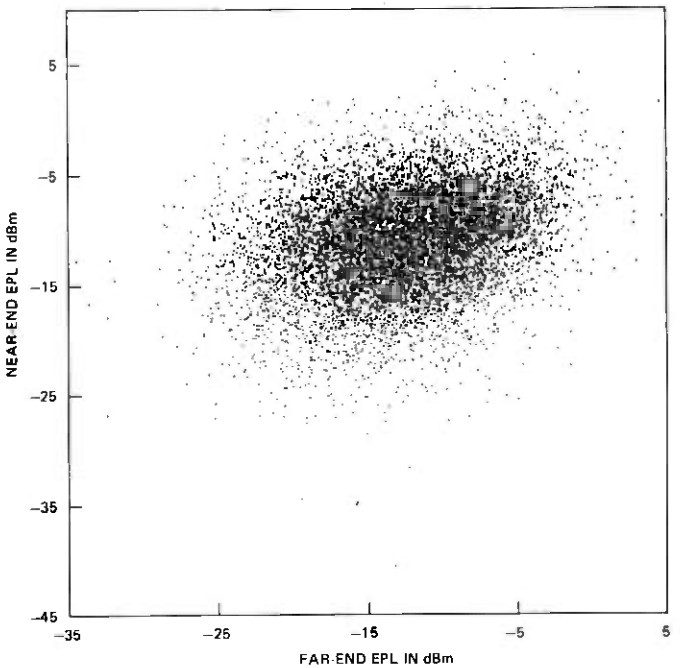


Fig. 5—Comparison of near- and far-end equivalent peak level.

and CDF for loop current are given in Fig. 9. The distribution is positively skewed, which means that it deviates from normality due to some large values of loop current associated with short loops. The distribution also deviates from normality at the lower tail because of a truncation of loop currents below 20 mA due to engineering limitations for signaling and transmission systems. Table II shows that the average loop current is 42.2 mA and the standard deviation is 12.8 mA.

Near-end EPL and average power are plotted as a function of loop current in Figs. 10 and 11, respectively. The scatter diagrams indicate that EPL and average signal power increase as loop current increases. Loop and telephone set characteristics suggest that a nonlinear relationship exists between loop current and the total loop and telephone set loss.[4] Nonlinear regression confirms this; however, the improvement in fit over the linear model, while statistically significant, is not of practical interest. The linear regressions of EPL and average power on loop current indicate that signal power increases about 0.13 dB per 1.0 mA increase in loop current. However, signal power varies substantially about the regression lines, indicating that loop current alone is not a good predictor of signal power. Visually, the variance appears to depend upon loop current; however, an analysis within loop current categories indicates that the variance is constant.

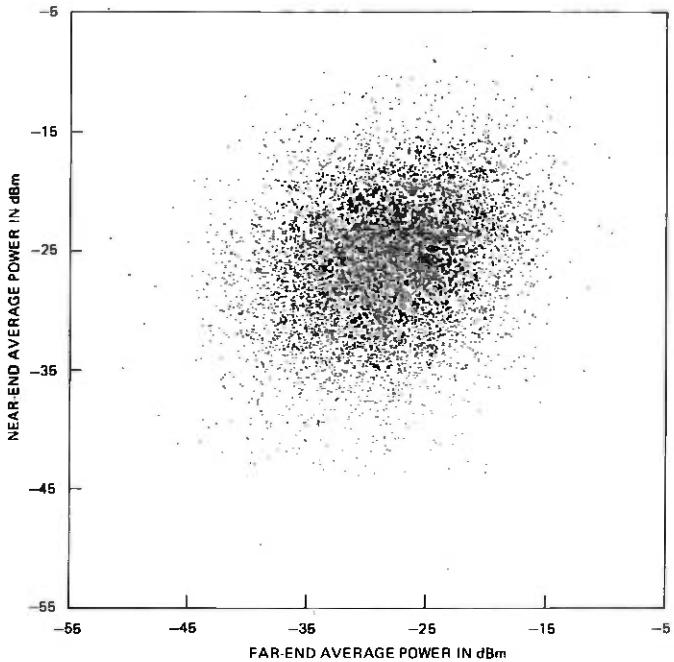


Fig. 6—Comparison of near- and far-end average power.

A more vivid illustration of the relationship between signal power and loop current is given in Fig. 12 by plotting the average EPL for each of the 36 MDFs in the sample as a function of the average loop current per MDF. The scatter shows a positive correlation, and the correlation coefficient is 0.82. A linear regression indicates that average EPL = −19.06 + 0.20 average loop current, and that the regression fit accounts for 67 percent of the variability in average EPL among MDFs.

2.3 Signal power and class of service

Class of service identifies the subscriber as a business or residential customer and identifies the station terminals as Bell or customer-provided equipment (CPE). The analyses discussed in this section deal with these service perspectives on the basis of originating class of service. The terminating customer class of service was not determined for the calls in this survey.

2.3.1 Business versus residential

The survey results for business- and residential-originated calls are summarized in Table III. Comparisons of the near-end EPL and average power results indicate that business-associated signal powers tend to be slightly higher than residential-associated signal powers, and that

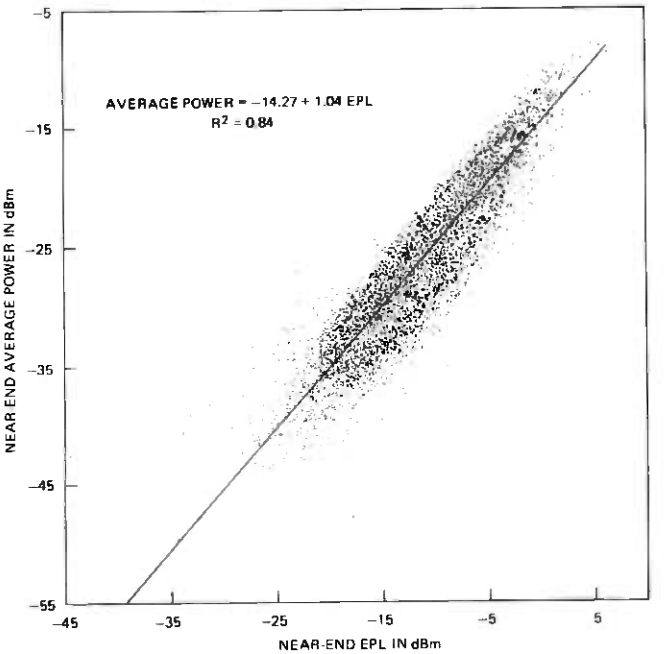


Fig. 7—Linear regression of near-end average power on equivalent peak level.

the variability among signal powers is about the same in both service categories. The 90-percent confidence intervals for the business and residential averages overlap, indicating that the differences are not statistically significant. Business loop currents are significantly higher and more variable than residential loop currents. The 5.3-mA difference in average loop current combined with the finding in Section 2.2, which indicates that EPL increases 0.13 dB per 1.0 mA increase in loop current, suggests that the business average EPL should be about 0.7 dB higher than the residential average. This difference agrees with the residence-business difference found for the near-end talker.

The far-end signal power results derived from the analysis by originating class of service are almost identical in the business and residential classifications. Since the originating parties in either category place calls to business and residential stations alike, the far-end speakers in each originating class of service category represent a mixture of business and residential customers. The far-end class of service mixture within each originating class of service category is sufficiently close to the overall traffic composition that the far-end results in each category are essentially the same as the far-end results for all telephone traffic listed in Table II. It is interesting to note that, although the average calling dis-

Table III—Originating class of service speech signal power results

Transmission Characteristic	Business				Residential			
	Mean	90% C.I.	Std. Dev.	Sample	Mean	90% C.I.	Std. Dev.	Sample
Near-end EPL (dBm)	−10.4	±0.7	4.6	6072	−11.0	±0.4	4.7	4179
Far-end EPL (dBm)	−12.8	±0.5	5.2	5228	−12.7	±0.4	5.2	3748
Near-end average power (dBm)	−25.0	±0.8	5.2	6072	−25.7	±0.4	5.4	4179
Far-end average power (dBm)	−28.4	±0.5	5.5	5228	−28.2	±0.4	5.6	3748
Near minus far-end EPL (dB)	2.5	±0.7	6.0	4916	1.7	±0.3	5.8	3562
Near minus far-end average power (dB)	3.4	±0.7	6.7	4916	2.5	±0.3	6.6	3562
Near-end EPL minus average power (dB)	14.6	±0.2	2.1	6072	14.7	±0.1	2.1	4179
Far-end EPL minus average power (dB)	15.7	±0.1	2.1	5228	15.5	±0.1	2.1	3748
Near-end loop current (mA)	45.0	±2.9	13.9	6384	39.7	±1.2	11.1	4365

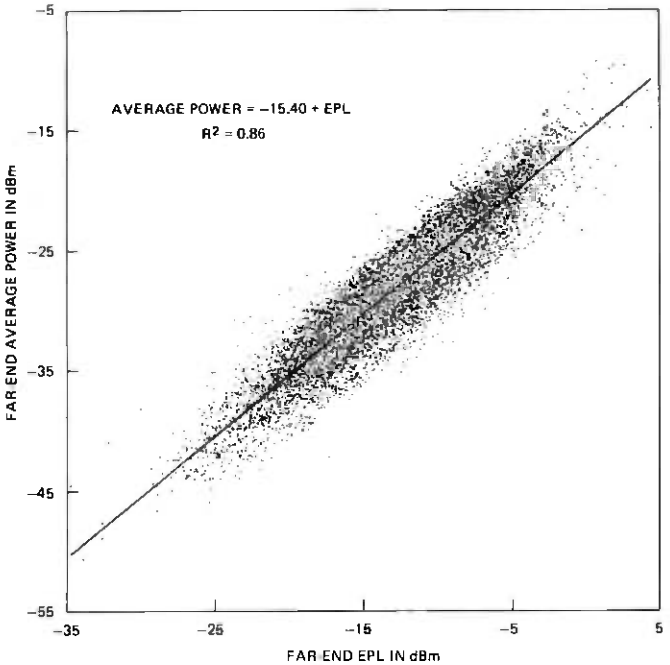


Fig. 8—Linear regression of far-end average power on equivalent peak level.

tance for the business-originated calls (50 ± 12 miles) is over 3.5 times the average for the residential calls (14 ± 4 miles), there is no noticeable call distance impact upon far-end talker received signal power. This does not imply that call distance has no influence upon network loss; it does imply that most of the data represent local calls or very short toll calls, and thus any potential call distance influence is not apparent.

Speaker speech activity during a telephone conversation is not affected by originating class of service. The EPL-average power differences have similar distributions for business- and residential-originated conversations.

The signal power distributions are all close to normal for business and residential calls. Therefore, the EPL and average power results listed in Table III completely define the signal power distributions for all practical applications. The business and residential loop current distributions differ significantly and are presented in Figs. 13 and 14, respectively. The business loop current distribution is comparable to the 1964 General Loop Survey[5] computed loop current distribution. The residential distribution has a greater proportion of lower current loops than the 1964 Survey result.

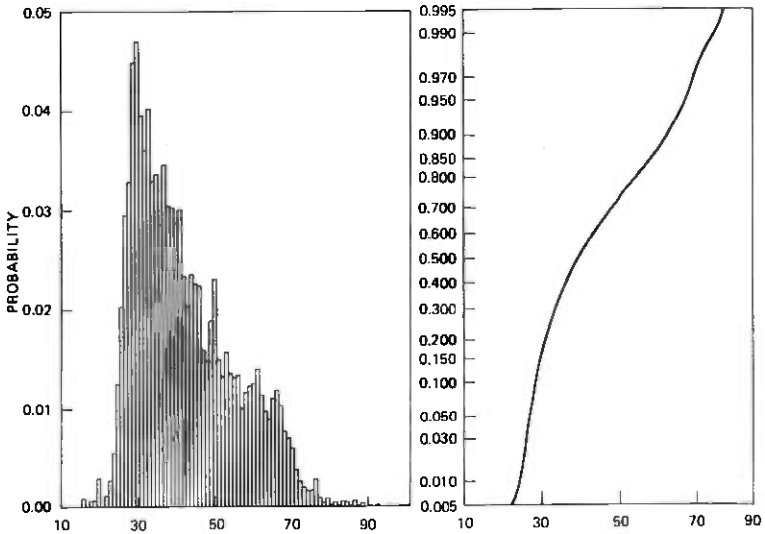


Fig. 9—Near-end loop current (mA) distribution.

2.3.2 Bell versus customer-provided equipment

Business calls are further classified on the basis of terminal equipment ownership in this section. One category contains those business calls which originated from subscriber lines with terminal equipment leased from the Bell System, and the second category contains those calls which originated from subscriber lines with customer-provided equipment (CPE). The results of this analysis are tabulated in Table IV. The near-end estimates show that the Bell signal powers on the average are more than 2 dB higher than the CPE signal powers, and that they are also somewhat less variable. The reason for this difference is suggested by examining the relationship between loop current and EPL for Bell and CPE loops, respectively. The correlation coefficients are 0.39 and 0.16, respectively, indicating that speech signal power on CPE loops is less strongly influenced by loop current than in the case of Bell loops. The reason for this is that the CPE station equipment battery is provided by a local power supply and not over the metallic loop facility. Thus, the electroacoustic efficiency of CPE station equipment is unrelated to the loop current observed in the central office, and the lower mean and higher variance in signal power may be attributable to the various local battery supplies and electroacoustic efficiencies of CPE terminals.

Comparisons of the far-end signal power estimates indicate that those far-end signals associated with CPE-originated calls have slightly lower signal powers than those associated with Bell-originated calls. The absence of detailed information about the far-end customers prevents further analyses to determine the cause of this difference.

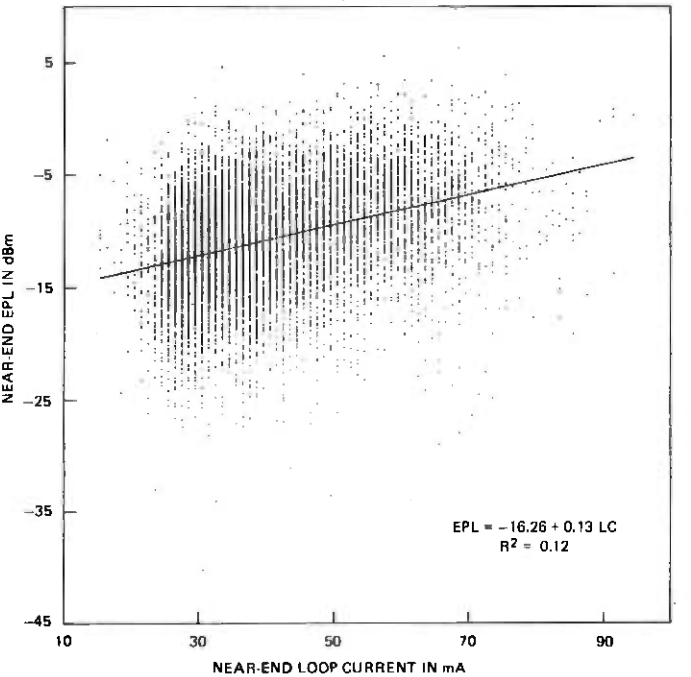


Fig. 10—Linear regression of equivalent peak level on loop current.

The signal power distributions again are all close to normal for both Bell- and CPE-originated business calls. The loop current distributions for both categories are comparable to the distributions given in the previous section for business calls in general.

2.4 Signal power and call destination

Four categories of call destination are considered in the following discussion; (*i*) intrabuilding local calls, (*ii*) interbuilding local calls, (*iii*) Home Numbering Plan Area (HNPA) toll calls, and (*iv*) Foreign Numbering Plan Area (FNPA) toll calls. The first two of these categories characterize local calls, and the last two characterize toll calls.

The trend lines in Fig. 15 summarize the relationships between signal power and call destination and between loop current and call destination. The near-end EPL and average power appear to increase slightly as the call destination becomes more remote from the originating office, with the exception of a slight drop in signal power for interbuilding local calls. The 90-percent confidence intervals for the four EPL estimates and for the four average power estimates overlap, which indicates that the differences among categories are not statistically significant. About half of the increase or decrease in signal power can be attributed to the call destination trend for loop current, which is plotted at the bottom of

Table IV—Bell and customer-provided equipment speech signal power results

Transmission Characteristic	Bell Business				CPE Business			
	Mean	90% C.I.	Std. Dev.	Sample	Mean	90% C.I.	Std. Dev.	Sample
Near-end EPL (dBm)	−10.4	±0.7	4.6	2857	−12.5	±1.2	5.1	2552
Far-end EPL (dBm)	−12.8	±0.5	5.2	2404	−14.0	±0.5	5.1	2228
Near-end average power (dBm)	−24.9	±0.7	5.2	2857	−27.5	±1.3	5.4	2552
Far-end average power (dBm)	−28.4	±0.5	5.5	2404	−29.3	±0.4	5.1	2228
Near minus far-end EPL (dB)	2.5	±0.7	6.0	2304	1.8	±1.3	6.2	2065
Near minus far-end average power (dB)	3.4	±0.7	6.7	2304	1.6	±1.6	6.6	2065
Near-end EPL minus average power (dB)	14.6	±0.2	2.1	2857	15.1	±0.1	2.4	2552
Far-end EPL minus average power (dB)	15.7	±0.1	2.1	2404	15.3	±0.2	2.0	2228
Near-end loop current (mA)	45.1	±2.9	14.0	2957	37.8	±3.2	11.3	2715

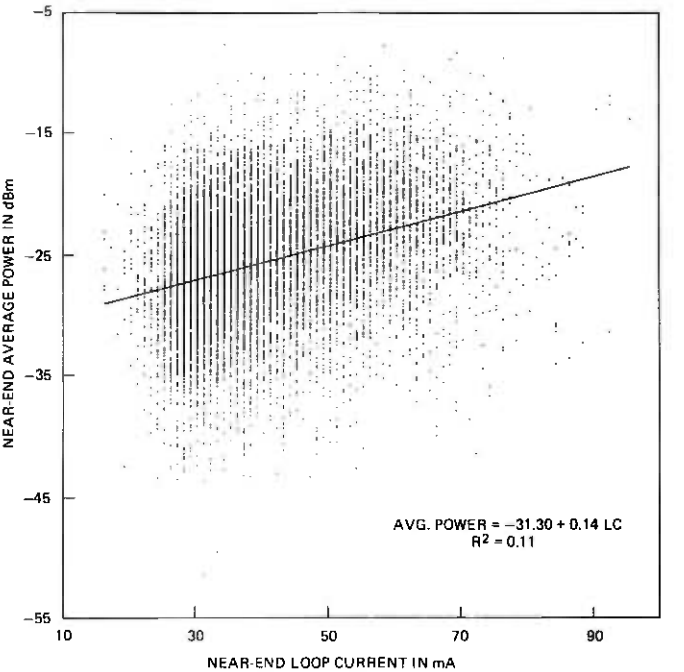


Fig. 11—Linear regression of average power on loop current.

Fig. 15. The correlation coefficients for near-end EPL and loop current are between 0.30 and 0.40 for all four destination categories. As the loop current decreases or increases, the EPL and average power trend lines follow. Since loop currents tend to be higher for business-originated calls (Section 2.3.1) and since the HNPA and FNPA categories of calls have increasingly more business-originated traffic (intrabuilding: 34 percent, interbuilding: 50 percent, HNPA: 59 percent, and FNPA: 69 percent), loop current tends to increase as the call destination becomes more remote. Interbuilding local calls, however, present an exception to this behavior which is not understood. It may be a real deviation from the overall trend, or it may be a random statistical phenomenon. Since the trends are so slight, further investigation of the interbuilding results is not warranted.

