

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING

ASPECTS OF ELECTRICAL COMMUNICATION

Volume 49

May-June 1970

Number 5

Copyright © 1970, American Telephone and Telegraph Company

SF Submarine Cable System

Foreword

I. HISTORICAL BACKGROUND

Prior to the mid-1950s, transatlantic voice communication was entirely by high frequency radio. Because the transmission quality and reliability did not come up to standards for domestic circuits, the use of the service was limited.

Completion of the first transatlantic telephone cable system¹ in 1956 brought consistently high-quality overseas transmission to the public for the first time. The telephone user responded to this improvement with a great increase in transoceanic calls.

II. TRAFFIC GROWTH

Figure 1 shows the growth in total bandwidth across the Atlantic from the United States which was provided to meet this increased use. This curve is an exponential whose increase averages $27\frac{1}{2}$ percent per year. A plot of message traffic for the same route would show a similar exponential increasing at 23 percent per year. The $27\frac{1}{2}$ percent figure includes channels to provide private wire and data service. In recent years these services have been growing even faster than the message load. The 23 percent annual growth of message service is approximately double the growth rate for domestic long-haul traffic within the United States.

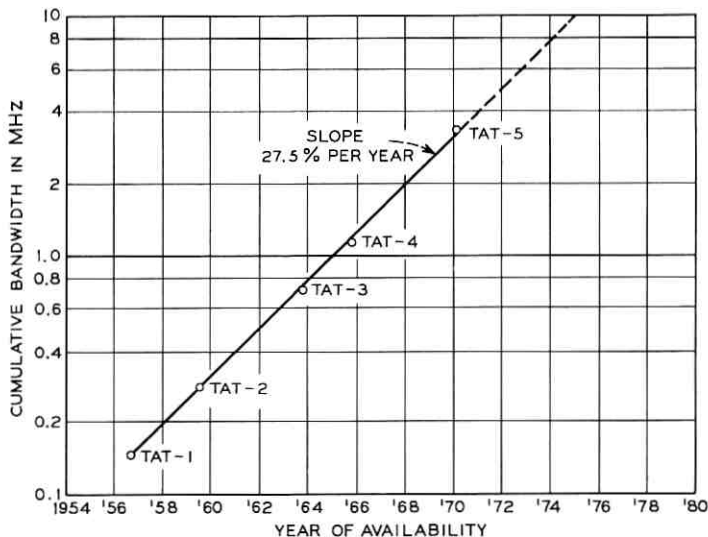


Fig. 1—Growth of transatlantic one-way bandwidth.

To date, there have been no indications of a saturation effect which would cause traffic growth to depart from this increasing exponential. In fact, reduced tariffs made possible by wider band systems will contribute to continuing growth in the future.

Today, overseas traffic goes by both cables and satellites with a small fraction by point-to-point radio. It is interesting to note what happens to the curve of Fig. 1 if one assumes an equal division between cables and satellites. So long as growth is exponential, sharing the load between these two media merely has the effect of shifting the cable bandwidth curve to the right by about three years, after which the growth curve is exactly as before.

III. BACKGROUND OF THE SF SYSTEM

In the latter half of the 1950s, Bell Telephone Laboratories carried on a limited exploratory program to see what could be done with devices and amplifier circuits for a new wideband system. This exploratory effort pioneered transistor characterization which later contributed significantly to the SF System. In the late 1950s, however, device technology was not yet ready for the extremely high reliability demands of submarine cables. Therefore, we decided to complete the development

of a second-generation electron tube system called SD.² This system carried 128 two-way channels on a single one-inch armorless cable.

The first SD System connected Florida and Panama Canal Zone via Jamaica in 1963. The TAT-3 System linked the U. S. and Britain later that year. A network of Pacific SD Systems joined Japan, the Philippines, and the United States in 1964. The TAT-4 System extended from the U. S. to France in 1965.

While the SD System was being designed, Bell Laboratories was developing semiconductor technology which could meet the reliability needs of undersea cables. Thus, in 1963 we were able to schedule detailed development of the SF System.

In the early stages of development, we aimed for 660 high quality 3-kHz channels. This was later revised to a goal of 720 channels. With the experience of installing and equalizing two SF Systems behind us, we now know that the system is capable of meeting objectives for over 800 channels.

Thus, it has been possible to install new capacity to keep pace with the rapidly mounting need.

IV. DESIGN CHALLENGES AND FIRST TRANSATLANTIC USE OF SF

The challenge of developing a new cable system is multidimensional. Diverse specialties must be organized and mobilized so that the final product is an economical system which will provide the maximum number of high quality, reliable channels permitted by current technologies. The range of disciplines and problems involved is great. It spans oceanography, complex mechanism design, sophisticated circuit design, and advanced semiconductor and component design. Because of the unusual environment, mechanical design of undersea units assumes an important role and must be closely interwoven with electrical design.

Throughout the design and manufacture, control and reliability are watchwords. Close control of cable and repeater characteristics is essential to equalize precisely a system which must provide successive stages of electronic amplification to make up for some 16,000 dB of signal attenuation at the top frequency. Reliability targets are set at one failure in 10^{10} hours for passive components and five failures in 10^9 hours for transistors. It simply is not feasible to test items and prove in advance that these levels of reliability have indeed been achieved. Therefore, we design and test system elements meticulously, pre-age and screen devices and components carefully, and manufacture and inspect the final product with extreme care.

This common thread of reliability runs through the various articles of this issue. It is an implicit part of all of the articles having to do with design of various components and subsystems. It is the dominant theme of the article on manufacture of repeaters and equalizers.

In March 1970, the TAT-5 SF System established a direct link between the United States and Spain. Installation and system alignment proceeded smoothly; performance has exceeded design expectations. These results are concrete testimony to the many people who so painstakingly performed the multitude of steps essential to the development, manufacture and installation of the SF System.

S. T. BREWER

REFERENCES

1. Special Transatlantic Cable Issue, B.S.T.J., 36, No. 1 (January 1957), pp. 1-348.
2. Special SD Submarine Cable System Issue, B.S.T.J., 43, No. 4 (July 1964), pp. 1155-1479.

An Overview: Requirements and Performance

By CLEO D. ANDERSON and ROBERT L. EASTON

(Manuscript received February 19, 1970)

This article describes the considerations which led to choice of the SF System design objectives and gives an overview of the system. It further discusses various system parameters and the techniques of laying and adjusting the system for optimum signal-to-noise performance. Finally, the article shows performance results of the recently completed TAT-5 System which extends from Rhode Island to Spain. These results show good agreement with calculated system performance.

I. INTRODUCTION*

At the time of the first transatlantic SD System installation in 1963, it was already clear that a new system with considerably more capacity would be required to meet the traffic demands of the early 1970s. Economic studies indicated that the large potential demand expected by that time would justify the greatest bandwidth achievable from available technology and consistent with high reliability.

Both the SB and SD Systems used vacuum tubes as the active device. As a result, the maximum terminal voltage which could be applied to power the systems without unduly affecting the reliability of the high voltage undersea components was a determining factor on the maximum feasible number of repeaters. This maximum limit, in turn, limited the bandwidth of these systems. The availability of transistors with their lower power requirements removed powering voltage as an important constraint. Instead, the need to obtain sufficient feedback to suppress the effect of device variations and nonlinear distortion played a dominant role in setting limits on achievable channel capacity. Halving the

* This article touches on all aspects of the SF System. For greater detail on any portion of the system or its installation, the reader is referred to the appropriate subsequent article.

repeater spacing from 20 nm to 10 nm* made it possible to quadruple the bandwidth. Increasing the cable diameter from 1.0 inch to 1.5 inches similarly increased the possible bandwidth by 50 percent. The combination of larger cable and reduced repeater spacing resulted in the SF System, with nearly six times the capacity of its predecessor, the SD System. The equivalent-four-wire configuration and the 192 nm spacing between ocean block equalizers were retained from the SD design. This article describes the general design and performance of the SF Submarine Cable System.

II. SYSTEM DESIGN CONSIDERATIONS

2.1 Load Assumptions

The design channel load statistics for the SF System are the same as those used in the SD System. These values and the resulting broadband load for the nominal 720 channels are shown in Table I. Although the

TABLE I—SF SYSTEM DESIGN LOAD STATISTICS

Average talker volume	-10.8 vu
Standard deviation	5.8 dB
Activity factor with TASI	0.75
Rms load per channel	-9.6 dBm0
Rms load per group	2.4 dBm0
Rms load per SG	9.4 dBm0
Rms system load per band (720 channels)	19.0 dBm0
Peak factor	13.9 dB
Instantaneous peak load per band*	32.9 dBm0

* This is the load which is exceeded only 0.01 percent of the time during the highest 1 percent of the busy hour.

per-channel-mile cost of the SF System is one-third that of the SD System, it still is economically sound to design for 3-kHz spaced channels and for a TASI load.

2.2 Noise Objectives

2.2.1 Thermal and Modulation Noise

The uncompandored noise objective for a 3500-nm[†] SF System is set at 43 dBm0 over the life of the system. This means that the great majority of voice channels are to contain a total of thermal and modula-

* In this and subsequent articles in this issue, the abbreviation nm indicates a length of 6087 feet or 1855.3 meters—that is, 1 nautical mile.

[†] Noise objectives are given for 3500 nm \approx 4000 statute miles usually used in specifying long haul objectives. This is not to be confused with the maximum design length (used to set high voltage requirements on system components) which is 4000 nm.

tion noise whose value does not exceed 43 dBrnC0 when all channels are carrying the design load described in the previous section. The value of 43 dBrnC0 corresponds to 3 picowatts/km and exceeds the objectives usually assumed for long haul land systems (40 dBrnC0 by the Bell System and 1 picowatt/km by CCITT). It is worth noting that at the per channel loads assumed for land systems, the SF System has no difficulty in meeting these land objectives. The higher noise associated with the design load is justified on the basis of the high cost of improving the performance and in light of the fact that this noise level will only be reached when these higher loads are actually present. Even when the system has been lined up to carry the design load, the noise will be about 2 dB lower at any time when the high level channels are carrying a few decibels less than the design load. This is expected to be the case during most of the operating life of the system since TASI will not be applied immediately and probably never to all the channels.

2.2.2 *Impulse Noise*

High-voltage direct current is used to power the system. With such an arrangement, insulation defects can result in corona impulses or "pops". Here we deal with the requirement on this type of noise.

It is data transmission requirements that set the controlling limits on permissible impulse noise. Because of the expected long life of the system, it was difficult to anticipate all of the possible data services that may be encountered. For this reason, a very conservative impulse noise objective was adopted. The approach was to require that noise impulses occur no more frequently than momentary peaks due to the statistical distribution of the gaussian noise generated by the thermal noise and intermodulation in the repeaters.

Quantitatively, it was assumed that we should hold the undersea system contribution to error probability in a group (48 kHz) data channel to less than 10^{-7} . Gaussian noise will exceed a level 14 dB above rms with about this probability. Thus, a data system which is to achieve a 10^{-7} error rate must be immune to noise peaks 14 dB above rms. Therefore, it could also withstand impulse noise of this magnitude occurring at the 10^{-7} rate.

It had been empirically observed that the corona pop rate changes much more slowly with a given change in threshold than does gaussian noise. For example, a tenfold increase in the probability of exceeding a threshold by corona noise would require dropping the threshold by about 15 dB. For gaussian noise, the threshold would have to be dropped only $\frac{1}{2}$ dB to increase the probability of exceeding it from 10^{-7} to 10^{-6} . Therefore, if corona noise contributed as many errors as gaussian

noise at a probability of 10^{-7} , it would contribute negligibly to the error at any larger probability. It also is true that if a 48 kHz channel met the above requirement, any narrower channel would have greater margin.

The above considerations lead to an objective that the undersea system group impulse noise exceed -21 dBm0 less than once every $3\frac{1}{2}$ minutes. This objective was used to obtain a requirement on all components subjected to high voltage which could contribute to impulse noise (for example, repeater power separation filters, seals, epoxy belly band capacitor, and so on) by adjusting for the transmission level at the location of the component, and by allocating the pop rate among all the possible contributors.

It was practicable to impose a sufficiently tight requirement on all contributors other than the power separation filter, so that practically the whole of the acceptable pop rate could be assigned to the latter. Thus, the permissible pop rate for a power separation filter was obtained by allocating the permissible rate for the system equally to a number of power separation filters corresponding to the 25 percent nearest the shores. It is only this number of power separation filters that are subjected to a sufficiently high voltage to make corona likely.

2.2.3 *Tone Interference*

The objective for single frequency interference from linear or non-linear crosstalk within the system was set at -65 dBm0. This objective is the same as that used for land systems.

2.3 *Repeater Performance Estimate*

Before the fundamental system parameters—cable diameter, bandwidth, and repeater spacing—could be optimized, it was necessary to estimate the obtainable repeater performance as a function of the highest transmitted frequency. The repeater performance in turn was determined by the following.

(i) *Amplifier configuration*—Early studies assumed three common-emitter stages using germanium transistors, shunt interstages, and a shunt feedback network. These could provide sufficient gain-bandwidth for a 5-6 MHz system. Simple resistive terminations were used in the coupling networks with insertion gain shaping in the feedback network. We assumed an output stage biased at one watt.

(ii) *Repeater noise figure*—7 to 8 dB.

(iii) *Feedback*—The top-frequency feedback could be limited by

either the available gain or by stability in approximately the following relation:

$$\mu\beta \simeq 81 - 3F - G_A \text{ dB (gain limited)}$$

or

$$\mu\beta \simeq 46 - 33.3 \log_{10} F \text{ dB (stability limited)}$$

where F = top frequency in megahertz and G_A = amplifier insertion gain in decibels. For any particular case, the expression resulting in the smaller $\mu\beta$ is the controlling one.

(iv) Transistor intermodulation—With a bias of 1 watt, the assumed device intermodulation coefficients were

$$2f/f \simeq -50 \text{ dB}$$

$$3f/f \simeq -80 \text{ dB}$$

for a 0 dBm, 1 MHz fundamental output into a 120-ohm load (60-ohm termination plus 60-ohm output circuit).

(v) Power output capability—At the top frequency it was assumed that the repeater would be capable of delivering +20 dBm of sine wave power into the cable before reaching the gain cracking point. Although it was realized that the system would be modulation noise-limited and that the system load would not normally approach this level, the resulting 7 to 10 dB overload margin is desirable because, as discussed below, it permits equalizing misalignment in a way that minimizes the associated noise penalty.

2.4 Selection of Top Frequency and Repeater Spacing

The layouts of SB and SD Submarine Cable Systems were made on the basis of thermal noise performance alone, which is the so-called overload limited case. In that case, the system signal-to-noise ratio is optimized by adjusting the repeater transmission levels for maximum average signal power, subject only to the constraint that no repeater is overloaded. For the SF System, unlike earlier submarine cable systems, the resulting transmission levels would cause the repeater modulation noise to exceed the thermal noise. In such a situation, the system is termed modulation limited and the total noise can be reduced by operating the repeater at a level lower than that dictated by overload considerations alone. Optimum noise performance results when the pre-emphasis or signal shaping is adjusted to make the total noise minimum and equal in all channels.

The fact that the SF System proved to be modulation limited also had an impact on the required allowance for misalignment noise penalty. In an ideal system the repeater gain would exactly correspond to the associated cable loss—in any real system this will not be the case. The resulting net gain or loss is called misalignment.

In an overload-limited system, only the highest level repeater can be operated at a transmission level corresponding to the level at which all repeaters would operate in an ideal unmisaligned system. All other repeaters must operate at a transmission level below this ideal level and as a result the thermal noise is increased by an amount (in decibels) corresponding to about $\frac{2}{3}$ the net gain or loss.

In a modulation-limited system, since the level for the ideal case is below that required by overload considerations, it is possible to adjust levels for the misaligned case so that the average repeater level is about the same as the ideal level. In this situation low level repeaters contribute more thermal noise and less modulation noise than a nominal repeater. On the other hand, high level repeaters contribute less thermal noise and more modulation noise. The final result is an increase in system total noise, but a much smaller increase than the corresponding noise penalty for a given magnitude of misalignment in the overload-limited case.

The fact that the SF System was modulation limited, led to the conclusion that a relatively small allowance for misalignment penalty would be adequate. A 2-dB margin was assumed. As an example of what magnitude of misalignment this permits, a 10-dB gain (uniformly distributed along the length of the system and flat with frequency) would result in a 2-dB penalty.

The above objectives, performance estimates, and transmission level considerations were the input to an extensive system study. Various cable diameters were assumed and the impact of moderate changes in the more uncertain performance parameters was studied. A typical output from the study is shown in Fig. 1 which gives required numbers of repeaters as a function of top frequency, assuming $1\frac{1}{2}$ inch cable and nominal repeater performance parameters. The results of the system computations were studied and the largest bandwidth consistent with stable, reliable performance without excessive sensitivity to moderate changes in uncertain performance parameters was selected. The final choice was a 6-MHz top frequency, $1\frac{1}{2}$ inch cable, and 10-nm repeater spacing. This is the case indicated by the dashed lines of Fig. 1.

2.5 Predicted Noise Performance

The computed noise for a nonmisaligned 3500-nm SF System is shown in Fig. 2. The computations are based upon prototype repeater data

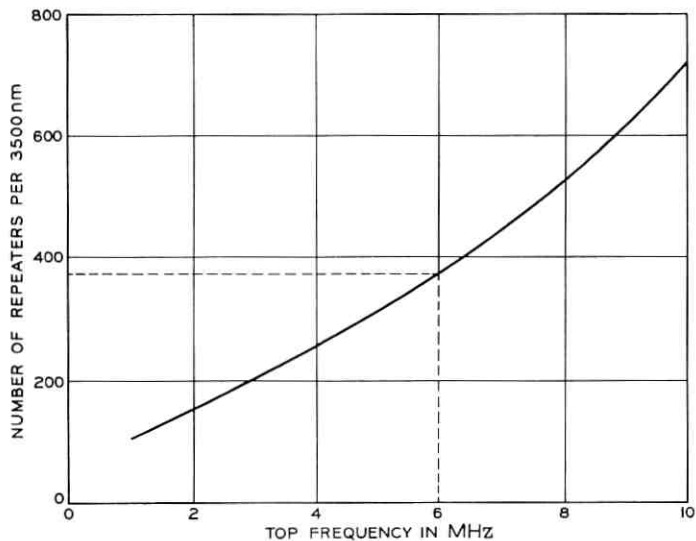


Fig. 1—Top frequency layout of SF System.

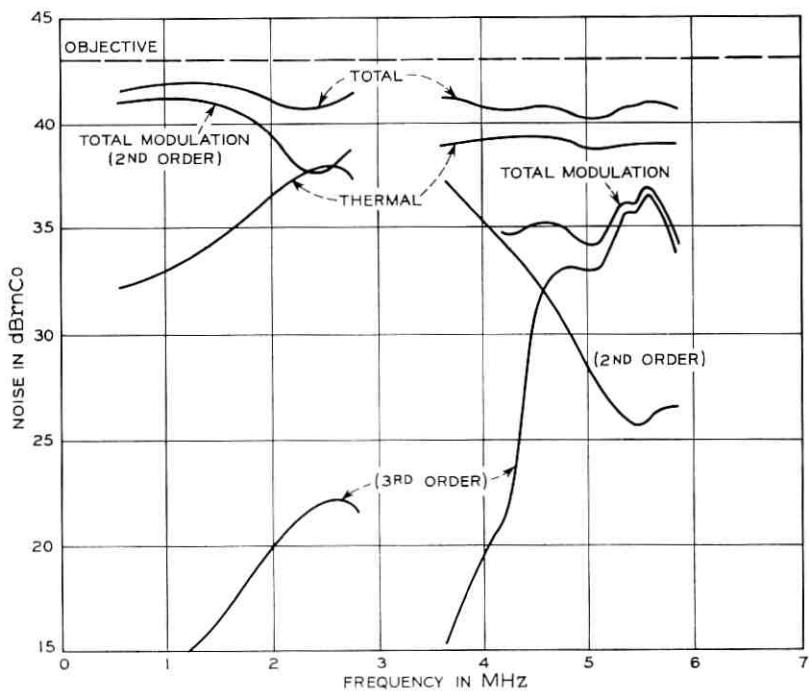


Fig. 2—SF System computed noise performance.

and include the load assumptions of Section 2.1, signal shaping, and the effect of delay distortion on the additive properties of intermodulation noise from repeater to repeater. The performance near the top of the high band indicates clearly that the system is limited by third-order modulation noise rather than overload. The repeater output levels which were obtained in the process of computing the noise are shown in Fig. 3.

Table II lists the nominal load statistics at the output of a repeater corresponding to the assumed loads and signal shaping. The results shown there correspond to a 10-dB margin against overload for an unaligned system whose levels have been adjusted to minimize total noise. It is also clear from Table II that the low band contributes negligibly to the total load.

2.6 Repeater Reliability Objectives

The SF repeater contains three transistors, four diodes and 120 passive electrical components in the main transmission path. The reliability to be achieved in these components was to be such that a repeater failure in one of the 365 repeaters making up a 3500 nm system would rarely occur. Failure rate objectives were set at:

5 in 10^9 Transistor Hours,

1 in 10^9 Diode Hours,

1 in 10^{10} Passive Component Hours.

Achievement of these objectives would result in a mean repeater failure rate of less than 0.1 failure per year over the useful life of the system. It is impractical to prove in such high component reliability levels by testing prior to installation; hence, the final evaluation of whether the objectives have been achieved or surpassed must await experience with the actual installed systems.

2.7 System Development

Once the main system parameters were selected, the final development of the system could proceed. Repeater circuits were refined to optimize performance, reliability, and manufacturability. Repeater gain was trimmed to match the loss of the associated cable as closely as possible. Protection circuits had to be developed to protect the active devices from the tremendous surges that occur in the power path in the event of a cable fault. Crystal-controlled oscillators were developed and incorporated in the repeater. Great care had to be exercised to insure that this additional circuitry would not jeopardize the basic

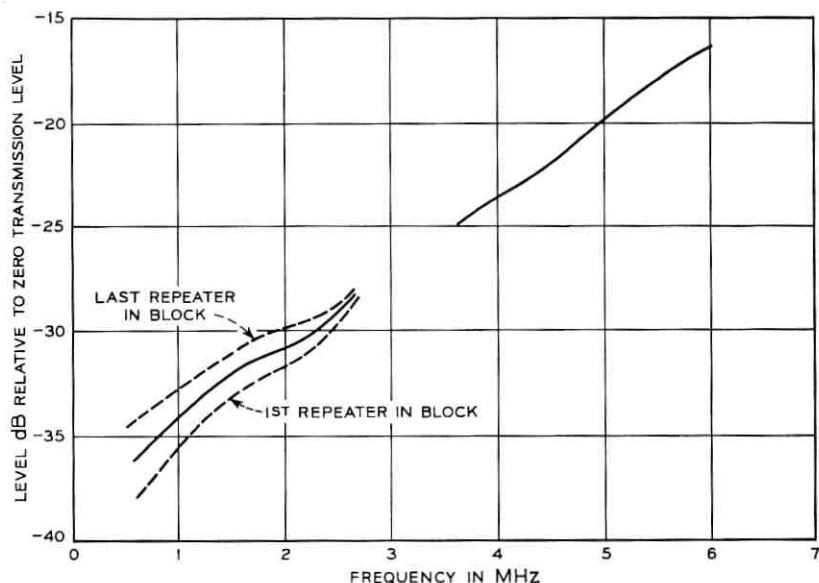


Fig. 3—SF System nominal repeater output levels.

performance of the amplifier. This required, for example, studying and setting limits on intermodulation in the output surge protection diodes. Similarly, very tight harmonic requirements were imposed on the low band supervisory oscillator. Furthermore, these "accessory" circuits had to be incorporated in such a manner that it would be unlikely for a failure in these circuits to cause a failure in the basic repeater.

At this point in the development, an unusual situation during the installation of an SD System led to the discovery of a new phenomenon. It was found that equivalent-four-wire-systems using a common amplifier for both directions of transmission were subject to a so-called "overload sing." Under certain conditions, a momentary overload could

TABLE II—MULTICHANNEL LOAD AT REPEATER OUTPUT

Band	Instantaneous Peak Load (dBm)	rms Load (dBm)
Both	13.8	- 0.7
High	13.5	- 1.1
Low	2.8	- 11.8

Repeater maximum permissible output power = 20 dBm rms sine wave.

produce a situation in which intermodulation shifts energy from high band to low band and low band to high band in such a way that the overload becomes self-sustaining even in the absence of any signal power. Analysis of the SF System indicated that it had inadequate margins against this type of instability. A method for obtaining adequate margins had to be discovered.

The problem was ultimately solved by including a diode limiter in the low band transmission path in the ocean block equalizer. This limiter prevented the power in the low band from ever reaching a level at which significant energy transfer from low band to high band could occur. It was possible to find a limiting point sufficiently low to achieve this and yet high enough so that normal signals are not significantly affected. The values of Table II which show a 20-dB margin between low band peak power and repeater load carrying capacity, make clear why this is practicable. The directional filters in the ocean-block equalizer prevented the harmonics generated by the limiting itself from entering the high band transmission path and contributing to the instability. The phenomenon, the method of analysis, and the effect of the limiter in increasing margins have been described in detail in a previous article.¹

Many of the proven SD System features were retained in the SF design.² The basic mechanical structure of the high pressure housing, the cordwood type assembly of electrical components, and many of the component types were carried over from the SD design. The directional filters followed the SD design scaled to the new frequencies. The deep sea coaxial cable for SF is $1\frac{1}{2}$ inches in diameter (as compared to the 1 inch diameter of SD Cable) and uses a lower conductance loss polyethylene, but has the same armorless structure that was used in the SD System. The composite center conductor of the SF Cable, composed of strength member and conductor, is identical to that used in the SD Cable.

Subsequent to the development of SD, burial of cable was developed as a technique for protecting cable on the continental shelves from trawlers. Armored cable extending hundreds of miles from the terminals would therefore no longer be necessary. For those limited exposures where burial would not be possible, an armored cable using a 1-inch coaxial was developed. This cable has $1\frac{1}{2}$ times the loss per nautical mile of the deep sea cable, but is much easier to handle than the unwieldy cable that would result from armoring the $1\frac{1}{2}$ inch cable.

III. SYSTEM DESCRIPTION

3.1 *Physical Layout*

The general layout and physical properties of the SF System can be summarized as follows:

(i) Signals are transmitted in both directions over a single light-weight $1\frac{1}{2}$ inch armorless cable. The cable has a 6 MHz loss of about 4 dB/nm at sea bottom and a characteristic impedance of 59.4 ohms.

(ii) At every repeater, the transmission signals are separated by directional filters into a low band originating at the A shore terminal, and a high band originating from the B shore terminal. (Equivalent-four-wire operation.)

(iii) The repeaters, which are spaced at nominal 10 nm intervals, are housed in rigid containers similar to those of the SD System. A brief summary of the achieved repeater performance at 6 MHz is given in Table III. In addition to the usual features of power separation and signal amplification, each repeater includes an oscillator for generating a unique supervisory frequency. These supervisory frequencies are transmitted from adjacent repeaters alternately in the high and low bands. They appear in the system spectrum above the high band channels and below the low band. Figure 4 shows a block diagram of the repeater.

(iv) Ocean-block equalizers are placed after every twentieth repeater (counting in the direction of lay). Thus, the length of an ocean block is about 200 miles (192 nm, to be precise) as it was for SD, but the reduced repeater spacing doubles the number of repeaters in a block. The repeaters adjacent to an equalizer are separated by only 2 nm of cable instead of the 10 nm for which the repeater gain is intended to compensate. The missing 8 nm, plus some excess gain incorporated in the repeater at low frequencies, provide the gain needed to permit the equalizer to perform its function of reducing the accumulating misalignment. Directional filters in the equalizer provide separate paths for the two transmission bands, permitting independent equalization of the two bands. This is a departure from the SD design in which equalizer networks were common to both bands.

(v) The shore terminals at each end of the undersea system contain

TABLE III—SF SYSTEM AVERAGE REPEATER PERFORMANCE
AT 6 MHz

Repeater insertion gain	40.146 dB
Repeater noise figure	7.6 dB
Repeater output power	20.0 dBm
Feedback	22.5 dB
Loss from output collector to cable	3.1 dB
Repeater modulation coefficients*	
M_{2R}	-65 dB
M_{3R}	-95 dB
Nominal repeater impedance	59.4 ohms

* Referred to the output collector at a fundamental tone power of 1 milliwatt.

multiplexing equipment, power plants for the undersea repeaters, and means to equalize the undersea system for flat end-to-end transmission and optimum noise performance. Automatic gain control has also been provided to compensate for the variations in the repeater compression observed when the system load varies around the design value.

The physical layout of the SF System is shown in simplified block diagram form in Fig. 5. Figure 6 shows the frequency allocation for the system. Table IV summarizes the dc power requirements for the undersea system.

IV. EQUALIZATION PLAN

4.1 General Objectives

In a 3500-nm SF System the total cable attenuation at the top frequency is about 15,000 dB. Basic equalization of the cable loss is accomplished by shaping the repeater gain to match the loss of its associated cable section. Any difference between repeater gain and cable loss at a particular frequency will cause a transmission deviation. These deviations can accumulate in the overall system, producing misalignments with an associated signal-to-noise penalty. To reduce the misalignment to tolerable limits equalizers are inserted after every 20 repeaters. Residual transmission deviations are equalized in the two terminals.

The repeater gain is shaped to match the loss of 10.07 nm of armorless cable at a nominal sea bottom temperature of 3°C. and depths of 2000 fathoms. In addition, the SF repeater has a specified excess gain in the low band for the purpose of providing additional equalization range at

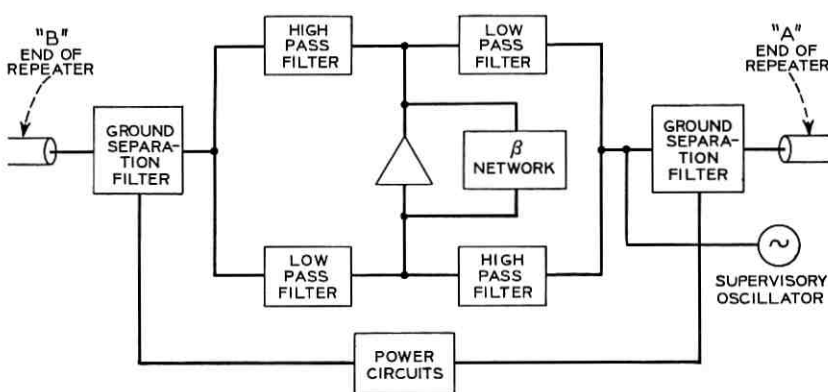


Fig. 4—SF repeater block diagram.

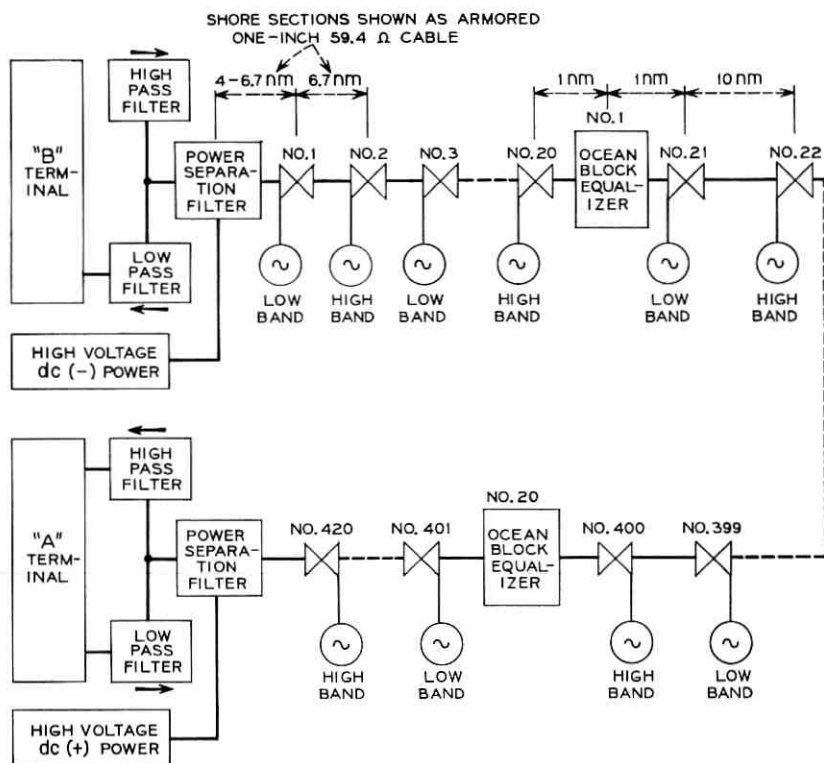


Fig. 5—Physical layout of SF Submarine Cable System.

ocean-block equalizers. The physical lengths of individual cable sections are trimmed at the time of manufacture to give a measured factory loss at 6 MHz that, when adjusted to the sea bottom conditions at the assigned location, will be equal to the repeater gain at that frequency. In this manner, misalignments produced by cable manufacturing variations and other-than-nominal sea bottom conditions can be reduced considerably. The maximum allowable repeater insertion gain deviation from design objectives is $\pm 0.05 (6/f)^{\frac{1}{2}}$ dB*, resulting in an allowable 6-MHz misalignment of 1 dB in a 20-repeater ocean block.