Examination of the near-end EPL and average power distributions within the individual call destination categories shows that they are close to normal in all categories except the FNPA category. In the FNPA category, both distributions modestly deviate from normality in the upper 10-percent tail due to a truncation of EPL around 0 dBm and a truncation of average power around −15 dBm. The reason for this truncation is not known; however, it represents a threshold above which speakers rarely drift. In the other call destination categories, 0 and −15 dBm signal

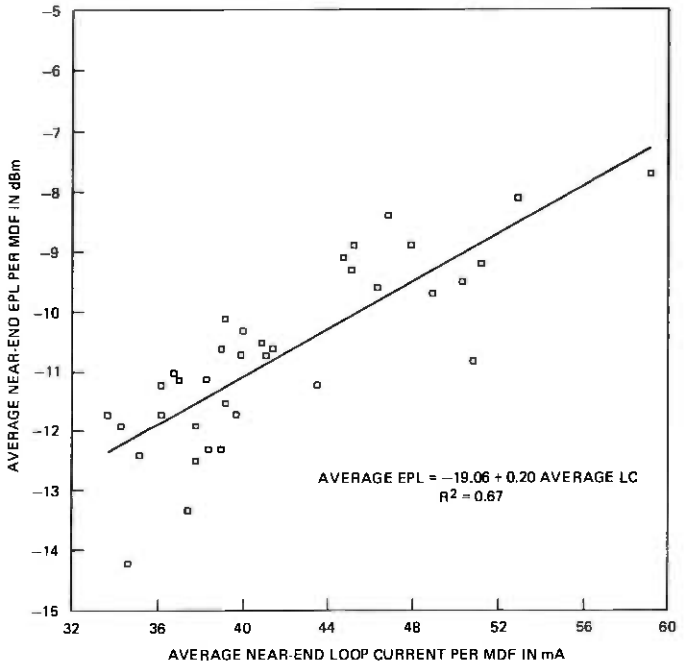


Fig. 12—Linear regression of equivalent peak level on loop current using MDF averages.

powers fall in the highest 1 percent of the EPL and average powers, respectively. The distributions for far-end EPL and average power are essentially normal in all categories.

The far-end signal powers tend to decrease as the call destination becomes more remote from the originating office due to increases in end-office-to-end-office network transmission loss. In the case of intrabuilding local calls where both parties are served by the same local switching office, the only additional network loss encountered by far-end signals is the switching office loss itself. As a result, the near- and far-end signal powers differ only slightly for intrabuilding local calls. These differences increase for interbuilding local calls and HNPA calls, which have similar far-end signal powers, due to an increase in the number of switching offices and trunks involved in the transmission path and the via net loss design[6] adopted for these arrangements of facilities. Likewise, an even greater difference between near- and far-end signal power is observed in the FNPA category. The detailed statistics associated with the trends illustrated in Fig. 15 are listed in Table V.

The correlation between near- and far-end signal powers also appears to depend upon call destination. A comparison of near- and far-end EPL provides correlation coefficients of 0.36, 0.27, 0.23, and 0.14 for intrabuilding, interbuilding, HNPA, and FNPA calls, respectively. The cor-

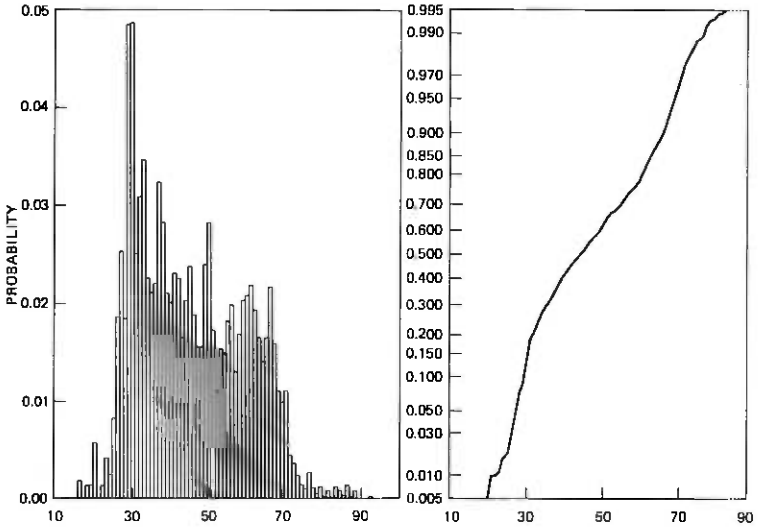


Fig. 13—Near-end loop current (mA) distribution for business.

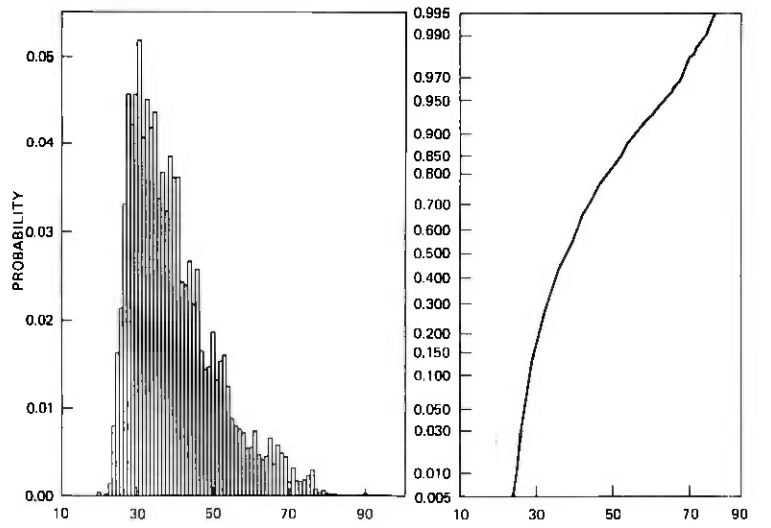


Fig. 14—Near-end loop current (mA) distribution for residential.

relation becomes poorer as the call destination becomes more remote because of the overall increasing and opposite impacts of network transmission loss and loop current on far-end and near-end signal powers, respectively.

The intrabuilding and interbuilding local call data were pooled to obtain overall local results, and the HNPA and FNPA data were pooled

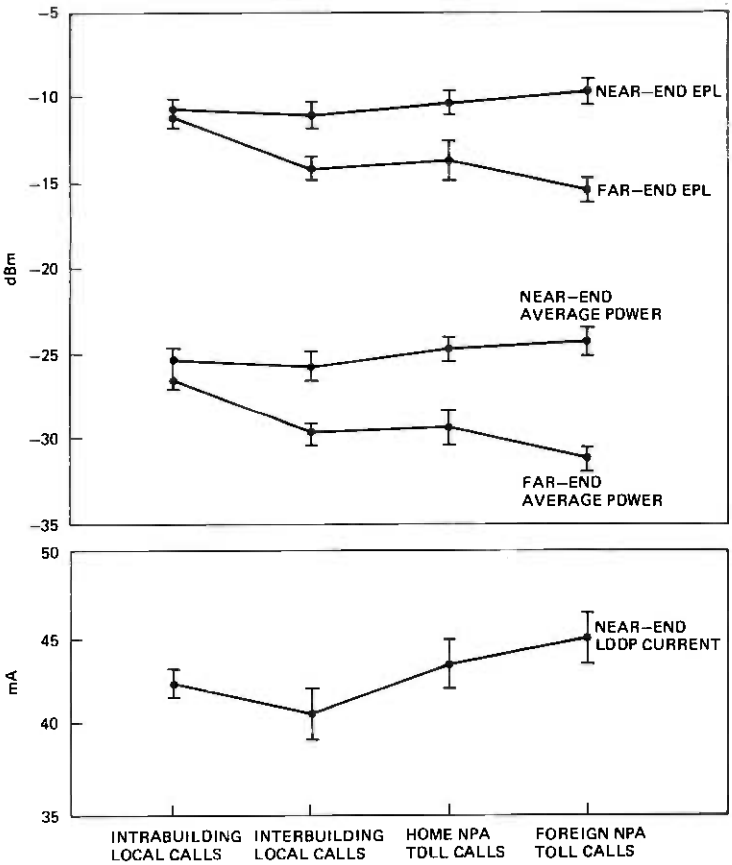


Fig. 15—Relationship of call destination to signal power and loop current.

to obtain overall toll results. Briefly, near-end toll signal powers are slightly, but not significantly, higher than near-end signal powers for local calls, and far-end toll signal powers are significantly lower than far-end powers for local calls. The reasons for these characteristics are discussed above. The only additional observation at this point is that the local loop current distribution resembles the residential distribution in Fig. 14 and the toll loop current distribution resembles the business distribution in Fig. 13. The dominance of residential and business traffic for local and toll calls, respectively, is responsible for these similarities.

2.5 Additional speech signal power analyses

The signal power data were also analyzed to determine the impact of call distance, local switch type, and several demographic factors upon

Table V—Call destination speech signal power results

Transmission Characteristic	Intrabuilding Local Calls				Interbuilding Local Calls				Home NPA Toll Calls				Foreign NPA Toll Calls			
	Mean	90% C.I.	Std. Dev.	Sample	Mean	90% C.I.	Std. Dev.	Sample	Mean	90% C.I.	Std. Dev.	Sample	Mean	90% C.I.	Std. Dev.	Sample
Near-end EPL (dBm)	−10.7	±0.5	4.6	3697	−11.0	±0.7	4.7	3704	−10.3	±0.6	4.5	995	−9.8	±0.7	4.3	763
Far-end EPL (dBm)	−11.0	±0.5	4.8	3348	−14.1	±0.4	4.8	3140	−13.6	±1.1	5.5	862	−15.4	±0.6	4.7	645
Near-end average power (dBm)	−25.3	±0.5	5.3	3697	−25.7	±0.8	5.5	3704	−24.7	±0.7	5.2	995	−24.2	±0.8	4.8	763
Far-end average power (dBm)	−26.5	±0.5	5.4	3348	−29.7	±0.5	5.2	3140	−29.3	±1.0	5.7	862	−31.1	±0.6	4.9	645
Near minus far-end EPL (dB)	0.2	±0.3	5.3	3170	3.2	±0.6	5.8	2980	3.6	±0.9	5.8	828	5.8	±1.0	6.1	617
Near minus far-end average power (dB)	1.0	±0.3	6.2	3170	3.9	±0.6	6.6	2980	4.9	±0.7	6.5	828	7.0	±1.1	6.7	617
Near-end EPL minus average power (dB)	14.6	±0.1	2.1	3697	14.7	±0.1	2.1	3704	14.4	±0.2	1.9	995	14.4	±0.3	1.9	763
Far-end EPL minus average power (dB)	15.6	±0.1	2.1	3348	15.6	±0.1	2.1	3140	15.7	±0.3	2.0	862	15.7	±0.3	2.1	645
Near-end loop current (mA)	42.4	±1.7	12.4	3875	40.7	±3.0	12.4	3864	43.6	±2.9	13.6	1029	45.1	±3.0	13.8	791

speech signal power. Call distance is defined as the airline distance between the originating and terminating local switching machines. Near-end signal power and loop current do not appear to be correlated with call distance. Far-end signal power is weakly correlated with call distance in a negative sense, due to the increase in network transmission loss which accompanies longer call distances as a result of the via net loss design.[6]

In the second of these analyses, the data were classified by originating local switching machine type. No significant relationship was found between machine type and near-end signal power.

Three demographic factors were considered in the third analysis. The first factor, geographical location, does not play an important role in determining speech signal power. While the average near-end signal power is highest in the northeast section of the country and lowest in the southwest, the range of the differences is only 2.7 dB, and the correlation between loop current and signal power accounts for about 40 percent of the difference between geographic areas. The second factor, city or town population, tends to mask rather than uncover relationships between signal power and population. A more appropriate measure is the population density of the exchange served by the local telephone office. The third demographic factor, locality type, was defined to capture the impact of population density upon speech signal power. Five locality types were considered: downtown areas of large and midsize cities, downtown areas of small towns, outer-urban areas, and suburban areas. Large cities were defined as cities with populations of 100,000 or more people; mid-size cities were defined as cities with populations ranging from 20,000 to 100,000 people; and small towns were defined as cities or towns with populations of 20,000 or less people. The outer-urban classification denotes areas with a mixture of residential dwellings and business establishments on the outlying fringes of large cities, and the suburban classification denotes areas which primarily contain residential dwellings. The average near-end EPL and loop current both exhibit the same trends with locality types. Both are highest for downtown MDFs in large cities and lowest for outer-urban areas. These results correlate with the fact that in the first case the population of customers is rather concentrated, and they tend to have relatively short loops, while in the second case the population of customers is rather widespread, and they tend to have relatively long loops. Between these extremes, the average EPL and loop current for small towns are higher than for mid-size cities, and both have higher averages than suburban areas. As illustrated in Figs. 16 and 17, the differences among the categories are not large; however, they do suggest a dependence of loop current and, as a result, EPL upon varying densities of populations.

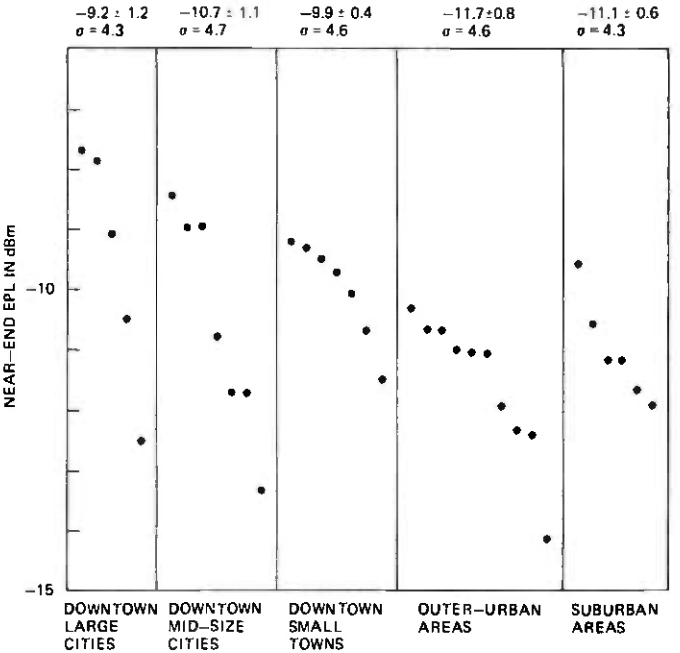


Fig. 16—Average signal power per MDF by locality type.

III. METHODOLOGY

3.1 Statistical survey sample plan

The Loop Signal Power Survey sampling plan consists of three major components—a precise definition of the target population and parameters, a scheme for the selection and measurement of a sample of calls, and the choice of the estimation formulas. Section 3.1.1 defines the target population and parameters, Section 3.1.2 describes the scheme used to select and measure a statistical sample of calls, and Section 3.1.3 describes the statistical estimation and confidence interval formulas used to estimate the target parameters.

3.1.1 Target population and measured parameter definitions

The target population consists of voice calls originating over the public switched network where the subscriber's loop is classified as business, single party residence, coin semipublic, Private Branch Exchange (PBX), or Centralized Exchange (centrex) service. The aggregate of subscriber loops in the target population are naturally partitioned according to the local MDF in which they terminate. In addition, the subscriber loops terminating in an MDF are naturally dichotomized into a customer-provided equipment (CPE) substratum and a Bell equipment substratum. A loop was identified as belonging to the CPE substratum when

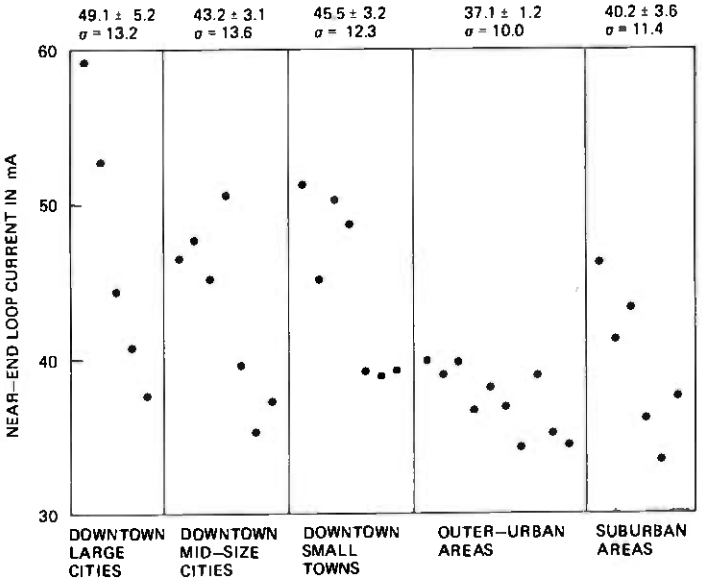


Fig. 17—Average loop current per MDF by locality type.

the local operating company billing records and a follow-up station verification identified the connection to the subscriber's loop of a protective connecting arrangement (PCA) listed in Table VI. A PCA is designed to interconnect non-Bell terminal equipment with the Bell System public switched network.

For potential statistical advantage, the MDFs were partitioned into 12 strata according to the average 1970 population census of the communities within the plant district where an MDF was located. The 12 strata were constructed so that they are approximately the same size with respect to the total number of business, residence, PBX, centrex, coin semipublic, and switched data telephone lines terminating on MDFs within the stratum. This form of stratification was suggested by the results of the 1960 Speech Volume Survey, which indicated a correlation of speech volume with city population. Stratification by city size offered the potential for reduction of the variability in speech signal power within

Table VI—Protective connecting arrangements (PCA)

PCA USOC*	Associated Non-Bell Terminal Equipment
STP	Key telephone system
STC	Single line set
C2ACP	Single line or key telephone system
CD8, CDH	PBX or centrex CU
CDA, CD1, CD7, CD9	Cord switchboard or console

* USOC—Bell System Universal Service Order Code.

each of the strata and, as a result, an increase in the precision of estimates of the mean signal power.

In general, the choice of the criterion for stratification is arbitrary and does not affect the validity of the final survey conclusions; however, a judicious choice of a stratification scheme can lead to an estimate of the mean with a smaller confidence interval than would be obtained otherwise.

Each loop associated with the target population is indexed by its stratum number, MDF number within a stratum, substratum number (e.g., 1 Bell, 2 CPE) and loop number within an MDF substratum.

The target population parameters estimated in the Loop Signal Power Survey are defined by the ratio

$$R = Y/X,$$

where

$$Y = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{a=1}^{D_{hi}} \sum_{j=1}^{M_{hia}} \sum_{k=1}^{Q_{hiaj}} Y_{hiajk},$$

X is defined similarly to Y with Y_{hiajk} replaced by X_{hiajk},

N_h is the number of MDFs located in class 5 offices in stratum h, for $h = 1, 2, \ldots, L$,

D_{hi} is the number of substrata into which the subscriber loops that terminate in the ith MDF of stratum h are partitioned ($D_{hi} = 2$),

M_{hia} is the number of subscriber loops that are in substratum a and terminate on the ith MDF in stratum h,

and

Y_{hiajk} and X_{hiajk}, $k = 1, 2, \ldots, Q_{hiaj}$, represent measurements associated with the Q_{hiaj} completed calls which originate from loop $(hiaj)$. Loop $(hiaj)$ is identified as the jth loop terminating in substratum a of the ith MDF in stratum h.

Some examples of applications of the ratio parameter R are given below.