The cable length adjustments mentioned above will correct manufacturing deviations in cable at all frequencies only if these deviations have the same frequency characteristic as the cable loss itself (that is, the deviations are a constant percentage of the attenuation across the

* In this and the following expressions, f is in MHz.

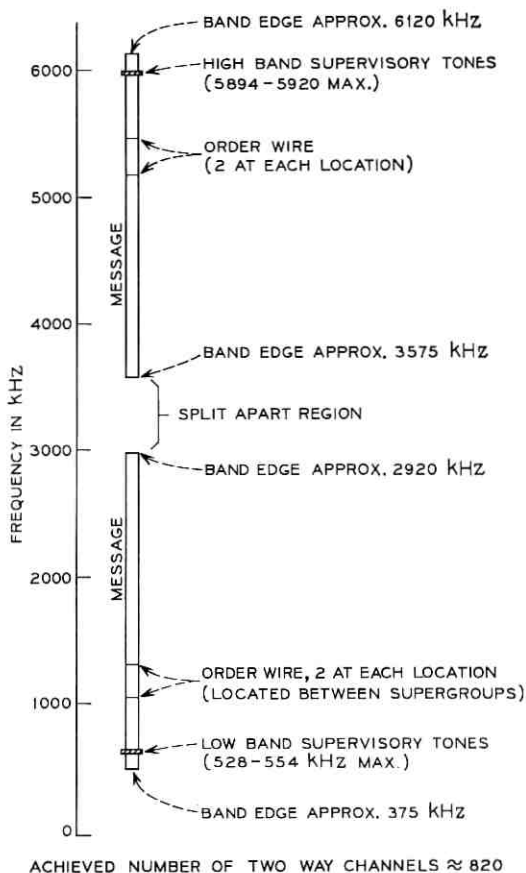


Fig. 6—SF System frequency allocation.

transmission band). To allow for this, as well as for small errors in the adjustment process, a tolerance of $\pm 1(6/f)^{1/2}$ dB of misalignment per block was allocated to the cable at the point where it leaves the factory.

Finally, there can be considerable error in predicting the change in attenuation experienced by cable from the time it is measured in the factory to the point where it is in place at sea bottom. This error is attributed to the effects of handling and possible small errors in the assumed sea bottom temperature and pressure. Errors as large as 2 dB per block at the top of the band are not uncommon—especially if there are several different factories doing the manufacturing. Adding up the various causes of misalignment itemized above gives the result that

misalignments as large as 4 dB per 200-mile block could occur. If these were allowed to accumulate across a 4000-nm system, they could accumulate up to 80 dB of net gain or loss at some frequencies. This is clearly unacceptable. Hence, the need for ocean block equalizers.

4.2 Ocean-Block Equalizer

The ocean-block equalizer contains directional filters to separate the two transmission bands. This, in addition to simplifying the design of many networks, increases the effectiveness of the equalizer, since each transmission band can be considered independently.

Most of the gain available to the ocean-block equalizer for equalization corresponds to the loss of the 8 nm of cable that the equalizer replaces. Since this is not sufficient at low frequencies where the cable attenuation per nautical mile is much less than at high frequencies, excess gain was built into the repeaters at low frequencies to provide additional gain for equalization. This excess gain is 0.2 dB per repeater at the bottom of the low band and tapers off to zero at the top of the low band approximately linearly with frequency.

Various categories of networks appear in the equalizer. The build-out network in each band has a shape such that an ideal block with no misalignment would require half of the total loss of all the switchable networks plus flat loss to be equalized perfectly. The flat loss is available for mop-up and factory option networks. In other words, the build-out network takes out the *shape* associated with 8 miles of cable, the low band gain bias, and one half of the total switched network loss.

TABLE IV—SUMMARY OF POWER REQUIREMENTS FOR UNDERSEA SYSTEM

Total repeater current	136	mA
Voltage drop across repeater	13.1	volts
Resistance of deep sea cable at 3°C	1.7	ohms/nm
Volt drop (0.136×1.7)	0.231	volts/nm

Voltage drop for 4000-nm system with 417 repeaters:

$$4000 \times 0.231 \approx 900 \text{ volts cable drop}$$

$$417 \times 13.1 \approx 5500 \text{ volts}$$

$$\text{Total} \quad \underline{\quad 6400}$$

$$\text{Earth potential allowance} \quad = 2000 \text{ volts}$$

$$\text{Total} \quad \underline{\quad 8400 \text{ volts}}$$

$$\text{Maximal volts per terminal} = \frac{1}{2} \times 8400 = 4200 \text{ volts}$$

In order to achieve the high reliability required of submarine cable systems, extended periods of aging and testing of electrical components are required. This results in a long interval between ordering of components and completion of the first repeaters. This means that some components for equalizer networks had to be ordered before a single repeater was completed by the factory. However, we knew that one of the largest sources of misalignment attributable to the repeater would be the difference between the insertion gain of laboratory prototypes and the average of the manufactured product.

In order to be able to compensate for this difference, a stockpile of submarine cable resistors, capacitors, and inductors was ordered in time to permit adequate aging and testing. Universal mounting plates were also manufactured in advance. Then, when a reasonable number of repeaters had been manufactured to permit estimation of the average, mop-up networks were designed to compensate for the difference between average repeater gain and cable loss.

These networks are restricted to simple series resonance, parallel resonance, low-pass, and high-pass bridged-T configurations. However, there is room for up to seven networks in each band per equalizer and different equalizers may contain different mop-up networks so that there is considerable flexibility available. The amount of flat loss available per equalizer is 8 dB in the low band and 13 dB in the high band, so flat loss transformations can be used to avoid impractical element values.

Provision is made to permit series or parallel connection of resistors and capacitors, allowing ten resistor codes and nine capacitor codes to achieve acceptable granularity. Adjustable inductors with 41 different nominal values complete the stockpile. This approach, coupled with massive use of machine computation, permits a three month interval between design and the completion of a finished unit containing that design. This compares to a period of one to two years for the other networks in the undersea system.

Certain sources of manufacturing deviation in the insertion gain of the repeaters had frequency shapes which could be anticipated. These were the deviations due to variations in the Q of directional filter coils and mu-beta effect. Networks to compensate for these shapes were designed and built in advance to be used, if necessary, in one of the mop-up slots. Similarly, networks to compensate for variations in the dissipation factor of the polyethylene were designed and built. The same shape can be used to compensate for the difference between $1\frac{1}{2}$ inch deep sea cable and 1-inch armored cable (whose length has been

shortened to make the top frequency loss correct). These cable networks, when used, are the only ones which are common to both transmission bands.

The last category of networks in the equalizer are the switchable networks. There are seven in each band, each one of which may be switched in or out of the transmission path just prior to the laying of an equalizer. This means there are 128 possible combinations in each band. This ship-board adjustment is the last opportunity to equalize the undersea system.

Most of the sources of misalignment between the time the mop-up networks are designed and the time of laying are associated with the cable. These misalignments are due to the inability to predict the sea bottom loss of the cable in a 200-mile ocean block to better than a couple decibels. Therefore, shapes associated with cable were selected for the switchable networks.

The most basic of these are networks whose loss in decibels is proportional to the square root of frequency. These correspond to a deviation in cable loss that is a percentage change in attenuation constant with frequency. It would be disastrous to run out of range in this shape so a very generous range of $\pm 7.5(f/6)^{\frac{1}{2}}$ dB in one-decibel steps is provided. Two other shapes are provided, both associated with the dissipation of the polyethylene in the cable. These shapes are combined with the $(f)^{\frac{1}{2}}$ to provide one network shape that has its maximum loss near the bottom and the minimum near the top of each band. The other network shape has a maximum near the center and minimum near the edges of each band.

These switchable shapes are shown in Fig. 7. It can be seen that though they are obtained from cause-associated functional forms, they also make a reasonable family for general purpose equalization. Switching circuits permit selection of the desired networks at the time of laying. A specially designed selector on board ship is attached to the switching pigtail brought out of the equalizer housing through a high pressure seal. The selector connects the networks in or out as desired by means of a series of dc pulses. The pigtails are then detached and insulated from the sea water prior to laying the equalizer.

In addition to the equalizing networks, the ocean-block equalizer also contains the previously described diode limiter for suppressing the overload sing.

The equalizing objective for the ocean-block equalizer is to hold the residual cumulative misalignment to an average value of less than ± 0.2 dB. This leads to a uniform misalignment of less than ± 4 dB in a 3500-

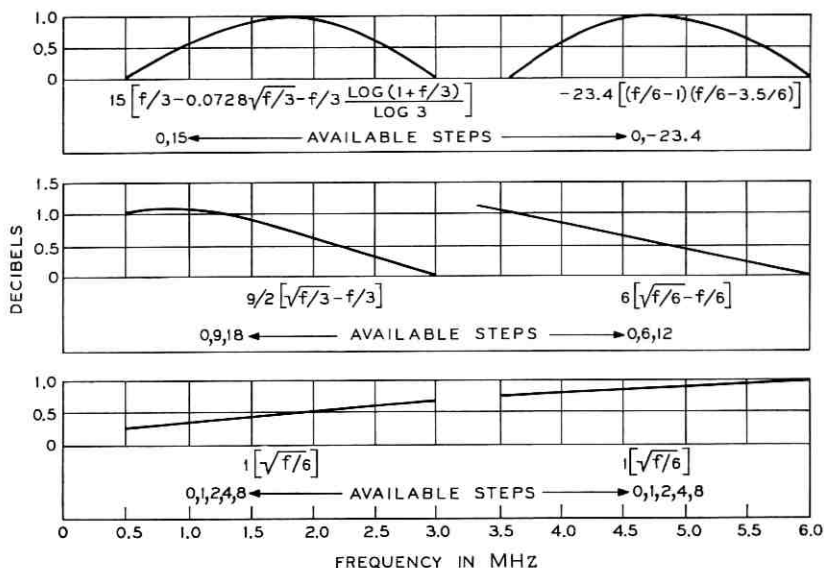


Fig. 7—Ocean block equalizer switchable shapes.

nm system and requires that the equalizer reduce the misalignment accumulated in 200 miles by a factor of 20. When the ± 4 dB uniform misalignment is superimposed on the within block misalignment of less than 4 dB, the resulting noise penalty will be below the allocated 2 dB.

4.3 Terminal Equalization

Once the system has been laid and misalignments are known, the signal-to-noise performance is determined by the operating signal levels on the system. It is the function of the equalizer networks in the transmitting terminal to adjust transmitted signal levels to optimize noise performance; that is, levels are adjusted to realize maximum and equal channel signal-to-noise ratios. The networks in the receiving terminal are then adjusted to deliver proper levels at the receiving multiplex. Figure 8 is a block diagram of the transmitting and receiving terminal equalizers.

4.4 Special Equalization Considerations

4.4.1 Temperature Misalignment

In the presence of temperature-associated misalignment, optimum noise performance requires that the transmitting equalizers be adjusted

so that the levels of the majority of the repeaters, namely those in deep water, are unchanged. To insure that such equalization is not constrained by overload of the off-shore repeaters, those shoreward cable sections in shallow water may be installed with their loss trimmed to an extreme rather than an average temperature condition. Since the high band signals are pre-emphasized about 10 dB above those in the low band only the high band direction of transmission need be considered. If the B-end cable sections, high-band transmit, are trimmed to nominal loss at the high temperature extreme and the A-end sections trimmed to the low temperature extreme, then transmission level of the shallow water repeaters will never be greater than that of repeaters in deep water. Such a scheme results in a negligible noise penalty but provides an increased margin against overload in case of subsequent change of system loss due to aging.

4.4.2 Compression

It has been observed that the gain of the SF repeater is very slightly dependent upon load. At nominal signal levels this "compression" is about 0.2 dB per block at 6-MHz. To compensate for this effect the TAT-5 System was laid with excess gain. With nominal system load, the excess gain compensates for the compression. With a lighter than normal load, there is a slight excess gain in the system, thus reducing the system thermal noise. The modulation noise, because of the assumed

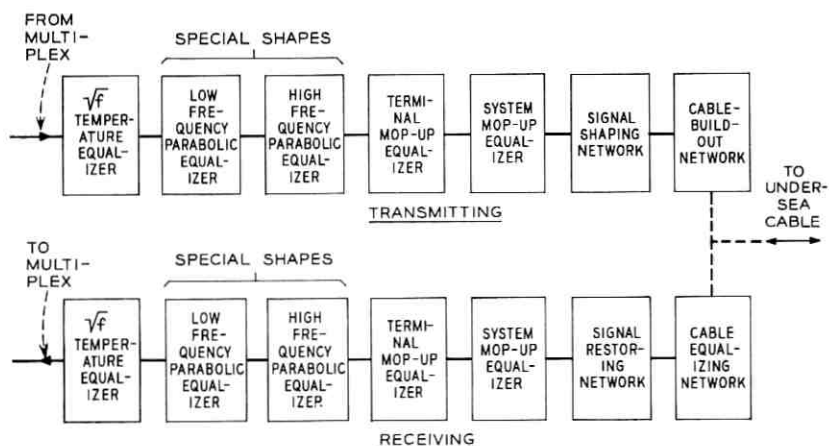


Fig. 8—Transmitting and receiving terminal equalizer.

reduced load, does not increase; consequently, the system noise should never be greater than it would be in the absence of compression.

V. INSTALLATION

Since the capacity of the C. S. LONG LINES is about 1100 nm of the $1\frac{1}{2}$ inch cable, multiple lays are necessary for the installation of trans-oceanic systems. The number of lays required is further increased by the multiplicity of cable factories on different continents, the constraints of manufacturing schedules, and the burying of cable on the continental shelves. The burial has so far been carried out by the Canadian cable ship JOHN CABOT which is ideally suited to the operation. A burial typically involves 50 to 150 nm whereas a deep sea lay usually consists of from 800 to 1200 nm.

Once the repeaters, equalizers, and cable have been manufactured in their respective factories, they meet aboard the cable ship, where they are spliced together and tested to verify that they are working properly. The ship then sails to the end of the previously layed cable. (This end has been left attached to a buoy.) The previously layed end is brought on board and spliced to the shipload. The buried portion and the whole shipload are then powered and laying commences. Testing is carried out continuously during the laying so that any trouble condition becomes evident immediately. Testing also yields data for setting the ocean-block equalizers. Each equalizer has a transmission pigtail coming out of the housing to permit measurements between shore and the next equalizer to be layed. Just prior to laying, the equalizer is stepped to a set of networks selected on the basis of the measurements.

Next, the stepping and transmission pigtails are overmolded to insulate them from the sea. Measurements are then made through to the subsequent equalizer. The data collected during laying is useful in the administration of the systems after installation and makes possible improvements in performance on subsequent systems.

VI. ACHIEVED PERFORMANCE

At this time, two SF Systems have been installed. One system extends between Florida and St. Thomas in the Virgin Islands. The other is TAT-5 between Rhode Island and Spain, which has just been completed. In both systems about 100 additional two-way channels beyond the nominal 720 were achieved.

System performance is measured and optimized during line-up by

noise loading the entire transmission band to the design load and then measuring the total noise in a number of cleared narrow slots. Thermal noise is measured conventionally on the unloaded system. Figure 9 shows the results of measurements on TAT-5 after level optimization has been completed, and Fig. 10 shows the corresponding transmission levels. In both cases, results are compared to the *a priori* computed values of Figs. 2 and 3. Discrepancies are due to the 100 extra channels, the misalignment, and the fact that, practically speaking, a unique best signal shape is hard to define. Most channels are seen to meet the 43 dB_{rnCo} objective with a margin of at least 1 dB.

Impulse noise measurements on a few groups while the system was unloaded, confirm that noise peaks obey the gaussian distribution typical of thermal noise. This verifies that corona impulse noise is not contributing.

Figure 11 shows an average within-block misalignment (misalignment of 20 repeaters and the associated cable, without the equalizer). The large shape across the low band is the previously mentioned excess gain which was purposely incorporated in the repeaters to provide enough low frequency gain for equalization by the ocean-block equalizer. The dashed curve is the average deviation per block after equalization by the ocean-block equalizer. An expanded scale plot of this is shown in Fig. 12. The results of Fig. 12 are average values. There is considerable variation from one block to the next. This is shown in

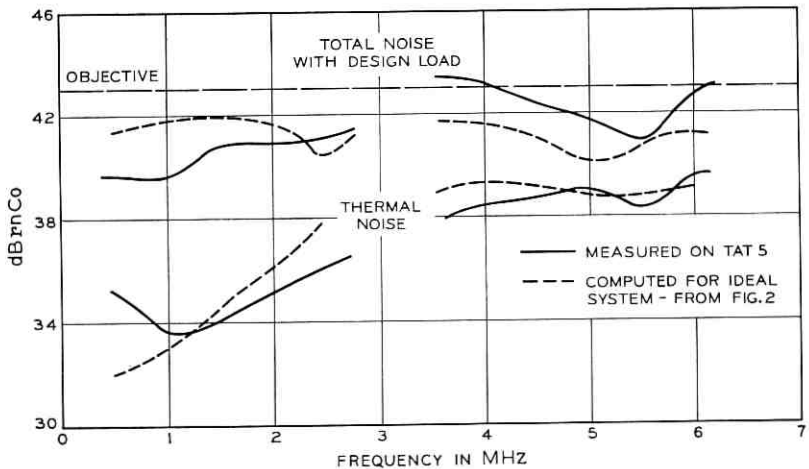


Fig. 9—Measured thermal and total noise for TAT-5.

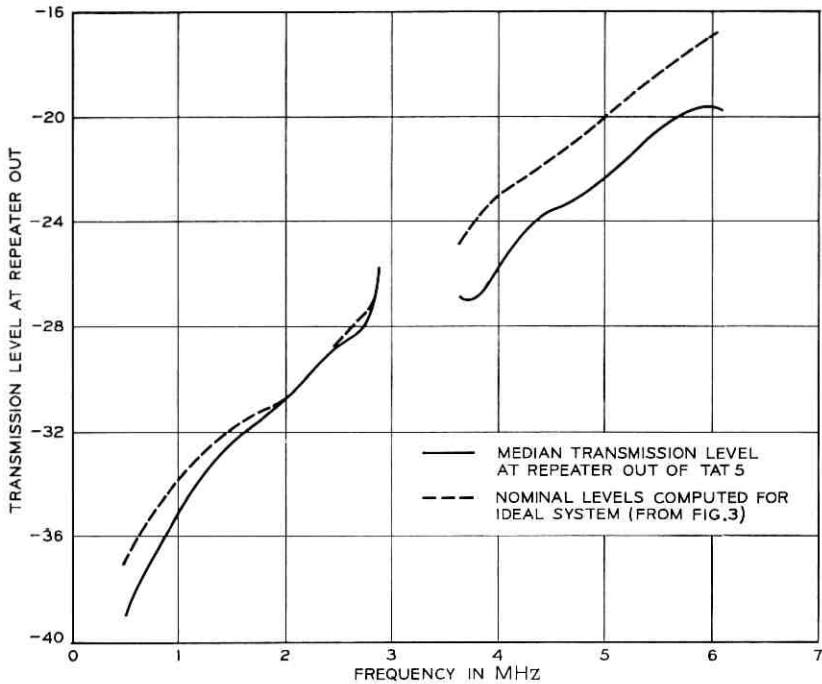


Fig. 10—Signal levels for TAT-5.

Fig. 13 which shows the relative levels of the first and last repeater in each block for a few typical frequencies.

Figure 14 shows the total misalignment of the undersea system. It is composed of the residual misalignment which accumulates from block to block plus the misalignment of the last block which has no ocean block equalizer associated with it. Figure 15 shows the residual misalignment after equalization by the high frequency lines in the terminals. In other words, this is the misalignment which remains to be mopped up by the multiplex. Examination of the equalization results discussed in this paragraph indicates that the equalization objective has been met or exceeded. Future systems can be expected to do somewhat better as the knowledge gained during the installation of TAT-5 will be applied to the design of new mop-up networks for the ocean-block equalizer.

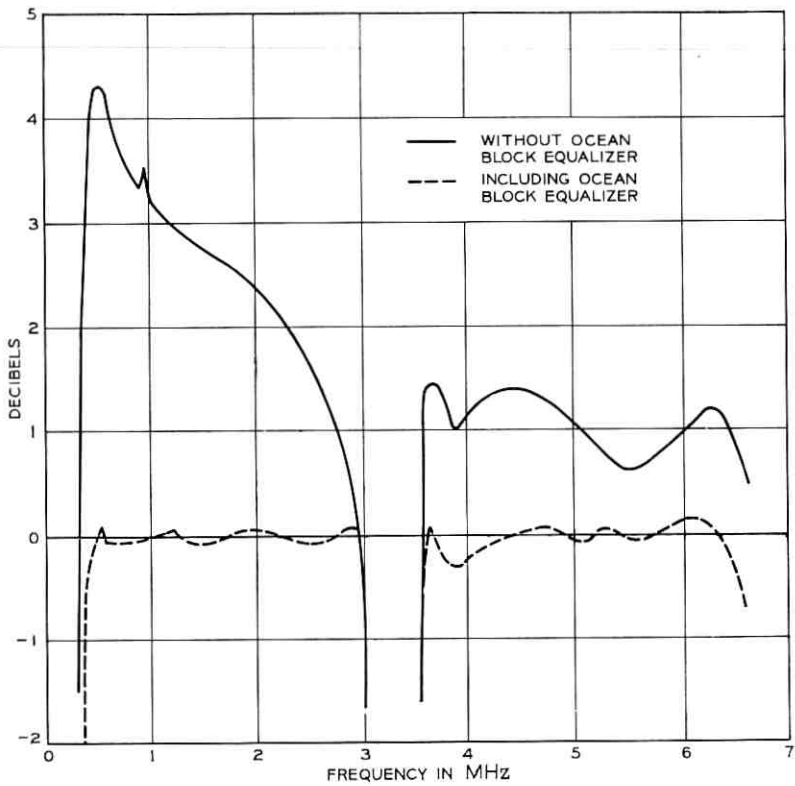


Fig. 11—Average within block misalignment.

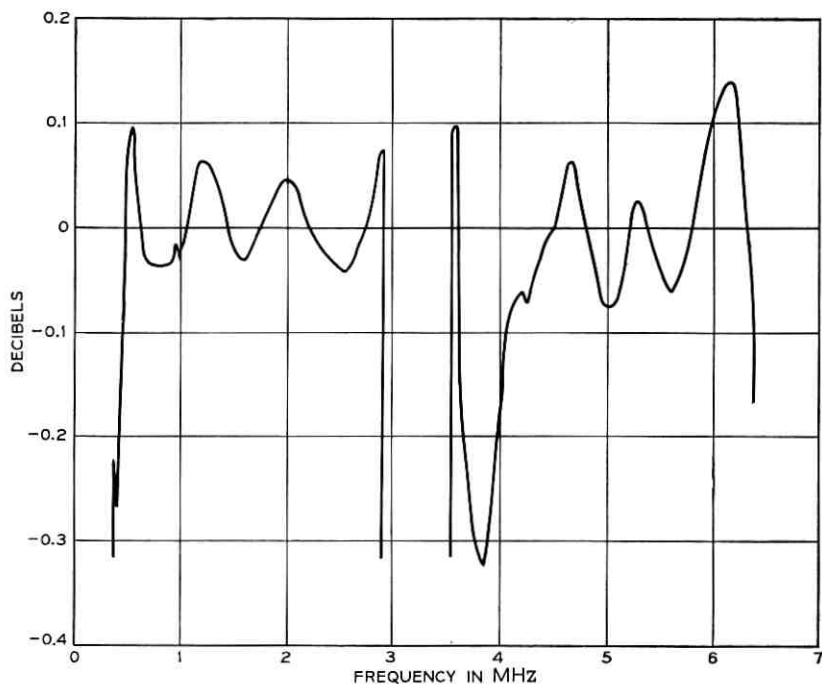


Fig. 12—Average misalignment after equalization by ocean-block equalizer.

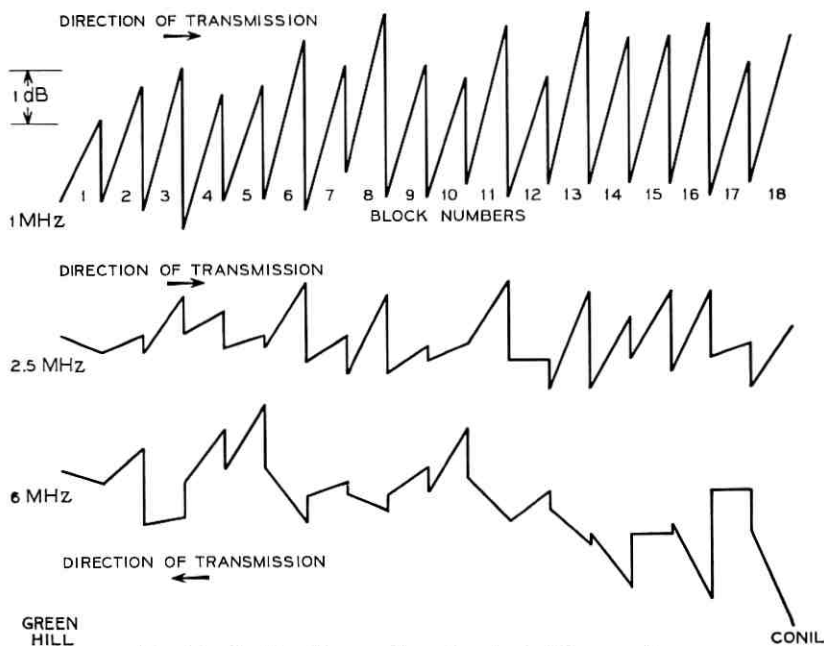


Fig. 13—Ladder diagram for a few typical frequencies.

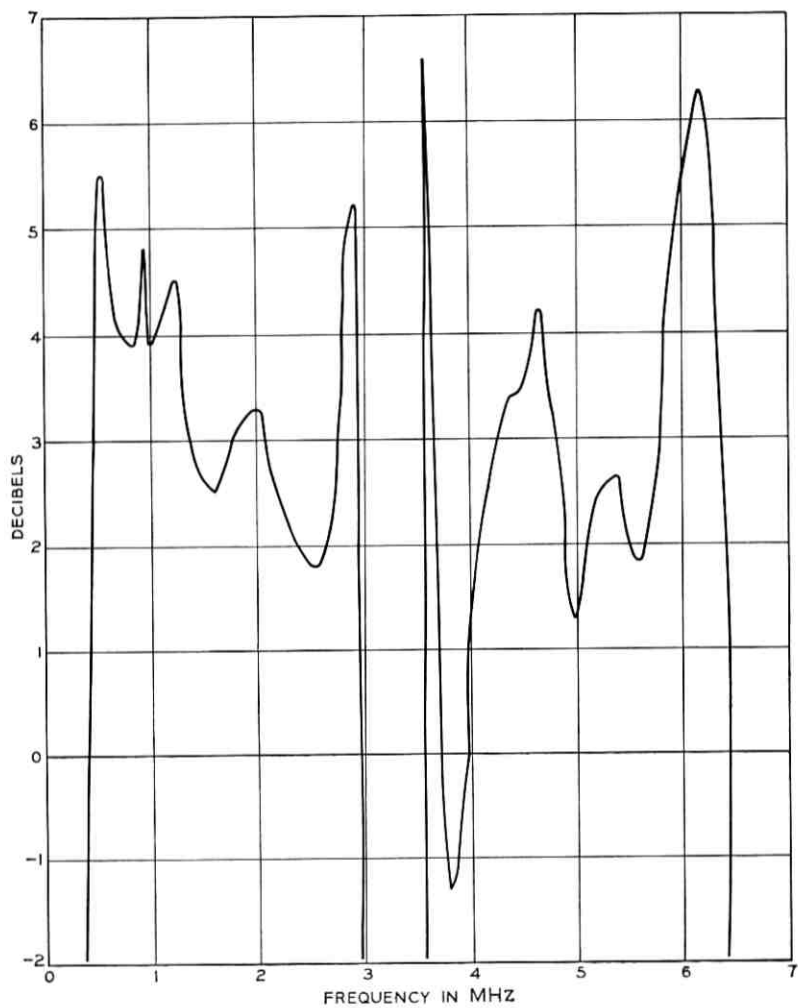


Fig. 14—Misalignment of undersea system.

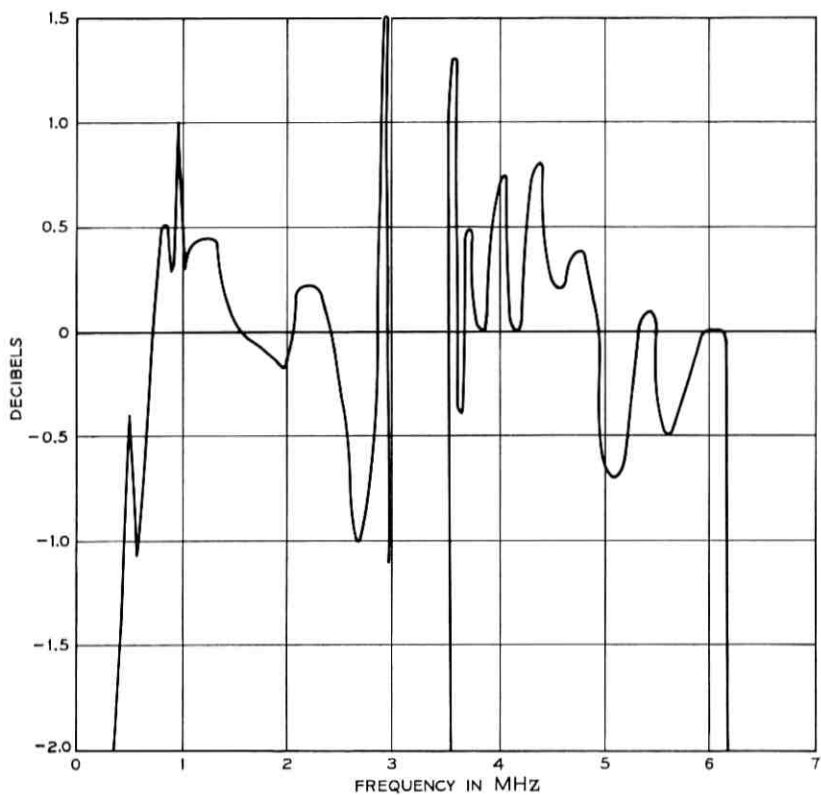


Fig. 15.—Residual deviation after terminal equalization (MX-out to MX-in).

REFERENCES

1. Anderson, C. D., "Overload Stability Problem in Submarine Cable Systems," *B.S.T.J.*, 48, No. 6 (July-August 1969), pp. 1853-1864.
2. Members of Technical Staff, "SD Submarine Cable System," *B.S.T.J.*, 43, No. 4, Part 1 (July 1964), pp. 1155-1479.

Repeater and Equalizer Design

By R. G. BUUS, J. J. KASSIG and P. A. YEISLEY

(Manuscript received October 1, 1969)

Long-distance submarine cable transmission systems require precise, highly reliable repeaters for amplification, and adjustable equalizers for reducing accumulated transmission deviations. This article describes the problems and solutions in designing suitable repeaters and equalizers for the high capacity SF System. This system uses a repeater with transistors to provide a capacity of 800-message circuits.

I. INTRODUCTION

The undersea portion of the SF System includes repeaters spaced at 18.52 kilometer (≈ 10 nautical mile) intervals along the undersea cable to compensate signal loss in the cable across the two nominal transmission bands of 564 to 2788 kHz and 3660 to 5884 kHz. This allocation permits 720* three-kilohertz two-way message channels. Since it is not possible to make the repeater gain characteristic exactly match the cable loss, an equalizer corrects the error accumulated after every 20 repeaters.

Physically both repeater and equalizer units are sealed, encapsulated, and shock mounted within a sealed beryllium copper pressure hull.

The repeaters of the SF System use transistor and semiconductor diode circuitry. The advantages gained include small size, improved power efficiency, and low aging. It should be noted that total repeater size and weight are little altered by the use of transistors, since most other repeater components are not scaled down in size. However, the smaller physical size of the transistors shortens the feedback loop and thus contributes to the wider bandwidth.

Careful attention to design detail, lengthy testing programs, and proven electrical and mechanical concepts refined to meet system re-

* The basic SF design objective was 720 channels. On the 1300-nautical mile system between Florida and the Virgin Islands, out of band misalignments are small enough to permit possible use of as many as 820 channels. The recently completed TAT-5 System is able to meet objective for over 820 channels.

quirements, were the guideposts in producing a satisfactory design. Specially devised digital computer techniques played a prominent role in circuit design and transistor characterization.

II. REPEATER REQUIREMENTS AND CHARACTERISTICS

The SF System requirements are discussed in a companion article.¹ Briefly, the SF System is an equivalent 4-wire (single cable) design with rigid repeaters. It uses armorless coaxial cable² as the transmission medium. The basic band of the SF System provides 720 three-kilohertz message circuits over routes up to 4000 nautical miles. It carries both regular telephone service and special services such as data.

2.1 Repeater Requirements

2.1.1 Electrical Performance

Important repeater electrical performance characteristics appear in Table I.

The nominal insertion gain of the repeater compensates the loss of 18.52 kilometers of 3.8 cm (1½ inch) diameter coaxial ocean cable for the two SF System frequency bands 564-2788 and 3660-5884 kHz. In the low-frequency band repeater gain exceeds cable loss by a small amount to provide additional loss range for the ocean block equalizer networks. The resultant repeater gain objectives are tabulated in Table II.

The allowable gain deviation from nominal has two parts—systematic and random deviations. The average repeater must match the objective of Table II to within $\pm 0.025(6/f)^{\frac{1}{2}}$ dB, with f the frequency in megahertz. The random deviation from this objective cannot exceed $\pm 0.1(6/f)^{\frac{1}{2}}$ dB on a 20-repeater block basis. Equalizers must remove the bulk of the residual deviations.

TABLE I—SF REPEATER NOMINAL PERFORMANCE PARAMETERS

At 5.884 MHz:	
Input/Output return loss (minimum)	28.0 dB
Noise Figure	7.6 dB
Power Output	20.0 dBm
Third Order Modulation Coefficient, M_{3R}	-95.0 dB
At 564. kHz:	
Input/Output return loss (minimum)	32.0 dB
Noise Figure	12.0 dB
Power Output	17.0 dBm
Second Order Modulation Coefficient, M_{2R}	-65.0 dB

TABLE II—REPEATER GAIN OBJECTIVE—dB

F(kHz)	Cable	Boost	Objective
500	11.339	0.200	11.539
600	12.449	0.192	12.641
700	13.463	0.184	13.647
800	14.403	0.176	14.579
1000	16.114	0.160	16.274
1200	17.662	0.144	17.806
1400	19.090	0.128	19.218
1500	19.767	0.120	19.887
1600	20.423	0.112	20.535
1800	21.679	0.096	21.775
2000	22.870	0.080	22.950
2200	24.005	0.064	24.069
2400	25.093	0.048	25.141
2500	25.621	0.040	25.661
2600	26.138	0.032	26.170
2650	26.394	0.028	26.422
2700	26.647	0.024	26.671
2750	26.898	0.020	26.918
2800	27.146	0.016	27.162
2850	27.393	0.012	27.405
3600	30.875	0	30.875
3650	31.092	0	31.092
3700	31.312	0	31.312
3750	31.529	0	31.529
3800	31.744	0	31.744
4000	32.593	0	32.593
4200	33.422	0	33.422
4400	34.233	0	34.233
4500	34.633	0	34.633
4600	35.028	0	35.028
4800	35.807	0	35.807
5000	36.571	0	36.571
5200	37.321	0	37.321
5400	38.058	0	38.058
5500	38.422	0	38.422
5600	38.783	0	38.783
5800	39.496	0	39.496
6000	40.199	0	40.199

Repeater gain deviations cannot include any sharp shapes or any which have many zero crossings, even if they fall within the above formula requirements. These kinds of mismatch cannot be equalized by available networks.

The repeater ports must match the cable impedance of 59.4 ohms to prevent the formation of undesirable ripples in the overall transmission characteristic. At the top of the band the repeater's return loss at both ports is better than 28 dB, as shown in Table I, while at lower frequencies it must be somewhat higher, due to reduction of cable loss. For this

reason minimum realized return loss at 0.5 MHz is 32.0 dB at both ports.

The repeater noise figure is slightly under 8 dB and nearly flat except for a slight degradation at the low end due to $1/f$ noise. The maximum output power of the repeater is 20.0 dBm. It also is nearly constant with frequency.

A direct current of 136.0 mA from the shore terminals powers the repeaters. The voltage drop is 13.1 volts per repeater. Including cable drop, a maximum length system comprising 400 repeaters requires shore terminal voltages of ± 3500 volts (plus an allowance of ± 1000 volts for peaks of earth potential). In order that most repeater components and devices be operated at low voltages, ground separation filters transfer the high voltage through the repeater and around sensitive circuitry.

With 4500 volts dc applied, the corona impulse noise generated by the high voltage components at each end of a fully assembled repeater must not exceed an average of 4 "pops" per 24 hours. A "pop" is any impulse larger than 31 microvolts in a 48 kHz band. This corresponds to a discharge of 10.8 picocoulombs into the repeater input circuit.

Directional filters give the repeater its two-way capability by splitting the band into two parts. An oscillator in each repeater injects a unique frequency onto the cable which can be monitored at the shore terminals to identify the repeater for fault location purposes and to check misalignment. Major electrical and mechanical parts of the repeater appear in the repeater cutaway of Fig. 1.

2.1.2 Mechanical Performance

The maximum permissible leak rate of the repeater under helium pressure test is 1×10^{-7} standard cc/sec at 844 kg/cm² (12,000 psi). This corresponds to a depth of 8,050 meters (26,400 ft.). Under this same test, individual seals and other components are allowed a maximum leak of 10^{-8} or 5×10^{-8} cc/sec., as required.

The temperature limits are from -18°C to 57°C (0°F - 135°F) for sustained exposure in storage. Operating temperature limits are in the range from 0°C to 30°C (32°F - 86°F) for a continuous period of 20 years.

Repeaters must remain undamaged after being subjected to 5 shocks to 60 g peak acceleration and 6 millisecond duration between half amplitude points along each of the three principal axes. Vibration limits are 1.27 cm sinusoidal peak-to-peak displacement from 5 to 11 Hz and sinusoidal vibration at 3 g maximum from 11 to 500 Hz.

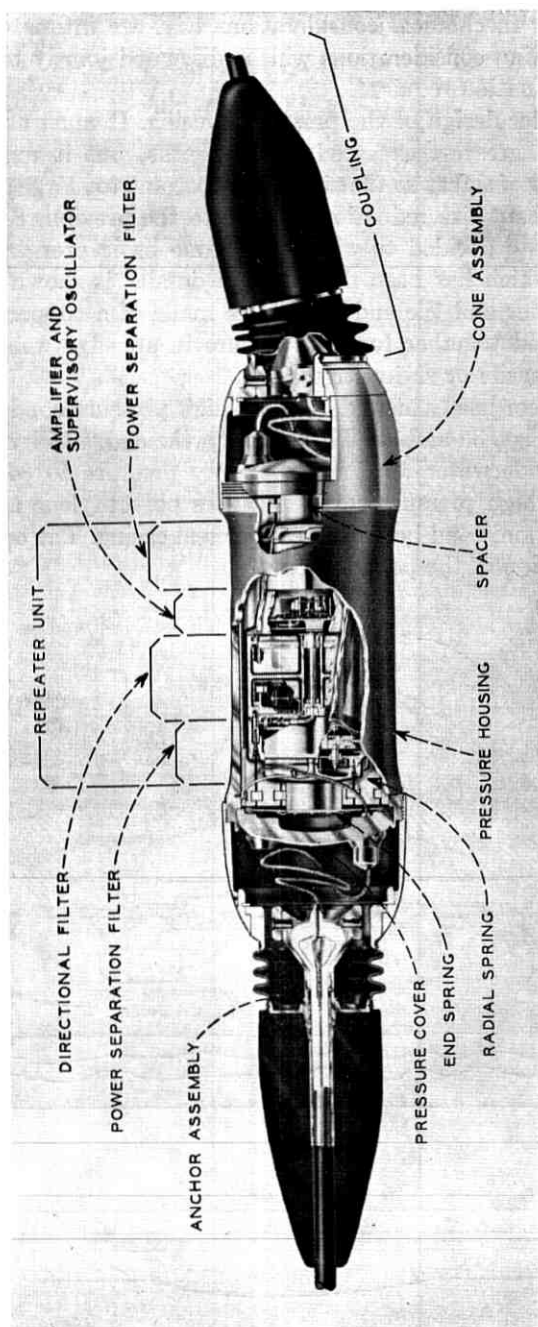


Fig. 1—SF repeater.

Several other mechanical considerations that are intimately related to electrical design considerations will be discussed jointly to illustrate the close coordination required.

Let us consider design of the pressure housing. It must not be overstressed at the greatest expected ocean depths, but it must have a minimum factor of safety so that it will not become too large and heavy. With no significant size reduction relative to the previous SD electron tube repeater, we decided to utilize the same basic pressure housing. This housing which has been described in detail,³ is shown in Fig. 2. Both the cylinder and the end covers are made of a copper beryllium alloy and welded together for final closure in an edge weld virtually free of either tensile or shear stresses.

The SF System uses a new design for high pressure conductor seals (Type 3). See Figs. 3a and 3b. This seal, a preloaded "Bridgman" type, designed for underwater application at any pressure to 844 kg/cm^2 , mounts in the high pressure cover and uses polyethylene as a gasket between the ceramic and beryllium copper seal casing. The ceramic disc prevents extrusion of the polyethylene.

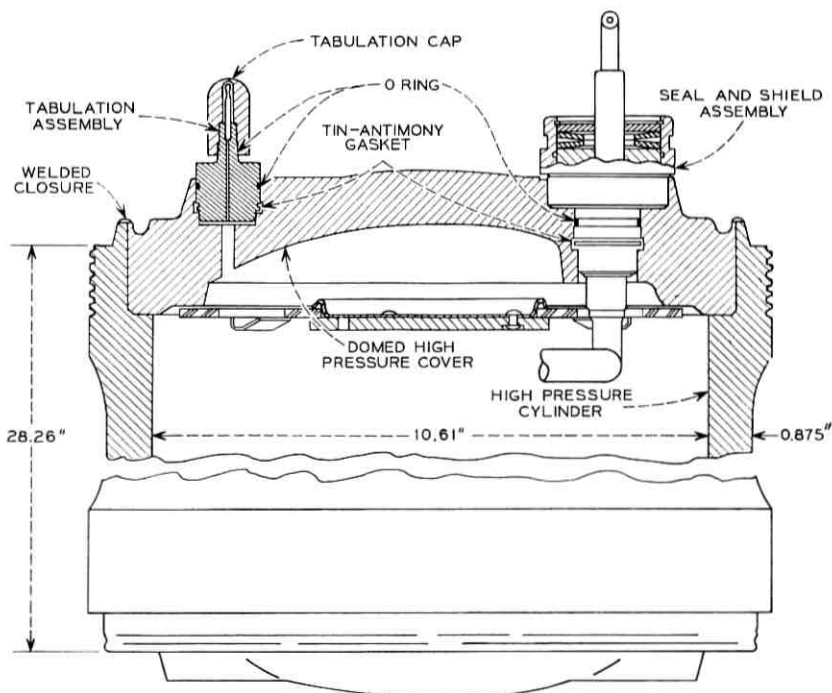


Fig. 2—Pressure housing.

The seal is exposed to the high voltage dc and must satisfy a requirement of no more than one "pop" per 24 hours at a test voltage of 7000 volts dc. This provides a substantial margin at 4500-volt operation.

2.2 *Troubles with 3A Seals and Pigtail Splices*

During the installation and early operation of the Florida-St. Thomas II System, two types of failures were encountered. The first involved improper seating of five seals in their seal casings. In each case, pressure on the type 3A (Fig. 3a) seal caused the outer sleeve to deform inward. This limited the plunger motion in the casing and prevented proper seating. A new seal, type 3C (Fig. 3b), is now being used for the SF System. It is a modified version of the 3A which provides a heavier cylinder plus a spring-actuated preload to establish and maintain proper seating of the plunger. Sixteen type 3C seals are operating satisfactorily in the Florida-St. Thomas II System. In addition, laboratory testing has confirmed the reliability of the 3C seal. All seals of the TAT-5 System are of the 3C design; no difficulty has been encountered with any of these.

The second failure mode involved four of the final repeater splices ("pigtail" splices) made on board ship. These faults occurred at the Jacksonville side of the Florida-St. Thomas II System at relatively low voltages within a short time of laying. The cause of these failures appears to have been a reduction in the prescribed mold block temperatures which occurred with a field modification of the temperature sensing device. Since that experience, we require periodic requalification of molds by molding a test sample instrumented with a thermocouple.

As a result of the pigtail problems, two further improvements were made in mold design and procedure:

- (i) elimination of a polyethylene preform which had been used as a filler, and
- (ii) provision of bleeder ports near the mold interfaces.

These two improvements provide more uniform heating and more complete mixing of the polyethylene materials, particularly at the critical interfaces. Severe over voltage tests have confirmed the large performance margin provided by this improved molding technique.

III. THE REPEATER UNIT, MECHANICAL

Certain repeater requirements, such as shock, vibration, electrical shielding, and the provision of a second barrier to water vapor diffusion, dictated the provision of a helium-tight metallic inner housing for the apparatus. This complete inner housing, with the repeater circuitry,

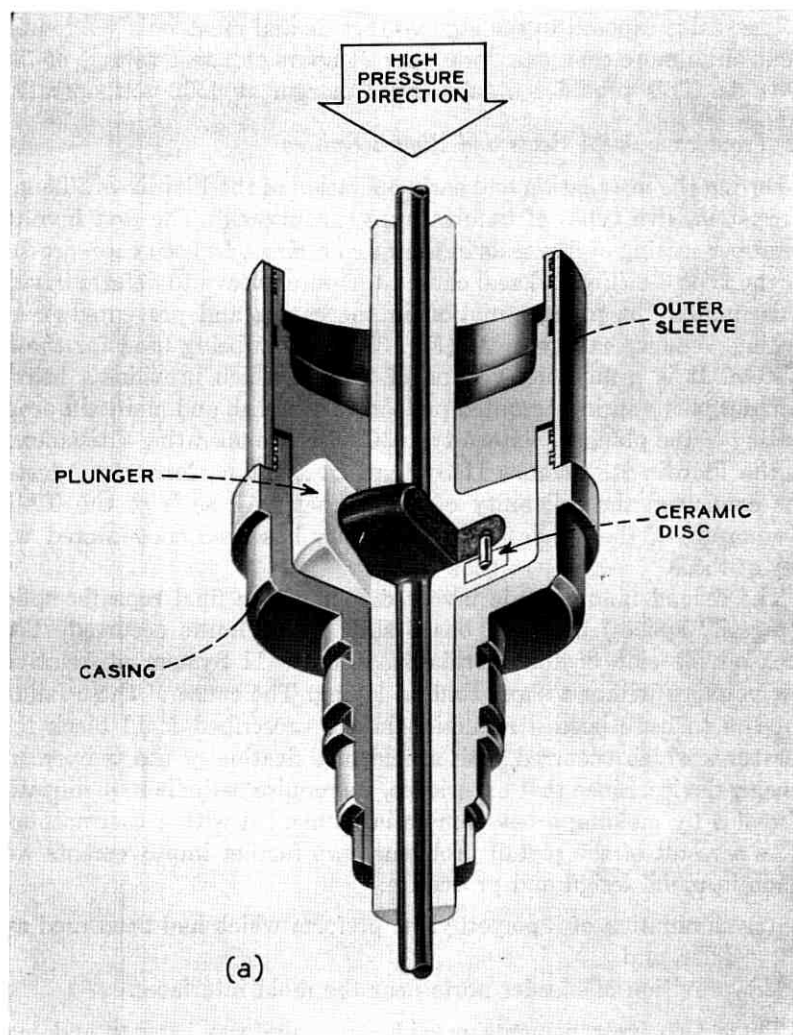


Fig. 3a—Type 3A seal.

is called the repeater unit to distinguish it from the complete repeater with pressure container.

The repeater unit housing is made of several aluminum plaster mold castings in a manner similar to that used for the SD repeater unit.³

The electrical connections to the circuitry within the repeater unit must be conveyed through the sealed walls of the repeater unit. This

requires low pressure seals. These seals are subjected to about 3.5 kg/cm^2 or 50 psi. Because we did not use a coaxial low pressure seal, the coaxial path is interrupted in coming through the repeater unit wall. Two separate seals are used and the path made coaxial outside the unit. Corona noise requirements argue against a double coaxial seal, and

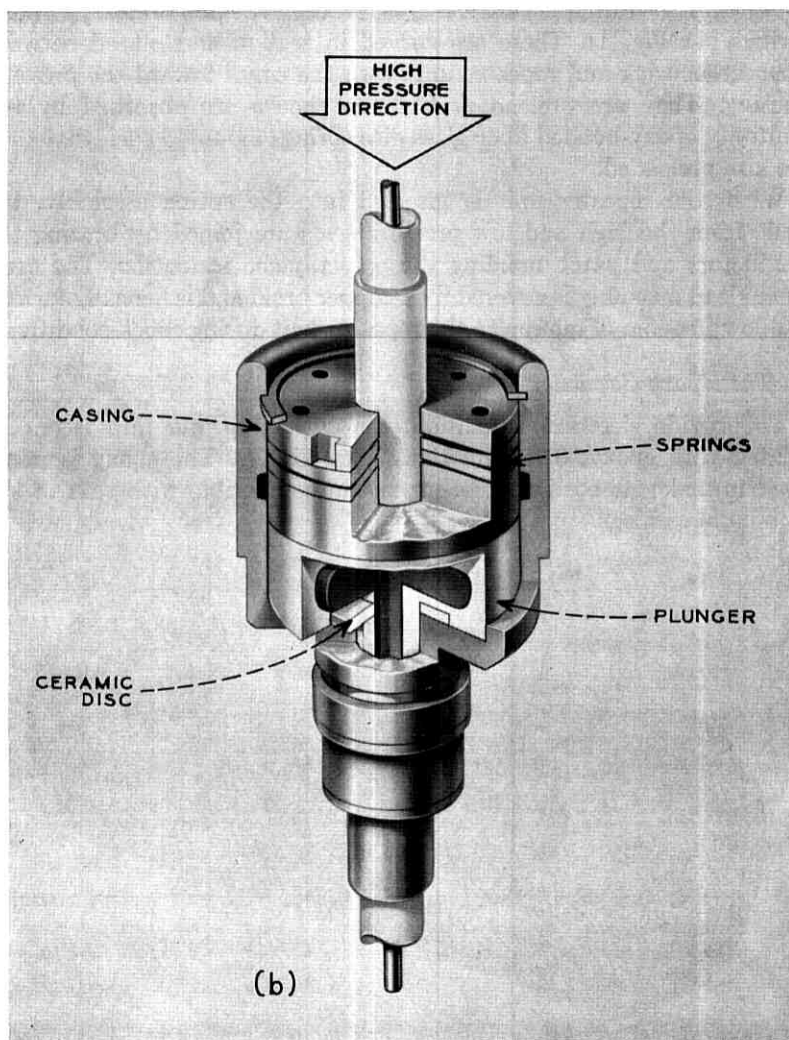


Fig. 3b—Type 3C seal.

impedance continuity is not a controlling argument in favor of such a seal.

3.1 Shock Mounting

To protect shock-sensitive elements in the repeater unit, the unit is spring mounted within the pressure housing. Twelve fiber glass and epoxy springs run the entire length of the repeater housing (Radial springs per Fig. 1). These are curved in section and placed between pressure housing and repeater unit with the edges toward the pressure housing. They are preloaded. Endwise shocks are absorbed by flat multi-ply epoxy-bonded fiber glass disc springs mounted such that they are also preloaded.

When the repeater unit is inserted into the spring assembly, the leads from the high and low pressure seals are joined by brazing the conductors and patch molding the polyethylene insulation. The completed lead assembly is covered by a copper braid and is helically formed to allow freedom of motion to the repeater unit during shock conditions.

3.2 The Epoxy Capacitor

To provide electrical insulation, the entire repeater unit is encapsulated with epoxy, 0.356 cm thick (See Fig. 4). The epoxy is mica-filled to make its coefficient of expansion compatible with that of the aluminum castings.

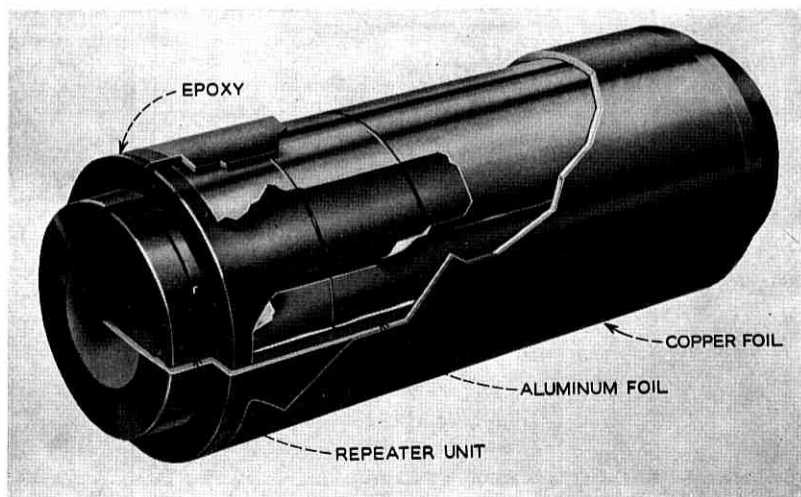


Fig. 4—Repeater unit.

To meet certain electrical requirements described in Section 4.3, it is necessary to provide a large-value capacitor between the repeater unit housing and the sea ground (Fig. 4). This capacitor must be free from corona impulse noise to an extent compatible with the assembled repeater objective of an average of no more than 4 "pops" per 24 hours at 4500 volts dc. Because of the surface discontinuities at the "O" ring between the cylindrical subassemblies, it is necessary to wrap the repeater unit with aluminum foil (0.05 mm thick) prior to epoxy coating to prevent voids and hence corona noise. This foil is stretched over the unit and rolled smooth. After epoxy encapsulation, the surface is machined smooth and an adhesive backed copper foil is wrapped over the machined epoxy surface. The foil is, in turn, electrically connected to the pressure housing through a foil wrapping over three of the fiber glass springs placed 120° apart. The net effect is that the epoxy coating provides over 3000 pF of capacitance between the repeater unit and the pressure housing. The high-frequency ground separation filter design relies on the fact that this capacitance has negligible parasitic series inductance.

IV. REPEATER CIRCUIT

The repeater block diagram, Fig. 5, shows the basic electrical sections of the repeater or repeater unit. After passing through ground separation filters, the transmission path is frequency split by the directional filters. The low-pass directional filters also serve to separate the dc power from the signal. This scheme uses a single common amplifier for both frequency bands.

4.1 *Amplifier*

4.1.1 *Configuration*

The amplifier is the heart of the repeater. Much of the design effort was spent on it. The extreme gain stability and linearity requirements dictate a feedback amplifier. As shown in the schematic of Fig. 6, the basic configuration consists of three common emitter connected transistors with the feedback network paralleling them. The transistors are germanium diffused-base mesa types having f_T 's in the 700 to 1000 MHz range. They were designed and fabricated solely for submarine cable application.⁴ No redundancy is provided. The output transistor, coded as the 38C, dissipates one watt. The remaining amplifier transistors, plus the one contained in the supervisory oscillator are coded separately as 38A, B, and D and biased in the 25 to 75 milliwatt range.

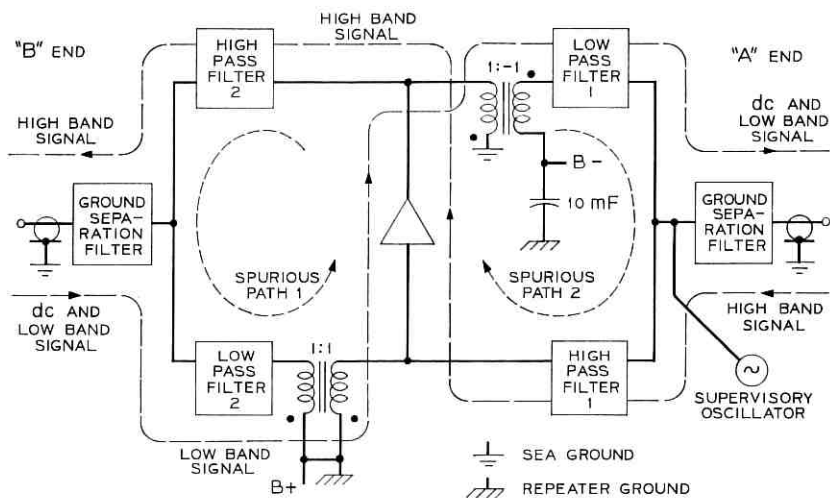


Fig. 5—SF repeater block diagram.

The repeater's signal carrying capacity depends on the power capability of the 38C, whose f_T of 900 MHz made it a state of the art device at the time the design was adopted. Noise figure requirements on the amplifier's input transistor led to operating this stage at 5.0 mA. To operate at this current and achieve high gain a separate transistor design was used. The same device is used for the middle transistor but operated at 15.0 mA, where its f_T and low frequency current gain are maximum, and its modulation properties preclude any significant degradation of amplifier linearity.

The shunt feedback connection produces very low active impedances at the amplifier's input and output ports. Thus we terminate these ports by means of simple series resistors. The resultant impedance of about 120Ω externally facing the input and output transistors is near the value needed for best noise figure and linearity.

4.1.2 Transmission Gain

All the gain shaping is accomplished in the amplifier's Beta network, which is a T -structure containing 16 components. A digital computer program using an iterative least-squares procedure adjusted the magnitude of the Beta network's transadmittance to achieve a theoretical repeater insertion gain match of better than ± 0.002 dB at all transmission band frequencies. In addition to compensating 10 miles of cable loss, the Beta network also provides for losses in the repeater's direc-

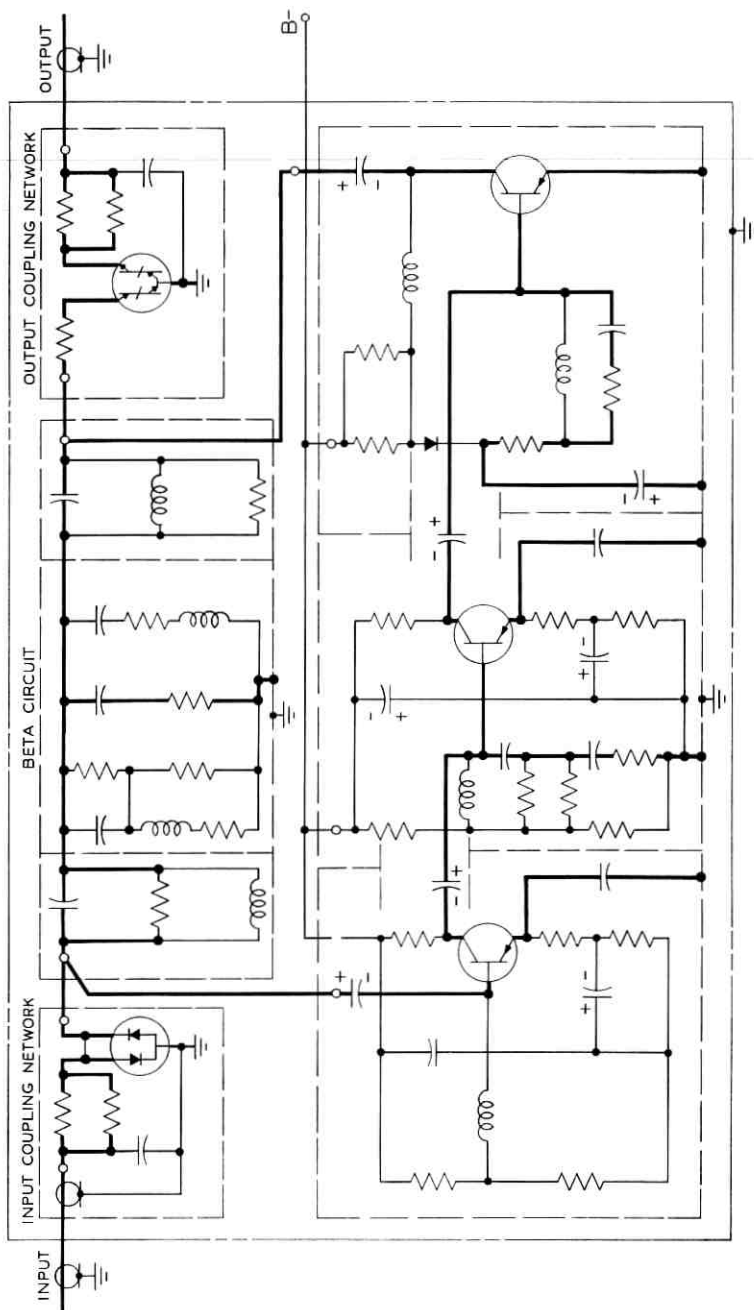


Fig. 6—SF amplifier schematic.

tional and ground separation filters, the amplifier's $\mu\beta$ effect, and a slight gain boost at low frequencies to provide additional flat loss for the ocean block equalizer.

Practical realization involved characterizing each component to an elaborate degree, extremely tight component tolerances, and precise adjustment of a slug-tuned inductor. A fixed inductor having the ± 0.05 percent tolerance required was not practical, but the same effective tolerance was obtained by adjusting it after its assembly as part of a low Q resonant circuit. This adjustment, made prior to amplifier assembly itself, is the sole trimming operation involved in the course of manufacture, and is instrumental in restricting amplifier gain deviations caused by Beta circuit component tolerances to about ± 0.03 dB.

The other major contributor to repeater transmission deviations is transistor variability. This effect, unlike Beta circuit variations, is suppressed by feedback. Besides referring to the differences in small signal characteristics that exist among transistors of the same code when biased identically, variability also involves the degree of accuracy with which the nominal transistor characteristics are known since any error here also translates into amplifier deviations.

To define the transistor's small signal characteristics, we adopted a device parameter equivalent circuit, following consultation with the device designers, experimental work on the transistors, and review of the equivalent circuit models used by others.⁵ We computer-fit the equivalent circuit model to two-port circuit parameters measured in the frequency range 0.5–250.0 MHz. The fitting operation validated the model shown in Fig. 7 within the accuracy of the test equipment (better than ± 0.2 dB and $\pm 2.0^\circ$). Although the device manufacturer's small signal measurements were not as comprehensive as those used to validate the equivalent model, they were designed to have sufficient scope to

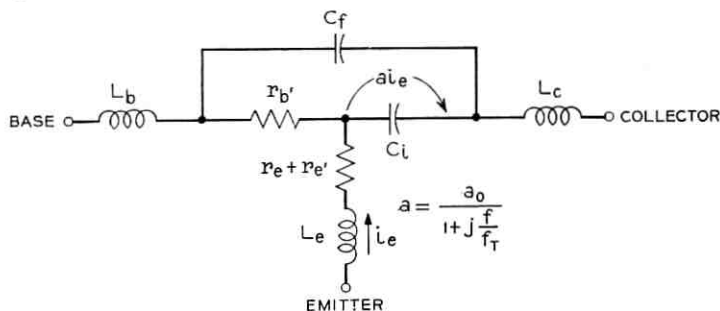


Fig. 7—Diffused base transistor equivalent circuit.

extract the element values of the model for each transistor destined for underseas use.

This procedure effectively yielded wide band characterization data necessary for establishing nominal or design center values, evaluating variability effects, and setting specifications. Development work indicated that random placement of transistors in amplifiers would produce transmission deviations in excess of 0.2 dB at the top frequency. This was unacceptable. Hence, a transistor grouping technique was used to sort transistors into amplifier sets on the basis of the manufacturer's small signal measurements.

The result of these efforts to control transmission deviations is summarized by the curves in Fig. 8, which indicate for the repeaters manufactured for the Florida-St. Thomas II Cable:

- (i) the closeness of match of average repeater gain to cable loss; and
- (ii) the degree of variability that exists among repeaters.

The ease and precision with which the Florida-St. Thomas II Cable was equalized are the best indications of how well these requirements were met.¹ The same transistor grouping was used in the TAT-5 SF System.