Application One: Fraction of Calls Where the Mean Transmitted Signal Power Exceeds Some Threshold

Suppose Y_{hiajk} is defined as 1 if the kth completed call on loop $(hiaj)$ is in the target population and the mean signal power exceeds some threshold T, and 0 otherwise. Second, suppose X_{hiajk} is defined as 1 if this call is in the target population, and 0 otherwise. R is then equal to the fraction of completed calls in the target population for which the transmitted mean signal power exceeds T. This form of the ratio parameter is applicable to target populations such as completed calls (toll and/or local) originating from the Bell and/or CPE subclasses of loops.

Application Two: The Mean Originating Signal Power Per Call

Suppose X_{hiajk} is defined as in Application One, and Y_{hiajk} is defined as a measure of signal power of the kth completed call originating on loop $(hiaj)$, then R is equal to the mean originating signal power per call.

3.1.2 Survey sampling scheme

The calls which were measured in the Loop Signal Power Survey were statistically selected in such a way as to permit precise estimates of the population parameters described in Section 3.1.1 and at the same time limit the costs of obtaining the measurements. The actual statistical sample selection scheme used was a classical three-stage sampling scheme with stratification and substratification. From each of the 12 strata described in Section 3.1.1, three MDFs were selected with probabilities of selection proportional to estimates of the total number of business, residence, PBX, centrex, and coin semipublic lines terminating on each MDF. The locations of the 36 sampled MDFs are illustrated in Fig. 18. A stratified random sample of CPE and Bell loops, which terminated on the 36 MDFs, was selected, specially designed measurement equipment was connected to these sampled loops, and signal power measurements were made on a sample of calls originating over the loops. The selection of the CPE loops was made from a billing records inventory of subscriber telephone numbers that were being billed for a PCA with one of the Universal Service Order Codes (USOC) listed in Table VI. A random sample of Bell loops was obtained by generating a list of random four-digit numbers and prefixing a local three-digit NNX code for each NNX associated with the MDF. These lists were forwarded to the local repair service bureau for determination of the class of service of each

Fig. 18—Locations of sampled MDFs.

telephone number and the location of the loop on the MDF. A stratified random sample of CPE and Bell loops, identified as members of the target population, was ordered according to the location on the vertical side of the MDF. Approximately 1 week prior to the scheduled arrival of the Bell Laboratories survey team, a verification was made by local operating company craft people to assure that each selected line was working, that the telephone number-cable-pair and horizontal frame assignments were correct, and that no bridged lines were present. From this verified list, a stratified sample of up to 30 CPE loops and at least 69 Bell loops (for a total of 99) were selected for connection to the survey equipment. The equipment included a device which, when activated, scanned the 99 loops for an originating off-hook signal. Following seizure of the loop and the establishment of a connection, the measurement process was started manually if a conversation ensued. Conversation was detected by utilizing an equipment operator's monitor channel which provided unintelligible speech during periods of conversation through the use of a low speech sampling rate. Because toll calls were relatively scarce, provision was made for the equipment operator to abort the measurement of local calls to obtain additional toll calls. The measurement period in a local office was 3 days.

The survey equipment provided peg count data from which the number of originated completed calls was estimated for each loop. These data formed the basis for traffic weights used to estimate the target population parameters.

3.1.3 Estimation formulas and confidence intervals

This section is devoted to a discussion of the statistical estimation formulas that are used to estimate the ratio parameter R. These formulas are tailored to the survey sample design discussed in Section 3.1.2. The form of the estimation formulas require the following information relative to the sampling plan:

n_h—the number of sampled MDFs in primary stratum h for $h = 1, 2, \ldots, L$. ($n_h = 3$ for $h = 1, 2, \ldots 12$).

z_{hi}—the probability of selection into the first stage sample of the ith sampled MDF in stratum h for $i = 1, 2, \ldots, n_h$ and $h = 1, 2, \ldots, L$.

m_{hia}—the number of measured subscriber loops that belong to the ath substratum of the ith sampled MDF in stratum h for $i = 1, 2, \ldots n_h$; $a = 1, 2, \ldots D_{hi}$, and $h = 1, 2, \ldots, L$.

q_{hiaj}—the number of calls associated with loop ($hiaj$) on which signal power measurements were made.

L, M_{hia}, D_{hi} and Q_{hiaj} are defined as in Section 3.1.1, and (x_{hiajk}, y_{hiajk}), $k = 1, 2, \ldots, q_{hiaj}$ represents a sample of q_{hiaj} values of (X_{hiajk}, Y_{hiajk}), $k = 1, 2, \ldots Q_{hiaj}$, where X_{hiajk} and Y_{hiajk} are defined as in the definition of R.

A three-stage estimator of the ratio $R = Y/X$ where Y and X are defined as in Section 3.1.1 is

$$r = y/x,$$

where

$$y = \sum_{h=1}^{L} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hia}} \sum_{k=1}^{q_{hiaj}} w_{hiaj} y_{hiajk},$$

$$w_{hiaj} = \frac{1}{n_h} \frac{1}{z_{hi}} \frac{M_{hia}}{m_{hia}} \frac{Q_{hiaj}}{q_{hiaj}},$$

and x is defined similarly to y with y_{hiajk} replaced by x_{hiajk}.

The mean squared error of r is defined as

$$\text{VAR}(r) = E(r - R)^2,$$

where $E(\cdot)$ denotes expected value.

A consistent estimator of VAR(r) is

$$v(r) = \frac{1}{x^2} \sum_{h=1}^{L} \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} \left[\frac{y_{hi} - r x_{hi}}{z_{hi}} - \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{y_{hi} - r x_{hi}}{z_{hi}} \right]^2,$$

where

$$y_{hi} = \sum_{a=1}^{D_{hi}} \sum_{j=1}^{m_{hia}} \sum_{k=1}^{q_{hiaj}} \frac{M_{hia}}{m_{hia}} \frac{Q_{hiaj}}{q_{hiaj}} y_{hiajk}$$

and x_{hi} is defined similarly to y_{hi} with y_{hiajk} replaced by x_{hiajk}.

An application of the Central Limit Theorem yields an approximate 90-percent confidence interval for R as the interval

$$(r - 1.645\sqrt{v(r)}\,,\ r + 1.645\sqrt{v(r)}).$$

3.2 Data acquisition plan

In this section, requirements pertaining to acquisition equipment capacity, compatibility, transparency, privacy, etc., are summarized, and a block diagram of the Loop Signal Power Survey acquisition equipment is discussed.

3.2.1 Requirements

As indicated in Section 3.1, the sample plan called for access to 99 customer loops in each of 36 class 5 offices and measurements of near- and far-end signal power on live calls. Determination of call destination required the detection of call originations on loop start and ground start lines, and the detection of dial pulse and *TOUCH-TONE*® address information. Because of the loop-to-loop and call-to-call variability in impedance at the MDF interface, the measurement of real power was required rather than bridged voltage. In the course of accessing and

measuring calls, no detectable impairment (loss or switching clicks) was to be added to the connection. Monitoring of intelligible speech was prohibited by privacy considerations. Speech signals are predominantly half-duplex in nature; however, both parties sometimes talked at the same time. Because the point of measurement was a two-wire point, it was necessary to devise a method to sort the speech signal data into two categories, near-end and far-end.

3.2.2 Data acquisition equipment

Figure 19 is a block diagram of the equipment used to acquire speech signal power data. The 99 customer loops were accessed at the protector socket of the MDF. Access cables connected the customers' loops to the acquisition console protector panel. This panel provided series access to 99 loops, circuit protection, and an electrical interface with the instrumentation switch. This interface contained current sensing resistors for the detection of metallic speech current and loop dc current. Modified service observing equipment was bridged across the tip-ring interface at this point to allow the detection of outgoing call seizures and the de-

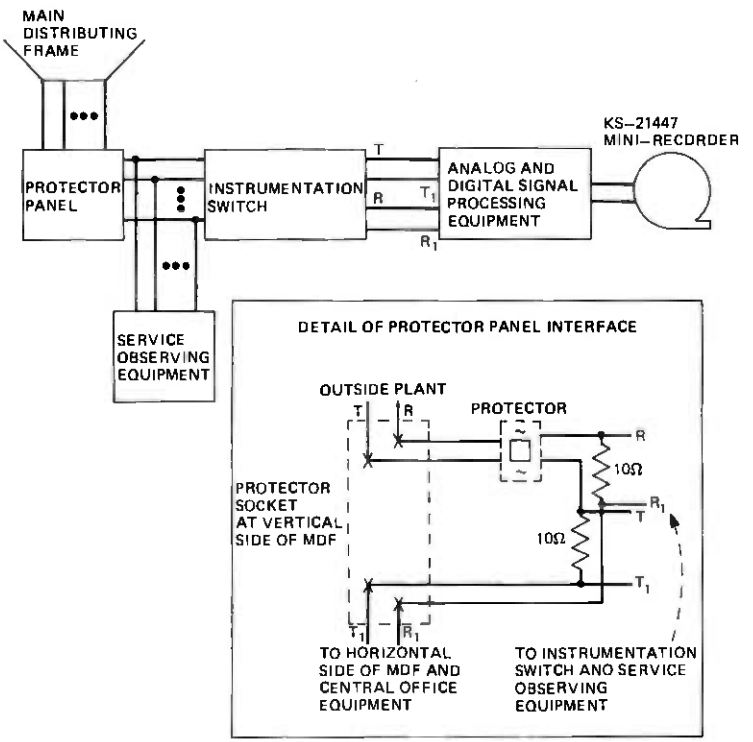


Fig. 19—Loop signal power survey data acquisition console.

tection of dial pulse/*TOUCH-TONE* address digits. The instrumentation switch connected the four leads associated with the current sensing resistors of one of the 99 loops to the analog signal processing equipment for the detection, amplification, and filtering of the metallic speech voltage and current.

The resulting voltage and current signals were simultaneously sampled at the rate of 200 samples per second using dual 12-bit A-D converters. The sampled data were stored in a buffer memory, combined with label information, and written in 16-kb blocks on a minirecorder magnetic tape unit. A paper-tape printer recorded off-hook event times for each of the 99 loops so that traffic weights referred to in Section 3.1 could be determined. In addition, the dialed area and office code were recorded on the tape. The digitally recorded speech signal data were subsequently analyzed in a manner described in the next section.

The loss due to the current sensing resistors and bridged equipment was negligible. This, combined with click supression circuitry, made the measurement equipment transparent from the customer's point of view. The low rate of sampling made the recorded speech signal unintelligible but allowed the recovery of pertinent signal power information. A low speech sampling rate was also used to make the equipment operator's monitor channel unintelligible, yet permit the identification of call progress signals. The acquisition of simultaneous speech voltage and current samples permitted the discrimination of the near-end from the far-end talker in a manner discussed in the next section.

3.3 Analysis of data

This section explains how voltage and current samples were processed to obtain measures of speech signal power for each talker in the two-way conversations.

3.3.1 Raw speech signal power data processing

The raw data upon which speech signal equivalent peak level (EPL) and average power estimates are based consisted of metallic speech voltage and current samples. The metallic speech voltage and current on the loop were amplified and filtered to exclude signals higher than 4 KHz and remove the effect of 60 Hz, its first two odd harmonics, and low frequency noise below 100 Hz. The resultant voltage and current analog signals were then simultaneously sampled at the rate of 200 samples per second using two 12-bit linear A-D converters. The digital sampled data were then recorded on tape cartridges, which were later reformatted onto standard computer tape.

The first step in computer processing of the digitally recorded signals consisted of removal of dc bias produced in the analog signal processing filters and computation of the instantaneous power (watts) associated

with each voltage-current sample pair. The equipment was designed so that the power values were positive (voltage and current in phase) when the signal source was the near-end talker and and negative (voltage and current out of phase) when the signal source was the far-end talker.

3.3.2 Discrimination of near-end and far-end talkers

Conversational speech is predominantly half-duplex, but brief periods occur when both talkers are active at the same time. The stream of instantaneous power samples is therefore positive or negative for half-duplex talk-spurts. However, during double talking, the sign of the power samples may change rapidly and the magnitudes of the power samples become useless for estimation of near-end or far-end talker power. To properly sort the power sample stream into two distinct "bins" corresponding to the near-end and far-end talkers, empirical algorithms were developed in laboratory simulations, and one algorithm (SGN algorithm) was chosen for use during the speech signal processing phase of the survey.

The SGN algorithm uses the sign and magnitude of the power in short subsequences of the stream of speech power samples to generate two sequences of speech power samples corresponding to near-end and far-end talkers.

Let $\{p\}$ be the sequence of instantaneous speech signal power values computed from the relationship: $p = v \cdot i$, where $\{v\}$ and $\{i\}$ are sequences of instantaneous, simultaneous samples of speech signal metallic voltage and current, respectively.

Let the sequence $\{p\}$ be divided into consecutive subsequences of length l. Associated with the ith subsequence is the average power:

$$\overline{p}_i = \frac{1}{l} \sum_{k=il-l+1}^{il} p_k.$$

$$\text{Let } \mathrm{SGN}(\overline{p}_i) = \begin{cases} -1 \text{ if } \overline{p}_i > 0 \\ \ \ 0 \text{ if } \overline{p}_i = 0 \\ +1 \text{ if } \overline{p}_i < 0. \end{cases}$$

The SGN algorithm depends on two conditions for every subsequence:

Condition 1: $\mathrm{SGN}(\overline{p}_i) = \mathrm{SGN}(\overline{p}_{i-1})$

Condition 2: $|\overline{p}_i| \geq \alpha |\overline{p}_{i-1}|$.

If either condition is true, then SGN $(\overline{p}_i)$ determines the sources of the speech signal for the ith subsequence. As stated earlier, the sign convention is such that a positive value indicates that the near-end talker is the source (far-end samples set to 0), and a negative value indicates that the far-end talker is the source (near-end samples set to 0). After the source is determined, the nonzero power samples are set positive and placed in the appropriate (near- or far-end) sequence.

If neither of the above conditions is true, then the direction is indeterminant and all power samples in the ith subsequence are set to 0. Laboratory investigations established that the values $l = 2$ and $\alpha = 10$ give good performance with the sample rate used in the survey (200 samples per second). The output from the SGN algorithm consists of two sequences of positive instantaneous signal power samples representing the near-end and far-end talkers.

3.3.3 Measures of speech signal power

Two measures of speech signal power are developed from each of the near-end and far-end sequences described above. The first measure is the average speech signal power defined over the observation interval (generally about a minute) as follows:

$$\text{Near-end average power} = 30 + 10 \log \frac{1}{n} \sum_{k=1}^{n} p_k \text{ near-end (dBm)}$$

$$\text{Far-end average power} = 30 + 10 \log \frac{1}{n} \sum_{k=1}^{n} p_k \text{ far-end (dBm)},$$

where p_k-end represents the elements in the sequence of instantaneous power samples for the direction of interest, and n is the total length of the power sample sequence.

The second measure used to characterize speech signal power is an estimate of the peak power in the distribution of samples of talker signal power. The estimator is the empirical equivalent peak level (EPL), developed by Brady. A complete discussion of the EPL and its properties is given by Brady in Ref. 2. The EPL is developed from the power sample sequence for the direction of interest as follows.

Let the instantaneous power of the kth sample be defined as:

$$p_k = v_k i_k \text{ watts.}$$

In logarithmic units,

$$p_k = 10 \log p_k \text{ (dBw).}$$

Define a threshold ϕ and multiplier δ_k so that:

$$\delta_k = \begin{cases} 1 & \text{if } p_k > \phi \\ 0 & \text{otherwise} \end{cases}.$$

The average power over threshold is defined:

$$\overline{p}_\phi = 10 \log \left(\frac{\sum_{k=1}^{n} p_k \delta_k}{\sum_{k=1}^{n} \delta_k} \right).$$

Now define $D = \bar{p}_\phi - \phi$ dB. From D compute Δ using the following empirical rule:

$$D \leq 6.75, \text{ then } \Delta = (D - 2.75)/0.4$$

$$6.75 < D \leq 13.5, \text{ then } \Delta = D/0.675$$

$$13.5 < D, \text{ then } \Delta = (D + 2.88)/0.819.$$

From Δ compute EPL as:

$$\text{EPL} = \Delta + \phi.$$

Some important properties of the EPL are that it is independent of the talker's activity since it is not affected by the silent periods in the conversation, and its estimate varies little over a wide range of threshold values. Some laboratory investigations indicate that a threshold of 10 to 20 dB below EPL gives good performance in the presence of noise; a threshold of 20 dB below EPL was selected as giving the best noise rejection without discarding an excessive number of samples. The EPL computation was iterated until the threshold was 20 ± 3 dB below the EPL value.

IV. COMPARISON WITH PREVIOUS DATA

In 1960, measurements of talker volume were made on live traffic using VU meters.[1] These measurements of talker volume are compared with the current survey results, which have been translated from EPL to VU using an empirical correction factor. These results are listed in Table VII together with the 1960 survey results.

The 1960 survey results differ substantially from the current results in that the toll volumes were substantially higher in 1960 and the ranges of volumes within the various call destination categories were substantially greater. There have been some substantial changes in the telephone plant since 1960 that may help to explain these differences. The proportion of toll grade battery has decreased substantially, resulting in a decrease in toll call speech volume. Loss plan improvements, the phasing out of the 300-type telephone set, and the growth of direct trunking have all tended to increase the uniformity of service in the network and make it more transparent to customers. The apparent result is a network with remarkable uniformity of speech signal power.

Table VII—Comparison with 1960 speech volume survey

	1960		1975–1976	
Call Destination	Average VU	Std. Dev.	Average VU	Std. Dev.
Intra-building	−24.8	7.3	−22.2	4.6
Inter-building	−23.1	7.3	−22.5	4.7
Toll	−16.8	6.4	−21.6	4.5

V. ACKNOWLEDGMENTS

Many individuals contributed to the Loop Signal Power Survey. P. W. Freeman, J. M. MacMaster, and E. J. Vlacich designed and assembled the test equipment, T. W. Thatcher, Jr. handled all phases of logistic support, and F. Grizmala, S. Vitale, and P. R. Wild developed software packages and processed the data with the help of C. Keinath. In addition, many individuals from Bell Laboratories and the Bell System operating companies participated in the data collection phase of the survey. We extend our thanks to all who have helped in this project.

REFERENCES

1. K. L. McAdoo, "Speech Volumes on Bell System Message Circuits—1960 Survey," B.S.T.J., *42*, No. 5 (September 1963).
2. P. T. Brady, "Equivalent Peak Level: A Threshold-Independent Speech-Level Measure," J. Acoust. Soc. Am., *44*, No. 3 (September 1968).
3. Bell System Center for Technical Education, *Telecommunications Transmission Engineering*, Vol. 1, Chapter 12, 1974.
4. D. H. Merchant, unpublished works.
5. P. A. Gresh, "Physical and Transmission Characteristics of Customer Loop Plant," B.S.T.J., *48*, No. 10 (December 1969).
6. *Notes on Distance Dialing*, Section 6, New York: American Telephone and Telegraph Company, 1975.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 57, No. 7, September 1978
Printed in U.S.A.

An Adaptive PCM System Designed for Noisy Channels and Digital Implementations*

By DEBASIS MITRA and B. GOTZ

(Manuscript received November 29, 1977)

We propose a new adaptive quantization scheme for digitally implementing PCM and DPCM structures. The arithmetics we develop for the digital processing are useful as well in the implementation of previously existing schemes for adaptive quantization. Two objectives are stressed here: (i) The system must be robust in the presence of noise in the transmission channel which causes the synchronization between quantizer adaptations in the transmitter and receiver to deteriorate. (ii) It must also minimize the complexity of the digital realization. In addition to the above objectives, we require, of course, good fidelity of the processed speech waveform. The problem of synchronization in digital implementations where the constraint of finite precision arithmetic exists has not been addressed previously. We begin by examining an existing, idealized adaptation algorithm which contains a leakage parameter for the purpose of deriving robustness. We prove that, to provide the necessary synchronization capability without impairing the quality of speech reproduction, it is necessary to use a minimum, unexpectedly large, number of bits in the machine words and, additionally, to carefully specify the internal arithmetic, as is done here.