Although no significant changes in the system's transmission characteristic are expected from component and transistor aging effects, this fact was not known at the time basic circuit design decisions were made. As insurance, we decided to retain the phase-controlled feedback characteristic employed in the SD amplifier to suppress electron tube g_m aging effects.³ This covers transistor Beta aging, the only aging phenomenon which had been considered physically significant. The latest data available from the long-term tests described elsewhere⁴ verify that aging is restricted to Beta changes only, but these changes are so small that the resultant changes in system transmission are insignificant over a 20-year period (<1.0 dB). However, maintaining the $\mu\beta$ characteristic close to 90° as is done between 2.0 and 6.0 MHz, where feedback is low, is instrumental in restricting transmission deviations due to Beta variability among transistors. Calculations indicate that the phase-controlled design suppresses the gain change due to Beta variability as much as an additional 8.0 dB of feedback.

An important added result of the transistor characterization program is that it confirms the absence of any unsuspected small signal behavior in the transistor. The Tee equivalent circuit used proves to be sufficient and complete in the sense that additional elements could not have been identified uniquely on the basis of the comprehensive two-port measure-

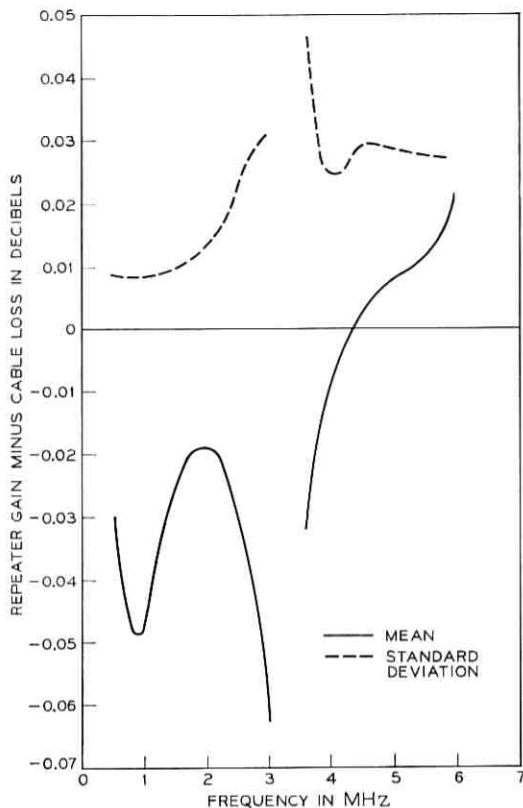


Fig. 8—Repeater gain characterization. Data applies to 112 production units.

ments in the frequency range of interest. (The word "elements" includes lumped components, frequency parameters, excess phase factors.)

The gain of this amplifier compresses slightly under signal load. The effect is measurable by loading all channels except for a narrow slot with gaussian noise and measuring the amplifier gain by application of a small amplitude sinusoid in the quiet slot. The result of such a measurement in the middle of the high band on a typical amplifier is shown in Fig. 9. Note that the gain with the full design load of -9.6 dBm0 per channel is 0.011 dB less than under light load conditions. For a transatlantic system, this could amount to over 4 dB of gain change with applied load and must be considered in the system layout.^{1,6}

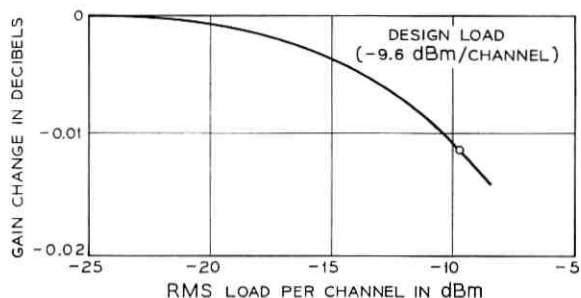


Fig. 9—Compression of SF amplifier at 4.715 MHz under noise load.

4.1.3 Feedback

Stability considerations limit the achievable feedback. The nominal feedback characteristic displayed in Fig. 10 shows that in-band feedback varies from 41.0 dB at 500.0 kHz, to 22.0 dB at 6.0 MHz. The major factors determining available feedback are high-frequency transistor response, length of the feedback loop, circuit parasites, and the margins established to guarantee stable operation. Except for the phase control exercised in-band, the $\mu\beta$ characteristic follows conventional lines. There are three major frequency regions to consider. To

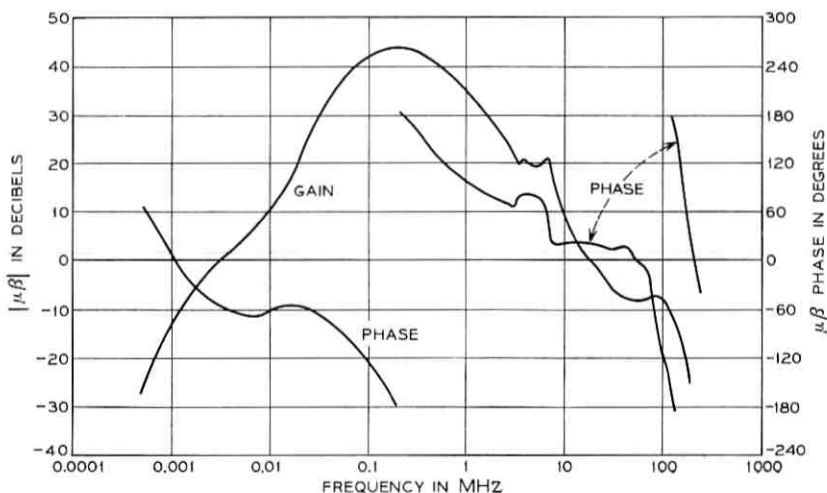


Fig. 10—SF amplifier $\mu\beta$ characteristic.

minimize deviations from the nominal in the high-frequency gain and phase margins, we pair the amplifier transistors with the input and output surge protector diodes; thus, variability in transistor high-frequency characteristics balances capacitance deviations of the diodes. This ensures that worst case margins are better than 4.0 dB and 15°. At the low-frequency cutoff, worst case margins are better than 10.0 dB and 30°. At the second high-frequency phase crossover (≈ 200.0 MHz) parasitic effects are stronger, and extra margin has been supplied there. Temperature effects in the operating range of interest (0° to 30°C.) are not significant.

As a means of verifying that measured margins were accurate, barely stable and unstable loops were constructed with the open loop measurements correctly forecasting the stability transition. Further checks proved the stability of the feedback loop under overload, and for the case in which an equalizer (whose return losses are very poor at high- and low-outband frequencies) terminates a repeater port. Pertinent computer-based calculations substantiated measurements, assessed variability effects, and evaluated parasites.

Two-terminal shunt interstage networks are the principal $\mu\beta$ shaping mechanisms. A special $\mu\beta$ computer program trimmed these networks and certain Beta circuit elements effective at high frequencies to achieve the desired feedback loop characteristic.

Amplifier construction uses "cordwood" configuration for network components and a compartmentalized cast aluminum framework for shielding individual amplifier stages and Beta circuit branches. The combination of three transistors, coupling, and Beta circuit bypass capacitors, plus inspectable soldered connections, produce a feedback loop length of about 36 cm (14"). The resultant inductive and capacitive parasites, plus transit time effects, are substantial contributors to the phase shift in the high-frequency asymptote. Although these parasites are relatively large for a high-frequency design, we keep them within narrow limits by close tolerances in fittings, piece parts and wiring strips, and by the extreme care taken in assembly.

To minimize the chance of unpleasant surprises in manufacture, we wanted to understand thoroughly the circuit design. To achieve such understanding, we characterized carefully each component and device, each network, the amplifier, filters, and finally the repeater. When we could accurately predict from computations how variations in components, devices, and parasites would affect pertinent response characteristics, we felt we had achieved the necessary confidence in the design.

4.1.4 Components⁴

A summary of the types of passive components used in the repeater is given in Fig. 11. All resistors are metal film types on sapphire substrates. Resistance tolerances range from a few tenths of a percent to 0.05 percent.

Open silvered mica capacitors serve for capacitances up to 6500 pF. The emitter bypass capacitors, in the tenth of a microfarad range, are paper capacitors, meeting tight requirements on low values of L and R at high frequencies. Solid tantalum capacitors in the 1 to 10 μF range serve as coupling capacitors.

The inductors are all air core, with values below 100 μH being single layer solenoids, and over 100 μH duolateral-wound solenoids. Tolerance and stability requirements are particularly tight for beta network values.

4.1.5 dc Design

The bias point of the output stage is maintained by a unique method.⁷ Because power for the entire system must be supplied from the shore ends, it is highly desirable to bias the transistor stages as efficiently as

RESISTORS

Type	Values (Ω)	Min Tol	Remarks
Metal Film	5-500	0.05%	Minimum L&C

CAPACITORS

Type	Values (μF)	Min Tol	Max V	Remarks
Mica	20-6500 pF	0.3%	15	Low L silvered
LV Paper	0.1-0.22	5%	15	Low L&R at high freq
Tantalum	1-10	7%	15	Low L&R at high freq
HV Paper	0.18	5%	3500	Low L

INDUCTORS

Type	Values (μH)	Min Tol	Q	Remarks
Air Core	0.087-15	5%	10-50	High stability, small size, 1 type adjustable
Air Core	0.2-65	0.5%	10-50	
Air Core	220-450	2%	10-50	Duolateral winding
Air Core	2-100	Adjustable $\pm 2\%$	150	High stability

TRANSFORMERS

Type	Remarks
Coaxial	4 turns/winding of 50 ohm coaxial cable
Phase Turnover	59.4 Ω 1:1 and 1:-1, $\pm 2^\circ$ phase match, electrostatic shield. For a 0.4 vrms, 1 MHz fundamental into 30 ohms $2f/f < -104$ dB and $3f/f < -131$ dB

Fig. 11—SF repeater passive components.

possible. For uniformity of performance, it is also desirable to maintain the bias points rather accurately over a wide range of transistor dc characteristics and component tolerances. This latter point suggests some form of dc feedback. Emitter feedback and base bleeder networks provide stable bias in the first two stages. This scheme is inefficient of power; hence, for the output stage another approach was created. This method places a zener diode between collector and base to establish a constant collector-to-base voltage. Since the emitter-base voltage is very low and reproducible, this scheme also establishes a near constant collector-to-emitter voltage. The current regulation of this stage occurs in the shore terminals and utilizes the fact that most of the dc cable current flows through the output emitter. At dc, the output stage serves as an efficient voltage regulator for the bias voltage of the previous stages. Under constant voltage conditions these earlier stages draw a constant current and thus leave a constant current for the output stage emitter.

4.1.6 *Equipment Design*

The physical design considers inspectability of completed assemblies, chemical and physical stability of all materials, and uniform thermal behavior under all conditions.

The circuit components are mounted in cavities in molded plastic end plates. Wiring strips spaced 0.2 cm away from the plates are used to interconnect the components. This spacing affords visibility for the inspection of both sides of the soldered joints. Amplifier networks are then mounted into individual cavities of a casting as shown in Fig. 12. Adding a cover casting completes the shielding of the amplifier.

Consistent thermal behavior is maintained by uniform heat conducting geometry. Realization of maximum reliability of the output transistor required that the device junction temperature be minimized while maintaining electrical isolation to ground. This is accomplished using a beryllium oxide mount for electrical insulation and thermal conductivity.

4.2 *Directional Filter*

The directional filter is a constant-impedance four-port network with two low-pass and two high-pass filters (See Fig. 5). Special high- Q , adjustable air core inductors meet the critical requirements of this filter. In-band transmission ripples are held to less than a thousandth of a dB.

The directional filter creates two spurious feedback paths around the

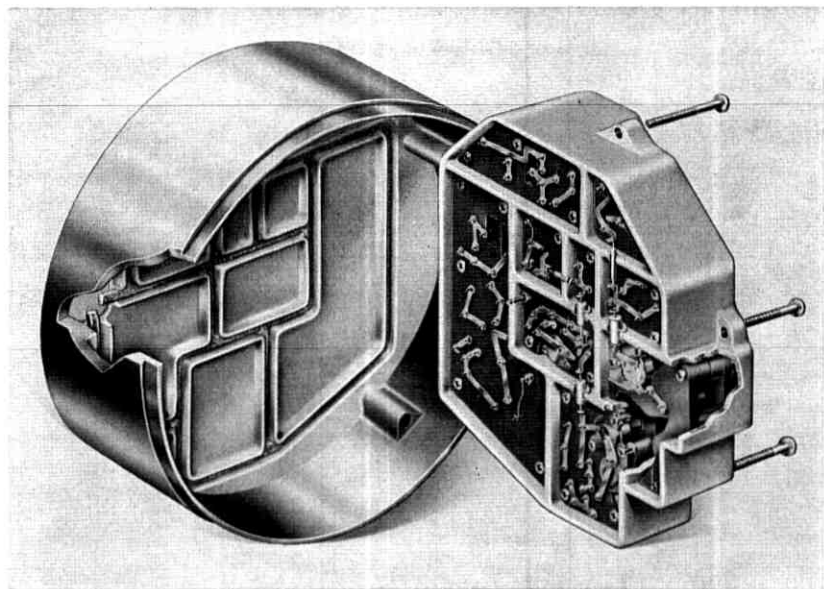


Fig. 12—SF repeater amplifier unit.

amplifier that could affect the overall gain characteristic if the loop losses were inadequate. These two spurious paths are symmetrical except for the presence of a 1 : 1 transformer in one path and a 1 : -1 transformer in the other. The two unwanted signals substantially cancel at the amplifier input port if an adequate balance is maintained. To aid in realizing this balance, return loss at each directional filter port must be maintained above 27 dB.

Referring to Fig. 5, the ground separation filters and the two low-pass filters conduct dc cable current to amplifier supply points marked $B+$ and $B-$. Since the $B-$ point must be at ac ground, it is bypassed with a $10 \mu\text{F}$ capacitor. Close control must be maintained on the series resistance of this capacitor since it directly affects the loss of the low-pass filters and hence the low-band repeater gain.

We avoid serious interactions among the filter sections of the directional filter by shielding each section separately.

Each section contains a large, high- Q , air core inductor which is surrounded by as much space as possible to minimize the loss of Q due to eddy currents in the shield walls. The design to meet these objectives is shown in Fig. 13 where ten sectors appear on each side of the filter. With the large inductor placed toward the base of the sector we realized

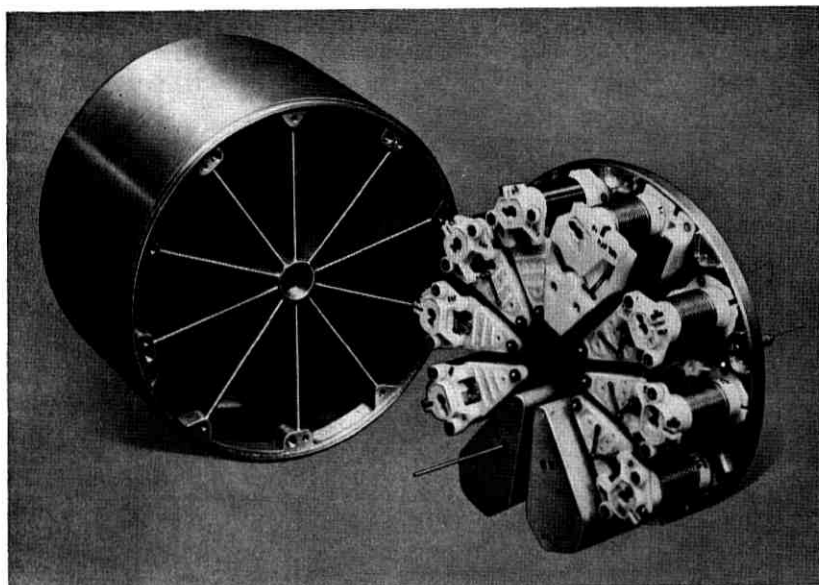


Fig. 13—Directional filter assembly.

maximum space to surrounding shields. The capacitors which make up the rest of the sector are placed toward the apex.

In the circular design the input low-pass filter section is adjacent to the output high-pass filter section, and vice versa. Crosstalk requirements are severe and are met through the use of additional copper shielding to double-shield each of the four critical sections.

Since the various filter sections used the same component codes, we modularized the filter construction. This modularization used three mounting plate designs instead of 28. With the quantities achieved thus, we exploited injection molding to reduce the cost. Tooling costs are considerably lower because of fewer designs. Material costs are reduced by using "celcon" (acetal plastic). This material also facilitates assembly by introducing controlled flexibility.

4.3 *Ground Separation Filter*

As illustrated by Fig. 14, the ground separation filters' function is simply to isolate at dc the cable's sea ground potential from the repeater ground's high potential, while coupling signals between the two. To achieve the dc isolation, capacitors with a voltage rating equal to the

highest voltage on the system (4500 volts in this case) couple the two grounds at each cable port. Unfortunately, any signal frequency impedance in this ground coupling path at one repeater port will cause a voltage drop which will appear across the opposite repeater port and this will affect the repeater gain. Thus, we must maintain the ground coupling impedance as low as possible. At the low end of the transmission band, this requires a large capacitance ($0.18 \mu\text{F}$) for the coupling capacitor. At the high end of the band, the impedance is determined by the inductance in the ground coupling path. We would like to minimize this inductance by keeping the ground coupling path as short as possible. However, the coupling path passes through a separate seal in the repeater unit. Also, the high-voltage capacitor is physically large. Hence, the inductance in the coupling path is large. The effects of this high coupling impedance at high frequencies are alleviated by the use of coaxial chokes at the input and output ports of the repeater. The chokes consist simply of 4 turns of coaxial cable wound around a ferrite core. This choke tends to force the current to be equal in the center conductor and the return sheath. As a consequence, additional isolation is achieved between repeater input and output. At still higher frequencies (60 MHz), such measures are still inadequate to assure adequate ground separation filter loop loss for repeater stability. At these frequencies, we use the capacitance between the repeater unit and the pressure housing. Fortunately, this capacitance has negligible series inductance. As previously mentioned, this capacitance is increased to 3000 pF by applying a metal electrode (connected to sea ground) around the epoxy coated repeater unit.

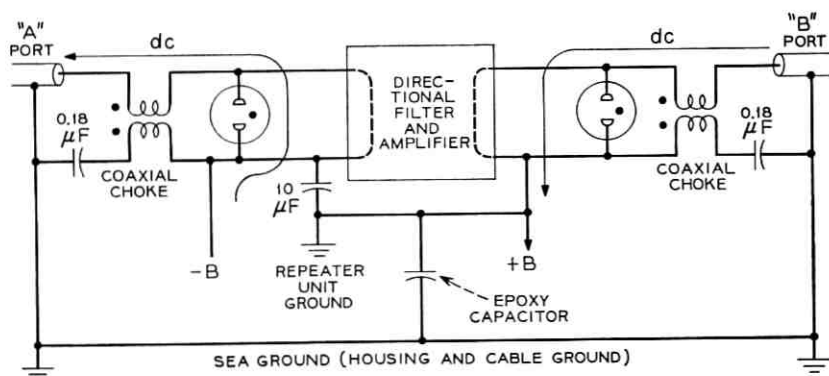


Fig. 14—Ground separation filter schematic.

4.4 *Surge Protection*

The high voltage present on the cable can produce a severe transient if the cable is accidentally shorted. The energy in the resulting transient is sufficient to damage a large number of repeaters in a system. To avoid such a catastrophe, we include surge protection circuitry in the repeater. We place gas tubes across the input and output ports of the repeater to dissipate much of the energy that could otherwise damage the capacitors in the directional filters. These gas tubes, of the same design as those used in the SD System, break down very rapidly when the voltage exceeds 75 volts. The gas tubes' protection is not adequate for the transistors in the amplifier but the tubes do reduce the subsequent surge appearing on these devices.

Upon passing through the directional filter, the components of the surge take different paths. A low-frequency component appears in the power path as a current transient having a peak value of as much as 60 amperes and a tail lasting for over 1 second. A zener diode across the amplifier power terminals protects the transistors from this surge component. This diode has a breakdown voltage of 15 volts, a few volts higher than the normal bias voltage.

High-frequency components of a surge which are not shunted by the gas diodes are passed by the directional filter to the amplifier input and output ports. The input port is protected by being shunted with a pair of oppositely poled silicon diodes. Signal levels at the point of application are sufficiently low so that intermodulation due to diode non-linearity is negligible. The normal signal level at the amplifier output is much higher, making this type of diode unsuitable at this point. Instead, the output port is protected by the shunt connection of a pair of oppositely poled silicon pnpn diodes. These devices break down at a peak voltage of about 18 volts and exhibit a negative resistance characteristic. It was found by experiment that these devices give adequate protection to the transistors used, without degrading modulation performance.

4.5 *Supervisor Oscillator*

Each repeater contains a single-transistor, crystal-controlled oscillator which injects a supervisory tone at each repeater. Half of the repeaters have tones falling just below the bottom edge of the low-frequency transmission band and the remainder of the repeaters have tones falling just above the top edge of the high-frequency transmission band. High-

and low-band tones are assigned to alternate repeaters. These tones can be monitored at the shore terminals to detect system misalignments and changes in repeater performance. In the event of a repeater failure which does not interfere with system powering, the received supervisory tones identify the failed repeater.

The oscillators have unique assigned frequencies spaced 100 Hz apart. To meet the lifetime stability requirement of ± 40 Hz the oscillators use quartz crystals.

Because the injection level of the supervisory tone is very low (-50 dBm in the high band and -60 dBm in the low band at the repeater output), we mount the oscillator in a shielded compartment and decouple the oscillator bias from the amplifier bias supply. This biasing isolation also assures repeater operation despite any type of fault in the oscillator circuit.

The low-frequency oscillator has stringent requirements on any harmonics that may fall in the transmission band. These requirements are met by filtering the oscillator output through a low-pass filter.

To assure accurate system performance monitoring, the injected signal level of each supervisory oscillator should not vary from its initial value by more than ± 0.5 dB at any operating temperature over the expected system lifetime. The initial value of this level is factory adjusted to ± 0.3 dB. To achieve the required level stability requires: (i) a diode amplitude limiter, and (ii) further temperature compensation. A thermistor is used in the oscillator output for this further compensation.

V. EQUALIZER

5.1 *General Description*

The function of the equalizer is to compensate most of the accumulated misalignment in an ocean block consisting of 20 repeaters and 356 kilometers (192 nautical miles) of cable. This misalignment arises from design and manufacturing deviations in both repeaters and cable, and also from the unpredicted effects produced on cable attenuation as the cable is layed (laying effect).

Each of the different causes of misalignment has a distinctive deviation shape versus frequency. At the end of any ocean block, the overall accumulated misalignment will consist of the sum of many such shapes. Equalization is accomplished by connecting several equalizer sections in tandem. Each section is designed to compensate a particular type of

deviation shape. In this respect, the SF equalizer is similar to the one used in the earlier SD System.³

The SF equalizer differs, however, in that the low and high transmission bands are equalized separately by using directional filters and separate sets of equalizer sections. A block diagram of the equalizer is shown in Fig. 15.

Each equalizer section contains passive constant-resistance, bridged-tee networks. The following basic types are furnished:

- (i) Buildout Equalizer,
- (ii) Repeater Direction Filter Q Equalizer,
- (iii) Mop-up Equalizer,
- (iv) Cable Manufacturing Equalizer (common to low and high bands),
- (v) Switchable Equalizer.

In addition to the coaxial input and output connections to the cable, the equalizer also has a transmission test lead and a switching control lead. These connect through high-pressure seals to test connectors on the exterior of the pressure housing. The transmission test lead goes directly to the signal path at the inboard end of the equalizer. Thus test room personnel may measure transmission between the cable ship and the shore terminal as the cable is being laid. These measurements show the amount of misalignment up to the equalizer.

Just before an equalizer overboards, the personnel switch it to a network combination which best compensates the measured deviation. This is done via the switching control lead on the outboard end of the equalizer, which sets 14 magnetically latched relays.

5.2 Mechanical Considerations

The equalizer unit design follows the pattern already described for the repeater unit. There are a few differences. The epoxy coating of the equalizer unit does not require any foil wrap since an external capacitor is not required. Internally there are differences in the mounting of components of the individual networks. A standard module or cavity serves much of the equalizer. This was especially attractive for this application because the many different shaping networks used in the equalizer are comparable in complexity. Thus, the space required for the most complex network can also be efficiently used for the somewhat simpler networks.

The cylindrical shape of the equalizer and the expected size of the networks dictated "pies" subtending 40° and thus allowing nine sections to a complete layer. Adoption of a standard module for the equalizer allows the use of "universal" end plates which can mount a wide variety of component types.

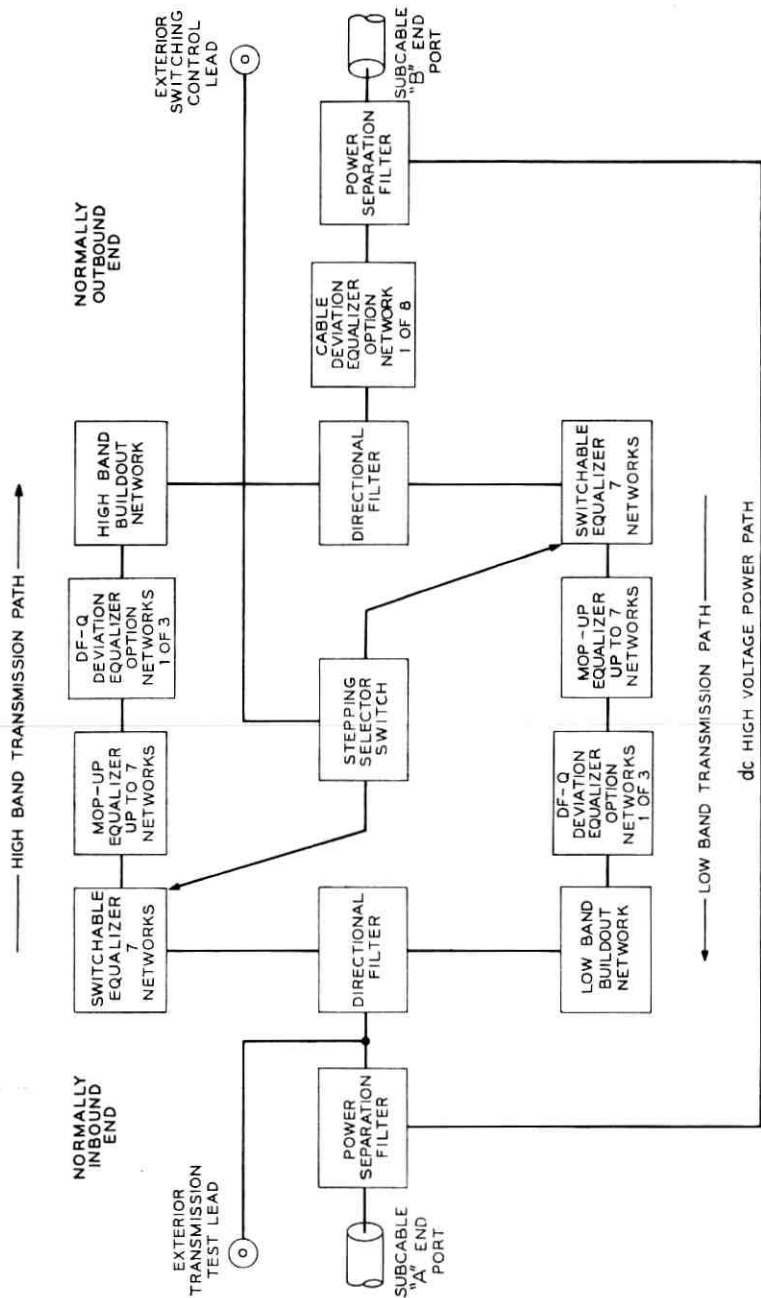


Fig. 15—Equalizer block diagram.

5.3 Equalizer Networks

5.3.1 Switchable Networks

Upon separation of the two directions of transmission by the directional filter, each band goes into a switchable equalizer network. The shapes provided by the switchable network are shown in Table III. The shapes were designed to compensate laying effects on cable attenuation and residual manufacturing deviations of repeaters and cable.

Seven bridged tee networks for each transmission band give the desired shape combinations. Each network contains a relay having two sets of transfer contacts. A positive pulse latches the contacts in one position, and bypasses the network. A negative pulse latches the contacts in the opposite position and inserts the network. Two states for each

TABLE III—SWITCHED NETWORKS—INSERTION LOSS FUNCTION

Band	Sequence Number	Approximate Insertion Loss Function
Low	1	$8\sqrt{f/6}$
	2	$4\sqrt{f/6}$
	3	$2\sqrt{f/6}$
	4	$1\sqrt{f/6}$
	5	$10\sqrt{f/3} - 3f$
	6	$5\sqrt{f/3} - 1.5f$
	7	$5f\left(1 - \frac{\log(1 + f/3)}{\log 3}\right) - 2.38\sqrt{f}$
High	8	$8\sqrt{f/6}$
	9	$4\sqrt{f/6}$
	10	$2\sqrt{f/6}$
	11	$1\sqrt{f/6}$
	12	$13.5\sqrt{f/6} - 2f$
	13	$6.8\sqrt{f/6} - f$
	14	$1.2 \text{ dB bump} + 2\sqrt{f/6}$

f in MHz.

of seven tandem networks yield a total of $2^7 = 128$ transmission shapes per band. The magnetically latched relays are energized sequentially via a stepping selector switch.

5.3.2 Mop-Up Networks

As blocked out in Fig. 15, in each transmission band the optional mop-up networks follow the switchable network. These mop-up networks are a special series of simple bump, slope, and flat networks which provide a highly flexible and accurate means for equalizing repeater design deviations.

From one to seven mop-up networks per band can be designed and assembled from a factory stockpile of piece parts and predetermined component values. Transmission measurements performed at the manufacturing locations on specific repeaters and cable sections of an ocean block determine the mop-up network designs. To achieve this flexible mop-up capability we have 8 dB of flat loss available in the low band, and 12 dB in the high band.

Fig. 16 shows the loss shapes of the four high band mop-up networks used in one ocean block equalizer of the Florida-St. Thomas II System.

5.3.3 Directional Filter Q Network

In each transmission band we follow the mop-up networks with a directional filter Q network. This factory option equalizes manufacturing Q variations of the repeater directional filter inductors. This deviation has a predictable shape whose amplitude depends on the average Q of directional filter inductors in the ocean block. The options available in each band allow either +1 dB, -1 dB, or 0 dB compensation at the cut-apart region band edge.

After all available flat loss, filter loss, and switched networks bias loss has been accounted for, buildout networks in each band add the loss shape required to make the total nominal equalizer loss match the loss of 8 nautical miles of cable. The low-band buildout also includes an overload limiting circuit. This circuit consists of a transformer and a pair of oppositely-poled diodes which clip abnormally high signal peaks. This limiter prevents buildup of a system overload singing condition which could develop as repeaters modulate highlevel noise components between low and high bands (See Ref. 8).

5.3.4 Cable Manufacturing Equalizer

Finally, we again combine the two bands passing through the equalizer via the other half of the directional filter. Here we have a cable manu-

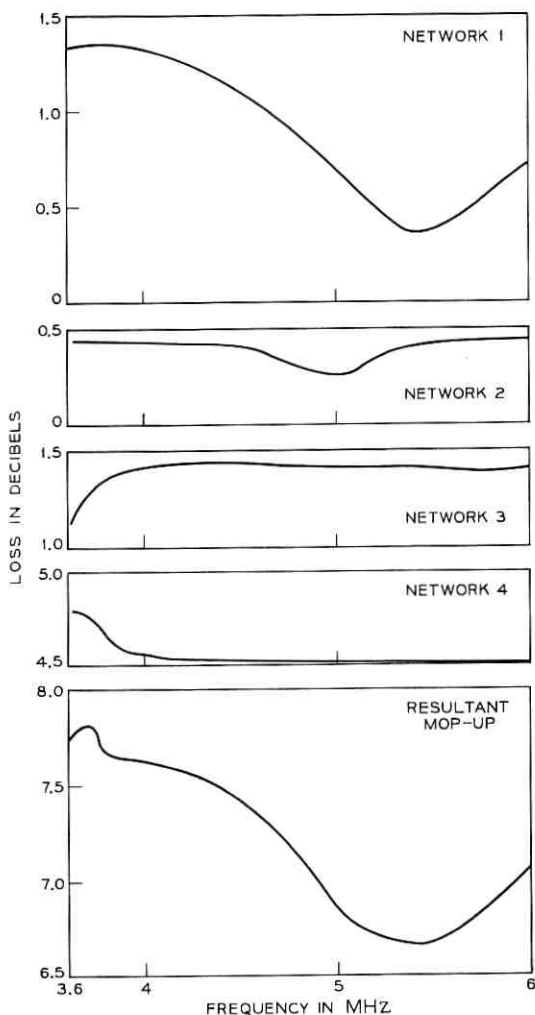


Fig. 16—Mop-up network shapes used in typical equalizer.

facturing equalizer which is common to both transmission bands. This cable equalizer is chosen during equalizer manufacture. It compensates for that part of block deviation caused by cable batches having more or less than nominal dielectric conductance loss. Any one of eight possible networks or none at all may comprise the cable manufacturing equalizer. The general loss function for all shapes is $\alpha_n[(f/6)^{\frac{1}{2}} - f/6] + \beta_n$. The eight possible sets of values for α_n and β_n appear in Table IV.

TABLE IV—CABLE MANUFACTURING EQUALIZER OPTIONS

n	α_n	β_n
1	1.95	0.52
2	3.95	1.01
3	5.96	1.01
4	-1.91	1.13
5	-3.89	1.56
6	-6.13	2.13
7	-8.15	2.63
8	-10.36	3.17

VI. CONCLUSION

The first SF Submarine Cable System went into service between Jacksonville, Florida, and St. Thomas, Virgin Islands, in August 1968. TAT-5, a second installation, is now in service between Green Hill, Rhode Island, and Conil, Spain. Equalization of both systems is outstanding due to the accurate gain match achieved in the repeaters and the high degree of flexibility provided in the underseas equalizers. With a nominal design capacity of 720 channels, these systems actually meet equalization and signal to noise objectives for over 820 channels.

The SF repeaters and equalizers form an effective partnership in transmitting high-quality telephone conversations over the SF Cable. Their success is due in no small measure to a coordinated effort among members of several Bell Laboratories groups and their Western Electric colleagues to realize extreme reliability, precision, and ruggedness in each undersea unit.

REFERENCES

1. Anderson, C. D., and Easton, R. L., "An Overview: Requirements and Performance," B.S.T.J., this issue, pp. 605-630.
2. Lebert, A. W., and Schaible, G. J., "Ocean Cable and Couplings," B.S.T.J., this issue, pp. 699-719.
3. Brewer, S. T., Dickinson, F. R., and Von Roesgen, C. A., "Repeaters and Equalizers for the SD Submarine Cable System," B.S.T.J., 43, No. 4 (July 1964), pp. 1243-73.
4. Wahl, A. J., McMahon, W., Lesh, N. G., and Thompson, W. J., "Transistors, Diodes, and Components," B.S.T.J., this issue, pp. 683-698.
5. Pritchard, R. L., "Electric-Network Representation of Transistors—A Survey," IRE Transactions of the Professional Group on Circuit Theory, CT-3, No. 1 (March 1956), pp. 5-21.
6. Lynch, R. L., Thomas, J. L., and Von Roesgen, C. A., "Shore Terminal Facilities and Fault Location," B.S.T.J., this issue, pp. 721-748.
7. Brewer, S. T., Buus, R. G., and Kassig, J. J., "Cascaded Transistor Amplifier Biasing Arrangement," U. S. Patent 3-254-303, applied for September 1963, issued May 1966.
8. Anderson, C. D., "Overload Stability Problem in Submarine Cable Systems," B.S.T.J., 48, No. 6 (July-August 1969), pp. 1853-1864.

Manufacture of Submarine Cable Repeaters and Ocean Block Equalizers

By ALAN T. CHAPMAN

(Manuscript received May 21, 1969)

In this article we describe the facilities and procedures used to manufacture SF repeaters and equalizers. Careful selection and training of production and inspection personnel play a vital role in achieving the desired standards in qualifying material and for assembly and test of units. Detailed engineering planning provides the best environmental conditions and manufacturing facilities and processes. Close collaboration by the manufacturer with Bell Telephone Laboratories during product design and specification and during manufacture assures a tightly controlled, well understood product.

I. INTRODUCTION AND OBJECTIVES

The responsibility for manufacture of SF Repeaters and the associated Ocean Block Equalizers is assigned to the Clark (New Jersey) Shop of the Western Electric Company. The organizational structure is comparable to those previously employed during manufacture of SB and SD Repeaters.^{1,2} Nine levels of Laboratory Technicians perform work directly on the product.

Manufacturing development work on the SF Repeaters started early in 1963 and culminated in the first delivery of the finished product near the end of 1966. This and later experience proved that a practicable minimum of three years is required to (i) develop and procure facilities for the Clark Repeater Shop, (ii) obtain and qualify Laboratory Technicians, (iii) establish other Western and outside supplier sources for components, parts and material, (iv) manufacture (including six months of life testing) a set of apparatus, and (v) resolve design and manufacturing difficulties before the first repeater of new design is fully assembled

and qualified for project application. The sequence of major steps in repeater and equalizer manufacture is shown on Fig. 1.

Engineering and manufacture of SF Repeaters and Equalizers provided new challenges to Western Electric in production of high reliability apparatus. In order to achieve a broader band system, we had to employ not only new devices, with less-than-desired time-proven integrity, but also approximately twice the number of repeaters, com-

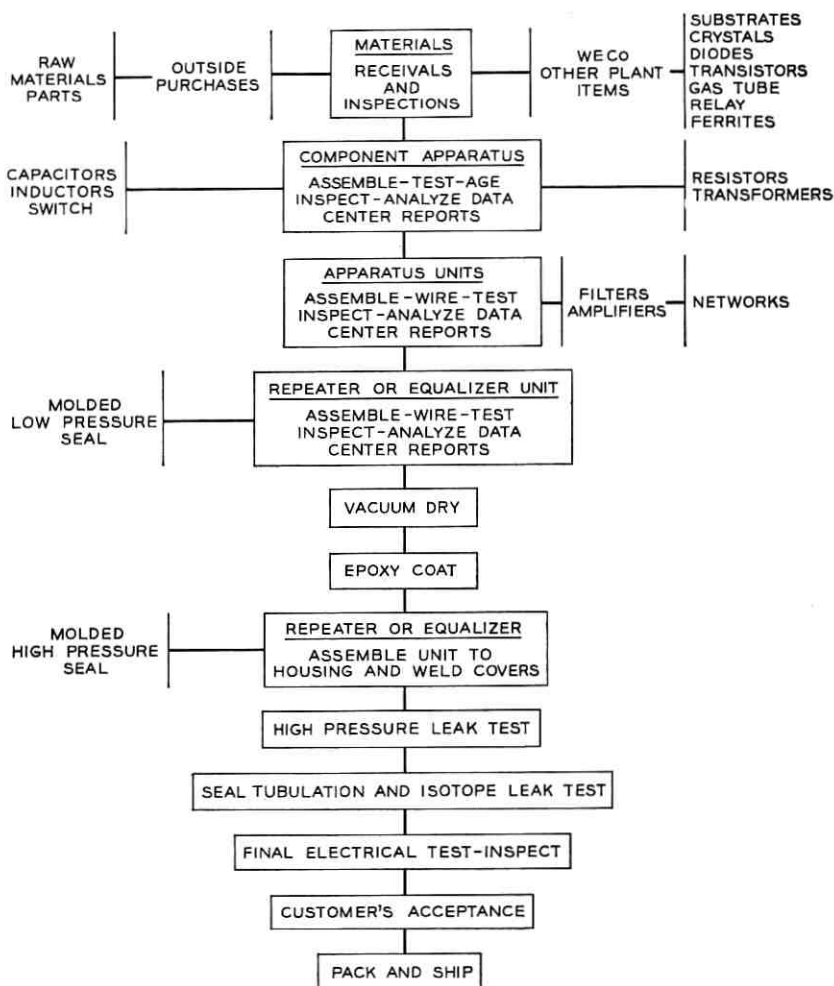


Fig. 1—Flow chart of major steps in the manufacture of repeaters and equalizers.

pared with SD Systems, in a given SF System to offset cable attenuation at the higher frequencies.

One essential key to success in manufacturing devices with exceptional reliability rests in the development and maintenance of a team with the integrity and dedication to strive constantly toward perfection and to be dissatisfied with less. By providing this prime essential, as well as superior environment and facilities, it is possible to realize the full reliability potential of the basic design.

1.1 *Production Objectives*

Although the initial production objective was to manufacture 417 SF Repeaters and Equalizers (sufficient for a Transatlantic System to be completed by the end of 1967), the program was revised downward to 152 units which were completed early in 1968. These units were provided for the Florida-St. Thomas System which was in service by August 1968.

After a production gap of about a year and a half, authorization was received to manufacture repeaters and equalizers for the TAT-5 System. To complete the units in time for an early 1970 in-service date, it became necessary to increase the manufacturing capacity to permit production at nearly double the previous peak rate. The Hillside (New Jersey) Shop (source of SB Repeaters) was reactivated by Western Electric Company as the Clark Annex. Operations temporarily transferred to this area included raw material inspection, inductor and transformer manufacture and directional filter section assembly work.

II. ENVIRONMENTAL CONTROL

"White" room conditions conform to an objective of less than 50,000 particles above 0.5 microns per cubic foot and temperature $75^{\circ} \pm 2^{\circ}\text{F}$. with relative humidity less than 40 percent. Humidity of less than 20 percent is maintained in the paper capacitor winding room. To monitor "dust" count effectively, we procured a Royco Particle Counter Model 200* and associated Royco Digital Recorder Model 122. This instrument permits identification and segregation of particles from 0.3 microns to 10 microns and larger. In case of difficulty, such information on particle sizes and concentrations can pinpoint the source of airborne contaminants.

Holding the environmental conditions to the desired levels during major moves and rearrangements and installation of new facilities was achieved through use of temporary dust-tight partitions, blocking

* Royco Instruments Inc., Menlo Park, California.

off registers to the air conditioning system and close examination of construction materials.

Laboratory Technicians wear special uniforms including caps and shoes. The uniforms and cleanliness rules are essentially those adopted for production of the SB and SD repeaters¹. Pleasant working conditions and amenities contribute indirectly to reliability by keeping turnover low and morale high.

2.1 *Transportation of Project Apparatus*

The necessity for opening the Clark Annex introduced a new problem of transporting process project material between two locations. Heretofore component material from other locations (electron and gas tubes from the Western Electric plant in Allentown, Pennsylvania; transistors and diodes from the Reading, Pennsylvania, Plant; crystal units from Merrimack Valley, Massachusetts; relays from Burlington, North Carolina) were shipped fully encapsulated and required primarily shock and some temperature protection. Therefore, proper packing and transportation by messenger sufficed.

In the case of SF, airtight containers, as shown in Fig. 2, were used. After being loaded with the product, they are fastened to the floor of an air conditioned station wagon for transport. Shock monitors nested in each container provide control of apparatus against excessive shock or vibration. Test runs with an impactograph aided in route development.

III. PERSONNEL QUALIFICATION AND INDOCTRINATION

Manufacture of high reliability devices demands basic characteristics of integrity, dependability and meticulousness on the part of every employee. All Laboratory Technicians receive a 10-hour (five two-hour sessions) indoctrination course emphasizing the need to report the slightest deviation in their own work from established written requirements. The course covers in considerable (but nontechnical) detail the nature of the product and applications, the point at which each employee's effort fits into the whole, and general practices including use of uniforms.

3.1 *Qualification of Laboratory Technicians*

On-the-job training ordinarily takes two to four weeks, the time required before a Laboratory Technician is qualified to work on project material. The final test requires proving the ability to repetitively perform an operation faultlessly; for example, wind five paper capacitors

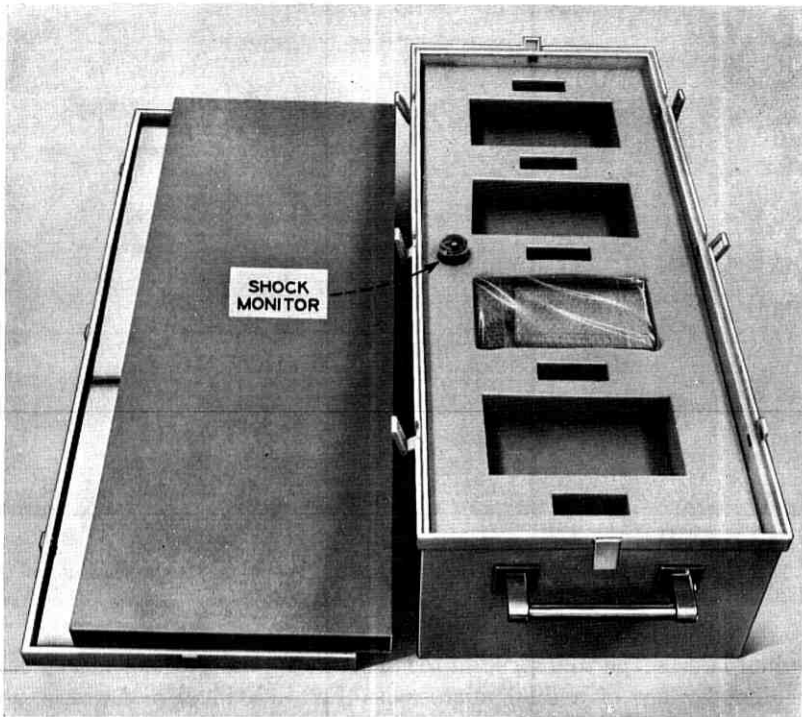


Fig. 2—Shipping container with product packaged and nested in plastic foam and protected by shock monitor.

in sequence perfectly. Only when an individual can demonstrate consistent mastery of all operations under his control by producing a predetermined number of pieces of required quality in a limited number of attempts can he be considered sufficiently qualified to work on project material. Furthermore, the Laboratory Technician is subjected to verbal interrogation to evaluate his understanding of the assignment. This inculcation procedure, though severe, is a basic prerequisite to built-in reliability. When a measurable deviation from what is acceptable (and within the control of the individual) is observed, the qualification procedure must be repeated. Individuals who fail to qualify as technicians are placed on other Company assignments without abasement. Disregard of requirements or taking short cuts for whatever reason can be very costly and is another basis for relieving a technician from submarine cable repeater work.

Modern training aids are used. Figure 3 shows an instructor using

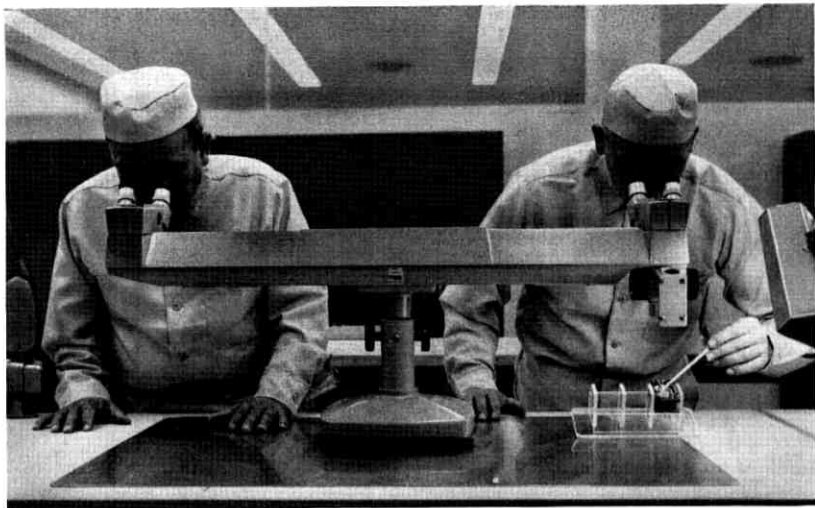


Fig. 3—Dual microscope used to instruct a Laboratory Technician as to the quality of a soldered connection.

a dual microscope to instruct a technician on quality of a soldered connection by use of a dual microscope. For individual and group use, certain operations are video taped in black and white or color. Figure 4 shows a video taping session with a laboratory technician stacking a mica capacitor with the recording being monitored on the TV screen.

3.2 *Indoctrination of Support Personnel*

Equal care is exercised in the selection and indoctrination of new supervisors, engineers, production people, and so on. The thoroughness, accuracy and promptness of their activities has a prime bearing on the effectiveness of the laboratory technicians.

The purpose of and the need for conforming to the special Clark Shop practices is explained to all support personnel, accountants, guards, employment, cafeteria and yard help. The caliber and size of the janitorial force is particularly important in order to maintain environmental conditions.

Clark Shop personnel working in the period 1963-1970 proved that they were capable of developing facilities and introducing manufacture of a new design (SF Repeaters), maintaining production on the previous design (SD Repeaters), restarting an old design (SB Repeaters), and beginning production of a new design for the Government (SD-C Re-

peaters), without relinquishing high reliability in manufacturing standards. The results add to the assurance that methods for selecting, indoctrinating and qualifying laboratory technicians and for obtaining support personnel are proper.

IV. MANUFACTURING FACILITIES AND PRODUCT DEVELOPMENT

Over 7 million dollars were expended to provide new and changed facilities required for SF Repeater and Equalizer manufacture at the Clark Shop and for associated components made at other Western Electric manufacturing locations, primarily Reading and Allentown. These facilities were necessary to produce, age and test new apparatus components and assemblies, improve precision and increase manufacturing capacity.

The following illustrations show typical development work required to prepare for manufacture.

4.1 *Mica Capacitor Lamination Silver Coating Machine.*

More stringent electrical requirements dictated the need for better registration between silver coatings on the mica lamination. Increased output of the machine was required to meet schedule requirements.

Precise registration of the screens for transferring silver paste to



Fig. 4—Stacking of a mica capacitor being videotaped and viewed simultaneously on TV screen.

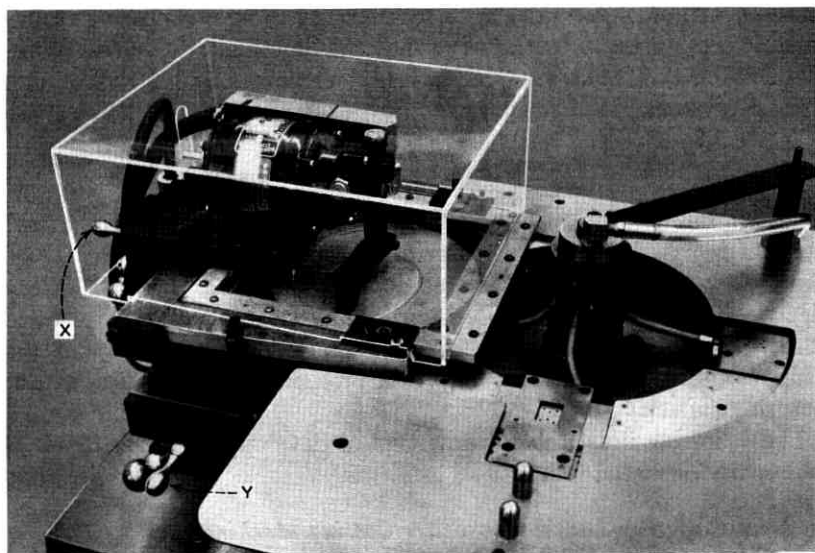


Fig. 5—Precision X and Y dimensional controls for Mica Silvering Machine are shown at left.

the mica lamination is achieved by the micrometer adjustments for X and Y dimensions on the silver coating machine shown in Fig. 5. The screens are made from 230 mesh stainless steel coated except for desired opening with polyvinyl chloride. Each screen is carefully inspected to avoid "petticoating" and to minimize possible variation in fringe capacitance. Consequently, by adding precision thickness sorting to the closely controlled registration, nearly every lamination can be stacked in a required mica capacitor.

4.2 Central Conductor Mold for 3-Type Seal

The techniques for molding the central conductor³ which comprises a black oxidized phosphor bronze central conductor encapsulated in high density polyethylene are an important factor in assuring the integrity of the 3-type seal. Figure 6 shows the principal details of the mold. Before injection, the polyethylene and the metal conductor are heated to 260°C. (500°F.). Injection is at 2000 pounds per square inch, a pressure which is maintained up to three minutes. As cooling starts pressure is increased to 5000 p.s.i. and held while the polyethylene cools through its transition temperature of 110°C (228°F.). Pressure is then

reduced to 2000 p.s.i. and held until the molded part is completely cooled.

4.3 *Corona Noise Data Acquisition Test Set*

The repeaters in submarine cable systems are exposed to voltages sufficiently high to cause corona noise. Such corona causes interference with communications, especially data transmission, even though it is of a magnitude well below that which could cause physical damage or be observed visually.

It was possible to attain design objectives at reasonable facility cost by recording noise pulses above a threshold of a 31 microvolt peak signal during continuous high voltage dc exposure of at least 21 hours on all seals, power separation filters, repeater units and repeaters. Figure 7 shows a Laboratory Technician loading a seal for corona noise testing under water.

Product developments to minimize corona noise included prestressing of 3-type seals at approximately 7500 p.s.i. pressure and back-filling of repeaters with 50 p.s.i. of dry nitrogen.

4.4 *Celcon Plate Engraving*

One equalizer is ordinarily required for each ocean block of twenty repeaters. Therefore, the quantity of plates used in the cordword con-

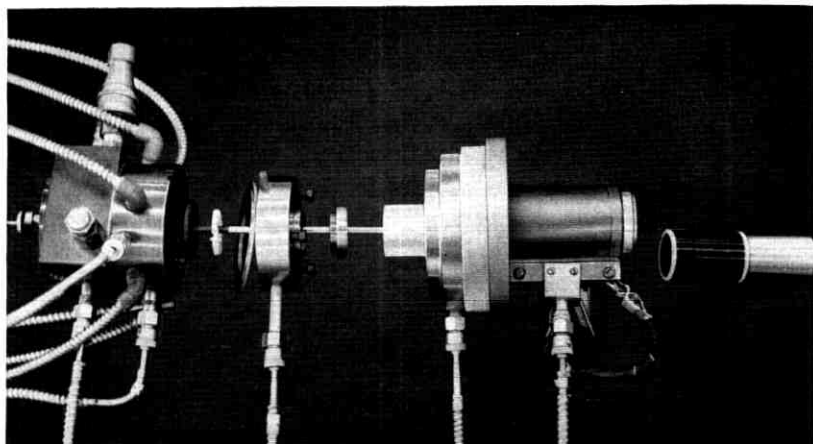


Fig. 6—3-Type Central Conductor Molding Tool. From left to right—Base with heating, cooling and thermocouple leads; central conductor and ceramic disc; vent ring and top mold plate; feeder plate; polyethylene slug extended from transfer tube with band heater; plunger with teflon "O" rings.

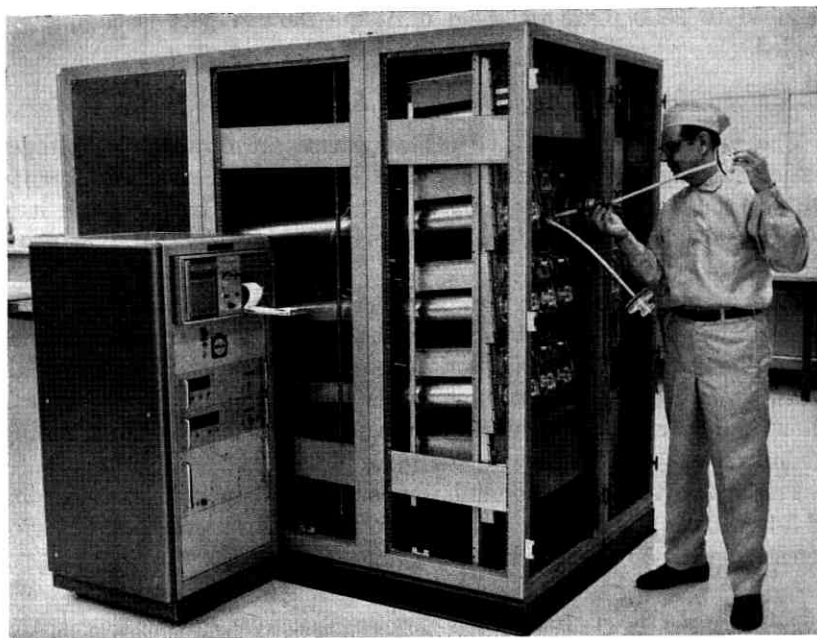


Fig. 7—3-Type Seals being inserted into water-filled cylindrical tubes for corona noise test.

struction³ of equalizer networks is limited. Unless production exceeds 300-400 of a given plate, it is less expensive to machine the plate from Celcon* material rather than to build a mold, and mold the plates from the thermosetting diallyl phthalate used in repeater plates. Templates were made for the engraving machines together with insert templates for all the various holes and cavities for apparatus components. Consequently a very wide selection of plates can be engraved with a relatively small number of templates. Figure 8 shows an engraver with template and Celcon plate in position for machining and Fig. 9 shows Celcon plates before and after detailed machining.

4.5 *Engineering Developments on Product*

Engineering development work on the product, done in collaboration with the Bell Laboratories, is thorough, rigorous and uncompromising in the pursuit of high reliability. Sometimes this work leads to use of very different processes to achieve the manufactured item. To illustrate:

* Trademark of Celanese Corp. for their acetal copolymer plastic material.

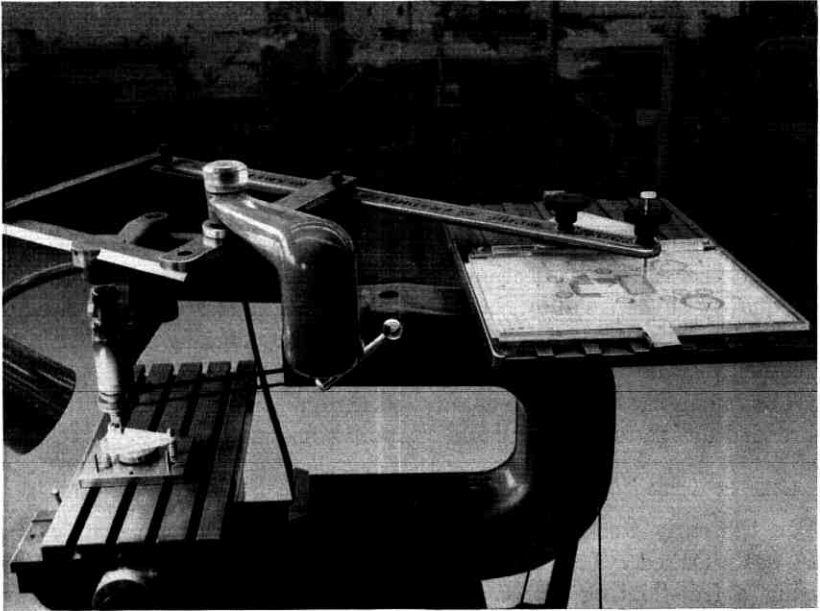


Fig. 8—Engraver with template at right and celcon part at left.

4.5.1 *Adoption of Butt Brazing*

The central conductors of pigtails were joined to seals by a crimped ferrule. Air entrapment in the ferrule (a source of corona noise) made molding difficult and, as a result, product rejections were high. Butt brazing was proven-in but only after special procedures were devised to protect the plastic from metal splashes, filings, and so on.

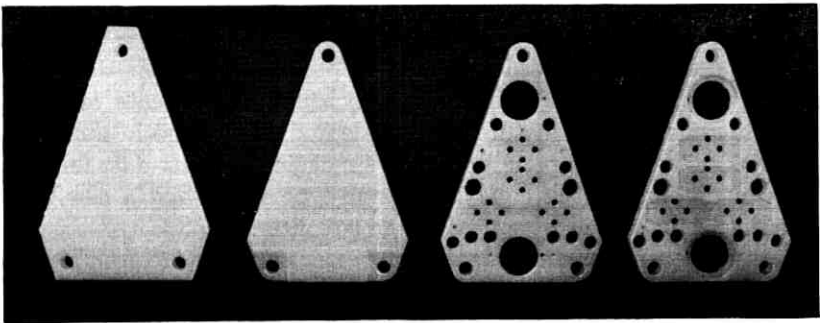


Fig. 9—Top and bottom celcon plates before and after engraving.

4.5.2 Bolt Elimination

Elimination of bolts as a design objective required the 3-type seal to be pressed into the covers, using a tin alloy gasket along with an O-ring. The alloy gasket is 95.5 percent tin, 3.5 percent silver and 1 percent cadmium. As the seal is pressed into position, the gasket flows into an annular cavity in the housing. This produces a hermetic fit which locks the seal into the cover. Also, tubulations have been changed to permit a press fit.

Improved X-ray inspection techniques were developed to control the quality of the tin alloy ring used to assure a tight pressed fit. A more recent development eliminates the press fitting by making the seal casing an integral part of the cover.

The remaining bolts were eliminated through use of a "major thread" threaded design for repeater and support housings.

4.5.3 Seal Modification

In the early weeks of the Florida-Virgin Islands SF System, a few 3A seals leaked³. The improved 3C seal had already been designed, using a free-floating piston with 2000 pounds preload. Introduction of the new seal was expedited to equip a few repeaters of the Virgin Islands system and all TAT-5 repeaters.

V. NEW MATERIALS, APPARATUS COMPONENTS AND ELECTRICAL TESTS

Increasing the repeater band width from one to six megahertz in order to accommodate 800 instead of 136 simultaneous two-way conversations necessitated application of materials and electrical components new to undersea units. Significant component changes included transistors and diodes⁴ for electron tubes, tantalum capacitors for paper capacitors and tantalum resistors for wire wound resistors. A new physically large low inductance capacitor was introduced (See Fig. 10). This capacitor is produced by foil wrapped externally on the epoxy coating of the repeater unit. Several types of components requiring long intervals to manufacture were made and stock-piled in numerous "post office" values to permit wide design choice at a late date for each tailor-made ocean block equalizer. Latch-type relays and a new, purchased selector switch were used in the equalizer.

Many of the major components in repeater manufacture are obtained from Western Electric locations: transistors and diodes from Reading; gas tubes and resistor substrates from Allentown; crystal units from Merrimack Valley; ferrites from Hawthorne, Illinois; relays from

Burlington, pigtail cable and braiding from Baltimore, Maryland; and thermistors from Kansas City, Missouri.

In view of the 6 MHz frequency, most components were necessarily of lower physical value than those used in SD Systems. Table I shows the value and accuracy ranges for apparatus used in SF Repeaters. Furthermore, precision test sets had to be developed with capabilities of measurements at frequencies as high as 500 MHz.

5.1 *Vendor Relationship*

Although many of the purchased materials and parts were similar to those used on the SD Repeater, in most cases new or modified tooling and gaging were required to assure conformance to specification. Gaps in production necessitated reestablishment and requalification of vendors, a tedious and costly procedure. Approved sources for new material were developed. Single lot material purchases were frequently resorted to in order to minimize variation; for example, a single large batch of polyethylene was procured to aid in uniformity of quality during seal

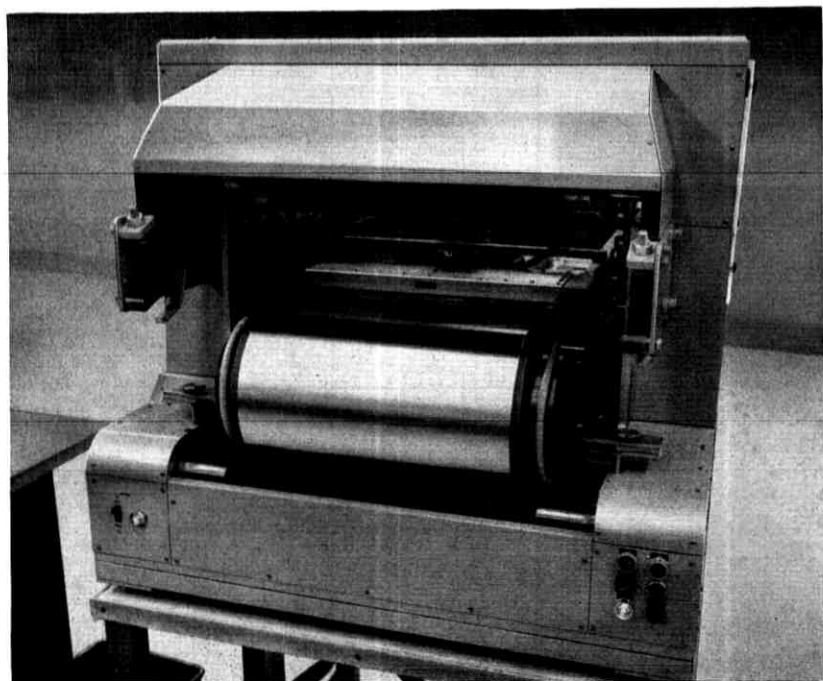


Fig. 10—Copper foil wrapped on epoxy coated repeater unit by machine.