The new scheme that we propose here uses an order of magnitude less memory in an ROM-based implementation. The key innovations responsible for the improvement are: (i) modification of the adaptation algorithm to one where leakage is interleaved infrequently but at regular intervals into the adaptation recursion; (ii) a specification of the internal machine arithmetic that guarantees synchronization in the presence of channel errors. A detailed theoretical analysis of the statistical behavior of the proposed system for random inputs is given here. Results of a simulation of a realistic 16-level adaptive quantizer are reported.

* A short version of this paper was presented at the International Conference on Communications, Toronto, June 1978.

I. INTRODUCTION

We propose a new scheme for adaptive quantization which is particularly well suited to the digital implementation of PCM and DPCM structures. In the course of this work, we have developed arithmetics for the digital processing that are useful as well in the implementation of previously existing schemes for robust quantization.

The exacting requirements on adaptive quantization stemming from the broad dynamic range and rapid transient behavior of speech are well known. Two additional objectives are given equal importance here: (*i*) To make the system robust in the presence of channel errors. Thus, while channel errors may cause the quantizer adaptations in transmitter and receiver to be put out of synchronization,* a mechanism must exist which acts to rapidly restore the synchronization during periods of error-free transmission. (*ii*) To minimize the complexity of the digital realization; specifically, to minimize the length of the internal words in the digital processors and to facilitate the multiplexing of the hardware.

Systems do exist in the literature for robust quantization in the presence of noisy channels; one such system is described below in some detail. However, the problem of synchronizing the quantizer adaptations in the transmitter and receiver in digital implementations, where the constraint of finite precision arithmetic exists, has not been addressed previously. We prove that, to provide the necessary synchronization capability without impairing the quality of speech reproduction, it is necessary to use an unexpectedly large number of bits in the internal words of the digital processors at both sites and, additionally, to carefully specify the internal arithmetic (which we do). If the digital processing is implemented using ROMs, as is being proposed, the long internal word length is reflected in large memory requirements and therefore costly implementations as well as exposure to new errors in the processing.

The scheme that we propose here uses an order-of-magnitude less memory in an ROM-based implementation in both the transmitter and receiver. This is for comparable performance with respect to loading characteristic, signal-to-noise ratio, and the synchronization capability. Another advantage not reflected in the above estimate is the fact that the essential costly digital component, the ROM, as distinct from other less costly components such as adders, is used only for a small fraction of the total operating time. Thus, further economies may be effected through multiplexing the ROM. The key innovations are: (*i*) the modification of the adaptation algorithm which allows the internal word length of the digital processors to be reduced significantly; and (*ii*) a specification of the internal arithmetic that guarantees synchronization in the presence of channel errors. As mentioned previously, the arithmetic is also applicable in digital implementations of previously existing adaptation algorithms.

* In our usage, synchronization is synonymous with tracking.

A byproduct of the work reported here is that it establishes a link between two hitherto unconnected areas, namely, finite-arithmetic digital signal processing and waveform quantization in the presence of a noisy channel. The problem of synchronizing two geographically separated digital processors gives rise to quite novel requirements on the processing, and we expect that the problem will be a subject of further investigation in the future.

The paper is organized as follows. In Section 1.1 we describe an existing quantizer adaptation scheme and the associated synchronization problem. Section II is devoted to the basic description of the new scheme. Section 2.1 introduces the key idea underlying the scheme. Section 2.2 considers the digital implementation of the system, and Section 2.3 considers the synchronization behavior of the resulting system. Section III is devoted to the probabilistic analysis of the behavior of the proposed algorithm. The basic notions of the bias functions, central log step sizes, and load curves are introduced, and the qualitative results proved in their connection are stated. In Section IV, some computational results are presented in the context of a realistic 16-level quantizer that has been proposed and investigated previously in connection with an industrial application. We try to illuminate the topics considered in Sections II and III through examples involving this particular quantizer. Four appendices to the paper present the detailed technical derivations.

On account of the length of the paper, we considered it desirable to include a final section, Section V, which summarizes and puts into perspective the key results obtained in the preceding sections.

We should mention that the digital implementation of adaptive DPCM systems is under investigation within Bell Laboratories in connection with TASI-D, subband voice coding, and new channel banks. The work reported here is a research study and not a description of a developed design.

1.1 Background and description of the problem

We begin by describing a system proposed in Ref. 1 which, unlike earlier systems upon which it is based,[2–5] possesses the capacity to recover from past channel errors during periods of error-free transmission.

1.1.1 An existing idealized scheme for robust quantization

Let $\Delta(i)$ (see Fig. 1) denote the *step size* of a quantizer, with $2N$ levels, at the ith sampling instant; $\Delta(i)$ is adapted according to the rule

$$\Delta(i+1) = \Delta(i)^{\beta} M(i), \qquad i = 0,1,2,\cdots \tag{1}$$

where $\beta, 0 < \beta < 1$, is the leakage constant and $M(i)$ is the *multiplier* at time i. $M(i)$ is selected from a prespecified collection of multipliers $\{M_1, M_2, \cdots, M_N\}$ according to the rule:

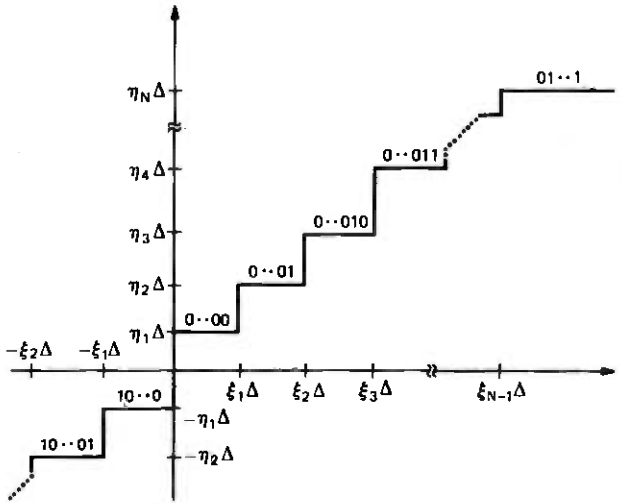


Fig. 1—The quantizer. A natural coding scheme is displayed. The step size is time-varying and the parameters (ξ_n) and (η_n) are prespecified and fixed.

$$\text{If } \xi_{r-1}\Delta(i) \le |x(i)| < \xi_r\Delta(i), \text{ then } M(i) = M_r, \tag{2}$$

where $x(i)$ is the input signal variable (speech or data) at time i and $0 = \xi_0,\xi_1,\cdots,\xi_{N-1},\xi_N = \infty$ are fixed, ordered parameters of the quantizer,* Fig. 1. The multipliers are also ordered, i.e.,

$$M_1 \le M_2 \le \cdots \le M_N.$$

It is widely recognized[6,7] that (1) is not in a form convenient for implementation, even analog implementation. To utilize conventional multipliers, it is necessary to work with the log-transformed version of (1).

Denote the *log step size* by $d(i)$, where

$$d(i) \triangleq \log_Q \Delta(i), \tag{3}$$

Q being a fixed number greater than 1, and the *log multipliers* by

$$m(i) \triangleq \log_Q M(i),\ m_r \triangleq \log_Q M_r,\ 1 \le r \le N. \tag{4}$$

Also let

$$\bar{\xi}_r \triangleq \log_Q \xi_r, \qquad 1 \le r \le N. \tag{5}$$

Thus, from (1) and (2),

$$d(i+1) = \beta d(i) + m(i), \qquad i = 0,1,\cdots \tag{6a}$$

where

* When the parameters $\{\xi_r\}$ and $\{\eta_r\}$ are spaced equal distances apart, the quantizer is usually referred to as a uniform quantizer and it is natural to call Δ the "step size." However, for nonuniform quantizers, the term "step size" is less natural and other candidates are "scale" and "range." However, since there is no reason for confusion, we retain the familiar term "step size."

$$m(i) = m_r \text{ iff } \bar{\xi}_{r-1} + d(i) \leq \log_Q |x(i)| < \bar{\xi}_r + d(i). \tag{6b}$$

The only information that is coded and transmitted at time i is that concerning the quantizer output which uniquely determines the selected log multiplier $m(i)$. A natural coding scheme is exhibited in Fig. 1. The recursion in (6) is implemented at both the transmitter and receiver. We let $m'(i)$ denote the log multiplier corresponding to the received code word at time i, and we employ the natural notation $d'(i)$ to denote the log step size in the receiver. The reconstruction, $R(i)$, at the receiver of the input signal variable is done according to the rule:

$$\text{If } m'(i) = m_r \text{ then } |R(i)| = \eta_r Q^{d'(i)}, \tag{7}$$

where η_r, $1 \leq r \leq N$, are also prespecified, fixed parameters of the quantizer, as shown in Fig. 1. The sign of the reconstructed value is obtained from the sign bit, usually the first and shown as such in Fig. 1, in the received code word.

The synchronization capability of the system, i.e., the capability possessed by the solutions of the recursions, $\{d(\cdot)\}$ and $\{d'(\cdot)\}$, at the transmitter and receiver to approach each other during error-free transmission is entirely due to the presence of the leakage parameter β. For if $d(0)$ and $d'(0)$ are two, possibly different, initial values of the log-step sizes at the commencement of an epoch of error-free transmission, then during the epoch

$$|d(i) - d'(i)| = \beta^i |d(0) - d'(0)|, \qquad i \geq 0. \tag{8}$$

The notion of introducing leakage as a mechanism for deriving robustness in the presence of a noisy channel is a well-known one in communication practice; witness, the leaky delta-modulator.[8]

As far as the synchronization of the transmitter and receiver adaptations is concerned, eq. (8) implies that decreasing β provides improved quality. However, there is an accompanying price. The data in Fig. 5 of Ref. 1 together with the theory developed here in Sections 3.2 and 3.3 on the load curves (which describe the statistical behavior of the step size for random inputs) show that the statistical dynamic range of the step size is reduced rapidly with decreasing β, with a concomitant deterioration of the quality of the reconstruction.* Recent subjective tests[10] have shown that it is very unlikely that β less than $^{63}/_{64}$ can provide acceptable quality speech reproduction.

Herein lies the gist of the problem: For good quality reproduction, the leakage parameter must necessarily be very close to 1, and this, on the other hand, makes it difficult to provide good quality synchronization. It is thus necessary to walk a narrow path between too small leakage and too large leakage. As we see next, the constraint of finite precision

* Numerous related topics are treated analytically in Ref. 9.

arithmetic imposed by a digital implementation compounds the design problem.

1.1.2 Digital Implementations

Equation (6) assumes continuous values of $d(\cdot)$ and infinite precision arithmetical operations, and hence it can only serve as an ideal in a digital implementation. An all-digital coder will have only a limited dictionary or total number (typically, ≥ 32, ≤ 128) of possible log step sizes. We will consider the log step sizes to be integers varying from 0 to $2^K - 1$; thus, typically, $5 \leq K \leq 7$. It is necessary to introduce the notion of an *internal machine word* with K integer bits and, say, F fractional bits (the need for fractional bits will become apparent shortly); the log step size is obtained from the internal machine word at time i, $y(i)$, by means of an external arithmetic, such as truncation. Although later we will consider other possibilities, for the purpose of this discussion let us assume that the external word at time i, which is the log-step size at that time, is simply the integer part of the internal word at time i, i.e.,

$$d(i) = [y(i)]_{\text{truncate}}, \qquad i = 0,1,2 \cdots . \tag{9}$$

The machine implementation of the ideal recursion in (6) is

$$y(i+1) = \langle \beta y(i) \rangle + m(i), \qquad i = 0,1,2, \cdots , \tag{10}$$

where $\langle \beta y(i) \rangle$ denotes some procedure, such as rounding, for taking $\beta y(i)$ into a $(K+F)$-bit word. It will turn out later that this operation is best viewed with greater generality as a mapping f of $(K+F)$-bit words, with F fractional bits into other such words. Thus we restate (10) as*

$$y(i+1) = f\{y(i)\} + m(i), \qquad i = 0,1,2, \cdots . \tag{10'}$$

It will be assumed that all the log multipliers $\{m_r\}$ have at most F fractional bits each, which ensures that if $y(i)$ is a $(K+F)$-bit word then so is $y(i+1)$.

Figure 2 shows an example of the most direct procedure for generating the discrete map $f(y)$, namely, by rounding βy to the nearest machine word. In the example, considered $F = 1$ so that the spacing between machine words is $2^{-F} = 1/2$. A feature common to such maps is that segments of unit slope are juxtaposed between other segments of zero slope which we call "breaks."

If, as before, we distinguish the quantities associated with the receiver by the superscript $'$, we see that the offset in the machine words behaves

* In (10) and (10′) we have not made allowances for overflow. This however can be done conventionally by employing saturation where:

$$y(i+1) = 0 \text{ if } \langle \beta y(i) \rangle + m(i) < 0,$$
$$= 2^K - 2^{-F} \text{ if } \langle \beta y(i) \rangle + m(i) > 2^K - 2^{-F},$$

and in every other case (10) holds. Saturation acts to attenuate the offset in the machine words at the two sites.

as follows during epochs of error-free transmission [i.e., periods in which $m(\cdot) = m'(\cdot)$]:

$$|y(i+1) - y'(i+1)| = |f\{y(i)\} - f\{y'(i)\}| \tag{11}$$

[compare with (8)].

The synchronization problem motivates us to impose the following two rather stringent requirements on the behavior of the offset.

Synchronization requirements:

(*i*) The offset is nonincreasing at all instants of error-free transmission.

(*ii*) The integer parts of the machine words at the two sites, and hence the respective log step sizes, differ in at most a finite (preferably small) number of time instants during error-free transmission.

We require the above to hold independent of the statistics of the input process. It is clear from (11) that these requirements imply restrictions on the discrete map f which are investigated below.

Let us digress to better motivate the second of the above requirements. If the integer parts of the machine words at the two sites at any instant are not identical, then the respective log step sizes differ by at least unity and, hence, the ratio of the two step sizes is at least Q [see eq. (3)]; this factor may be unacceptably large since values of Q as high as 1.5 are being

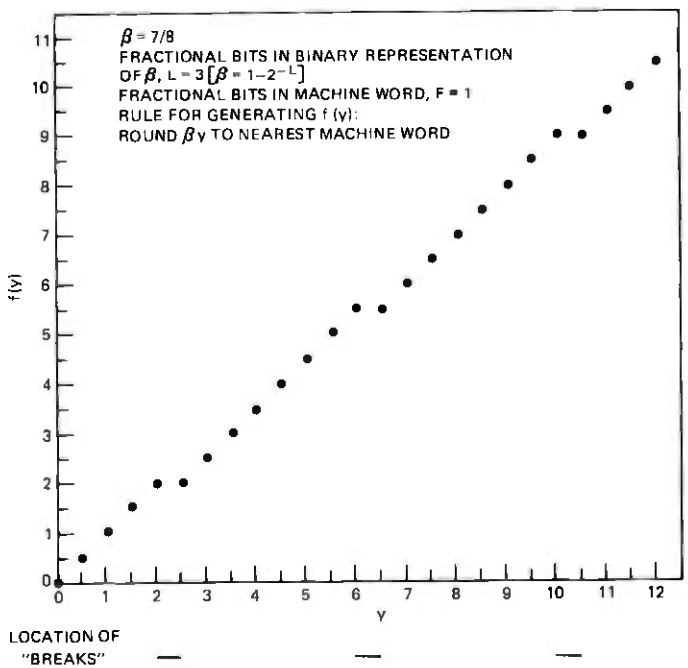


Fig. 2—An example of a naive machine arithmetic.

considered in practical designs.* To illustrate another facet of the second requirement, consider the case where, at a particular instant, the transmitter and receiver machine words are rather close, say, 1.9375 and 2.0625 ($F = 4$). Yet the integer parts are 1 and 2, respectively. Thus the step sizes are Q and Q^2, rather far apart. This example serves to illustrate that the mere proximity of the two machine words is not enough to guarantee that the log step sizes are identical.

In the following discussion, we will need to know the value of L, an integer, which is such that

$$1 - 2^{-L+1} < \beta \leq 1 - 2^{-L};$$

if $\beta = 7/8$, as in Fig. 2, then $L = 3$ and if $\beta = 63/64$ then $L = 6$. To simplify the following discussion, we shall assume that

$$\beta = 1 - 2^{-L}, \tag{12}$$

i.e. $\beta \in \{1/2, 3/4, 7/8, \cdots\}$; with this form for β, L is the minimum number of fractional bits required for the binary representation of β. The assumption on the form of β is unessential, and later in Section 2.2 we indicate that no difficulties are presented if β is not of the assumed form.

We give two different but connected reasons which separately lead to the rather consequential conclusion that $F \geq L$ if the resulting system is to have certain essential properties, including the synchronization capability. The first reason stems directly from the synchronization requirements. We show that the latter requires the map f to incorporate certain contraction properties which in turn can be possible only if the internal machine word has at least L fractional bits. The second related reason is that fewer than L fractional bits gives rise to rounding errors in each iteration of the recursion which makes it hard to predict the effective value of the leakage parameter. Recall from Section 1.1.1 the stringent requirements on the leakage parameter.

Below we amplify both the above arguments. This discussion will motivate a more exact treatment in Section 2.2, which will also provide answers to the questions raised here.

Consider (10′) in conjunction with the synchronization requirements (*i*) and (*ii*). For the first of the synchronization requirements to be satisfied, it is apparent that it is necessary and sufficient that

$$|f(y) - f(y')| \leq |y - y'| \tag{13}$$

for all machine words y and y'. We refer to the above property of the map f as the *weak contraction everywhere* property. The map f shown in Fig.

* This is the case if K is 5 or 6. If K is larger, then it is possible to relax the second requirement by requiring that the offset in the integer parts of the machine words be reduced to a small number (instead of requiring them to be identical). Thus it is possible to trade a higher K for a lower F while keeping $K + F$ fixed. In any case, only minor modifications to the framework that is developed here will allow such cases to be handled.

2 possesses this property by virtue of the fact that the slope of the graph of f is everywhere either 0 (at the breaks) or 1.

For the second of the synchronization requirements to be satisfied, we claim that it is necessary and sufficient that the map f have the following property:

If y and y' are any machine words with different integer parts, then

$$|f(y) - f(y')| \leq \delta |y - y'| \text{ for some } \delta < 1. \tag{14}$$

We call the above the *strong contraction across integer boundaries* property. Sufficiency is clear, since we have that during epochs where the machine words do not have identical integer parts and error-free transmission exists,

$$|y(i) - y'(i)| \leq \delta^i |y(0) - y'(0)|. \tag{15}$$

Conversely, if (14) is not true, then it is easy to construct examples where the integer parts of the two machine words are different at an unbounded number of time instants. Referring to Fig. 2 we see that the graph of f does *not* possess the strong contraction property (14). To illustrate, suppose that initially the two machine words have different integer parts and that both words occur in the range [2.5,6]; we see from the figure that no mechanism exists to prevent the two words from indefinitely remaining in this range and simultaneously having different integer parts.

We will now argue that the above two contraction properties, together with any weak fidelity criterion relating $f(y)$ to βy, implies that $F \geq L$. Observe that the strong contraction property, (14), requires a "break" (see "breaks" in Fig. 2) in the graph of $f(y)$ just prior to *every* integral value of y. Reason: $y = k - 2^{-F}$ and $y = k$, k integral, have different integer parts. Further, if the local slope of the graph of $f(y)$ is not zero, then by virtue of the weak contraction property it is either 1 or -1. Finally, if F fractional bits are used, then each unit interval of y is composed of 2^F intervals of equal length corresponding to that many distinct machine words. These three considerations show that the

$$\text{average slope of the graph of } f(\cdot) \leq \frac{2^F - 1}{2^F} = 1 - 2^{-F}. \tag{16}$$

But $f(y)$ is supposed to approximate βy, $\beta = 1 - 2^{-L}$. Thus, just about any weak fidelity criterion will give that the smallest value of F, which allows the map f to have the properties required of it, is L.

Our second reason is closely related to the aforementioned fidelity criterion. Implicit in a choice of a leakage parameter β with a large number of fractional bits, L, in its binary representation (e.g., $\beta = {}^{63}/_{64}$) is the requirement that the absolute rounding error in each iteration of (10′), $|f\{y(i)\} - \beta y(i)|$, be not larger (at least not by much) than an error in the least significant bit of β, i.e. 2^{-L}:

$$|f(y) - \beta y| < 2^{-L}, \qquad \text{for all machine words } y. \tag{17}$$

Otherwise, there is no *a priori* need to specify β to that degree of precision. (Our experience with the idealized system, discussed previously, shows that it is indeed necessary to specify β to a high degree of precision.) A little thought will convince the reader that for such a bound, (17), on the rounding error to be valid it is necessary that the internal machine word have at least L fractional bits.

In Section 2.2 we show that it is possible to obtain maps f with the weak and strong contraction properties that satisfy the fidelity criterion with the minimum possible number of fractional bits, i.e., $F = L$. We show that, in fact, the maps obtained are *unique*. The results will show that, for our maps, the offset in machine words during error-free transmission decreases exponentially fast to a value less than unity, after which there may be at most $(2^L - 1)$ occasions at which the integer parts differ.

Let us now consider in broad terms what the preceding results imply in terms of the cost and complexity of the digital implementation of the scheme for adaptive quantization discussed in Section 1.1.1. Consider the fairly typical case where the total number of integral log step sizes is 64 and $\beta = {}^{63}/_{64}$, i.e. $K = 6$ and $L = 6$. We now know that the total word length should be at least 12 bits. Consider the implications on the associated ROM size. The table stored in the ROM will have 2^{12} addresses, each address containing 12 bits, giving a total memory size in the transmitter and receiver of about 50K bits each! Moreover, with each additional bit in the internal word, the memory requirement more than doubles.*

In the next section, we propose a new adaptation algorithm and specify the required arithmetic. The new algorithm requires significantly fewer fractional bits in the machine words while possessing the necessary synchronization capability.

II. THE PROPOSED SYSTEM

2.1 Idealized description

We propose the following *interleaved-leakage algorithm* (ILA) as the basis for the machine adaptation of the log step size. For fixed parameters I and γ, $I \geq 2$ and $0 < \gamma < 1$ [see eq. (6)]:

$$\left.\begin{aligned} d(i+1) &= \gamma d(i) + m(i) \\ d(i+2) &= d(i+1) + m(i+1) \\ d(i+I) &= d(i+I-1) + m(i+I-1) \end{aligned}\right\} i = 0, I, 2I, \cdots. \tag{18}$$

* We have considered the possibility of exploiting the idea due to Croisier et al. (Ref. 11) and Peled and Liu (Ref. 12) wherein the ROM size may be reduced at the cost of increased processing time. The processing times available and the relative costs do not make this approach particularly promising at the present time. However, it is an approach worth keeping in mind.

Here γ is the leakage constant, and leakage is introduced only once in every I iterations. Thus we refer to I as the *interleaving interval*. The $m(\cdot)$ terms are the log multipliers, $m(\cdot) \in \{m_1, \cdots, m_N\}$, and the selection rule is as in (6b). However, in general, the optimum values of the multipliers may be different from the ones in the scheme described in Section 1.1.1 (we refer to the latter scheme as the uniform-leakage algorithm, or sometimes only as ULA).

We observe that for two geographically separated implementations, $\{d(\cdot)\}$ and $\{d'(\cdot)\}$, of the recursion in (18) subject to possibly different initial values, $d(0)$ and $d'(0)$, but identical $\{m(\cdot)\}$ sequences, as is the case during error-free transmission, we have for the offset,

$$|d(i) - d'(i)| = (\gamma^{1/I})^i |d(0) - d'(0)|, \qquad i = 0, I, 2I, \cdots. \tag{19}$$

Comparing (19) with the similar expression in (8) for the offset in ULA, we find that the capability for recovery from channel errors is comparable in the two schemes if

$$\gamma^{1/I} = \beta. \tag{20}$$

The above is a key relation. Table I tabulates typical values of β and the corresponding choices of γ and I which give comparable recovery capabilities. There are small, inconsequential errors in the table which has been obtained from the approximation $\gamma = [1 - (1 - \beta)]^I \approx 1 - I(1 - \beta)$ for small values of $(1 - \beta)$.

The important point about the table is that, for given β, the fractional bits required for a binary representation of the equivalent value of γ is reduced by an additional bit for every doubling of the interleaving interval, I, in ILA. This simple fact is at the heart of the system that is proposed.

Table I — Leakage parameters (β,γ) and interleaving intervals (I) for comparable synchronization capabilities in the uniform and interleaved leakage algorithms*

	γ (ILA)				
β (ULA)	$I = 2$	$I = 4$	$I = 8$	$I = 16$	$I = 32$
$127/128$	$63/64$	$31/32$	$15/16$	$7/8$	$3/4$
$63/64$	$31/32$	$15/16$	$7/8$	$3/4$	
$31/32$	$15/16$	$7/8$	$3/4$		

* We have stopped short of using $\gamma = 1/2$ for two reasons. First, there may be no advantage in reducing γ beyond $3/4$ because two fractional bits may be required in any case on account of the specification of the log multipliers, m_r. Second, the change in the step size may be too drastic, and this may be reflected in the subjective quality. However, it is a possibility worth keeping in mind.

A slight generalization of the proposed scheme would have the multiplier set in the iteration where leakage γ is inserted to be different from the common multiplier set in all other iterations. This generalization provides no gain when the midpoint of the input signal intensities ($\hat{\sigma}$ of Section IV) is scaled to be unity, which is the case considered in the simulations reported in Section IV. Goodman[10] has suggested that, when $\hat{\sigma} \pm 1$, the log multipliers in the leaky iterations be $m(\cdot) + (1 - \gamma) \log_Q \hat{\sigma}$, where $\{m(\cdot)\}$ are the log multipliers in the nonleaking iterations.

2.2 The digital implementation

We now consider the digital implementation of the idealized recursion (18).

Here we let L, an integer, be such that $1 - 2^{-L+1} < \gamma \leq 1 - 2^{-L}$. We make the simplifying, and inessential, assumption that $\gamma = 1 - 2^{-L}$; in this case, the binary representation of γ requires L fractional bits. (Later we indicate through an example that it is easy to make the modifications which allow other values of γ to be used.) Assume K integer and L fractional bits for the internal machine words. Thus, following the discussion on the synchronization requirements in Section 1.1.2, we are assuming that the fractional bits in the machine words are the minimum necessary for the system objectives to be satisfied. Finally, assume that the log multipliers $\{m_r\}$ are specified to L fractional bits.

The internal description of the machine is

$$\left.\begin{aligned} y(i+1) &= f\{y(i)\} + m(i) \\ y(i+2) &= y(i+1) + m(i+1) \\ &\cdots\cdots\cdots\cdots \\ y(i+I) &= y(i+I-1) + m(i+I-1) \end{aligned}\right\} i = 0, I, 2I, \cdots, \qquad (21)$$

where $y(\cdot)$, the internal machine word, is a $(K + L)$-bit word with L fractional bits. In (21), f maps $(K + L)$-bit words with L fractional bits into other such words. The mapping f may be implemented most easily using ROMs; the characterization of the map f that we give below is a recipe for the programming of the ROMs.*

The integral log step size $d(\cdot)$ is obtained from the internal word $y(\cdot)$ by a rule determined by an *external arithmetic*. We consider two natural and simple external arithmetics, rounding and truncation. Thus,

$$\text{Rounding: } d(\cdot) = [y(\cdot)]_{\text{round}} \qquad (22a)$$

$$\text{Truncation: } d(\cdot) = [y(\cdot)]_{\text{truncate}}. \qquad (22b)$$

We mean that if, for integral k, $k - 0.5 < y \leq k + 0.5$, then $[y]_{\text{round}} = k$; if $k \leq y < k + 1$ then $[y]_{\text{truncate}} = k$.

* Observe that the specifications of the maps given here and in Appendix A apply as well to the uniform leakage algorithm described in Section 1.1, provided β replaces γ and the appropriate value of the parameter L associated with the leakage parameter β in ULA is substituted.

We consider first the truncating external arithmetic. Following the discussion in Section 1.1.2, we impose the following requirements on the map f. (It is understood that all arguments of the map have L fractional bits.)

$$(i)\quad \forall\,\sigma_1,\sigma_2, \qquad |f(\sigma_1) - f(\sigma_2)| \le |\sigma_1 - \sigma_2|:$$

"weak contraction everywhere." (23)

$$(ii)\quad \sigma_1 \epsilon [k, k+1),\ \sigma_2 \epsilon [k+1, k+2),\ k \text{ integral} \Rightarrow \frac{|f(\sigma_1) - f(\sigma_2)|}{|\sigma_1 - \sigma_2|} \le \delta < 1:$$

"strong contraction across integer boundaries." (24)

$$(iii)\quad \forall\,\sigma, \qquad |f(\sigma) - \gamma\sigma| < 2^{-L}:$$

"fidelity of discrete map to continuous map." (25)

Recall from Section 1.1.2 that the first two properties are equivalent to the synchronization requirements. We also know that these two conditions together with almost any weak fidelity criterion relating $f(\sigma)$ to $\gamma\sigma$ implies that the number of fractional bits in the machine words is at least L. We find that we can construct maps f which satisfy in addition the fidelity criterion in (*iii*) without incurring the penalty of using more than L fractional bits. Also, as discussed previously, the fidelity criterion in (*iii*) is important in itself.

In Appendix A we give the complete specification of a map for each value of L. In Fig. 3a, we show the graph of the map f for the example of $\gamma = 3/4$, where $L = 2$. In Appendix A we also show that there is only one such map f for any given L which satisfies conditions (*i*) to (*iii*), (23) to (25). Further, for this unique map the value of the contraction parameter δ in (24) is $2\gamma/(1+\gamma)$.

When the external arithmetic is the rounding arithmetic (22a), the

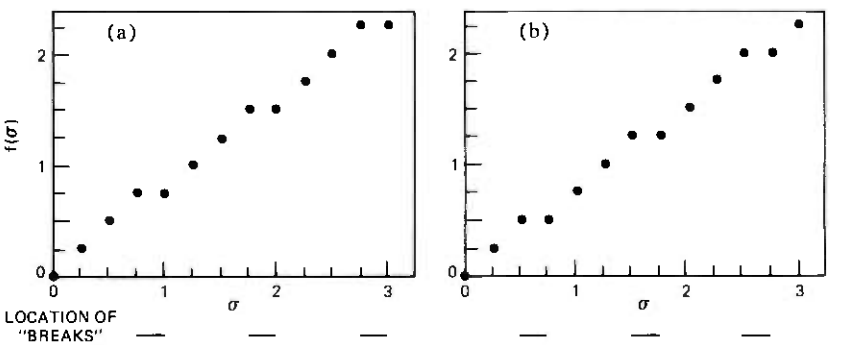


Fig. 3—Machine arithmetics incorporating contraction properties and fidelity criterion for (a) truncating and (b) rounding external arithmetics. $\gamma = 3/4$ and $L = 2$ (see Section 2.2).

resulting map f is somewhat different. Appendix A gives the complete specifications of the maps for all values of L; these maps are also unique. Figure 3b shows the graph of one such map.

Recall that earlier we made the simplifying assumption that $\gamma = 1 - 2^{-L}$. In general, L is defined to be such that $1 - 2^{-L+1} < \gamma \leq 1 - 2^{-L}$. Figure 4 illustrates a map f for the case of $\gamma = 5/8$ ($L = 2$) and the truncating external arithmetic. It may be verified that all the requirements in (23) to (25) are satisfied. We may similarly generate maps satisfying the requirements for arbitrary rational values of γ.

Note that the maps obtained are rather special and quite distinct from the usual maps encountered in digital signal processing.

Another point to note is that while we have specified arithmetics which use the minimum number of fractional bits, $F = L$, additional fractional bits, if they are available, may be put to use by incorporating more than one break in the graph of $f(\sigma)$ per unit interval of σ. The net effect is to give superior synchronization capability.

Finally, note that the implementation of (21) requires by way of hardware only the ROMs, for implementing the map f, and adders. However, the ROMs are used only once in every I iterations. This provides an ideal opportunity for multiplexing the ROMs between different channels and different frequency bands in subband coding[13] applications.

2.3 Synchronization in the digital implementation

We give some bounds on the offset between transmitter and receiver during periods of error-free transmission.

By y and y', two machine words, having different integer parts we mean in the following that $[y]_{\text{round}} \neq [y']_{\text{round}}$ or $[y]_{\text{truncate}} \neq [y']_{\text{truncate}}$, depending on the external arithmetic chosen. Thus, depending upon whether the two machine words have identical or different integer parts, the corresponding log step sizes are identical or different, respectively.

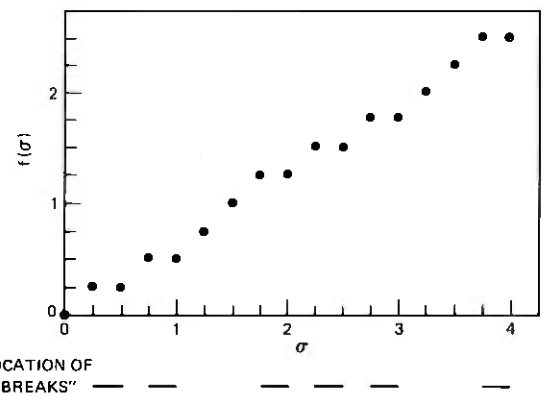


Fig. 4—Machine arithmetic for $\gamma = 5/8$ ($L = 2$) for two fractional bits in machine word and truncating external arithmetic. The contraction requirements and fidelity criterion are satisfied.

Suppose the machine implementations of the recursions in (18) in the transmitter and receiver during error-free transmission are: $i = 0, I, 2I, \cdots$

$$\begin{aligned} y(i+1) &= f\{y(i)\} + m(i) \\ y(i+2) &= y(i+1) + m(i+1) \\ &\vdots \\ y(i+I) &= y(i+I-1) + m(i+I-1) \end{aligned}$$

$$\begin{aligned} y'(i+1) &= f\{y'(i)\} + m(i) \\ y'(i+2) &= y'(i+1) + m(i+1) \\ &\vdots \\ y'(i+I) &= y'(i+I-1) + m(i+I-1). \end{aligned} \quad (26)$$

Observe that

$$|y(i+I) - y'(i+I)| = \cdots = |y(i+1) - y'(i+1)| = |f\{y(i)\} - f\{y'(i)\}|.$$

Now from (23) and (24),

$$|f\{y(i)\} - f\{y'(i)\}|$$

$$\leq |y(i) - y'(i)| \text{ if } y(i) \text{ and } y'(i) \text{ have identical integer parts,} \quad (27)$$

$$\leq \delta |y(i) - y'(i)| \quad \text{if } y(i) \text{ and } y'(i) \text{ have different integer parts.} \quad (28)$$

By repeated application of (28) we see that, if $|y(0) - y'(0)| > 1$, then

$$|y(j) - y'(j)| < 1 \text{ for all } j > I \log \{|y(0) - y'(0)|\}/\log(1/\delta). \quad (29)$$

Thus, once the offset is reduced to less than unity it subsequently remains thus.

Now consider the case where $|y(0) - y'(0)| < 1$. Consider the time instants j which are integral multiples of I. There can be at most $(2^L - 1)$ such time instants at which the integer parts differ. This is because a reduction of 2^{-L} in the offset is guaranteed by (28) in every such time instant. However, at time instants which are not integral multiples of I, the convergence of the integer parts is not quite as strong and is a penalty (which we believe to be insignificant) of ILA.

III. ANALYSIS: PROBABILISTIC ASPECTS

In this section, we investigate the probabilistic behavior of the log step sizes, $\{d(\cdot)\}$, when the input signal variables, $\{x(\cdot)\}$, are random and channel errors are absent. Clearly such an analysis is called for if we are to be able to guarantee certain qualitative features of performance that are basic and necessary in adaptive PCM systems.[4,5] The key notions of the bias function, central log step sizes, and load curves are introduced and their qualitative behavior pinned down.

For our purposes here, the defining equations for the log step sizes are

in (18); the selection rule for the multipliers are in (6b). The key assumption that is made throughout this section is that $\{x(\cdot)\}$ is a sequence of independent, identically distributed random variables with mean zero and standard deviation σ. We sometimes refer to σ as the signal intensity. In keeping with the characteristics of speech, we are interested in σ in the range of $\sigma_{\max}/\sigma_{\min} = 100$, or even 400 (40 and 52 dB ranges, respectively).

3.1 The bias function

Define the bias function $B(\cdot|\sigma)$ to be

$$B(d|\sigma) \triangleq E[d(i+I)|d(i) = d] - d, \qquad i = 0, I, 2I, \cdots. \tag{30}$$

A little thought will show that the right-hand side of (30) does not depend on i—a consequence of the iid assumption on the input signal variables. Different values of σ will generally yield different bias functions, which explains the notation. In engineering parlance, $B(d|\sigma)$ measures, for initial log step size d, the mean drift of the log step size after one cycle of updating of the log step size.

We are able to show for a wide range of values of σ that the bias functions consistently have a distinctive form, depicted in Fig. 5, of considerable significance. In particular, we show that $B(d|\sigma)$ is positive when d is sufficiently small, and negative when d is sufficiently large. Further, under a rather mild restriction, we can prove the consequential result that $B(d|\sigma)$ is monotonic, decreasing with increasing d. The above results in their precise forms are proven in Appendix B. The restriction that is mentioned above is interesting in itself and, roughly, it calls for a propensity for the expected log step sizes after one iteration to be ordered in the same way as the initial log step sizes. This turns out to require, roughly, that $(m_N - m_1)$ be not too large.

The importance of the above results is on account of the following corollary which we state in qualitative terms:

If $(m_N - m_1)$ is not too large, then there exists a unique root, or zero-crossing, of the bias function $B(\cdot|\sigma)$.

Without the monotonicity of the bias function, the possibility exists of

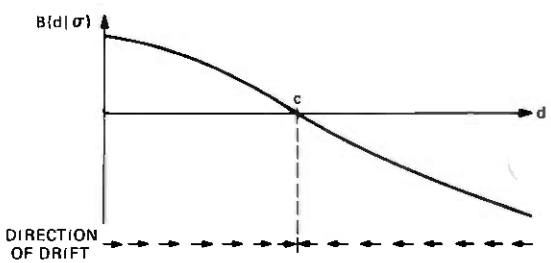


Fig. 5—Sketch of a bias function.

there being many roots with a consequent dilution of the importance that we attach to the root.

Let c denote such a root for a fixed value of σ, Fig. 5:
Definition of c:

$$B(c|\sigma) = 0. \tag{31}$$

We refer to c as the *central log step size* (for signal intensity σ). For a different value of σ and hence a different bias function, the root will generally be different, and to make this dependence quite clear we use the notation $c(\sigma)$.