TABLE I—APPARATUS USED IN SF REPEATERS AND EQUALIZERS

Part	No. of Different Codes	No. of Units per Repeater	No. of Units per Equalizer	Value Range	Accuracy Range
Capacitors					
High-voltage paper	2	2	3	0.02-0.18 μ F	2-5%
Low-voltage paper	2	2	—	0.100-0.2200 μ F	5%
Tantalum	4	16	—	1.0-10.0 μ F	7%
Mica	136	46	131	9.96-9500 pF	0.25-5%
Inductors					
Air Core Solenoid					
Type-adjustable	60	21	55	0.227-139.0 μ H	0.8-2.5%
Air Core solenoid type	87	10	80	0.087-1900 μ H	0.5-15%
Transformers ferrite core	4	4	1	—	—
Resistors					
Tantalum	159	46	188	5.00-7000 ohms	0.5-5%
Bifilar wire wound	4	0	4	1.870-4.580 ohms	1.0-5%
Diodes	6	5	1	—	—
Transistors	4	4	—	—	—
Gas tubes	1	2	2	—	—
Crystal	2	1	—	—	—
Thermistor	1	1	—	—	—

and pigtail manufacture, and a single lot of Celcon for equalizer plates was also procured. Engineering development work with material and parts suppliers is continuous. Comprehensive tests indicate whether each lot meets specifications. Variations are studied for pre-trouble drift or means of material improvement.

VI. CONTROL OF PRODUCT RELIABILITY

To remain competitive, broader band, more sophisticated systems must be made available with a high degree of reliability and at reasonable cost. To realize SF, transistors and resistors are essential. These items lack long-term performance history. There is also added risk due to many more repeaters being required for a higher frequency system (10 miles separation on SF System compared with 20 miles on SD Systems). Therefore, emphasis is added on the care with which the design is manufactured and tested.

The engineers translated the design into tools, fixtures, gages, test sets, ordering descriptions and detailed manufacturing layouts. All manufacturing facilities and procedures were subject to intensive check for conformance to requirement. Two pilot repeaters were made to prove that the design was manufacturable and that all tools performed as intended.

Laboratory Technicians are trained with non-conforming material. Product examiners patrol assembly operations and sample or detail product depending on the nature of the operation.

6.1 *Control of Deviations*

Whenever a deviation from the expected is detected, the deviating material must be set aside for detailed analysis. Product material and components are serialized which permits segregation and examination of the records to determine precisely what materials, processes and personnel were involved in the unit.

Stopping the job until the item is resolved generally is preferable to mating questionable material with other product material which might have to be discarded at a later, more costly stage. The questionable items number 100-150 per month and cause appreciable disruption in production flow. To assure that questionable items are dealt with properly and not eased through, work is controlled on a time rather than piece work basis. Many of the deviations are disposed of within a day or two; few take as long as a week to decide. Two or three Bell Laboratories designers are resident at the Clark Shop to review and evaluate observed variations.

Not all deviations involve possible degradation of reliability. Frequently, an exceptionally good batch of material is procured. Does the better material produce significantly better product? Could the supplier produce consistently to an upgraded specification?

6.2 Control of Data

Electrical test data on components, networks, amplifiers, repeater units and repeaters must be collected, recorded and analyzed particularly for trends. Not only do these trends forecast out-of-limit conditions in time to make necessary adjustments but also they aid in evaluating the extent of equalization required for the repeaters during laying operations.

Most of the electrical data are transmitted directly to a Datex Receiver which produces cards for subsequent processing by an IBM 1620 Computer on a daily basis^{2,5}. In spite of excellent performance, certain improvements were desired (*i*) faster feedback, (*ii*) shared time, (*iii*) less handling and sorting of cards prior to printout.

(i) This episode showed the need for faster feedback:

- (a) A *repeater unit* supervisory level tested out of limits.
- (b) Next, analysis of earlier *supervisory oscillator* data showed beyond- 3σ level deviation (but in limits).
- (c) Errant oscillator was replaced.
- (d) A defective thermistor problem was revealed and corrected.

We could have avoided rework *c* and sped remedy *d* by faster data analysis and feedback.

(*ii*) Frequently special data and analyses have been desired by Bell Laboratories and Clark Shop engineers to establish acceptability of components deviating slightly in one parameter. Without shared-time capacity, the delay in performing such analyses diminished the value of the results and slowed the introduction of constructive action. Two examples are shown. First is Fig. 11, the Leakage Current for KS-19458, List 2, Tantalum Capacitors, and second, Fig. 12, the Insertion Gain of a series of repeater units. Both printouts show nominal value, maximum and minimum limits, the actual average and 3 sigma deviations.

(*iii*) Extensive handling of punched cards has been the practice. Data plotting of apparatus and network characteristics required transfer of primary data to compressed cards which in turn were processed to produce plotting cards which were then translated into a chart.

To provide the desired flexibility with adequate capacity, an IBM 1800 Computer has been procured and installed. The new system has a disc storage capacity capable of providing information on conformity

of transmitted data with apparatus requirements in a matter of seconds. The "On Line Printer" will print a record of each day's transmission for shop use without the extensive sorting and handling which cards previously required. Furthermore, the computer has a program which indicates priorities to operations to be performed upon receipt of data.

As valuable as the computer is, it does not perform the whole operation. It is so important that accurate data be transmitted and certain visual requirements be properly recorded, that product examiners do

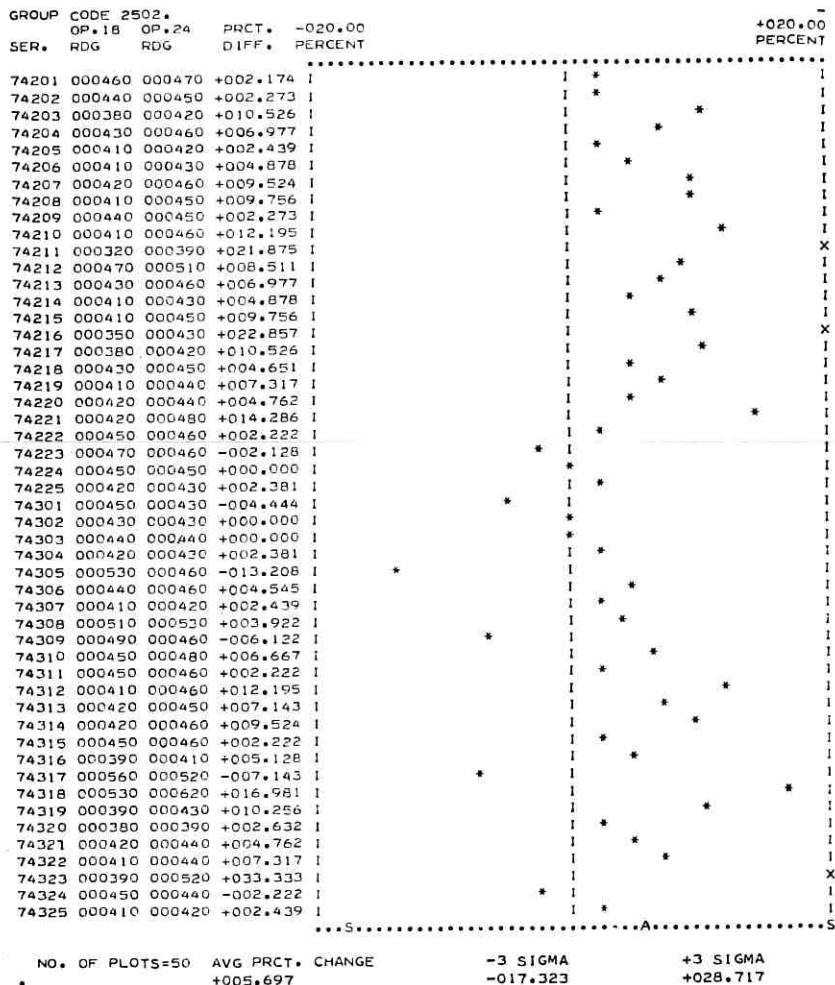


Fig. 11—KS 19458, List 2, Tantalum Capacitor, Leakage Current.

SER	READING	FREQUENCY KC - 2850.		
		MIN LIMIT 27.222	NOM LIMIT 27.328	MAX LIMIT 27.434
34147.	27.332.		I*	
14541.	27.290.		I	
14542.	27.330.		I*	
14543.	27.299.		I	
14544.	27.334.		I*	
14545.	27.320.		I*	
14546.	27.333.		I*	
14547.	27.321.		I*	
14548.	27.359.		I	
14552.	27.346.		I*	
14646.	27.337.		I*	
24162.	27.279.		I	
24164.	27.334.		I*	
24166.	27.349.		I	
24242.	27.337.		I*	
24243.	27.330.		I*	
24244.	27.312.		I*	
24246.	27.352.		I	
24253.	27.251.		I	
24254.	27.338.		I*	
24257.	27.334.		I*	
24259.	27.348.		I*	
24266.	27.327.		I*	
24268.	27.315.		I*	
24343.	27.325.		I*	
24344.	27.382.		I	
24345.	27.335.		I*	
24348.	27.382.		I	
24351.	27.335.		I*	
24352.	27.359.		I	
24354.	27.340.		I*	
24355.	27.326.		I*	
24362.	27.382.		I	
24363.	27.363.		I	
24364.	27.322.		I*	
24366.	27.299.		I	
24441.	27.366.		I	
24442.	27.318.		I*	
24443.	27.357.		I	
24444.	27.292.		I	
24445.	27.293.		I	
24446.	27.306.		I	
24447.	27.278.		I	
24448.	27.352.		I	
24451.	27.350.		I	
24453.	27.317.		I*	
24454.	27.329.		I*	
SER	READING		I	
		1	AVG = 27.3262	1

3 SIGMA 0.0712

I

I

Fig. 12—304 AL Repeater Unit, Insertion Gain.

this process checking work, and in turn are subject to check by quality auditors. The auditors are responsible for all data being properly recorded and verified.

6.3 Additional and Final Controls

The Quality Assurance Organization independent from the Engineer of Manufacture prepares and carries out quality surveys in collaboration with their counterparts at Bell Laboratories. The surveys cover

in detail specific areas such as raw materials, paper capacitors, ground separation filters, and so on.

It might seem that controls to assure product reliability were extensive and expensive. This is true; however, not only must the "freaks" and "sports" be eliminated but also the minor mistakes made by human beings. Each step in the process of manufacture presupposes that all requirements have been met prior to the current step. If this is not so, the deviation may not be caught until a much more costly assembly has been made.

All final data is compiled by Western Electric in a mechanical and an electrical data book for each repeater and presented to the Bell Telephone Laboratories' representative who has been designated as the agent to accept each unit for the Long Lines Department of the American Telephone and Telegraph Company.

REFERENCES

1. Lamb, H. A., and Heffner, W. W., "Repeater Production for the North Atlantic Link," *B.S.T.J.*, 36, No. 1 (January 1957), pp. 103-138.
2. Johansson, S. G., "Manufacture of Rigid Repeaters and Ocean-Block Equalizers," *B.S.T.J.* 43, No. 4 (July 1964), pp. 1275-1310.
3. Buus, R. G., Kassig, J. J., and Yeisley, P. A., "Repeater and Equalizer Design," *B.S.T.J.*, this issue, pp. 631-661.
4. Wahl, A. J., McMahon, W., Lesh, N. G., and Thompson, W. J., "Transistors, Diodes and Components," *B.S.T.J.*, this issue, pp. 683-698.
5. Tobey, S. B., "Semi-Automatic Data System," *W. E. Engineer*, VII, No. 1 (January 1963), pp. 12-16.

Transistors, Diodes and Components

By A. J. WAHL, W. McMAHON, N. G. LESH and W. J. THOMPSON

(Manuscript received January 2, 1969)

In this paper we describe the active and passive components used in the underwater portion of the SF Cable System. We present reasons for choosing certain types and, where appropriate, outline methods of manufacture, screening, aging and selection. Measures taken to safeguard the components before, during, and after assembly in repeaters are discussed briefly.

I. INTRODUCTION

The choice of passive components and active devices for use in a submarine cable system must often try to satisfy two conflicting demands. On the one hand, high system performance objectives may call for the use of the most advanced device technology and components. Reliability requirements, however, demand the use of those components and devices which have a proven record of dependable operation in conventional applications, and whose failure modes are understood. Where such conflicts arise, the choice is of necessity weighted in favor of reliability considerations.

In this section on components, both active and passive, for the SF System, we cover: (i) the basic types and their functions, with some of the reasons for their choice; (ii) procedures for providing them; (iii) screening, aging, and selection; (iv) results from experience in providing them; and (v) precautions for assembly into circuits.

II. BASIC TYPES AND FUNCTIONS

2.1 *Semiconductor Devices*

The evolution of transistor technology was especially timely for the development of a submarine cable system of substantially greater bandwidth than that of the SD System which had very successfully used vacuum tubes. At the time when final development of devices for the

SF System began, the frequency capability of germanium (but not of silicon) transistors had advanced to the stage where they could, with reasonable design configuration, provide the required bandwidth. Redesigned vacuum tubes, however, would entail mechanical configurations which would be unacceptable for submarine cable use. Furthermore, with the decreased repeater spacing necessitated by the higher operating frequencies, the supply of power to vacuum tubes would involve prohibitively high voltages at the ends of a long cable system; whereas, transistors could be powered with manageable voltages under similar circumstances. Considerations such as these made the use of transistors almost mandatory for the SF System. In an accompanying role, the semiconductor diodes could easily perform the limiter and secondary surge protective functions for the transistor circuits.

In a system which can accept the mechanical fragility of vacuum or gas tubes, the mechanical requirements in terms of shock and vibration presented no problems for semiconductor devices.

The failure rate objectives had been estimated to be 0.0001 and 0.0005 percent per thousand hours of operation for diodes and transistors respectively. Past experience with germanium diffused base transistors of similar design had indicated that the reliability objectives could probably be met. This experience, however, had been in the realm of far less stringent operational requirements and conditions, particularly in regard to power dissipation. Past experience with the reliability of high quality silicon diodes had also given favorable indications for submarine cable use. Feasibility for manufacture had been demonstrated for all necessary processes with the exception that the processing for the protective oxide and the Ni-Pt-Rh-Au contacting surfaces for the diode elements was new at the time the choice had to be made. To compensate for the missing experience, the diode elements were handled, mounted, and encapsulated as if they had no protective oxide over the junctions.

In order to simplify the manufacture, all semiconductor devices were encapsulated in containers of one design and configuration.

2.1.1 *Transistors*

Two basic codes of transistors are used. Both are diffused base germanium with alloyed emitters, differing only in size of the active element to accommodate appropriate power handling capability. One basic code (the L2287) is split into three actual codes for operation in three places in the SF Repeater: (i) first stage amplifier as 38A, (ii) second stage amplifier as 38B, (iii) supervisory oscillator as 38D. Table I shows the primary electrical characteristics.

TABLE I—SOME MAJOR ELECTRICAL CHARACTERISTICS
OF LOWER POWER TRANSISTOR (L-2287)

Parameter	Bias Condition	Typical Value
Collector-base breakdown voltage	100 μ A	40 V
Collector-base leakage current	15 V	0.4 μ A
Low frequency common base current gain	5 V, 15 mA	0.985
Frequency of unity common emitter current gain	5 V, 15 mA	950 MHz
Base resistance		60 ohms
Collector-base capacitance	5 V	3.5 pF
Noise figure at 5 MHz	5 V, 5 mA	3.0 dB

The other basic transistor code (the L-2288) is used only in the amplifier output stage of the SF Repeater as code 38C. Table II shows the primary electrical characteristics.

2.1.2 Diodes

The four basic diode codes are all of diffused silicon with junctions protected by a hard oxide grown in high pressure steam and with mounting surfaces of Ni-Pt-Rh-Au applied in that order outward from the silicon. The basic code L2317 is an n-p junction diode (zener) used as a surge protection device in the power path circuit of the SF Repeaters, and is designated as a 467A. Its primary electrical characteristics are shown in Table III.

The basic code L2318 actually consists of two pn junction diode elements in parallel and oppositely poled so that only forward characteristics are presented at the terminals. It is used in two places in the SF Repeater. As code 468A, it serves as a surge protection device for the amplifier input; as code 468B it serves as a limiter in the supervisory oscillator circuit. Table IV shows the primary electrical characteristics.

TABLE II—SOME MAJOR ELECTRICAL CHARACTERISTICS
OF POWER OUTPUT TRANSISTOR (L-2288)

Parameter	Bias Condition	Typical Value
Collector-base breakdown voltage	100 μ A	50 V
Collector-base leakage current	15 V	2.5 μ A
Low frequency common base current gain	11 V, 90 mA	0.982
Frequency of unity common emitter current gain	11 V, 90 mA	750 MHz
Base resistance		40 ohms
Collector-base capacitance	11 V	9 pF

TABLE III—SOME MAJOR ELECTRICAL CHARACTERISTICS OF THE POWER PATH PROTECTION DIODE (L-2317)

Parameter	Bias Condition	Typical Value
Reverse breakdown voltage	120 mA	16 V
Reverse leakage current	13 V	30 nA
Forward voltage	500 mA	0.808 V

The basic code L2319 also consists of two elements in parallel and oppositely poled, but each element is a four-layer pnpn diode. The characteristics seen at the terminals are, therefore, the typical pnpn forward characteristics in either direction of polarity. This device, as code 469A, serves as a surge protector at the output of the amplifier. The pnpn forward characteristic was required in order to have a very high impedance across the amplifier output over the maximum signal voltage swing while still having the capability of breaking into a low impedance shunt path if this voltage swing is significantly exceeded. The primary electrical characteristics are shown in Table V.

The basic code L2320 is a pn junction (zener) diode used as code 467B to bias the third stage of the amplifier in the repeater. Primary electrical characteristics are shown in Table VI.

2.1.3 Gas Tubes

Primary signal path surge protection is accomplished by the WE 458A-gas tubes type used in the SD Submarine Cable System.¹

2.2 Resistors

Prior submarine cable systems employed wire wound resistors.² Because the top frequency of the SF System is 5884 kHz³ compared to 1052 kHz for the SD System, use of wire wound resistors would have required very special winding techniques to meet the out-of-band as well as in-band impedance characteristics. It, therefore, seemed prudent to

TABLE IV—SOME MAJOR ELECTRICAL CHARACTERISTICS OF INPUT PROTECTION AND OSCILLATOR DIODE (L-2318)

Parameter	Bias Condition	Typical Value
Forward voltage	1 mA	0.518 V
Forward voltage	200 mA	0.794 V
Capacitance	0 V	40 pF

TABLE V—SOME MAJOR ELECTRICAL CHARACTERISTICS OF OUTPUT PROTECTION DIODE (L-2319)

Parameter	Bias Condition	Typical Value
Forward breakover voltage		22 V
Forward leakage current	12 V	22 nA
Capacitance	0 V	45 pF

select film type resistors with their inherently lower parasitic effects.⁴ As the number of channels which can be provided depends, not only on the gain-bandwidth product, but also on the stability of this product, the contribution of the resistance elements to the gain is critical and stability must be at least as good as the best wire wound types. Extensive laboratory tests led to the conclusion that tantalum nitride resistors could provide the desired characteristics.⁵

To achieve the reliability objectives for the cable system,³ the average failure rate for the passive components must be of the order of 0.00001 percent per thousand hours. It is not practical to confirm this level by life tests. Therefore, exhaustive screening tests to weed out early failures and derating to insure long life are required to meet this goal.

While tantalum thin film resistors had not been widely used heretofore,⁴ they have been under extensive laboratory study since 1957. A moderately large-scale study program comprising approximately 2500 resistors grouped in seventeen combinations of variables was initiated in 1964.^{5,6,7} The resistors were subjected to a series of tests that included oven heating, temperature cycling, and operation at various power overload conditions. Measurements of resistance and current noise were made at appropriate intervals during the test after which samples of each group were subjected to a long-term intermittent rated load test, long-term continuous load tests, and humidity tests with and without intermittent loads. Other samples were subjected to long-term overload tests.

TABLE VI—SOME MAJOR ELECTRICAL CHARACTERISTICS OF BIAS REGULATOR DIODE (L-2320)

Parameter	Bias Condition	Typical Value
Reverse breakdown voltage	2 mA	11 V
Reverse leakage current	8 V	10 nA
Forward voltage	200 mA	0.850 V

The superiority of sapphire over glass or glazed high alumina as a substrate material became evident early in the program. Other variables had less effect on resistor performance, but the tests did indicate that films having a sheet resistance between five and ten ohms per square were somewhat less stable than films having a higher sheet resistance. The final design consisted of films with resistivities between 5 and 50 ohms per square deposited on sapphire substrates and with line widths not less than 0.008 inch as shown on Fig. 1.

Resistors above 5 ohms in the SF cable system are of one type: tantalum thin film. Four wire wound codes (246 type) complete the requirements except for one power wire wound resistor in the Ocean Block Equalizer. Selection of a single resistor style for essentially all repeater and equalizer functions greatly simplifies the reliability prediction problem and aging considerations. Sapphire provides the support or substrate on which the tantalum nitride film is deposited by reactive sputtering. After evaporating nichrome-gold terminations, the tantalum film is patterned using photolithographic techniques, anodized** to 95 percent of the final desired value and heat stabilized. After attachment of clip leads, the resistors are trim anodized to the final value and placed in a protective Kel-F sleeve and secured with O-rings. These resistors serve all resistance functions including biasing of the transistors and diodes, current limiting, use in feedback networks, wave shaping networks and equalizers, impedance and load matching and level adjustment.

2.3 Capacitors

Repeaters in the SF Cable System contain three general types of capacitors: castor oil impregnated paper, silvered mica and solid tantalum. The first two were used in previous transoceanic cable systems with very satisfactory results, but solid tantalum capacitors were new to this application. They were introduced because of the need in the SF design for comparatively large value coupling capacitors having small physical volume.³ Experience with this type of capacitor, both in actual service and in laboratory studies, established that it is fundamentally very reliable but that, because of the complexity of its manufacture, any given population was apt to contain a number of potential early failures. It was imperative, therefore, to devise and apply very exacting screening procedures to eliminate these.

* Anodizing consists of passing a direct current from an external cathode through an electrolyte in contact with the tantalum film connected as the anode, thereby converting a portion of the tantalum to its oxide.

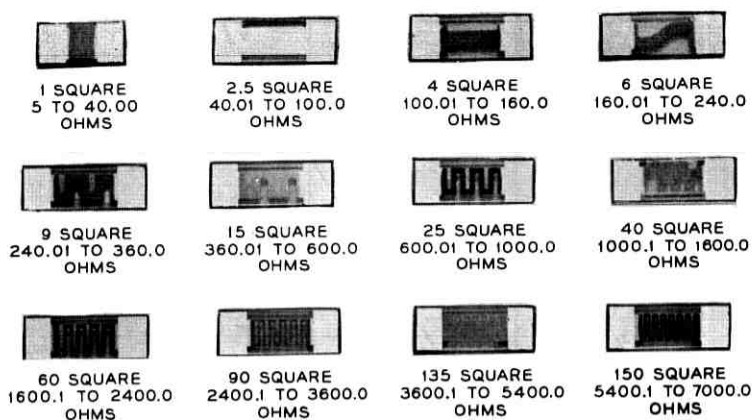


Fig. 1—Resistor patterns.

The reasons for choosing oil impregnated paper capacitors and unencased silvered mica capacitors are given in an earlier article by M. C. Wooley.² Little more need be said here about this except to point out that time has borne out the wisdom of these decisions. During the thousands of component years of operation, no capacitor failures have occurred. The addition of the solid tantalum capacitor necessitated obtaining some measure of its intrinsic durability. Accordingly, a large number of capacitors were tested under accelerating conditions of elevated temperatures and voltages. These tests have been in progress for about six years and have provided much information on the dependence of failure rate on voltage and the failure rate at sea-bottom temperature at elevated voltage. Life tests at 85°C indicate a failure rate dependency on voltage to the eighth or ninth power as shown in Fig. 2. The failure rate at 5°C (sea bottom temperature) and 62.5 volts is 3×10^{-6} failures per hour. Extrapolating to 17 volts which is somewhat above the use voltage with the aid of the expression:

$$\frac{FR_u}{FR_t} = \left(\frac{V_u}{V_t}\right)^8$$

the computed failure rate at 5°C — 17 volts is 0.00001 percent per thousand hours which meets the general reliability objective for components in the system.

2.4 Inductors and Transformers

The broader frequency range and the large number of repeaters in the SF System requires that the inductors and transformers that affect

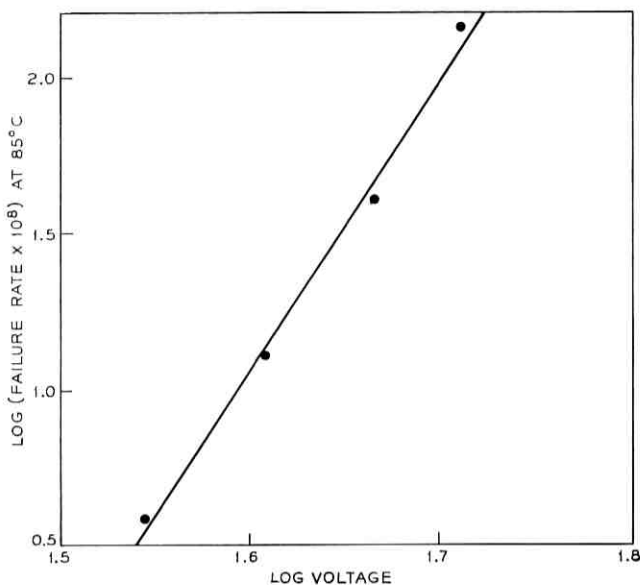


Fig. 2—Dependence of solid tantalum capacitor failure rate on voltage.

the frequency characteristic be more accurate and stable than those of earlier systems. Most of the frequency shaping inductors have solenoidal windings on forms of molded glass bonded mica. The winding surface of this dimensionally stable material is manufactured to a fraction of a mil tolerance so that, with precise control of the winding pitch, the inductor can be manufactured to high accuracy. This makes unnecessary even slight shifting of turns that could affect the stability. By the additional control of the winding tension, the long-time stability is further assured and by closer than normal control of wire and insulation, the effects of dissipation and parasitic capacitance are made very reproducible.

The winding form has a temperature coefficient of expansion that permits the use of molded-in metal terminals. This makes it possible to solder the fine wire of the winding directly to the terminal to avoid subsequent flexing. As an added feature, the form provides thermal isolation between the winding connection and the circuit wiring connection so that soldering the second cannot disturb the first.

The inductors in the directional filters have a small range of adjustment to tune precisely the sharply resonant circuits. This adjustment is accomplished by positioning a small slug of low loss magnetic material

more or less in the field of the winding. The threaded carrier for this slug has an interference fit with a parallel thin wall tube of low-friction plastic so that the adjustment is smooth and sufficiently stable that there is no perceptible change when the carrier is later cemented in place.

In the ocean block equalizer, a somewhat lower accuracy of inductance can be tolerated. Advantage is taken of this by increasing the range of adjustment to minimize the number of inductor designs that must be manufactured to cover a wide range of inductance values.

The broad band transformers are used primarily to reduce unwanted feedback around the amplifier. In one case, there are parallel feedback paths through the directional filters and around the amplifier, so that, by using a transformer in one path that differs by 180° of phase shift³ from a transformer in the other path, the signals will cancel. This imposes relatively severe requirements for similarity of the two transformers particularly over the normal frequency range of the repeater. Again, this is attained by controls on winding and on the ferrite core properties.

A second unwanted feedback path exists around the repeater because of practical problems associated with separating the signal from the dc power. The result is a longitudinal mode of transmission both below and well above the normal signal band that would cause instabilities if uncorrected. A ferrite core transformer wound with turns of a coaxial conductor⁴ effectively attenuates the longitudinal mode without loss to the signal in the transverse mode.

III. PROVISION OF DEVICES AND COMPONENTS

3.1 *Manufacture*

In order to assure that the best possible quality was built into the devices and components, the manufacturing operation was conducted under continuous and detailed scrutiny in an environment which minimized contamination.⁴ Incoming material was certified before entering the production stream, and all operations were pedigreed by a system of records whereby the history of any completed individual device could be traced back to its beginning. Gaining the maximum benefit of this scrutiny required the continuous active participation by development engineers of Bell Laboratories and manufacturing engineers of Western Electric.

Due to the extensive technology and the complexity of equipment required to produce tantalum thin films, the basic tantalum resistor was produced at the Western Electric plant in Allentown, Pennsylvania,

and shipped to the Kearny Works Clark Shop in New Jersey for finishing, screening, aging and final selection. To maintain identity the resistance element fabricated at Allentown was coded as the 247 type resistor.⁴ It consists of the tantalum nitride element and terminations deposited on the sapphire substrate, patterned into the desired configuration, preanodized and heat treated to be between 95 and 100 percent of the final desired resistance value (see Fig. 3).

Upon receipt of 247 type resistors at Clark, terminals are attached, as shown in Fig. 4. The resistors are then trimmed to final values and given rigid visual inspection. Finally, the resistors journey through the screening, aging, and selection processes.⁴ The completed resistors are designated 243-type resistors.

IV. SCREENING, AGING AND SELECTION

4.1 *Semiconductor Devices*

4.1.1 *Screening*

The screening operation first involved electrical testing to assure that operational specifications were met. It also involved subjecting the devices and components to a series of tests with the purpose of finding and identifying those individuals which did not conform to the behavior of the general population. Such tests include leak tests, mechanical

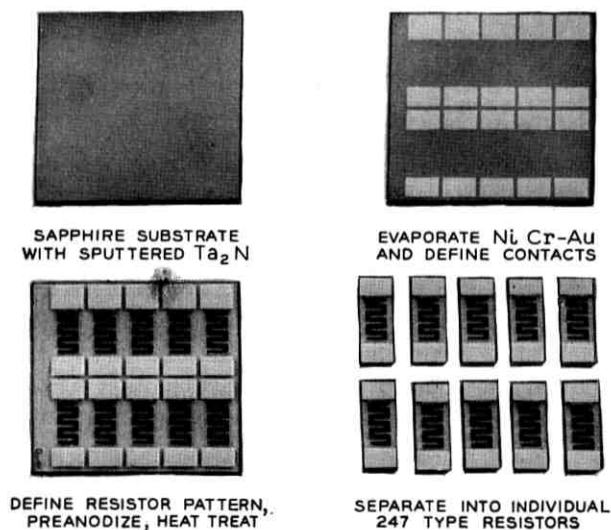
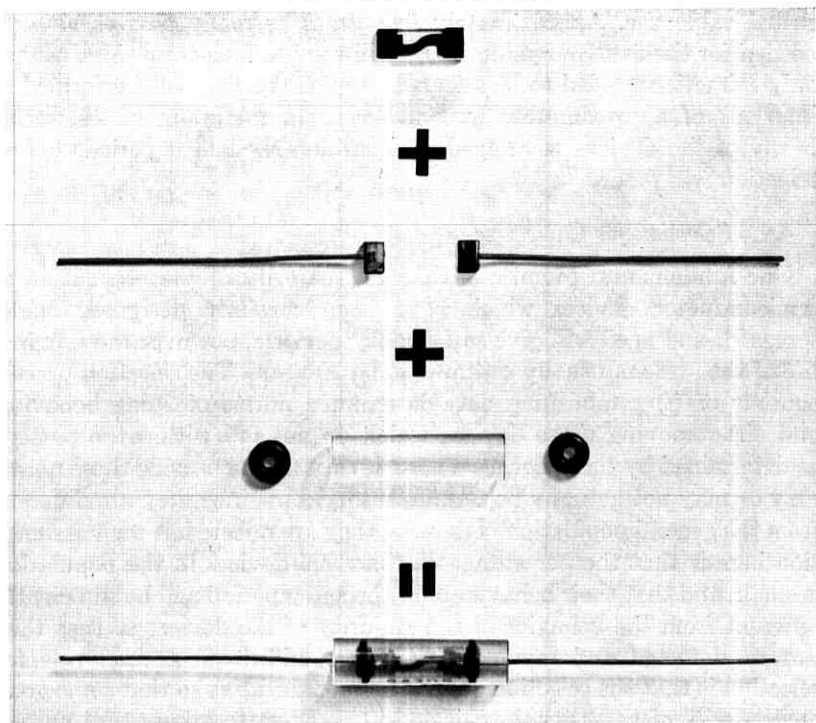


Fig. 3—247 type resistor fabrication.

247 TYPE RESISTOR



243 TYPE RESISTOR

Fig. 4—243 type resistor assembly.

tests, temperature cycling and storage, surge tests, and short-term aging under electrical conditions which exceed normal operating levels by as wide a margin as possible without causing damage to the normal devices. Interspersed among the screening tests were the appropriate number of electrical tests and visual inspections with records maintained on individual devices.

In order to be effective in locating the nonconforming devices, the screening operations must be done on all individual devices instead of just on a sample.

4.1.2 Aging

All candidate semiconductor devices were subjected to long-term aging under simulated worst-case use conditions of power and temperature. This operation provides information on how the devices might be

expected to behave under the conditions of power and temperature of actual cable use. Approximately five times as many devices as were needed for the first SF System were subjected to long-term aging so that only a fraction needed to be selected. In-place testing was performed at intervals of approximately four weeks for a minimum of 24 weeks. Many of the devices were aged for considerably longer periods before selection was made.

4.1.3 *Selection for Cable Use*

One fundamental premise is that the majority of the population of semiconductor devices which have been carefully designed, made, screened, and aged will give satisfactory performance over the nominal life of the system, usually assumed to be 20 years. The selection process consists of (i) establishing data-determined norms of aging behavior, and (ii) removing those devices which do not fall within the pattern as determined by these norms. Those devices falling outside these norms may or may not actually be unsatisfactory, but since they are different from the general population of devices, they are rejected. A tacit assumption here is that the percentage of abnormal devices in the population is small, and that their behavior over the aging period will be sufficiently different from the behavior of the majority of the devices so that they can be detected and removed. A further additional ground rule for selection is that the selection criteria are conservative so that some good devices will, of necessity, be rejected in an effort to ensure that no bad devices are accepted. A basic challenge then is to develop the statistical methods and criteria which will minimize the rejection of good devices.

The data upon which selection is based contains variations due both to the devices and to the data collection system. Systematic data analysis is necessary to assess and separate these sources of variation, and then the development of special statistical techniques is required to adjust for or reduce these effects. Since these techniques could not be developed prior to shipment of devices for the first SF System, extremely conservative selection procedures were used in an effort to ensure the exclusion of abnormal devices. More sophisticated selection procedures are now being developed.

4.2 *Resistors*

4.2.1 *Screening*

In the reliability assessment of any component, it is the control of the manufacturing process and uniformity of these processes that pro-

vides the greatest contribution to ultimate reliability. Screening and aging tests provide a measure of the effectiveness of process control and additionally locate nonconforming devices. To be effective, screening must be done on 100 percent of the product.

Of the 247 type resistors shipped to Clark from Allentown and made into 243 type resistors, approximately 75 percent passed the screening tests and were placed on long-term aging tests.

4.2.2 Aging

All screened 243 type resistors were subjected to a power of 0.25 watts at 25°C for 4000 hours. These conditions are not considered harmful to the resistors even though they represent accelerated aging over the use conditions of approximately 0.0625 watts maximum at an average temperature of 4°C. An acceleration factor of 18.2 was calculated by taking the ratio of the change at aging conditions versus the change at use conditions.

In addition to the aging performed routinely on all resistors, representative samples have been aged at a power of $\frac{1}{2}$ watt and 25°C for periods up to 20,000 hours with an average change of approximately 0.008 percent.⁶ This is in good agreement with changes calculated by using the formula for time and temperature dependence.⁸

A variety of initial resistance tolerances are used ranging from 0.04 percent to 2.5 percent.³ In all but one case, the initial tolerance was set at one-half the end-of-life tolerance (EOL), that is, the tolerance at which circuit performance would be degraded to an unacceptable level. In this case the EOL tolerance was 0.1 percent permitting a change of 0.06 percent from an initial tolerance of 0.04 percent. To obtain an estimate of the average mean life, the mean life for each EOL tolerance was calculated and weighted by its usage in a repeater. Applying a 90 percent confidence interval, the lower limit of the estimate yields a failure rate of 0.0000024 percent per thousand hours.

Of special note was the additional screening provided on certain critical codes. These are resistance values on which dc voltage is impressed during service. By virtue of the dc application, they are the most susceptible to whisker growth emanating from the soldered termination area. All codes were desleaved at the end of the 4000 hours of aging and examined for whisker growth. The sleeves were reinstalled and the critical codes aged for an additional 250 hours at $\frac{1}{2}$ watt and 25°C. They were then re-examined for possible whisker growth and remeasured to ascertain whether they still met electrical requirements. Terminations having less than complete solder coverage are most sus-

ceptible to whisker growth. For this reason, all critical codes have now been changed to require complete solder coverage of the termination.

4.2.3 Selection

After aging, resistors are selected for use by utilizing both fixed and variable limits. Resistance readings at 1000, 2000 and 4000 hours of aging are compared to the initial readings and resistance changes, average and sigma calculated. Resistors may be rejected at any point if they fail to meet the following fixed limits:

Elapsed Aging	Percent Change	
	≤ 50 ohms	> 50 ohms
First 1000 hours	0.05	0.05
Second 1000 hours	0.05	0.04
Total 4000 hours	0.08	0.06

In addition to these fixed limits, they had to meet variable limits. For 0.1 percent EOL resistors (codes 243 AC, AD & AE) only those falling within \pm two standard deviations were selected. For other tolerances, the variable limit was \pm three standard deviations from the mean.

4.3 Capacitors

While there is good reason for confidence in the adequacy of the solid tantalum capacitor intrinsically, it is necessary to take elaborate measures to screen out potential "early failures." This is evidenced by the fact that at least one early failure occurred in each of the accelerated tests. To eliminate these among capacitors to be used in the system, all Ta solid capacitors are exposed to 35 volts at 85°C for six months. During this period parameters such as capacitance, dissipation factor and leakage current are carefully monitored. A significant deviation of any of these properties from the norm is cause for rejection.

Both paper and mica capacitors are subjected also to lengthy screening and aging. However, the incidence of early failures of paper capacitors in these tests is rare and among micas it is virtually absent. This is undoubtedly attributable to the great care exercised in the manufacture of these units from the selection of raw materials through each process step to the final product.⁴

V. RESULTS FROM EXPERIENCE IN PROVIDING THE COMPONENTS

5.1 Semiconductor Devices

Previous experience had shown that the only semiconductor device characteristic that would exhibit an aging effect was transistor gain,

which would increase at a slowly decreasing rate. The repeater circuits were designed to accommodate an average common base current gain change of 0.002 per transistor over a 20-year period. Long-term aging results on some of the candidate transistor devices which have remained on aging for 3.5 years indicate that actual aging will be less than the 0.002 objective. Figure 5 illustrates this behavior, which is equivalent to a total excursion of only 0.0001 in the common base current gain. Note also that most of the change occurred during the first half of the aging period.

It is always desired to compare these long-term aging results with destructive step-stress results performed during manufacture of these devices. One possible method for making this comparison, although recognized as being inadequate or possibly misleading, is in terms of the exponential or constant failure-rate model. At the time when the first SF Submarine Cable was being laid, the total accumulated device hours of aging (transistors and diodes) was over 140 million without either a catastrophic failure or failure due to aging beyond initial specification limits. Under the assumptions of the constant failure rate model, a range of 0 to 0.0017 percent failures per thousand hours is consistent with a chance of more than 9 out of 10 of actually observing no failures in 140 million total device hours. This upper bound of 0.0017 percent per thousand hours is of the right magnitude and the best that can be obtained from 140 million device hours of aging.

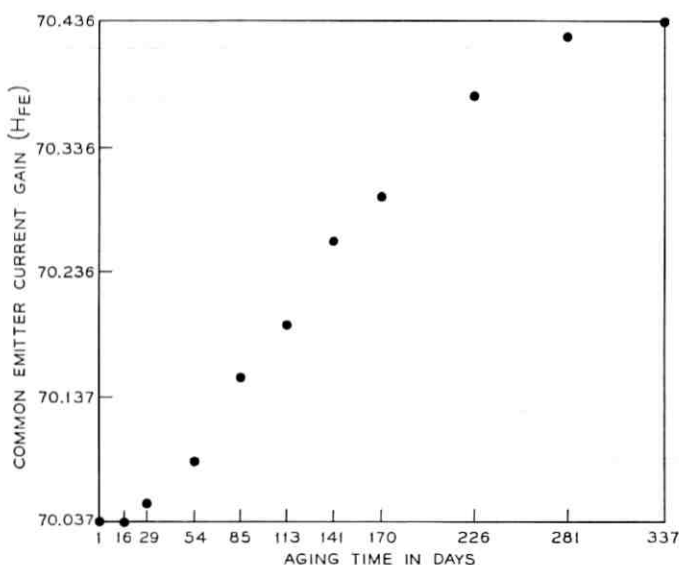


Fig. 5—Aging of transistor common emitter current gain.

VI. PRECAUTIONS FOR ASSEMBLY OF DEVICES AND COMPONENTS INTO CIRCUITS

6.1 *Semiconductor Devices*

In order to protect the semiconductor devices from the possibility of mechanical, electrical, and thermal damage from the time they leave the device manufacturing area until they are completely incorporated into the circuit, a fixture is attached at the base of each device. This fixture has floating tab conductors through which each lead of the device is threaded so that no force of any kind is exerted on the device leads. After the fixture is rigidly attached to the body of the device and each lead is rigidly attached to its floating tab, the tabs are rigidly cemented in place to the fixture. Later, in the SF Repeater Circuit, electrical connection is made only to the tabs. In this way the possibility of mechanical damage to the glass seals around each lead of the device is avoided. The possibility of thermal damage to the device due to soldering into the circuit is thus also much more remote. To protect against the possibility of electrical damage the external leads of each device have a short circuit, which is not removed until the device has been electrically and mechanically secured in its circuit.

6.2 *Passive Components*

To protect against corrosion during storage or handling, all metal cases and leads were gold plated. In the assembly, no component was mounted by its leads. Mounting fixtures were provided which minimized the possibility of damage to the component, its leads and connections due to shock or vibration. Components which were not in hermetically sealed containers were protected from moisture during manufacture and storage by maintaining a fairly low ambient humidity. The finished repeater was flushed with dry nitrogen gas and then sealed.

REFERENCES

1. Holdaway, V. L., VanHaste, W., and Walsh, E. J., "Electron Tubes for the SD Submarine Cable System," *B.S.T.J.*, 43, No. 4 (July 1964), pp. 1311-1338.
2. Gleichmann, T. F., Lince, A. H., Wooley, M. C., and Braga, F. J., "Repeater Design for the North Atlantic Link," *B.S.T.J.*, 36, No. 1 (January 1957), pp. 81-87.
3. Buus, R. G., Kassig, J. J., and Yeisley, P. A., "Repeater and Equalizer Design," *B.S.T.J.* this issue, p. 631-661.
4. Chapman, A. T., "Manufacture of Submarine Cable Repeaters and Ocean Block Equalizers," *B.S.T.J.*, this issue, pp. 663-682.
5. Berry, R. W., Jackson, W. H., Parisi, G. I., and Schafer, A. H., "A Critical Evaluation of Tantalum Nitride Thin Film Resistors," Proc. 1964 Elec. Components Conf., pp. 86-96.
6. Schafer, A. H., unpublished work.
7. Greenidge, C. H., unpublished work.
8. Berry, R. W., Hall, P. M., and Harris, M. T., *Thin Film Technology*, New York: D. Van Nostrand Co., Inc., 1968, p. 351.

Ocean Cable and Couplings

By A. W. LEBERT and G. J. SCHAIBLE

(Manuscript received February 19, 1970)

An SF armorless coaxial cable design having lower loss than its predecessor, the SD design, has been developed. Its basic construction is similar to that of the SD Cable. However, its larger diameter, along with lower loss polyethylene and higher conductivity inner conductor copper, provides a substantial decrease in attenuation.

A new coupling design has also been developed. The armorless version of the coupling design provides an improvement in return loss between cable and repeater and smaller deflection of the plastic parts under load. The armored version of the coupling provides improved isolation of the steel and beryllium copper parts.

I. INTRODUCTION

This article covers the design of the cable and of the coupling used to join the cable to the repeater or equalizer. Other articles in this issue give the details of the SF System, repeater, and power circuitry.

To provide 800 two-way, 3-kHz voice-frequency message channels, frequencies in the range from 500 kHz to 6 MHz are used. The cable design must provide a transmission medium in this frequency range whose loss, stability and uniformity allow the repeated system to meet modulation and noise requirements for a span of 4000 nautical miles.

If SD Cables were used for the broader band SF System, its 6 MHz loss at 10°C. and zero pressure would amount to 24,000 dB (6.0 dB/nm). To compensate for this loss, the large number of repeaters required would degrade the modulation performance and increase both the overall system cost and the possibility of failure. Higher voltage would also be needed. To minimize the number of repeaters required for the desired increased channel capacity and to optimize system cost, a lower loss cable is dictated.

II. ARMORLESS DEEP SEA CABLE

Traditionally, deep sea cable designs have been made as small as possible to permit the maximum amount of cable to be loaded on board the cable ships. For example the C. S. LONG LINES, designed specifically for laying armorless cable, can carry approximately 2000 nautical miles of SD Cable which has a diameter over its jacket of 1.25 inches.

To arrive at the desired cable design, therefore, the following general objectives were established at the start of the development program.

- (i) Decrease the loss of the cable as much as possible.
- (ii) Increase the size of the structure with respect to the SD design as little as possible.
- (iii) Obtain the maximum loss advantage from component materials.
- (iv) Minimize changes in existing SD production facilities.
- (v) Provide adequate mechanical strength for laying in water as deep as 4,000 fathoms.
- (vi) Provide sufficient strength for recovering cable with the additional weight of one repeater from depths of 3000 fathoms.
- (vii) Provide resistance to mechanical and environmental hazards equivalent to that of earlier designs.

The deep sea SF design, shown in Fig. 1, meets these objectives. It consists of virtually the same components as its SD Cable predecessor; namely, an inner steel strand strength member, encapsulated by a copper inner conductor; a low density solid polyethylene dielectric; a longitudinally overlapped copper outer conductor; and a high density poly-

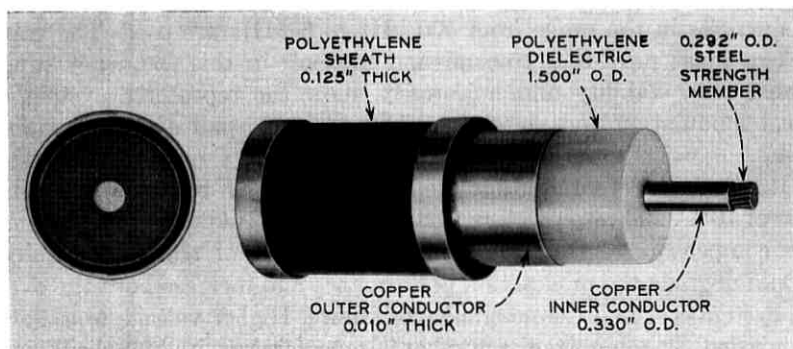


Fig. 1—Armorless ocean cable.

ethylene outer jacket. The major differences between the two deep sea cable designs are that the diameter over the dielectric of the SF Cable is $1\frac{1}{2}$ inches instead of 1 inch; the conductivity of the copper inner conductor of the SF Cable is higher; and a polyethylene dielectric material having a lower dissipation factor at 6 MHz is used. Because of its larger diameter, only about 1100 nm of SF cable can be carried by the c. s. LONG LINES.

The SD inner conductor was retained based on the following considerations. The weight in water of the SF Cable (2000 pounds per nautical mile) is virtually equivalent to that of the SD Cable. Although the drag forces of the SF Cable are larger than those for SD, the strand has sufficient strength for recovering SF Cable at depths up to 3000 fathoms at normal slow recovery rates.

To provide a higher strength cable which could be reliably recovered in depths exceeding 3000 fathoms, it would have been necessary to increase the number and/or diameter of the strand wires because the practical limit in tensile strength of the individual wires had been reached. Either of these approaches would enlarge the strand diameter and require extensive modification of existing manufacturing facilities.

Since there are relatively few locations in the Atlantic or Pacific Oceans where the depths exceed 3000 fathoms, it appears reasonable to accept the restriction that the cable can only be recovered from these depths when the weather conditions are ideal and the recovery rate is kept very low to minimize the drag force.

Having established the diameter over the dielectric as $1\frac{1}{2}$ inches, the inner conductor could have been enlarged slightly to achieve the optimum conductor diameter ratio for minimum attenuation. However, the reduction in loss achieved by this action was judged insufficient to offset the cost of modifying existing manufacturing facilities. The possibility of achieving some material cost savings by reducing the inner conductor copper thickness was also considered. The basis for this approach is the fact that 500 kHz is the bottom frequency for the SF System and 100 kHz is the bottom frequency for the SD System. Because of the skin effect phenomenon, the signal current at the higher frequency flows closer to the outer surface of the copper conductor. The conductor therefore does not have to be as thick to provide satisfactory transmission at 500 kHz. A thinner conductor, however, is mechanically weaker. A comparison of the mechanical difficulties likely to be encountered in producing and handling such an inner conductor, with the savings in materials cost indicated that the disadvantages outweighed the advantages.

III. ARMORED SHORE END CABLE

In most ocean cable systems about one or two percent of the total length is located in depths where protection against chafing on a rocky bottom or against breakage by trawlers or ships' anchors is needed. Because of these potential hazards a plow has been developed to bury the shallow water portions of cables. Where plowing is not feasible, armored cables are used. These designs are similar to the SD armored shore end cables. Figures 2a and 2b show single- and double-armored SF Cable. In these cables, the composite inner conductor is replaced by a solid copper wire and the strength of the cable is provided by outer armor wires.

The armored shore end cables are subject to many of the requirements of the deep sea cable designs. In this case, the jacketed cable diameter is $1\frac{1}{4}$ inches instead of $1\frac{3}{4}$ inches to minimize the overall armored cable size. However, the inner conductor diameter is reduced so that both

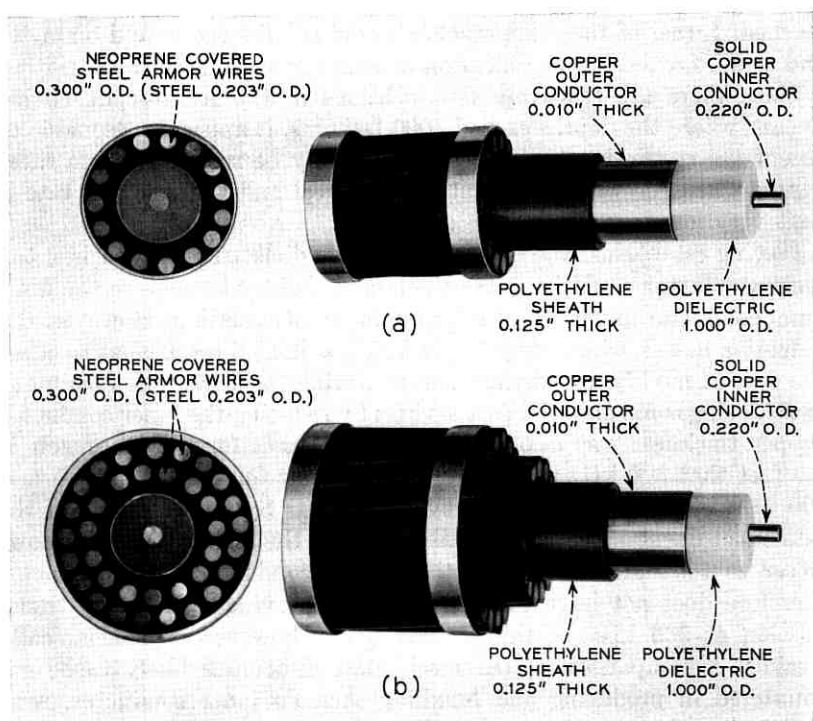


Fig. 2—(a) Single-armored ocean cable, and (b) double-armored ocean cable.

the shore end and deep sea SF Cables have the same characteristic impedance. One or two layers of neoprene-jacketed armor are placed over the basic coaxial structure, depending on the type of ocean bottom, the environmental conditions and the mechanical hazards anticipated.

IV. SHIELDED CABLES

Cable immediately adjacent to a terminal station requires shielding protection to reduce the effect of external electromagnetic disturbances on the system. Figures 3a and b show the two designs of shielded, armored cable. Normally the single armored structure is used, but where anchor damage is expected the double armored structure is used.

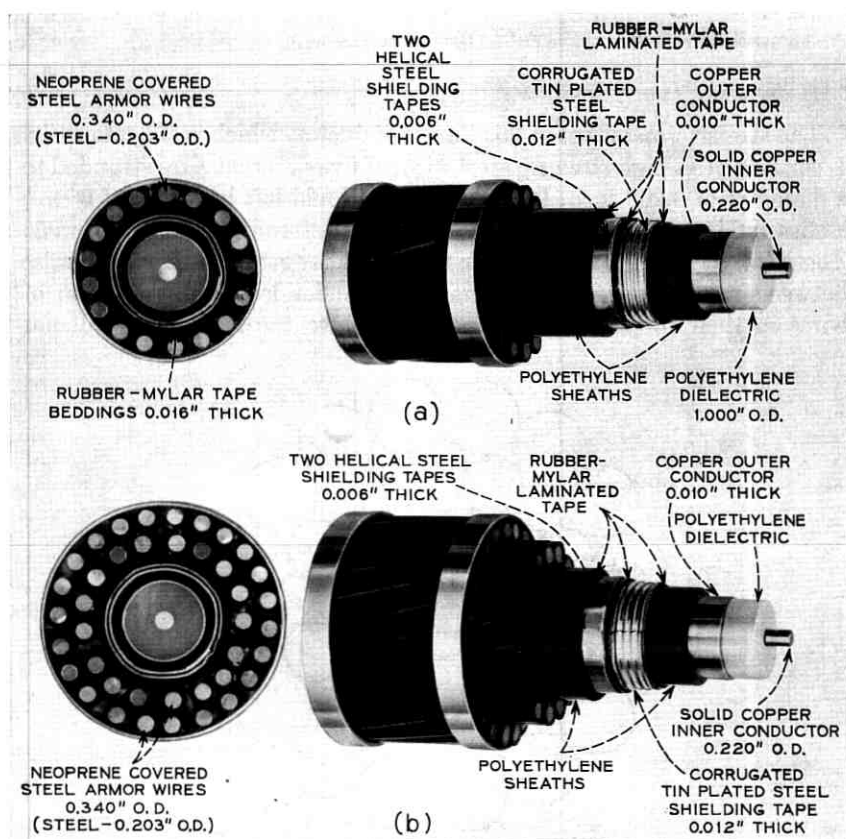


Fig. 3—(a) Shielded single-armored ocean cable, and (b) shielded double-armored ocean cable.

The earlier SD shielded design used four helically applied and one longitudinal soft iron tape. Although this design provides satisfactory electrical shielding, potential wrinkling of the steel tapes during mechanical handling of the cable and a possible subsequent reduction in the life of the cable were considered undesirable. As a result, a modified design (see Fig. 3) was used for the SF Cable. It consists of a corrugated longitudinal steel tape with a soldered seam replacing the unsoldered longitudinal tape used in SD. Over this structure two helically applied steel tapes with opposite directions of lay are applied. Cushioning of the coaxial is provided by beddings of rubber-mylar tape applied under, between and over the steel tapes. The change in shielding structure results in considerable improvement in cable-handling characteristics.

V. ARMORLESS CABLE CONSTRUCTION DETAILS

5.1 Composite Inner Conductor

The strength member for this inner conductor, which is shown in Fig. 4, consists of 41 high strength steel wires of five different sizes stranded to a diameter of 0.292 inch. It has a unidirectional left hand lay of $6 \pm \frac{1}{4}$ inches and is a compact fill, achieving a tensile strength of 16,000 pounds. The relatively long 6 inch lay for this diameter strand is a compromise between minimization of torsional twist under load and retention of wires in their proper stranded position. Since the lay length will not

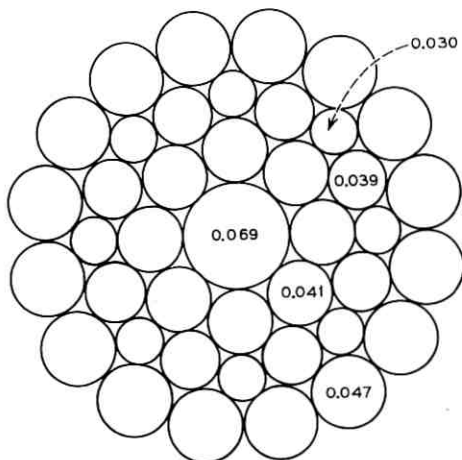


Fig. 4—41-wire steel deep sea cable center strand (average ultimate individual strength is 300,000 psi and the approximate breaking strength is 16,000 pounds).

remain uniform unless enclosed while under stranding tension, the copper shell which is the coaxial inner conductor is formed around the strand in a tandem operation.

The uniformity of the cable strength is controlled by placing a tolerance of ± 0.0005 inch on the five sizes of wire diameters comprising the center steel strand and a ± 7 percent tolerance on their tensile strength. In addition, the diameter of the completed strand is closely maintained and adequate fill of the strand cross section is achieved by placing a tolerance of ± 0.008 inch on the algebraic summation of diameter deviations for the individual wires comprising the strand. To assure that the inner conductor copper and the steel strand act as a single unit the minimum depth of penetration of the copper between the individual steel outer layer strand wires is 0.005 inch.

The 0.023 inch thick oxygen-free copper shell is formed from annealed tape into an oversize tube, the longitudinal seam of which is sealed by a continuous weld in an inert gas atmosphere. The oversize tube is then swaged down tightly onto the steel strand, partially filling its outer interstices. Oxygen-free copper is specified for ease and reliability of welding. The diameter of the finished composite inner conductor is 0.330 inch. The average conductivity of the copper raw material is 101.7 percent as compared to the International Annealed Copper Standard which is generally accepted as 100 percent. For SD Cable, the average conductivity was 101.4 percent.

5.2 Dielectric

To achieve the desired cable attenuation characteristics, a dielectric material must possess good electrical properties which are not degraded by processing, mechanical handling, or severe environmental conditions. The material must also be capable of adhering to the inner conductor surface when extruded. This adhesion assists in preventing interfacial voids and assures adequate longitudinal shear transfer from external forces to the center strength member.

Although the existing SD Cable polyethylene had good mechanical properties, its dielectric loss at 6 MHz was considered excessive for use in the SF System. Since in general, the dielectric loss is a direct function of frequency, the magnitude of the loss is more important at the upper end of the band where the contribution of dielectric loss to the total loss of the coaxial is significant. Through the cooperation of the polyethylene raw material suppliers, a low density, high molecular weight material having a dissipation factor of less than 80 microradians at 6 MHz was obtained. A contributing factor to this low loss was the substitution of

Ethyl Antioxidant 330 [1, 3, 5 trimethyl-2, 4, 6 tri (3, 5-ditert butyl-4 hydroxy-benzyl) benzene] in place of Santonox^R [4, 4' thiobis-2 tertbutyl-5 methyl phenol] for the thermal antioxidant. The dielectric material was closely controlled during manufacture to assure uniformity of product. The result was a material which met the dissipation factor requirements shown in Table I. The dielectric constant also was closely controlled to a requirement of 2.285 (+ 0.0002 or - 0.0003).

TABLE I—DISSIPATION FACTOR REQUIREMENTS

Frequency in MHz	Dielectric Loss in Microradians		
	Minimum	Nominal	Maximum
0.1	35	45	55
1.0	56	61	66
6.0	74	79	84

The low density, high molecular weight polyethylene is pressure extruded over the inner conductor slightly oversize and is then shaved to a precise dimension of 1.500 ± 0.001 inch at a temperature of 20°C. The eccentricity of inner conductor is limited to 0.005 inch.

5.3 Outer Conductor

Experience with the overlapped longitudinal outer conductor on SD Cable indicated that this type of structure is electrically superior to helical tapes and is adequate mechanically with proper restrictions placed on bending radii. Consideration was given to reducing the copper thickness for SF Cable below the 0.010 inch thickness previously used, since from an electrical standpoint thinner copper would suffice. Theoretical and experimental evidence, however, showed that reducing the thickness would aggravate the problem of buckling of the outer conductor even though the cable was bent around larger sheaves and cable drums than were used for the SD Cable. The 0.010 inch thick copper outer conductor was therefore retained because it performs satisfactorily when completed cable is bent around a minimum diameter of 9 feet. To assure that the cable will provide the desired mechanical performance, short lengths of each 10 nm section of cable are required to withstand 50 reverse bends at the 9-foot diameter without experiencing outer conductor failure.

As shown in Fig. 5, an annealed electrolytic, tough pitch copper tape 5.000 ± 0.005 inches wide and 0.010 ± 0.0002 inch thick is formed

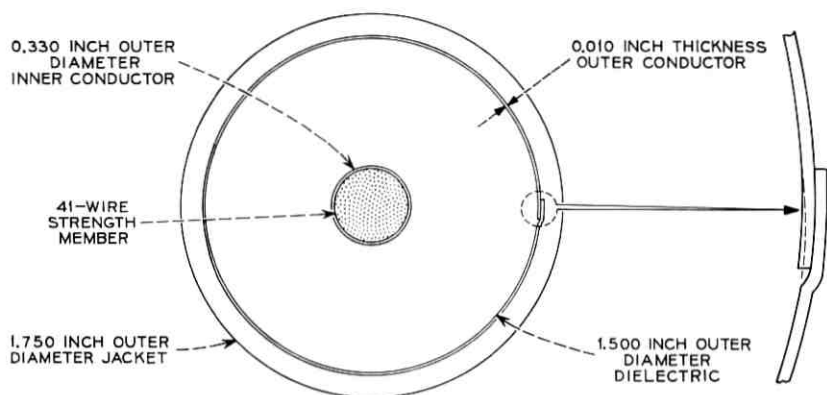


Fig. 5—Cross section of armorless ocean cable.

tightly around the shaved core by a forming mill to provide an overlapped longitudinal seam. The air space between the core and tape at the seam overlap is practically eliminated by the formation of a knee in the overlapping tape edge by the final forming roller. Since the outer conductor is not constrained after formation, this operation is performed in tandem with the jacket application which provides the necessary constriction.

5.4 Jacket

Several jacket materials were investigated including the high density black polyethylene jacket used in the SD Cable. Since none of the newer materials examined were significantly better than the SD jacket polyethylene in limiting outer conductor buckling when the cable was bent over sheaves or in abrasion and environmental stress crack resistance, it was decided to use the SD type jacket for SF Cable. This material is a high density, high molecular weight black polyethylene which is extruded over the outer conductor of the coaxial to a thickness of approximately $\frac{1}{8}$ inch. The carbon black is added to minimize deterioration from photo-oxidation during periods of storage and shipment. The eccentricity of the jacket is restricted to 20 mils and the outside diameter tolerance is 1.750 ± 0.010 inch.

5.5 Cable Loss Characteristics

In the SD Cable design, emphasis was placed on providing a structure having a loss which is not only predictable but stable. In the case of the SF Cable design, both of these requirements are of even greater im-

portance. For example, in a 4,000 nm system, the SF Cable loss at 6 MHz is about 16,000 dB (4.0 dB per nautical mile). An overall loss deviation of 0.1 percent is 16 dB, which is sufficient to cause system problems. Therefore to minimize these problems, the attenuation is controlled within close tolerances during manufacture and laying by providing equalizers at 192 nautical mile intervals in the system. If, however, the loss of a cable section and the gain of the associated repeater has frequency characteristic deviations which are not readily compensated and are not stable with time, serious system misalignments can occur. Consequently, the effect of all parameters on the cable loss must be evaluated and tight tolerances on material and finished cable characteristics established to avoid system problems.

Nineteen repeater sections are connected together to form a 192 nm ocean block. Each repeater section consists of 10 nm of cable and a repeater. Between any two ocean blocks, two one-nautical-mile lengths of cable and an equalizer are installed to provide means for reducing the misalignment of the system. The equalizer primarily corrects for unpredictable changes in loss encountered during laying. In the cable factory, an objective of ± 1 dB at ± 6 MHz, 10°C . and zero pressure is placed on the cumulative loss of the nineteen sections manufactured for an ocean block when compared with specified cable loss for the ocean block.

In general, variations in conductor dimensions, dielectric constant or copper conductivity are reflected in a percentage change in loss characteristic which is constant over the transmission frequency band. A variation of outer conductor thickness however has a small effect on the cable loss up to about 1.3 MHz and no effect at higher frequencies as shown in Fig. 6 for a ± 0.0002 inch thickness change. This results from the conductor being electrically thin at the lower frequencies.

Another parameter which has a sizeable effect on the cable loss is the dissipation factor. The effect of a ± 5 microradian change is shown in Fig. 7. Here the increased importance of dissipation factor at higher frequencies is apparent.