As the terminology implies, we expect the probability distribution of the log step size to have a concentration of mass around $c(\sigma)$ whenever the signal intensity is σ. The reason for expecting this (see direction of drift indicated by arrows at bottom of Fig. 5) is that, whenever the log step size is not at $c(\sigma)$, the mean drift of the log step size is toward $c(\sigma)$.

The above conclusion is amply borne out by computational results (see Section IV). We find, for instance, that the fit between $c(\sigma)$ and the mean log step size in steady state is extremely good for a rather broad range of values of σ.

In summary, the dual properties of the central log step size (namely, that it predicts so well the mean log step size and that it is so much more tractable and easily obtained) explain the emphasis that we place on the notion of the central log step size.

3.1.1 Method for generating the bias function

The following recursive formula which is developed in Appendix B is the most effective method we know for obtaining the bias function. First, it is necessary to define the following functionals:

$$b_r(\tau) \triangleq 2 \int_{\xi_{r-1}Q^\tau}^{\xi_r Q^\tau} p(\mu)d(\mu), \qquad 1 \le r \le N, \tag{32}$$

where $p(\mu)$ is the common pdf of the input signal variables $\{x(\cdot)\}$. (It is slightly simpler to make as we do the inconsequential assumption that $p(\cdot)$ is symmetrical about 0.) Then $B(d|\sigma)$ is obtained as the solution of the following functional recursion:

$$B_0(d|\sigma) = 0, \ \forall d$$

$$B_k(d|\sigma) = \begin{cases} \sum_{r=1}^{N} b_r(d)\{B_{k-1}(d + m_r|\sigma) + m_r\}, & 1 \le k \le I-1 \\ -(1-\gamma)d + \sum_{r=1}^{N} b_r(d)\{B_{k-1}(\gamma d + m_r|\sigma) + m_r\}, & k = I. \end{cases} \tag{33}$$

Finally, $B(d|\sigma) = B_I(d|\sigma)$.

The above formula is used in the following manner: Assume that the function $B_{k-1}(d|\sigma)$ is known for all values of d. Use (33) to generate next the complete function $B_k(d|\sigma)$. After I such iterations, the resulting function $B_I(d|\sigma)$ is in fact $B(d|\sigma)$.

The reader is referred to eq. (50), Appendix B, for the probabilistic interpretations of the ancillary functions $B_k(\cdot|\sigma)$.

The above formula is used in the analysis presented in Appendix B to determine the previously mentioned qualitative properties of the bias function $B(d|\sigma)$.

Figure 6 is a plot of the bias function $B(d|1)$ for a 16-level quantizer and normally distributed input signal variables. The interleaving interval, I, is 16. Observe in the figure that the graph is for d in the range $[-200,800]$. Values of d outside this range are not of much interest, since the maximum range of the log step sizes in this example is $[\mathrm{Im}_1/(1-\gamma), \mathrm{Im}_N/(1-\gamma)] = [-163,828]$.

3.2 Load curves

The load curves provide information regarding the manner in which the log step sizes depend on the input signal intensity, σ. We use the term to describe a graph of $\log_Q \sigma$ vs. $\bar{d}$, where $\bar{d}$ is the mean log step size in steady state for signal intensity σ. Naturally, the range of σ should cover the range of values expected in the specific application.

From our previous discussion on bias functions and their roots, the

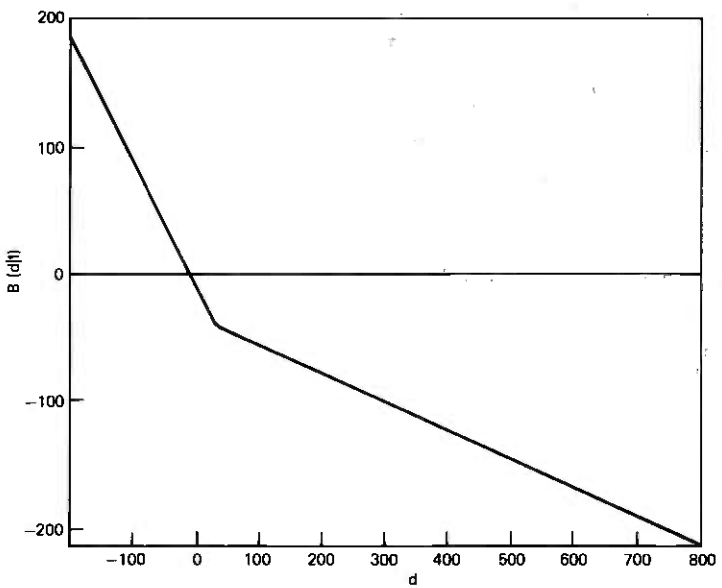


Fig. 6—The bias function for uniform 16-level quantizer and normally distributed input signal variables, $\sigma = 1$. Interleaving internal, $I = 16$ and $\gamma = 0.777$. The log multipliers are given in (39).

central log step sizes, we expect a plot of $\log_Q \sigma$ vs. $c(\sigma)$ to be a rather good fit to the load curves.

The utility of the load curve derives from the fact that it may be visually compared with a plot of the ideal log step size with respect to σ. This information may be obtained from solving a variational problem as is done by Max,[14] who has also tabulated the solutions for the case of normally distributed input signal variables. In any case, the solutions to the variational problem for the *optimum log step size* $\hat{d}(\sigma)$ have the following form

$$\hat{d}(\sigma) = \log_Q \sigma + \hat{D}, \tag{34}$$

where $\hat{D}$ is a constant which depends on the fixed parameters of the quantizer and, importantly, on the common pdf of the input signal variables.

Figure 7 is a plot of the load curve obtained for the 16-level quantizer.

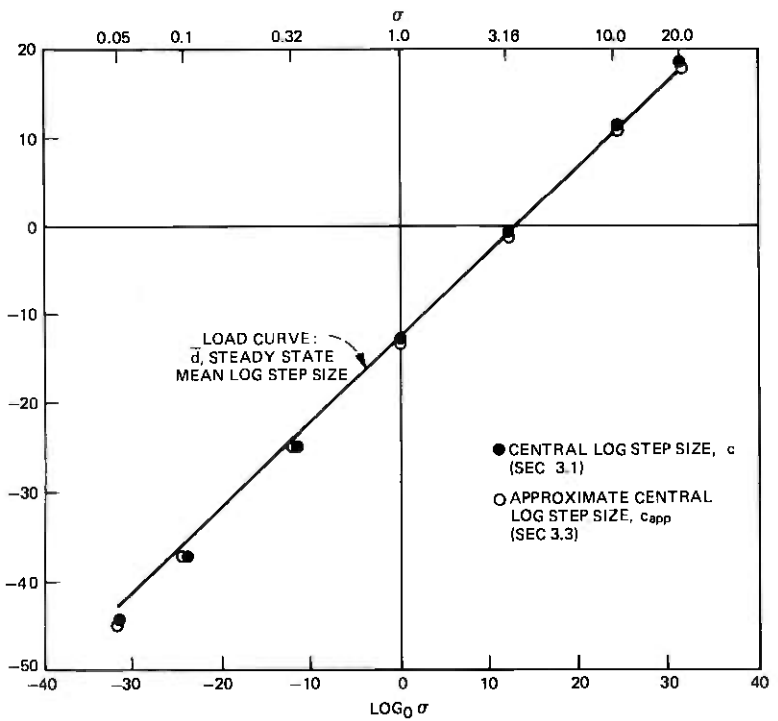


Fig. 7—Load curve ($\bar{d}$), central log step size (c), and approximate log step size (c_{app}) for uniform 16-level quantizer and Gaussian, zero-mean, input signal variables of variance σ^2. The log multipliers are given in (39) and $Q = 1.1$. Interleaving interval, $I = 16$ and leakage, $\gamma = 0.777$.

3.3 The almost-linear dependence of the central log step sizes on signal intensity

Even though a plot of $\log_Q \sigma$ vs $c(\sigma)$ may be expected to be a rather good approximation to, and certainly simpler to obtain than, the load curve ($\log_Q \sigma$ vs $\bar{d}$), it is an unfortunate fact that it is not a very simple matter to obtain $c(\sigma)$. However, our graphs of $c(\sigma)$ have consistently displayed a most remarkable trait, namely, the almost-linearity of $c(\sigma)$ with respect to σ. Intrigued by this feature, we found in an earlier study[9] that it could be explained if the following rather unusual approximation is effective:

$$\int_0^y p(\mu)d\mu \approx \alpha_1 \log y + \alpha_2, \tag{35}$$

where α_1 and α_2 are constants and $p(\cdot)$ is the common pdf of the input signal variables scaled to have unit variance.

Certainly, the above cannot be a good approximation when either y is very small or y is very large. But, as we see in Appendix C, we need the above to be a good approximation only for a limited range of y; specifically, the range of y is required to include the range encountered by $\xi_1 Q^{d(\cdot)}$ at one end, and $\xi_{N-1} Q^{d(\cdot)}$ at the other end, where $d(\cdot)$ is the typical log step size. It turns out that in the important cases where $p(\cdot)$ is either Gaussian or Laplacian, the range of validity of (35) is adequate, at least for the analysis of quantizers with up to 16 levels ($N = 8$). Further details may be found in Ref. 9. For both these distributions, we have found (35) to be an effective approximation in the range $\frac{1}{3} \leq y \leq 2$. For the former distribution, we have found good fits to be obtained if

$$\alpha_1 = 0.44 \text{ and } \alpha_2 = 0.34.$$

(Below, we find it more convenient to express the rhs of (35) as $\alpha_1 \log_Q y + \alpha_2$.)

With (35) as the sole approximation, in Appendix C we go through the involved and tedious process of approximating the bias function and thence deriving its root. The final result, however, is the following remarkably informative formula ($c_{\text{app}}(\sigma)$ is the *approximate central log step size* for signal intensity σ):

$$c_{\text{app}}(\sigma) = S \log_Q \sigma + D, \tag{36}$$

where

$$S = \frac{1}{1 + \dfrac{(1-\gamma)\{1 - 2\alpha_1(m_N - m_1)\}^{I-1}}{1 - \{1 - 2\alpha_1(m_N - m_1)\}^I}} \tag{37}$$

and

$$D = \frac{m_N - 2\sum_{r=1}^{N-1} (m_{r+1} - m_r)(\alpha_1 \bar{\xi}_r + \alpha_2)}{2\alpha_1(m_N - m_1)} S. \tag{38}$$

Let us remark on certain features of the formula. Observe that, on account of α_1 being small, $1 - 2\alpha_1(m_N - m_1) > 0$ almost certainly; for example, $\alpha_1 = 0.018$ when Q [see eq. (3)] is 1.1 and the input signal variables are Gaussian. Consequently, we observe, from the formula in (37) for the slope S, that $S < 1$. Now the ideal slope is 1 [see (34)]. Thus eq. (37) expresses the undesirable but expected fact, alluded to earlier in Section 1.1, that decreasing the leakage parameter γ has the effect of driving the load curve away from the ideal, as sketched in Fig. 8.

As a digression, note that when $\gamma = 1$, the slope S is unity. This is, of course, known to be the case.[4,5] We may also compare the expression for S with a similar expression for ULA derived in Ref. 9—the two expressions are practically identical when $\gamma = \beta^I$ [eq. (20)] and β is close to unity. This important fact, also confirmed in simulations in the example of Section IV, shows that in terms of the loading we expect the behavior in ILA and ULA to be roughly equivalent.

One of the uses that formulas (36) to (38) can be put to is in the optimum choice of the multipliers. The approach we take is that γ and $(m_N - m_1)$ are determined *a priori* on the basis of requirements arising from the quality of synchronization and transient response, respectively. This then fixes the value of S, eq. (37). However, there is still considerable freedom in the choice of the quantities $(m_{r+1} - m_r)$, $1 \leq r \leq N - 1$, and thereby in the choice of the value of D, eq. (38). This degree of freedom may be exploited to determine the point of intersection of the graph of $c_{app}(\sigma)$ and the ideal graph, which are shown in Fig. 8. A sensible choice for the point of intersection is at the signal intensity, σ, that is most likely to be encountered. Usually,[1] this is at the midpoint of the range of signal intensities expected to be encountered in the application.

IV. COMPUTED RESULTS

Throughout this section, the input signal variables $\{x(\cdot)\}$ are independent, Gaussian, random variables with mean zero and standard deviation σ. The signal intensity σ is varied about a central value of 1.0.

The quantizer is a 16-level, uniform quantizer, i.e., $N = 8$, $\xi_r = r$, $1 \leq r \leq N - 1$, and $\eta_r = r - \frac{1}{2}$, $1 \leq r \leq N$. Throughout, the log base for the step sizes and multipliers, Q, is 1.1.

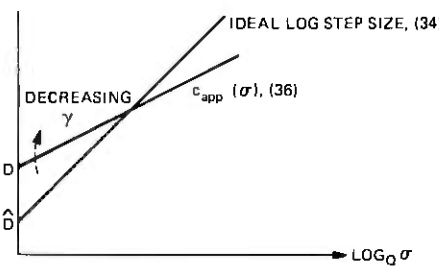


Fig. 8—The behavior of the central log step size compared to the ideal. See eqs. (34) and (36).

For the uniform-leakage algorithm, ULA, we used as the leakage constant $\beta = {}^{63}/_{64}$. The multipliers for ULA are approximately those used by Rosenthal et al.[15] after correction, in the manner suggested in Ref. 1, for the following specifications: In the notation of Ref. 1, $\hat{\sigma}$ = midpoint of signal intensities = 1.0, the ideal loading factor = ideal step size/signal intensity = 0.257. This procedure gave the following values for the log-multipliers for ULA,

$$m(1) = m(2) = m(3) = m(4) = -2.25;\ m(5) = m(6) = 2.50;$$
$$m(7) = 7.25;\ m(8) = 11.50. \quad (39)$$

The multipliers used for the interleaved algorithm, ILA, were also selected to be those given above. We are aware of the advantages of fine tuning the multipliers and Q to take advantage of the special features of ILA, but decided on balance to keep the multipliers and Q unchanged. We found that, as it stands, the transient behavior for ILA is slightly superior to that of ULA; reducing Q in ILA equalizes the transient behavior in the two schemes and yields s/n ratios slightly better than those reported here for ILA.

4.1 Computed load curve, central log step sizes, and their approximation

We illustrate the above notions for the interleaved leakage algorithm for the case of the interleaving interval, $I = 16$. We set $\gamma = \beta^I = 0.777$. Figure 7 plots three quantities with respect to $\log_Q \sigma$: (*i*) $\bar{d}$, the steady-state, mean log step size. This was obtained from 10,000 iterations; (*ii*) $c(\sigma)$, the central log step size defined in (31); (*iii*) $c_{\text{app}}(\sigma)$, the approximate central log step size as given by (36) to (38).

For the given specifications,

$$c_{\text{app}}(\sigma) = 0.99 \log_Q \sigma - 13.20.$$

To clarify Fig. 7, we have also tabulated in Table II the values of the above variables at seven values of σ.

Table II — Computed load curve, central log step sizes, and their approximation ($I = 16$ and $\gamma = 0.777$)

σ, signal intensity	0.05	0.10	0.3162	1.0	3.162	10.0	20.0
$\bar{d}$, steady state mean log step size	−42.40	−35.53	−24.14	−12.46	−0.84	10.81	17.88
$c(\sigma)$, central log step size	−44.35	−37.07	−25.00	−12.93	−0.85	11.24	18.51
$c_{\text{app}}(\sigma)$, approximate central log step size	−44.63	−37.36	−25.28	−13.20	−1.12	10.96	18.23

4.2 S/N ratios and load curve for ULA and ILA

Table III compares signal-to-noise ratios for the two schemes for a variety of interleaving intervals. The signal energy is simply the energy of the variables $\{x(\cdot)\}$. The noise is exactly the difference between the input signal variable and its reconstruction at the receiver, *assuming error-free transmission.* Thus, the reported s/n ratios reflect the effect of the step-size adaptation algorithms but do not measure synchronization capabilities of the systems—the latter is measured separately in Section 4.3.

Note the almost identical s/n ratio performance for the two algorithms, ULA and ILA.

Tables IV and V compare the mean and standard deviations of the log step sizes. Again, note the uniformity of the results for the ULA and ILA; the loading characteristics of the two approaches are almost identical.

Table III — Signal-to-noise ratios (dB)

σ	ULA $\beta = 63/64$	ILA; $I = 2$ $\gamma = \beta^2 =$ 0.969	ILA; $I = 4$ $\gamma = \beta^4 =$ 0.939	ILA; $I = 8$ $\gamma = \beta^8 =$ 0.881	ILA; $I = 16$ $\gamma = \beta^{16} =$ 0.777
0.10	14.89	14.92	14.90	14.70	14.16
0.3162	14.55	14.57	14.56	14.48	14.17
1.0	14.19	14.16	14.18	14.14	14.13
3.162	13.80	13.77	13.63	13.84	13.76
10.0	13.37	13.30	13.36	13.31	13.24

Table IV — Steady-state mean log step sizes

σ	ULA $\beta = 63/64$	ILA; $I = 2$ $\gamma = \beta^2 =$ 0.969	ILA; $I = 4$ $\gamma = \beta^4 =$ 0.939	ILA; $I = 8$ $\gamma = \beta^8 =$ 0.881	ILA; $I = 16$ $\gamma = \beta^{16} =$ 0.777
0.10	−35.75	−35.78	−35.72	−35.66	−35.53
0.3162	−24.12	−24.11	−24.10	−24.13	−24.14
1.0	−12.47	−12.54	−12.47	−12.54	−12.46
3.162	−0.88	−0.90	−0.87	−0.82	−0.84
10.0	10.74	10.81	10.84	10.78	10.81

Table V— Standard deviation of log step size in steady state

σ	ULA $\beta = 63/64$	ILA; $I = 2$ $\gamma = \beta^2 =$ 0.969	ILA; $I = 4$ $\gamma = \beta^4 =$ 0.939	ILA; $I = 8$ $\gamma = \beta^8 =$ 0.881	ILA; $I = 16$ $\gamma = \beta^{16} =$ 0.777
0.10	4.48	4.50	4.57	4.75	5.26
0.3162	4.56	4.63	4.64	4.74	4.97
1.0	4.70	4.69	4.74	4.74	4.85
3.162	4.80	4.80	4.81	4.81	4.80
10.0	4.87	4.90	4.88	4.89	4.97

4.3 The steady-state mean offset in the transmitter and receiver log step sizes

Here we present some computational results connected with the *steady state*, joint distribution of the transmitter and receiver log step sizes assuming, as we have done throughout Section IV, that the input signal variables are independent, normally distributed.

The channel is assumed to be memoryless; further, the event that a transmitted "1" is received as a "0" and the event that a transmitted "0" is received as a "1" have the common probability p. Thus, p is the *bit error probability*. In the numerical results presented below, the following typical value for the bit error probability is assumed: $p = 10^{-4}$.

Two geographically separated implementations of the interleaved leakage algorithm, (18), are assumed to be occurring: $i = 0, I, 2I, \cdots$

$$\begin{aligned} d(i+1) &= \gamma d(i) + m(i) \\ d(i+2) &= d(i+1) + m(i+1) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ d(i+I) &= d(i+I-1) + m(i+I-1) \end{aligned}$$

$$\begin{aligned} d'(i+1) &= \gamma d'(i) + m'(i) \\ d'(i+2) &= d'(i+1) + m'(i+1) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ d'(i+I) &= d'(i+I-1) + m'(i+I-1) \end{aligned} \qquad (40)$$

The information regarding the log multipliers $m(\cdot)$ are assumed to be coded in the manner shown in Fig. 1 and transmitted through the channel described above. The log multipliers $m'(\cdot)$ are the log multipliers corresponding to the received code word.

By the "steady state mean offset in the transmitter and receiver log step sizes" we mean the quantity $\bar{e}$ where

$$\bar{e} = \lim_{i \to \infty} E\{d(i) - d'(i)\} \qquad (41)$$

In Appendix D we show that $\bar{e}$ is given by the following expression:[9]

$$\bar{e} = \frac{I}{1-\gamma} \sum_{r,s=1}^{N} (m_r - m_s) T_{sr} p_r, \qquad (42)$$

Table VI — Steady state mean offset in transmitter and receiver log step sizes. Bit error probability in channel, p = 10^{-4}

σ	ULA $\beta = {}^{63}/_{64}$	ILA; $I = 2$ $\gamma = \beta^2 =$ 0.969	ILA; $I = 4$ $\gamma = \beta^4 =$ 0.939	ILA; $I = 8$ $\gamma = \beta^8 =$ 0.881	ILA; $I = 16$ $\gamma = \beta^{16} =$ 0.777
0.10	−0.025	−0.025	−0.025	−0.026	−0.026
0.3162	−0.022	−0.022	−0.022	−0.023	−0.024
1.0	−0.020	−0.020	−0.020	−0.021	−0.022
3.162	−0.018	−0.018	−0.018	−0.018	−0.020
10.0	−0.015	−0.015	−0.015	−0.016	−0.017

where $T = \{T_{sr}\}$ is the *channel transition matrix* given below and p_r is the steady state probability that the rth code word is transmitted (00· · ·0 is the first code word, 11· · ·1 is the last, Nth, code word; the sign bit is ignored).

The channel transition matrix T is defined thus:

$$T_{sr} \triangleq \Pr[s\text{th code word recd.} \mid r\text{th code word trans.}].$$

In the special case where the codes are as shown in Fig. 1, the elements of the matrix are obtained in a simple manner from the Hamming distance between the code words. Thus, if $d(s,r)$ is the Hamming distance between the sth and rth code words, then

$$T_{sr} = p^{d(s,r)}(1-p)^{\log_2 N - d(s,r)}, \qquad 1 \le s,r \le N. \tag{43}$$

In the example under consideration where $N = 8$, $T_{11} = (1-p)^3$, $T_{12} = p(1-p)^2$, etc.

The formula given in (42) for $\bar{e}$, the mean offset in log step sizes, is extremely useful. To see this, recall that $\bar{e}$ is defined in (41) in terms of the joint behavior of the transmitter and receiver in steady state, yet (42) provides the means for calculating $\bar{e}$ provided only that the transmitter log step size distribution is known, since the quantities $\{p_r\}$ are statistics of the latter distribution. Thus, the considerably harder task of evaluating the joint distribution of the log step sizes at the two different sites is circumvented.

Table VI enumerates the computed steady-state mean offset in transmitter and receiver log step sizes for various signal intensities and designs; note the almost identical performance.

V. SUMMARY

We consider it important that digitally implemented adaptive quantization systems possess two properties which, regardless of the statistics of the input signal, ensure that synchronization in the step-size adaptations at the transmitter and receiver is restored during periods of error-free transmission: The offset in step sizes is monotonic and nonincreasing and the step sizes differ in at most a finite number of sampling time instants. A detailed examination of the uniform-leakage algorithm (ULA) shows that a necessary and sufficient condition for the synchronization requirements to be satisfied is that the internal machine arithmetic, given by the nonlinear map f, possesses certain contraction properties. It is further shown that these contraction properties may exist only if the number of fractional bits (F) in the internal machine word is at least L where the leakage parameter β is such that $1 - 2^{-L+1} < \beta \le 1 - 2^{-L}$. Thus, if $\beta = 1 - 2^{-L}$ then L is the number of fractional bits required for the binary representation of β. We proceed to show that it is actually possible to obtain internal machine arithmetics which satisfy

all the requirements with the minimum possible number of fractional bits, i.e., $F = L$. The arithmetics that we obtain are moreover unique. With these arithmetics the offset in machine words during error-free transmission decreases exponentially fast to a value less than unity, after which there may be at most $(2^L - 1)$ occasions in which the step sizes differ.

We give a complete specification of the unique maps f. Thus, in the case where truncation is used to obtain the log step size from the internal machine word, the formula that generates f is:

If $\sigma = k + j2^{-L}$, where k and j are integral and $0 \leq j \leq 2^L - 1$, then

$$f(\sigma) = k(1 - 2^{-L}) + j2^{-L}.$$

Figure 3a is the graph of the map f for the example of $L = 2$.

Even the minimum length of the machine words translate into large memory requirements in ROM-based implementations. Thus, in the fairly typical case where the total number of step sizes is 64 and the leakage parameter $\beta = {}^{63}/_{64}$, we find that the minimum word length is 12 bits, which translates into a ROM size of about 50K bits.

We propose a new adaptation algorithm which is considerably more efficient in terms of the memory used in the implementation. In this algorithm, ILA, leakage is interleaved infrequently but at regular intervals into the recursion for the step-size adaptation. Thus, this scheme has as parameters γ, the leakage parameter, and I, the interleaving interval. We find that, for comparable synchronization capabilities in ULA and ILA, the parameters are related thus:

$$\gamma^{1/I} = \beta.$$

Thus for β close to unity, $\gamma \approx 1 - I(1 - \beta)$. Table I shows that for given β the fractional bits required for the binary representation of the equivalent value of γ is reduced by an additional bit for every doubling of the interleaving interval.

To illustrate, consider the example given above where $\beta = {}^{63}/_{64}$; the new scheme provides the option of interleaving leakage once in 8 iterations ($I = 8$) with a leakage parameter $\gamma \approx {}^{7}/_{8}$, which has three fractional bits. Thus, for the same total number of step sizes, the total word length required is 9 bits, which translates into an ROM size of about 5K bits and an order-of-magnitude reduction in memory size. Furthermore, the essential costly element of the system, the ROM, is used only once in 8 iterations, thus allowing for the additional multiplexing of the ROM.

The internal machine arithmetic that is proposed for ILA is identical to that specified for ULA, except that the machine word in the former system is of shorter length.

A detailed theoretical analysis of the statistical behavior of the step sizes for independent random inputs is undertaken. Perhaps the most

insightful result obtained is a simple formula giving the approximate dependence on the input signal intensity, σ, of the central log step size, $c(\sigma)$, which is the particular log step size about which the distribution of log step sizes is concentrated. The formula depends on only two parameters, α_1 and α_2, of the input signal distribution; in the case of Gaussian input distributions, $\alpha_1 \approx 0.44 \log Q$ and $\alpha_2 \approx 0.34$. This simple formula is given in (36) to (38).

The idealized adaptation algorithms were simulated for a representative 16-level quantizer and independent, Gaussian inputs. In the simulations, the multipliers in ILA were selected to be identical to those used in ULA, although *in general we expect the optimal multipliers to be different for the two schemes.* The results of the simulations show that the performances of the systems are almost identical.

APPENDIX A

Specification of the Machine Arithmetics

We describe first the maps f corresponding to the truncating external arithmetic in (22b) which satisfy conditions (*i*) to (*iii*) given in (23) to (25), Section 2.2. In the example shown in Fig. 3a, observe that the breaks, i.e., zero slope segments between pairs of points, occur just prior to the integral values of σ. This is also the rule by which f is obtained for general values of L.

The following formula generates f for general values of L:

$$\text{If } \sigma = k + j2^{-L}, k \text{ and } j \text{ integral and } 0 \le j \le 2^L - 1, \tag{44}$$

then $f(\sigma) = k(1 - 2^{-L}) + j2^{-L}$.

Condition (*i*), (23), is trivially verified. For condition (*ii*), (24), note that for all integral k

$$f(k + 1 - 2^{-L}) - f(k + 1) = 0. \tag{45}$$

Thus a strong contraction across integer boundaries exists and, in fact, for σ_1 and σ_2 with different integer parts

$$\frac{|f(\sigma_1) - f(\sigma_2)|}{|\sigma_1 - \sigma_2|} \le \frac{2\gamma}{1 + \gamma}, \tag{46}$$

so that we may take

$$\delta = 2\gamma/(1 + \gamma) < 1. \tag{47}$$

For the final condition (*iii*), we find that

$$0 \le f(\sigma) - \gamma\sigma \le 2^{-L}(1 - 2^{-L}), \tag{48}$$

where the two inequalities become equalities at $\sigma = k$ and $\sigma = k - 2^{-L}$, respectively, whenever k is integral.

We can also show rather easily that the map f given by (44) is unique,

i.e., there does not exist any other map satisfying the requirements (*i*) to (*iii*). Uniqueness follows from the following two reasons: (*a*) Condition (*ii*) requires that there be a break in the graph of f between $\sigma = k - 2^{-L}$ and $\sigma = k$, k integral, i.e., $f(k - 2^{-L}) = f(k)$. Reason: $\sigma = k - 2^{-L}$ and $\sigma = k$ have different integer parts. (*b*) In order to satisfy at once both the fidelity condition (*iii*) and the weak contraction (*i*) there can be at most one break in the typical integer interval $[k, k + 1]$.

We now describe the slightly different map f which is obtained for the rounding external arithmetic, (22a). For the requirements on f, the only difference is in condition (*ii*) which now reads as follows:

$$(ii')\ \begin{matrix}\sigma_1 \epsilon (k - \tfrac{1}{2}, k + \tfrac{1}{2}] \\ \sigma_2 \epsilon (k + \tfrac{1}{2}, k + \tfrac{3}{2}]\end{matrix} \Rightarrow \frac{|f(\sigma_1) - f(\sigma_2)|}{|\sigma_1 - \sigma_2|} \leq \delta < 1. \qquad (24')$$

The graph of f shown in Fig. 3b is obviously similar to the one displayed in Fig. 3a, the main difference being the locations of the breaks which are here positioned immediately following the midpoint of the integer intervals.

We rapidly summarize the key features of f. The formula for generating f for general L is:

If $\sigma = k - \frac{1}{2} + j2^{-L}$, $\quad k$ and j integral, $1 \leq j \leq 2^L$,

then
$$f(\sigma) = k(1 - 2^{-L}) - \tfrac{1}{2} + j2^{-L}. \qquad (44')$$

The weak contraction condition (*i*) is trivially satisfied as well as the strong contraction condition (*ii'*), (24'), with the same value of δ that was previously obtained:

$$\delta = 2\gamma/(1 + \gamma) < 1. \qquad (47')$$

Finally,

$$|f(\sigma) - \gamma\sigma| \leq 2^{-L-1}, \qquad (48')$$

and hence condition (*iii*) is also satisfied. It is noteworthy that in keeping with the familiar properties of rounding and truncating, the above error bound is generally smaller than the corresponding bound in (48) for the truncating external arithmetic.

The arguments used previously for establishing uniqueness apply as well for the above construction.

APPENDIX B

On the Bias Function

We give here the derivations of the results on the bias function that are stated in Section 3.1, accompanied by more detailed insights and interpretations. It is convenient to drop the adjunct σ in $B(\cdot|\sigma)$, the bias function, with the understanding that here σ is arbitrary, but fixed.

B.1 Generating the bias function

We derive (33), which is a functional recursion yielding the bias function,

$$B(d) = E[d(i)|d(0) = d] - d. \tag{49}$$

Define the ancillary functions

$$B_k(d) \triangleq E[d(I)|d(I-k) = d] - d,\ 0 \le k \le I, \tag{50}$$

so that

$$B(d) = B_I(d).$$

Observe that

$$\begin{aligned} E[d(I)|d(I-k) = d] &= \sum_s s\Pr[d(I) = s|d(I-k) = d] \\ &= \sum_t \Pr[d(I-k+1) = t|d(I-k) = d] \\ &\quad \times E[d(I)|d(I-k+1) = t], \end{aligned} \tag{51}$$

where the Markov property has been used to obtain (51). Now t can take only N possible values. In fact, from (18), we see that if $k < I$, then $t \epsilon \{d + m_r | r = 1, \cdots, N\}$, and if $k = I$ then $t \epsilon \{\gamma d + m_r | r = 1, \cdots, N\}$. Further, the respective probabilities are easily given in terms of the functionals $b_r(y)$, $1 \le r \le N$, defined in (32), of the common pdf of the input signal variables. Thus,

$$\begin{aligned} b_r(d) &= \Pr[\xi_{r-1}Q^d \le |x(\cdot)| < \xi_r Q^d] \\ &= \begin{cases} \Pr[d(i-k+1) = d + m_r | d(I-k) = d], \\ 1 \le k \le I-1 \\ \Pr[d(I-k+1) = \gamma d + m_r | d(I-k) = d], \\ k = I. \end{cases} \end{aligned} \tag{52}$$

Substituting in (51), we arrive at the relations

$$\begin{aligned} &E[d(I)|d(I-k) = d] \\ &\quad = \begin{cases} \sum_{r=1}^{N} b_r(d) E[d(I)|d(I-k+1) = d + m_r], & 1 \le k \le I-1 \\ \sum_{r=1}^{N} b_r(d) E[d(I)|d(I-k+1) = \gamma d + m_r], & k = I. \end{cases} \end{aligned} \tag{53}$$

Substituting in the expressions in (50) for the functions $B_k(\cdot)$, we obtain the recursive formula given in the main text:

$$B_0(d) \equiv 0,$$

$$B_k(d) = \begin{cases} \sum_{r=1}^{N} b_r(d)\{B_{k-1}(d + m_r) + m_r\}, & 1 \le k \le I - 1 \quad (54) \\ -(1-\gamma)d + \sum_{r=1}^{N} b_r(d)\{B_{k-1}(\gamma d + m_r) + m_r\}, & k = I, \quad (55) \end{cases}$$

and $B(d) \equiv B_I(d)$.

B.2 The range of the bias function

Note that, as $d \to -\infty$, the values of all the probabilities $b_1(d), \cdots, b_{N-1}(d)$ approach 0, while $b_N(d) \to 1$. Similarly, as $d \to \infty$, the values of all the probabilities $b_2(d), \cdots, b_N(d)$ approach 0, while $b_1(d) \to 1$. Thus, from (54) we have that

$$\text{As } d \to -\infty, B_1(d) \to m_N, \qquad \text{and as } d \to \infty, B_1(d) \to m_1. \quad (56)$$

Iterating, we obtain that

$$\text{As } d \to -\infty, B_{I-1}(d) \to (I-1)m_N, \qquad \text{as } d \to \infty, B_{I-1}(d) \to (I-1)m_1. \quad (57)$$

Finally, for the bias function we obtain from the above and (55) that

$$\begin{aligned} d \to -\infty, \qquad & B(d) \approx -(1-\gamma)d + \mathrm{Im}_1 > 0 \\ d \to \infty, \qquad & \approx -(1-\gamma)d + \mathrm{Im}_N < 0. \end{aligned} \quad (58)$$

The above is the basis for the claim that at least one zero-crossing of the bias function is guaranteed from observing the values of the function at the two limits.

B.3 The monotonicity of the bias function

We establish here sufficient conditions which imply the rather important monotonicity property of the bias function. Equations (54) and (55) provide the working definition of the bias function. Observe from (54) that for $1 \le k \le I - 1$,

$$\begin{aligned} B_k'(d) &= \sum_{r=1}^{N} b_r'(d)\{B_{k-1}(d + m_r) + m_r\} + \sum_{r=1}^{N} b_r(d)B_{k-1}'(d + m_r) \\ &= -\sum_{r=1}^{N-1} F_r'(d)\{B_{k-1}(d + m_{r+1}) - B_{k-1}(d + m_r) + m_{r+1} - m_r\} + \text{''}. \end{aligned} \quad (59)$$

We have found it convenient to introduce

$$F_r(d) \triangleq \sum_{s=1}^{r} b_s(d), \qquad 1 \le r \le N. \quad (60)$$

The reason for this is that $F_r'(d)$ is positive since

$$F_r(d) = 2\int_0^{\xi_r Q^d} p(\mu)d\mu. \quad (61)$$

At this point, it is worth noting from (59) that $B_{k-1}' < 0$ is not enough

to establish that $B_k' < 0$; it is necessary in addition that B_{k-1}' be not excessively negative. This motivates the bounding of the derivative of B_{k-1} from both below and above. We therefore introduce the quantities

$$\alpha_k \leq \min_y B_k'(y); \quad \max_y B_k'(y) \leq \beta_k, \tag{62}$$

where it is understood that we are only interested in y having values in the finite dynamic range of the log step size. Further, let

$$\delta(d) \triangleq \sum_{r=1}^{N-1} F_r'(d)(m_{r+1} - m_r) \tag{63}$$

and

$$0 < \delta_{\min} \leq \delta(d) \leq \delta_{\max}. \tag{64}$$

From (59) we obtain

$$\begin{aligned} B_k'(d) &\leq -\delta(d)(\alpha_{k-1} + 1) + \beta_{k-1} \\ &\leq -\delta_{\min}(\alpha_{k-1} + 1) + \beta_{k-1}, \textit{ assuming } \alpha_{k-1} \leq -1. \end{aligned}$$

Thus, we may take

$$\beta_k = -\delta_{\min}(\alpha_{k-1} + 1) + \beta_{k-1}, \tag{65}$$

provided $\alpha_{k-1} \leq -1$. In identical fashion, we also obtain

$$\alpha_k = -\delta_{\max}(\beta_{k-1} + 1) + \alpha_{k-1}, \tag{66}$$

again assuming $\alpha_{k-1} \leq -1$.

Summarizing, we have at this stage a coupled pair of recursions for the upper and lower bounds on the derivatives of the functions B_k, $1 \leq k \leq I-1$, provided $\alpha_{k-1} \geq -1$, $1 \leq k \leq I-1$. Finally, we also have from (55) that

$$B'(d) = B_I'(d) \leq (1 - \gamma) - \delta_{\min}(\alpha_{I-1} + 1) + \gamma\beta_{I-1}. \tag{67}$$

We may now solve the linear recursions in (65) and (66) for (α_k, β_k) with the initial conditions $\alpha_0 = \beta_0 = 0$. The following solution is obtained: $1 \leq k \leq 1$,

$$\alpha_k = \frac{1}{2}\{(1 + \bar{\delta})^k + (1 - \bar{\delta})^k\} - \frac{1}{2} \cdot \frac{\delta_{\max}}{\delta} \cdot \{(1 + \bar{\delta})^k - (1 - \bar{\delta})^k\} - 1. \tag{68}$$

$$\beta_k = \frac{1}{2}\{(1 + \bar{\delta})^k + (1 - \bar{\delta})^k\} - \frac{1}{2} \cdot \frac{\bar{\delta}}{\delta_{\max}} \cdot \{(1 + \bar{\delta})^k - (1 - \bar{\delta})^k\} - 1. \tag{69}$$

We have denoted by $\bar{\delta}$ the geometric mean of $\delta_{\max}$ and $\delta_{\min}$, i.e.,

$$\bar{\delta} = (\delta_{\min}\delta_{\max})^{1/2}. \tag{70}$$

The reader will recall that the recursions (65) and (66) were contingent

upon $\alpha_{k-1} \geq -1$. We find, upon examining the "solutions," that we can ensure its validity over the range $1 \leq k \leq I-1$ provided $\alpha_{I-1} \geq -1$, i.e.,

$$\delta_{\max} \leq \bar{\delta} \cdot \frac{(1+\bar{\delta})^{I-1} + (1-\bar{\delta})^{I-1}}{(1+\bar{\delta})^{I-1} - (1-\bar{\delta})^{I-1}}. \tag{71}$$

The above is a key relation. The first observation on it is that the relation implies not only that $\alpha_{I-1} \geq -1$ but also that $\beta_{I-1} \leq 0$, which is of primary interest. This may be verified either directly from the expression in (69) or, more conveniently, from the recursion in (65) for β_k and the fact that $\beta_0 = 0$. But, as an examination for the bound on $B'(d)$ in (67) shows, these two conclusions, namely, $\alpha_{I-1} \geq -1$ and $\beta_{I-1} \leq 0$, are sufficient to guarantee that $B'(d) < 0$. We have thus arrived at the main result of this section:

If $\delta_{\max}$ satisfies the inequality (71), then $B'(d) < 0$. (72)

Some insight into the nature of the inequality (71) may be gained by considering the case of $\bar{\delta} \ll 1$. In this case, the rhs of (71) reduces to $1/(I-1)$. Further, we observe from (68) and (69) that $\alpha_k \approx -k\delta_{\max}$ and $\beta_k \approx -k\delta_{\min}$. Thus, summarizing, we have that

$$\text{If } \bar{\delta} \ll 1 \text{ then } \alpha_k \approx -k\delta_{\max}, \qquad \beta_k \approx -k\delta_{\min}, \qquad 1 \leq k \leq I-1$$
$$\text{and (71) requires that } \delta_{\max} \leq 1/(I-1). \tag{73}$$

Thus, we have demonstrated that the monotonicity of the bias function is implied if the quantity $\delta(d)$ defined in (63) is uniformly small.

Let us now examine the probabilistic import of the condition in (71), namely, that

$$\delta(d) = \Sigma F_r'(d)(m_{r+1} - m_r)$$

be not large. First, recall from the definition of $F_r(d)$ in (61) that

$$F_r'(d) = 2(\ln Q)(\xi_r Q^d)p(\xi_r Q^d), \qquad 1 \leq r \leq N-1. \tag{74}$$

Thus,

$$\delta(d) = 2(\ln Q) \sum_{r=1}^{N-1} (m_{r+1} - m_r)(\xi_r Q^d)p(\xi_r Q^d)$$

$$= 2 \sum_{r=1}^{N-1} \ln (M_{r+1}/M_r)(\xi_r Q^d)p(\xi_r Q^d). \tag{75}$$

Requiring that $\delta(d)$ be not too large is tantamount to requiring that the ratios of the multipliers, M_{r+1}/M_r, be not too large. To make this connection quite transparent, we see that

$$\delta(d) \leq 2 \ln (M_N/M_1) \left[\max_y yp(y) \right]. \tag{76}$$

For $p(\cdot)$ Gaussian with variance σ^2, observe that

$$\max_y yp(y) = p(\sigma) = 0.242, \tag{77}$$

so that, in this case, (76) states that

$$\delta(d) \leq 0.484 \ln (M_N/M_1). \tag{78}$$

The above is not a particularly good bound, relative to the expression in (75), but it does illuminate the manner in which $\delta_{\max}$ depends on the ratios of the multipliers.

Finally, in summary let us recall in purely qualitative terms the reasons for requiring that $\delta(d) = \Sigma F_r'(d)(m_{r+1} - m_r)$ be not large. This condition is tied in a natural way to the conditions that $B_k'(d) \geq -1$, $1 \leq k \leq I - 1$, which is at the core of the above analysis since it follows rather easily from these conditions that $B_k'(d) \leq 0$, also. The conditions "$B_k'(y) \geq -1$" have an entirely natural, underlying probabilistic interpretation. It merely states that, for two starting log step sizes, $d(0) = d$ and $d(0) = d'$, where, say, the ordering is $d < d'$, the respective expected log step sizes after k iterations should also be ordered in the same way. A little thought is enough to convince one that such a condition can only be guaranteed by requiring that $\delta(d)$ be not too large, since $\delta(d)$ itself measures the potential for initial orderings to be reversed in one iteration.

APPENDIX C

Approximate Formula for the Central Log Step Sizes

The object here is to derive the following approximate formula for the dependence of the central log step size on the signal intensity, σ:

$$c_{\text{app}}(\sigma) = S \log_Q \sigma + D, \tag{79}$$

where S and D, given in (37) and (38), are obtained from the fixed parameters of the system. The sole approximation that is made is in approximating the distribution of the input signal variables in the following manner:

$$\int_0^y p(\mu)d\mu \approx \alpha_1 \log_Q y + \alpha_2, \tag{80}$$

where $p(\cdot)$ is the pdf of the input signal variables normalized to have unit variance.

The procedure that is followed consists of first deriving the approximation to the bias function, using the recursive formula in (33), and subsequently deriving the root of the approximate bias function. Observe that the recursive formula in (33) calls for the quantities $b_r(\cdot)$, $1 \leq r \leq N$. We find it essential to work with the partial sums

$$F_r(d) = \sum_{s=1}^{r} b_s(d), \qquad 1 \le r \le N-1$$
$$= 2\int_0^{\xi_r Q^d} p(\mu)d\mu$$
$$\approx 2\alpha_1 \log_Q(\xi_r Q^d/\sigma) + 2\alpha_2, \text{ from (80)},$$
$$= 2\alpha_1 d - 2\alpha_1 \log_Q \sigma + (2\alpha_2 + 2\alpha_1 \bar{\xi}_r), \tag{81}$$

where σ^2 is the variance of the input signal variables. Note that $F_N(d) = 1$.

Examining (33), we find that we may also write it as follows [for notational simplicity, we drop the adjunct σ in $B_k(d|\sigma)$]:
for $1 \le k \le I-1$

$$B_k(d) = B_{k-1}(d + m_N) + m_N - \sum_{r=1}^{N-1} F_r(d)\{B_{k-1}(d + m_{r+1}) - B_{k-1}(d + m_r) + m_{r+1} - m_r\}. \tag{82}$$

Now suppose that $B_{k-1}(d)$ may be expressed in the form

$$B_{k-1}(d) = (f_{k-1} - 1)d + g_{k-1} \log_Q \sigma + h_{k-1}, \tag{83}$$

where $(f_{k-1}, g_{k-1}, h_{k-1})$ do not depend on either d or σ. Certainly, $B_0(d)$ may be expressed in this form since $B_0(d) \equiv 0$. We now show that $B_k(d)$ may also be expressed in the above manner.

Upon substituting the above expression for $B_{k-1}(d)$ and the expression in (81) for $F_r(d)$, in (82) we find that

$$B_k(d) = (f_k - 1)d + g_k \log_Q \sigma + h_k, \tag{84}$$

where

$$f_k = \{1 - 2\alpha_1(m_N - m_1)\}f_{k-1},$$
$$g_k = g_{k-1} + 2\alpha_1(m_N - m_1)f_{k-1},$$
$$h_k = h_{k-1} + \left\{m_N - 2\sum_{1}^{N-1}(m_{r+1} - m_r)(\alpha_1\bar{\xi}_r + \alpha_2)\right\}f_{k-1}. \tag{85}$$

Certainly, the newly defined quantities are independent of d and $\log_Q \sigma$. Thus, the basis exists for an inductive construction. Further, the coupled recursions in (85) are trivial to solve for the initial conditions $f_0 = 1, g_0 = 0, h_0 = 0$; thus, we obtain $(f_{I-1}, g_{I-1}, h_{I-1})$.

As is apparent from (23), the final iteration in the recursion for generating the bias function differs from all the others. In fact,

$$f_I = \{\gamma - 2\alpha_1(m_N - m_1)\}f_{I-1}$$
$$g_I = g_{I-1} + 2\alpha_1(m_N - m_1)f_{I-1}$$
$$h_I = h_{I-1} + \left\{m_N - 2\sum_{1}^{N-1}(m_{r+1} - m_r)(\alpha_1\bar{\xi}_r + \alpha_2)\right\}f_{I-1}. \tag{86}$$

The complete solution for the approximation to the bias function is:

$$B(d) = (f_I - 1)d + g_I \log_Q \sigma + h_I, \tag{87}$$

where

$$f_I = -(1-\gamma)\{1 - 2\alpha_1(m_N - m_1)\}^{I-1} + \{1 - 2\alpha_1(m_N - m_1)\}^I, \tag{88}$$

$$g_I = 1 - \{1 - 2\alpha_1(m_N - m_1)\}^I,$$

$$h_I = \frac{\left\{m_N - 2\sum_1^{N-1}(m_{r+1} - m_r)(\alpha_1\bar{\xi}_r + \alpha_2)\right\}\{1 - \{1 - 2\alpha_1(m_N - m_1)\}^I\}}{2\alpha_1(m_N - m_1)}$$

Recall that the central log step size is the root of the bias function $B(d)$. Thus, denoting by $c_{\text{app}}(\sigma)$ the root of the function in (87), we obtain

$$c_{\text{app}}(\sigma) = \frac{g_I}{1 - f_I}\log_Q \sigma + \frac{h_I}{1 - f_I} \tag{89}$$

$$= S \log_Q \sigma + D, \tag{90}$$

where S and D, trivially identified by comparing the two expressions, are as given in the main text, (37) and (38).

APPENDIX D

Formula for the Steady State, Mean Offset in Transmitter and Receiver Log Step Sizes

We derive the formula for $\bar{e}$ given in (42). First, it is necessary to define certain quantities in connection with (40), which describes the step-size adaptations at the two sites.

$$e(\cdot) \triangleq d(\cdot) - d'(\cdot), \quad \text{the offset at time } \cdot, \tag{91}$$

and $u(\cdot) \triangleq m(\cdot) - m'(\cdot)$, the offset in the log multipliers at time $\cdot$. From (40) we obtain

$$\left.\begin{aligned} e(i+1) &= \gamma e(i) + u(i) \\ e(i+2) &= e(i+1) + u(i+1) \\ &\cdots\cdots\cdots\cdots \\ e(i+I) &= e(i-I-1) + u(i+I-1) \end{aligned}\right\} \qquad i = 0, I, 2I, \cdots. \tag{92}$$

Thus,

$$e(i+I) = \gamma e(i) + \{u(i) + u(i+1) + \cdots + u(i+I-1)\}. \tag{93}$$

Taking expectations of both sides of the equation,

$$\bar{e}(i+I) = \gamma \bar{e}(i) + \{\bar{u}(i) + \bar{u}(i+1) + \cdots + \bar{u}(i+I-1)\}, \tag{94}$$

where the bar has been used to denote mean values.

Consider $\bar{u}(i)$, the first term inside the parentheses. Observe that $u(\cdot) \in \{m_r - m_s \mid 1 \le r,s \le N\}$. Also,

$$\bar{u}(i) = \sum_{r,s=1}^{N} (m_r - m_s) \Pr\left[\begin{matrix} r\text{th code word transmitted } \textit{and } s\text{th} \\ \text{code word received at time } i. \end{matrix}\right]$$

$$= \sum_{r,s=1}^{N} (m_r - m_s) \Pr\left[\begin{array}{l|l} s\text{th code word recd.} & r\text{th code} \\ & \text{word trans.} \end{array}\right]$$

$$\times \Pr\left[\begin{array}{l} r\text{th code word trans.} \\ \text{at time } i \end{array}\right] = \sum_{r,s=1}^{N} (m_r - m_s) T_{sr} p_r(i), \quad (95)$$

where T_{sr} is simply the (s,r)th element of the channel transition matrix, and $p_r(i)$, $1 \le r \le N$, is simply obtained from the pdf of the transmitter log step size at time i.

Expressions for $\bar{u}(i+1), \cdots, \bar{u}(i+I-1)$ may similarly be derived. Thus, for $i = 0, I, 2I, \cdots$

$$\bar{u}(i) + \cdots \bar{u}(i+I-1) = \sum_{r,s=1}^{N} (m_r - m_s) T_{sr} \{p_r(i) + \cdots + p_r(i+I-1)\}. \quad (96)$$

To proceed further, it is necessary to assume ergodicity, i.e., more specifically, convergence in the mean for the time-evolving distributions of the transmitter log step size. With this assumption, as $i \to \infty$

$$\bar{e}(i) \to \bar{e} \quad (97)$$

and

$$p_r(i) + \cdots + p_r(i+I-1) \to I p_r, \qquad 1 \le r \le N, \quad (98)$$

where $\bar{e}$ and p_r have the interpretations mentioned in the main text. Substituting in (94) and (96) yields

$$\bar{e} = \frac{I}{1-\gamma} \sum_{r,s=1}^{N} (m_r - m_s) T_{sr} p_r, \quad (32)$$

which is what we set out to establish.

REFERENCES

1. D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," IEEE Trans. Communication (Correspondence) (November 1975), pp. 1362–1365.
2. N. S. Jayant, "Adaptive Quantization with a One-Word Memory," B.S.T.J., *52*, No. 7 (September 1973), pp. 1119–1144.
3. R. M. Wilkinson, "An Adaptive Pulse Code Modulator for Speech," Proc. Int. Conf. Commun., Paper 1C (June 1971), pp. 1.11–1.15.
4. D. Mitra, "New Results From A Mathematical Study of an Adaptive Quantizer," B.S.T.J., *54*, No. 2 (February 1975), pp. 335–368.
5. D. J. Goodman and A. Gersho, "Theory of an Adaptive Quantizer," IEEE Trans. Commun., *COM-22* (August 1974), pp. 1037–1045.
6. P. Cumminskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," B.S.T.J., *52*, No. 7 (September 1973), pp. 1105–1118.
7. S. Bates, unpublished work.
8. R. Steele, *Delta Modulation Systems*, New York: John Wiley, 1975.
9. D. Mitra, "An Almost Linear Relationship Between the Step Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization," to be presented at the National Telecommunications Conference, 1978.
10. D. J. Goodman, private communication.

11. A. Croisier, D. J. Esteban, M. E. Levilion, and V. Rizo, "Digital Filter for PCM Encoded Signals," US Patent 3,777,130, December 3, 1973.
12. A. Peled and B. Liu, "A New Hardware Realization of Digital Filters," IEEE Trans. Acoust., Speech, Sig. Proc., *ASSP-22* (December 1974), pp. 456–462.
13. R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital Coding of Speech in Subbands," B.S.T.J., *55*, No. 8 (October 1976), pp. 1069–1085.
14. J. Max, "Quantization for Minimum Distortion," Trans. IRE, *IT-6* (March 1960), pp. 7–12.
15. L. H. Rosenthal, L. R. Rabiner, R. W. Schafer, P. Cummiskey, and J. L. Flanagan, "A Multiline Computer Voice Response System Utilizing ADPCM Coded Speech," IEEE Trans. ASSP, *ASSP-22* (October 1974), pp. 339–352.

Contributors to This Issue

Anthony S. Acampora, B.S.E.E., 1968, M.S.E.E., 1970, Ph.D., 1973, Polytechnic Institute of Brooklyn; Bell Laboratories, 1968—. At Bell Laboratories, Mr. Acampora initially worked in the fields of high power microwave transmitters and radar system studies and signal processing. Since 1974, he has been studying high capacity digital satellite systems. His current research interests are modulation and coding theory, time diversion multiple access methods, and efficient frequency re-use techniques. Member, Eta Kappa Nu, Sigma Xi, and IEEE.

William C. Ahern, B.S.E.E., 1956, Newark College of Engineering; M.S.E.E., 1961, New York University; Bell Laboratories, 1959–1977. Mr. Ahern first worked on air defense and communication systems. In 1966 he became supervisor of a network planning group. From 1968 to 1971 he supervised a government network simulation and analysis group. From 1971 until his death in 1977, he supervised surveys of loop signal power and quality of network service.

Corrado Dragone, Laurea in E.E., 1961, Padua University (Italy); Libera Docenza, 1968, Ministero della Pubblica Istruzione (Italy); Bell Laboratories, 1961—. Mr. Dragone has been engaged in experimental and theoretical work on microwave antennas and solid-state power sources. He is currently concerned with problems involving electromagnetic wave propagation and microwave antennas.

Francis P. Duffy, B.A., 1965, King's College; M.S., 1968, Stevens Institute of Technology; Bell Laboratories, 1965—. Mr. Duffy has been involved in conducting statistical surveys to determine telephone network performance and customer behavior characteristics. Currently, he is involved in studying customer trouble reports.

David D. Falconer, B.A.Sc., 1962, University of Toronto; S.M., 1963, and Ph.D., 1967, Massachusetts Institute of Technology; post-doctoral research, Royal Institute of Technology, Stockholm, 1966–1967; Bell

Laboratories, 1967—. Mr. Falconer has worked on problems in communication theory, and high-speed data communication. During 1976–77 he was a visiting professor of electrical engineering at Linköping University, Linköping, Sweden. He presently supervises a group working on digital signal processing for speech bit rate reduction. Member, Tau Beta Pi, Sigma Xi, IEEE.

G. S. Fang, B.S.E.E., 1967, National Taiwan University; Ph.D. (E.E.), 1971, Princeton University; Computer Sciences Corporation, 1971–72; Bell Laboratories, 1972–1977. At Bell Laboratories, Mr. Fang worked on high-speed digital transmission, protection switching, and microprocessor applications. Since the fall of 1977, he has been with the Department of Electrical Engineering, National Taiwan University, Taipei, Republic of China.

Ben Gotz, B.E.E., 1966, The City College of New York; M.E.E., 1968, and Ph.D., 1971, New York University; Bell Laboratories, 1966–1969, 1971—. Mr. Gotz is engaged in studies related to speech coding for bit rate compression.

James A. Maher, B.S., 1962, Seton Hall University; M.S., 1969, Ph.D., 1973 (Applied and Mathematical Statistics), Rutgers University; instructor, Rutgers University, 1970–1973; Bell Laboratories, 1973—. Mr. Maher has been involved in planning statistical surveys designed to characterize network performance. He is presently involved in studies dealing with characterizing customer expectations and perceptions of network service. Member, ASA, ASEE, and Sigma Xi.

E. J. Messerli, B.A. Sc. (E.E.) 1965, University of British Columbia; M.S. (E.E.) 1966, Ph.D. (E.E. and C.S.) 1968, University of California, Berkeley; E.E. & C.S. faculty, Berkeley, 1968–69; Bell Laboratories, 1969—. Mr. Messerli has been primarily involved in systems analysis and network planning. His work includes studies on the demand assignment of capacity for a domestic satellite system, on the impact of faulty trunks on customers and the network, and on the worth of more accurate data for trunk provisioning. He is currently supervisor of a group concerned with planning for the integrated measurement of network performance; Member, IEEE, ORSA.

Debasis Mitra, B.Sc. (E.E.), 1964, and Ph.D. (E.E.), 1967, London University; Bell Laboratories, 1968—. Mr. Mitra has worked on stability analysis on nonlinear systems, semiconductor networks, computer memory management, analysis of queues in communication systems, growth models for new communication services, and adaptive systems. Most recently he has been involved in the analysis of speech coders and digital filters. Member, IEEE and SIAM.

V. Ramaswamy, B.Sc., 1957, Madras University, India; D.M.I.T., Madras Institute of Technology, Chromepet, Madras, India; M.S., 1962, and Ph.D., 1969, Northwestern University; Zenith Radio Corporation, 1962–65; Bell Laboratories, 1969—. Mr. Ramaswamy's previous work has included microwave components, diode parametric amplifiers and wave propagation in semiconductor plasmas. His present research interests are thin-film optical waveguides and devices and polarization effects in single mode optical fibers. Member, Sigma Xi, IEEE.

R. D. Standley, B.S., 1957, University of Illinois; M.S., 1960, Rutgers University; Ph.D., 1966, Illinois Institute of Technology; USASRDL, Ft. Monmouth, N.J., 1957–1960; IIT Research Institute, Chicago, 1960–1966; Bell Laboratories, 1966—. Mr. Standley has been engaged in research projects involving microwave, millimeter wave, and optical components. He is presently concerned with electron beam lithography as applied to fabrication of integrated optic devices. Member, IEEE, Sigma Tau, Sigma Xi.