Table II presents the tolerances on important characteristics and indicates the magnitude of their combined accumulated effect on total cable loss in percent. On a root sum square basis the deviations at 6 MHz are ± 0.24 percent, which represents a maximum misalignment of about 2.0 dB for an ocean block. It, therefore, is apparent that in order to satisfy the desired overall requirement every effort must be made to achieve the nominal value and to keep the excursions from that value to a minimum. The average percent deviation of the loss of manufactured

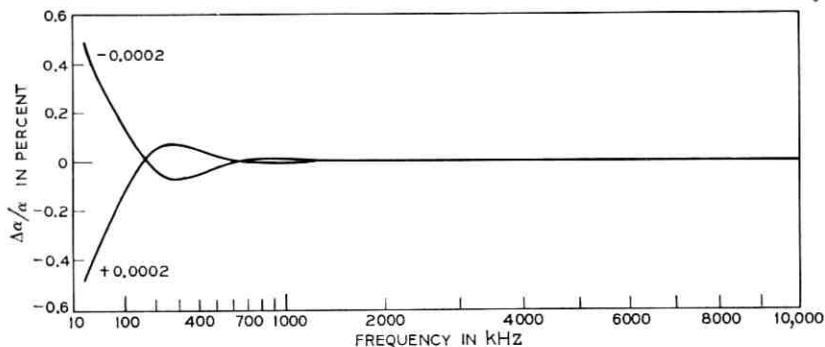


Fig. 6—Effect of a 0.0002 inch change of outer conductor thickness on attenuation of 1.5 inch cable (nominal thickness is 0.0100 ± 0.0002 inch).

product from design for the Florida-St. Thomas and TAT-5 SF Systems are shown in Fig. 8. Minor adjustments in section lengths are made in the factory when required to keep the total loss of an ocean block well within the objective.

5.6 Derivation of the SF Design Loss Characteristic

The loss characteristic of the SF Cable was predicted by: (i) reduction of average SD Cable loss data to primary and (ii) secondary constants, and recalculation of the loss after modification of parameters affected by the changes in inner conductor conductivity, core diameter, and dissipation factor of the dielectric.

Measurements made on experimental cables produced in the pilot

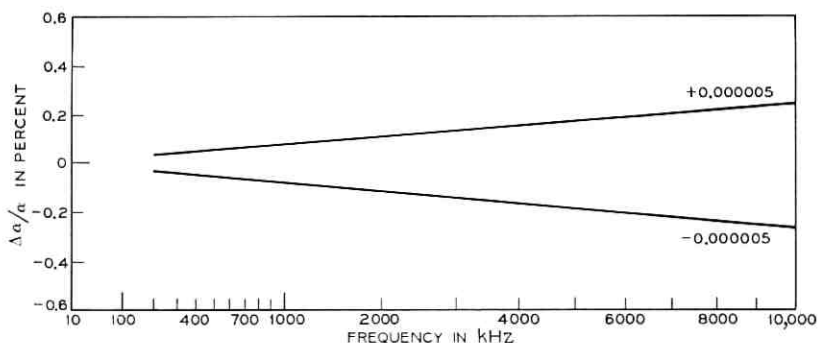


Fig. 7—Effect of a 0.000005 radian change in dissipation factor on attenuation of 1.5 inch cable. (nominal value is 0.000079 ± 0.000005 radians).

TABLE II—SF DEEP SEA CABLE—PERCENT CHANGE IN TOTAL CABLE LOSS RESULTING FROM TOLERANCES ON CABLE CHARACTERISTICS

Characteristic	Dimensions & Permissible Tolerance	Effect of Tolerance on Cable Loss in Percent	
		0.5 MHz	6 MHz
Inner conductor diameter	0.330 inch \pm 0.001	\pm 0.051	\pm 0.048
Diameter over dielectric	1.500 inch \pm 0.001	\pm 0.055	\pm 0.054
Thickness of inner conductor	0.023 inch \pm 0.0005	0.00	0.00
Thickness of outer conductor	0.010 inch \pm 0.0002	\pm .023	0.00
Dielectric constant	2.285 { + 0.002 - 0.003 }	+ .044 - .066	+ .044 - .066
Dissipation factor 0.1 MHz	0.000045 \pm 0.000010		
0.5 MHz*	0.000055 \pm 0.000010*	\pm .111	
1 MHz	0.000061 \pm 0.000005		
6 MHz	0.000079 \pm 0.000005		\pm .188
Conductivity of inner conductor	101.7 \pm 0.3%	\pm .121	\pm .118
Conductivity of outer conductor	101.2 \pm 0.3%	\pm .024	\pm .026
	Algebraic summation	\pm .440	\pm .489
	RSS	\pm .192	\pm .241

* Interpolated from values specified at 0.1 and 1.0 MHz.

laboratory were then used to refine the predicted loss characteristic and to project the loss data to 6 MHz. The temperature coefficient of attenuation and the pressure coefficients were also determined from measurements made on the experimental cables.

The dissipation factor of the experimental product was derived from the loss measurements. The following formula for dissipation factor

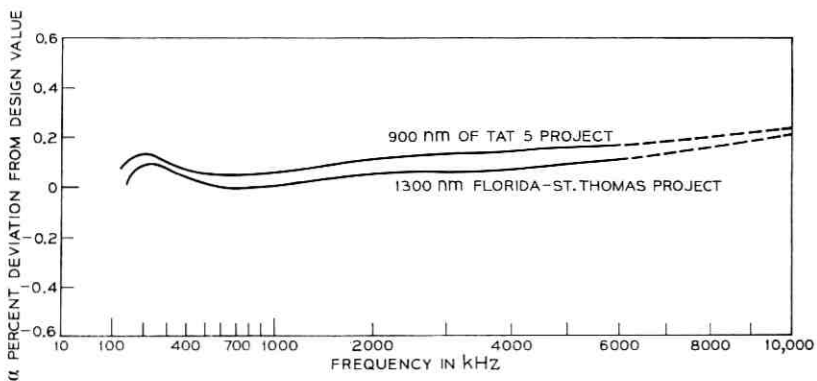


Fig. 8—Per cent deviation of attenuation from design value.

was derived from data at 10°C, zero pressure

$$F_p = 56.36 + 46.07 \log (1 + f_{\text{MHz}}/3) \text{ microradians.}$$

5.7 Other Areas Investigated

The increase in diameter of dielectric from 1.0 to 1.5 inches created some uncertainty as to whether the large volume of polyethylene could be extruded in one operation at a reasonable line speed without producing retraction voids or excessive deformation. Since these defects are dependent upon the rate at which the extruded core is cooled, the relationship between line speed, water trough lengths and temperatures was investigated and a suitable time-temperature curve was developed. The overall length of the experimental extrusion line used to establish these parameters was such that good cores could only be achieved at a line speed of about 5 to 8 feet per minute. Based on these experimental data, it was estimated that a speed of 12 to 15 feet per minute could be achieved if the SD manufacturing facilities in the existing cable plants were used to produce SF Cable. This estimate proved correct in actual practice.

A review of the minimum diameter to which the cable can be bent repeatedly without rupturing the outer conductor was necessary because of the larger cable diameter. The minimum diameter to which SD Cable could be bent is 72 times the diameter of the cable core. Applying the same factor to SF, the minimum-bend diameter is 9 feet. Tests as discussed previously, confirmed that this requirement will provide equivalent outer conductor life expectancy to that obtained on SD Cable. Therefore, the minimum diameter of all reels, coils, cones, sheaves and drums was specified as 9 feet.

VI. CABLE COUPLINGS

Since the cable must be connected to a repeater, means for accomplishing this had to be devised. The device had to provide a mechanical transition from the cable strength member to the repeater housing, and a waterproof connection for the coaxial transmission path. Both had to be accomplished without degrading the inherent mechanical strength and electrical performance of the cable.

A special coupling assembly was therefore designed and requirements compatible with those for the cable were established, as given below:

(i) Metallic materials were to be beryllium-copper compatible with the repeater housings.

(ii) All beryllium-copper parts were to be capable of an in-line pull of 50,000 pounds.

(iii) A gimbal was to be provided, with a minimum angular displacement capability of 45° from its normal axis in any direction of bending.

(iv) The coupling was to be capable of carrying a repeater through 180° excursions over specified bow sheaves and cable drums at laying and recovery tensions.

(v) The return loss measured in the transmission path extensions (pigtailed) midway between the cable and the repeater was to be greater than 30 dB over the frequency range between 0.5 and 6 MHz.

(vi) The inception of corona noise was to occur at a voltage greater than 4500 volts dc when measured in a 48 kHz band with a 6.8 microvolt threshold.

(vii) Reliable bonds between the parent and injected polyethylene material were to be achieved in dielectric restorations at junctions of cable and the pigtail extension to the coupling components.

6.1 Armorless Coupling

The initial coupling design used in the Florida to St. Thomas SF System is shown in Fig. 9. During the manufacture of this system, the

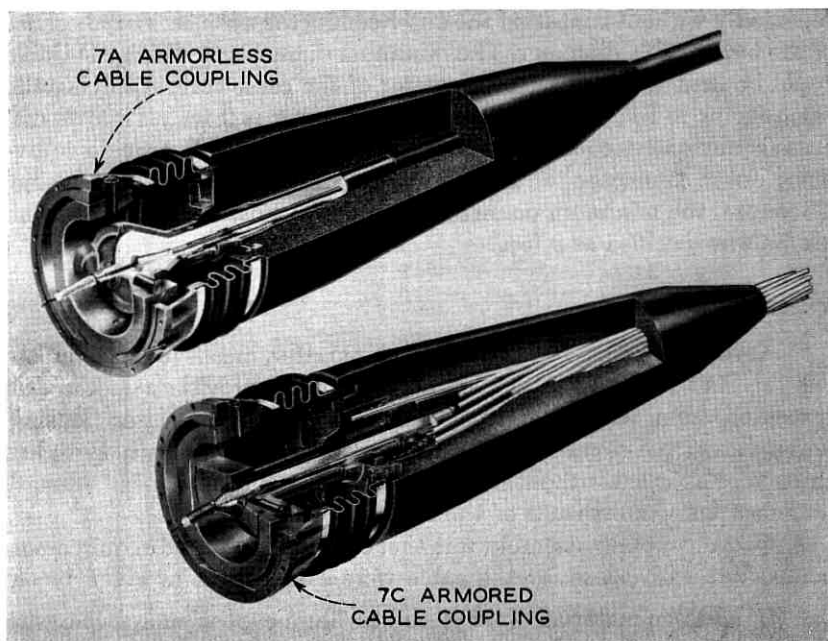


Fig. 9—7A and 7C coupling.

design was modified to improve the reliability of the outer conductor and jacket connection. The latest design is covered by Figs. 10 through 15 inclusive.

The coupling is supplied as piece parts which are assembled and joined to the cable in the cable manufacturing plant. Terminated cable sections and repeaters are then loaded separately on shipboard and joined together prior to laying the cable.

The heart of the deep sea cable coupling is a beryllium-copper termination cone assembly and a ceramic anchor which are assembled and molded with polyethylene, as shown in Fig. 11, to form the anchor molding assembly. This assembly is attached to the cable by inserting the inner conductor with an end portion of strand exposed into the extension tube until the steel wires flare out in the termination cone. Epoxy is then injected into the cone through the small fill tube and cured, thus anchoring the strand. The copper inner conductor is soldered to the end of the extension tube and this junction is bridged by several turns of tinned copper wires soldered at each end to provide a back-up connection. The cable and anchor mold assembly dielectrics are joined by molding. The tapered portion of the anchor mold assembly seats in the matching taper of the beryllium-copper cone housing shown in Fig. 12, and is pinned to prevent rotation of the core under load. The supported tapered cone provides a surface area which results in low unit pressures on the polyethylene during laying and recovery. In addition to the mechanical advantage offered by this shape, it has an electrical advantage since the return loss of the unit is better than 40 dB over the frequency band involved.

Calculations based on the projected surface area indicate that under maximum load conditions the average polyethylene stress is 2400 psi which is well below the stress found acceptable for the earlier SD System 5A coupling. Laboratory tests of deflection, defined as combined movement of ceramic anchor and anchor molding assembly, under constant tensions of 3000 and 12,000 pounds are shown in Fig. 13. The deflection noted after 100 hours at the lighter load is less than 0.010 inch which is considered negligible. At the 12,000 pound load, which simulates expected maximum sustained cable recovery loads, an acceptable deflection of 0.027 inch is noted after 21 hours, which is considerably longer than required for the normal recovery operation.

The outer conductor is restored by brazing a preformed copper insert to the outer conductor of the cable and welding it to the cone housing. Beryllium-copper pins (Fig. 14) in the housing prevent a preformed jacket flange from rotating. A collar (Fig. 15) prevents the

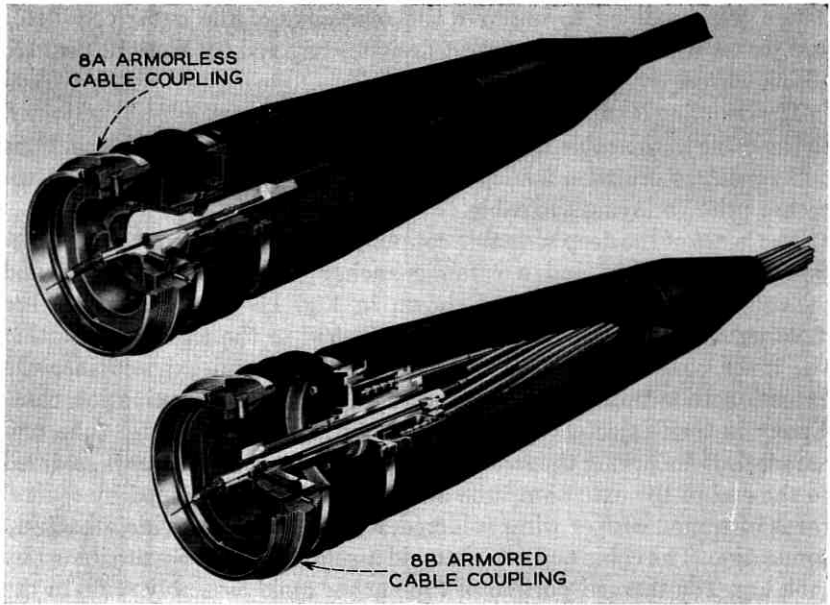


Fig. 10—8A and 8B coupling.

flange from retracting when the cable is under tension. The jacket between the flange and the cable is then restored by molding.

The pigtail is joined to the anchor mold assembly by crimping and soldering the center conductor to the fill tube and restoring the dielectric by molding (Fig. 14). The cone housing cap, pigtail cap and braid are then assembled on the coupling, and the pigtail is formed into a coil which is heat treated to give it a permanent set. The coupling is completed by assembling the gimbal ring, gimbal housing, rubber boot and bellows. The boot serves to minimize stresses in the molded joints and the bellows serve to restrict the entrance of shells, sand and rocks into the area of the moving parts.

The initial outer conductor and jacket restoration designs of the coupling differed somewhat from the present one. During installation of the Florida to St. Thomas SF System, jacket retraction, or "milking," was observed with resultant outer conductor breakage. Although satisfactory temporary remedies were employed for this system, the design was modified to assure that the core, outer conductor and jacket act as a single unit when the cable is subjected to torsional and tensile forces.

The use of bolts to connect the coupling to the repeater was carried

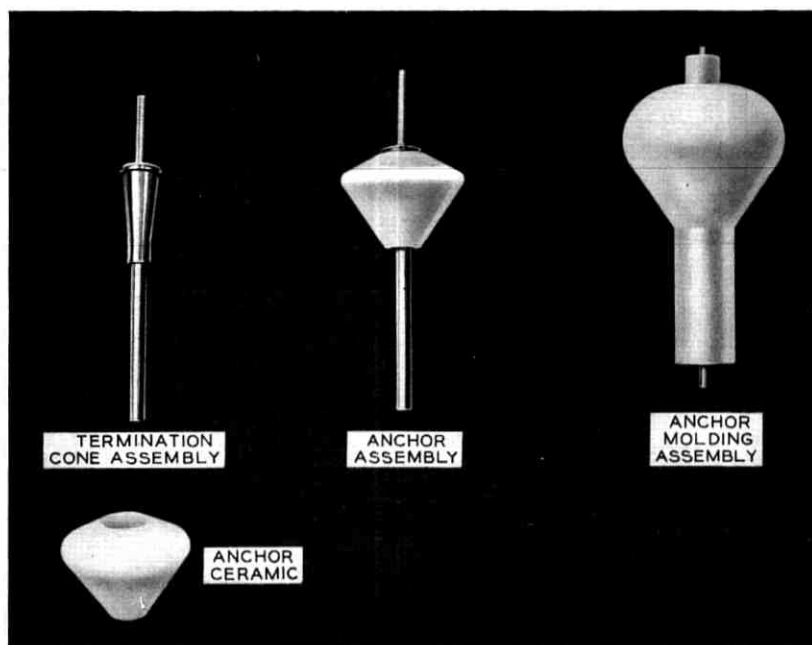


Fig. 11—Anchor molding assembly.

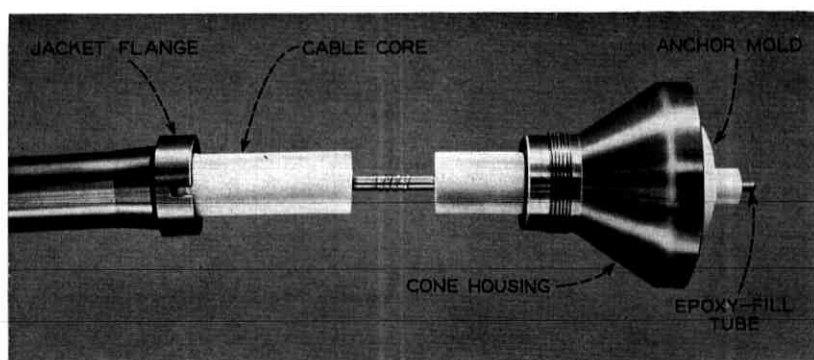


Fig. 12—SA coupling inner conductor restoration.

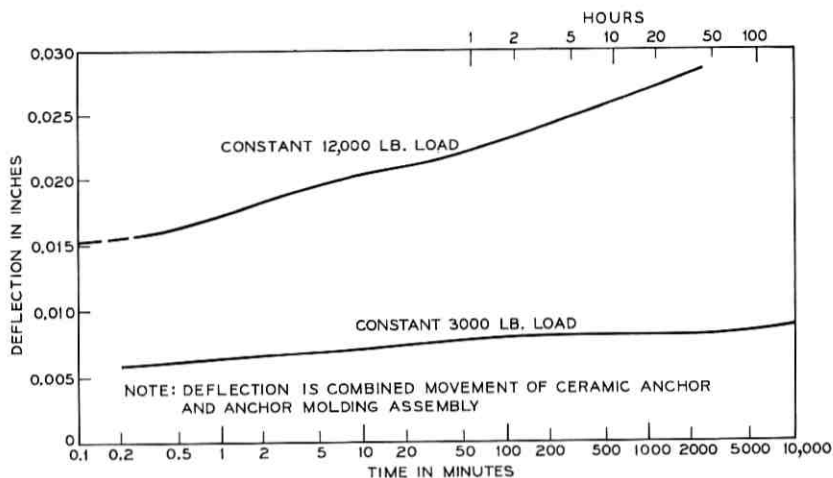


Fig. 13—Tension-loading 8A coupling anchor.

over from the SD coupling design to the initial SF design. Because of possible stress corrosion, these have been replaced by a major thread on the coupling and support housing. The large thread bearing surface minimizes stress concentrations and consequently any potential stress corrosion. Tests in the laboratory and at sea have shown the validity of the final design. The results during the installation of TAT-5 confirm this.

6.2 Armored Coupling

For this coupling assembly (see Figs. 10, 16 and 17), the single or inner layer of armor wires are unlaid and the end of the cable is prepared

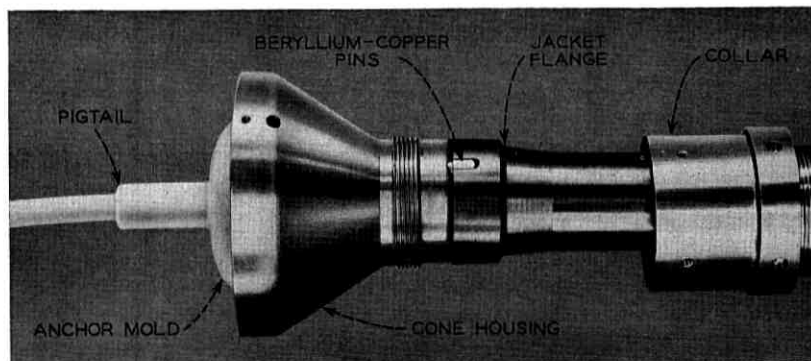


Fig. 14—8A coupling pigtail mold.

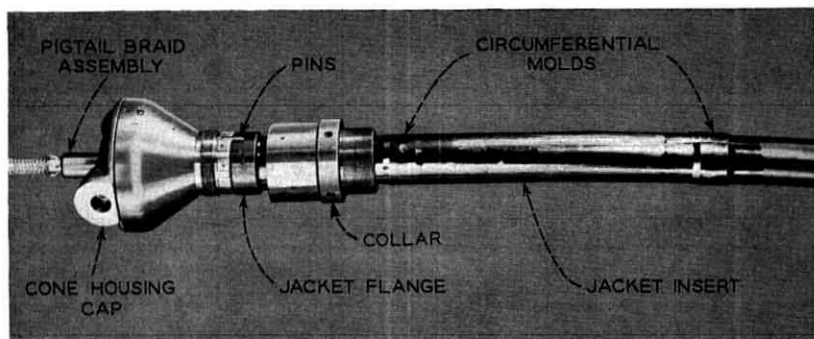


Fig. 15—8A coupling jacket restoration.

to expose the solid copper inner conductor. The core diameter is shaved; the pigtail conductor is attached directly to the cable conductor; and the insulation is restored by molding. A round copper nut with pigtail cap and braid attached is slid over the pigtail onto the shaved portion of the cable dielectric. The pigtail is then formed into a coil and given a permanent set by the application of heat. Outer conductor restoration is accomplished by slitting it into narrow longitudinal strips and brazing the strips to the round nut. A preformed jacket flange extending from the parent jacket to the end of the round nut is molded to the parent jacket. A collar screwed onto the round nut prevents longitudinal jacket movement (Fig. 16). A sub-assembly (Fig. 17) consisting of the gimbal armor support and slotted armor ring previously placed on the cable is positioned near the cable end. The armor wires are relayed and their

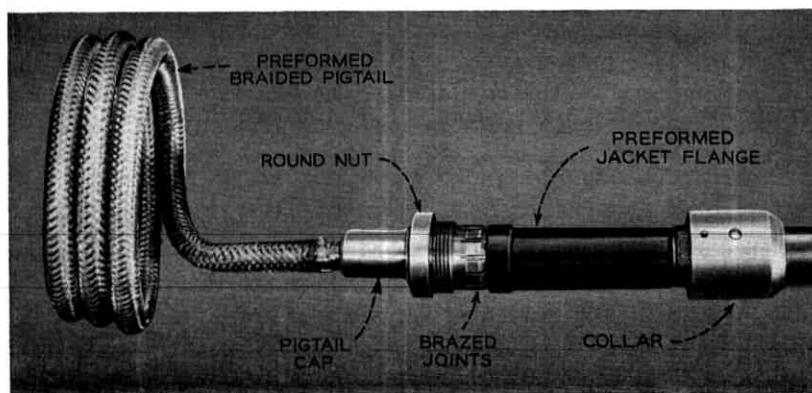


Fig. 16—8B coupling attachment of outer conductor and positioning.

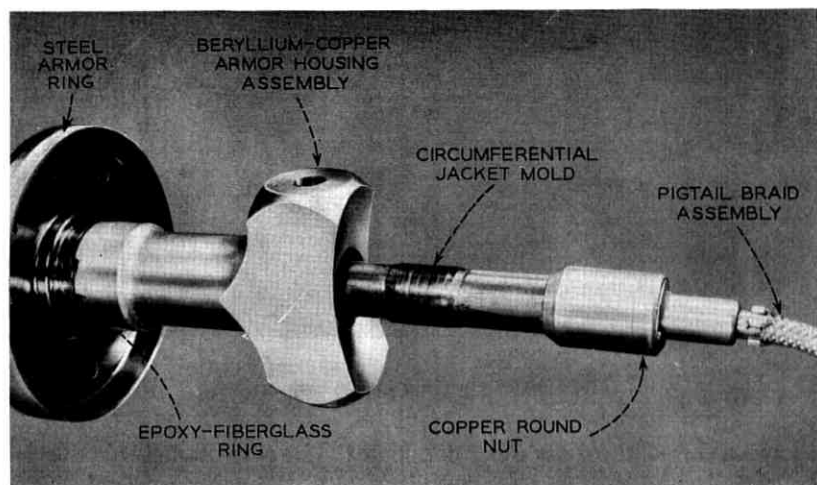


Fig. 17—SB coupling completed assembly of jacketed coaxial.

sleeved ends are placed in the proper slots on the armor ring and lashed down. The completed assembly is shown in Fig. 10.

The possible galvanic cell between the beryllium copper piece parts and the steel is eliminated by having the armor wires in direct contact with the steel armor ring, isolating this ring from the beryllium copper strength members of the coupling by the use of a nylon sleeve and epoxy-fiberglass rings. Additional corrosion protection is obtained by filling the space which the armor wire connections occupy with polyethylene-polybutene compound.

In the original coupling design, corrosion isolation was achieved by individually isolating the steel armor wires from a beryllium-copper armor ring. This proved to be only partially successful. An interim modification improved this condition for the Florida to St. Thomas SF System. The latest design described above, however, was used in TAT-5.

6.3 Dielectric Molding

Dielectric restorations are required at times during cable manufacture and also when couplings are attached to the cable. When joining two polyethylene parts, it is of prime importance that reliable bonds between injected and parent polyethylene material be obtained. To achieve these bonds it is necessary to have the parent polyethylene, the injectate, and the conductor in the molding area considerably hotter than the melt point of the polyethylene. These temperatures must be

sustained a sufficient length of time to permit the bond to take place. Cooling must be gradual and injection pressure must be high enough and maintained long enough to prevent retraction voids and undesirable levels of internal stress. The mold must be designed to vent air, have dual diametrically opposite injection ports to minimize conductor eccentricity, and be coated with a material such as Teflon to prevent sticking. Purgings the mold with the injectate improves the possibilities of good bonding because the surfaces of parent polyethylene will be washed and the temperature of the conductor will be raised to an adequate level.

All molds must be instrumented so that the complete molding cycle, including preheat, injection, and cooling is recorded. By providing the proper control equipment, it is then possible to repeat the process once a satisfactory cycle is established, since the cycle can be confirmed by chart recordings.

Operators and machines are subjected to qualification routines to assure the consistent reproduction of the proper molding cycle. The quality of a joint is determined by microtoming it to produce a thin wafer which is examined, tested in tension and heat treated. Flexing the joint around a small mandrel is also used as a criterion.

VII. SUMMARY

The purposes of this development were to provide a cable having a lower loss than the SD Cable, to better accommodate the increased frequency range of the SF System, and to provide an improved repeater to cable coupling arrangement. The tight overall system transmission requirements made it necessary to achieve the predicted cable loss frequency characteristic within narrow limits. To assure this result, tight limits had to be placed on allowable variations in material characteristics and rigid in-process controls had to be imposed. It was also necessary to design couplings which would permit the system to be laid and recovered without difficulty. The experience extending through TAT-5 indicated that the efforts were successful.

VIII. ACKNOWLEDGMENTS

The authors wish to acknowledge the many contributions made by members of all the Bell Laboratories groups who participated in this cable and coupling design. In addition, we acknowledge with thanks the contributions of the members of staff of Standard Telephone and Cables, Ltd., and the Western Electric Company, Baltimore Works.

Shore Terminal Facilities and Fault Location

By R. L. LYNCH, J. L. THOMAS, and C. A. VON ROESGEN

(Manuscript received August 13, 1969)

A special shore terminal transmission facility has been developed for the SF Cable System. The multiplex portion differs from similar domestic plant in that the frequency allocation follows international standard, and can operate with 3-kHz as well as conventional 4-kHz spaced message channels. High reliability is achieved through duplication of equipment combined with added failure detection circuitry and protection switching. Additional features include broadband flexible equalization capability to optimize transmission performance, a separate 4-channel order wire, and automatic pilot measuring equipment. From the point of view of physical design, two departures from conventional practice are the use of shorter than normal bay frameworks and enclosures to cover them, both of which are compatible with foreign design practice.

Newly designed test equipment enables monitoring undersea performance and fault locating from the shore terminal. Inclusion of a supervisory oscillator in each repeater is a major change from previous submarine cable systems.

I. INTRODUCTION

1.1 General

Shore terminals at each end of an SF Submarine Cable System provide an interface between the undersea portion of the system and inland transmission facilities. Conventional frequency-division multiplexing techniques are used to translate message signals received from the domestic network to the appropriate frequencies for transmission over the ocean cable; the inverse process is performed on signals received from the cable.

The high-frequency portion of the shore terminal provides broadband signal shaping (pre-emphasis) and equalization to optimize the

signal-to-noise performance of an SF System. Final adjustment of equalization for a particular system takes place after the cable has been laid and its actual transmission characteristic has been determined.

Once a submarine cable system is installed, information concerning its performance can be obtained only from measurements made at the shore. Thus, another function of the terminal is to provide means for gathering information necessary for system maintenance and for locating faults that might occur either in cable or repeaters.

These functions of the SF Terminal are described in more detail in the sections that follow. Emphasis is placed primarily on those features unique to the SF System and, secondarily, on those features unique to a submarine cable terminal. Generic categories of terminal equipment are as follows:

- (i) Multiplex and carrier supply,
- (ii) High-frequency line,
- (iii) Pilot monitoring and measuring circuits,
- (iv) Order-wire, and
- (v) Maintenance and fault locating test sets.

DC power for operating the undersea repeaters is supplied from the shore terminal over the cable. The power plant is described in a companion article.¹

1.2 Frequency Allocation

The line frequency allocation of the SF System is shown in Fig. 1. Since the undersea portion of the system is equivalent 4-wire, the overall

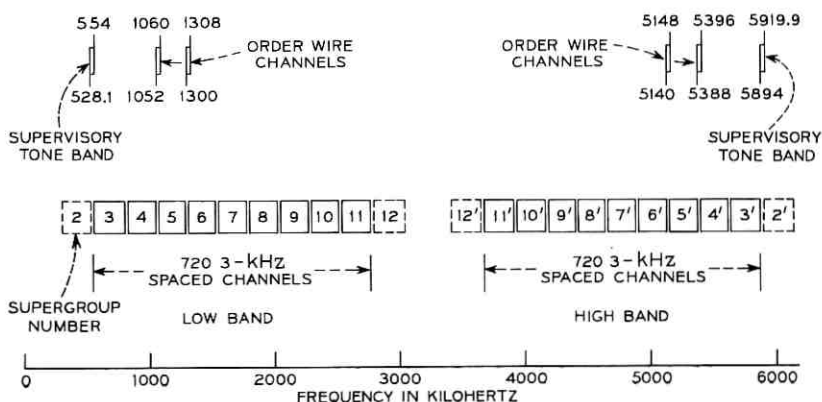


Fig. 1—Frequency allocation—SF System.

transmission spectrum is divided into low and high bands. The terminal that transmits in the low band and receives in the high band is designated an A Station; the B Station, at the other end of the system, is complementary. The actual usable system bandwidth will vary from installation to installation, depending principally on system length and equalization.^{2,3} Each band consists of nine complete 240-kHz-wide supergroups (numbered 3 through 11) plus whatever additional capacity is obtainable in the spectrum of supergroups 2 and 12.

The low band is obtained as a direct output of the multiplex equipment. Its formation from voice frequency is shown schematically in Fig. 2. The supergroup frequency assignments conform to CCITT* recommendations. Efficient use is made of the available bandwidth, since guard bands only 8 kHz wide are between all supergroups except 2 and 3, which have 12-kHz separation. The high band is formed by amplitude modulating a 6448-kHz carrier with the multiplex output and selecting the lower sideband. Because of its relationship to the direction of signal transmission, this step is often referred to as directional modulation.

1.3 Reliability

High reliability has always been an essential design objective of the undersea portion of submarine cable systems. Shore terminal reliability is also important. Although shore-based equipment is much more accessible for repair, loss of service because of terminal failure is indistinguishable, from the point of view of the customer, from undersea failure. The reliability criterion established for the SF Terminal is that no single failure will result in the loss of more than the equivalent of one supergroup. As illustrated in Fig. 3, which is a simplified block schematic of the SF Terminal transmission arrangement, duplication of equipment is the major means used to meet this requirement. Supergroup banks and high-frequency lines are duplicated with automatic protection switching between units, under control of pilot monitoring circuitry. A spare group bank is available on a patch-in basis.

II. MULTIPLEX AND CARRIER SUPPLY

2.1 Channel Banks

Signals to be transmitted over the SF System can be obtained from the domestic plant via group connectors (for signals already translated to the group frequency band), or from channel bank equipment at the

* The International Telegraph and Telephone Consultative Committee.

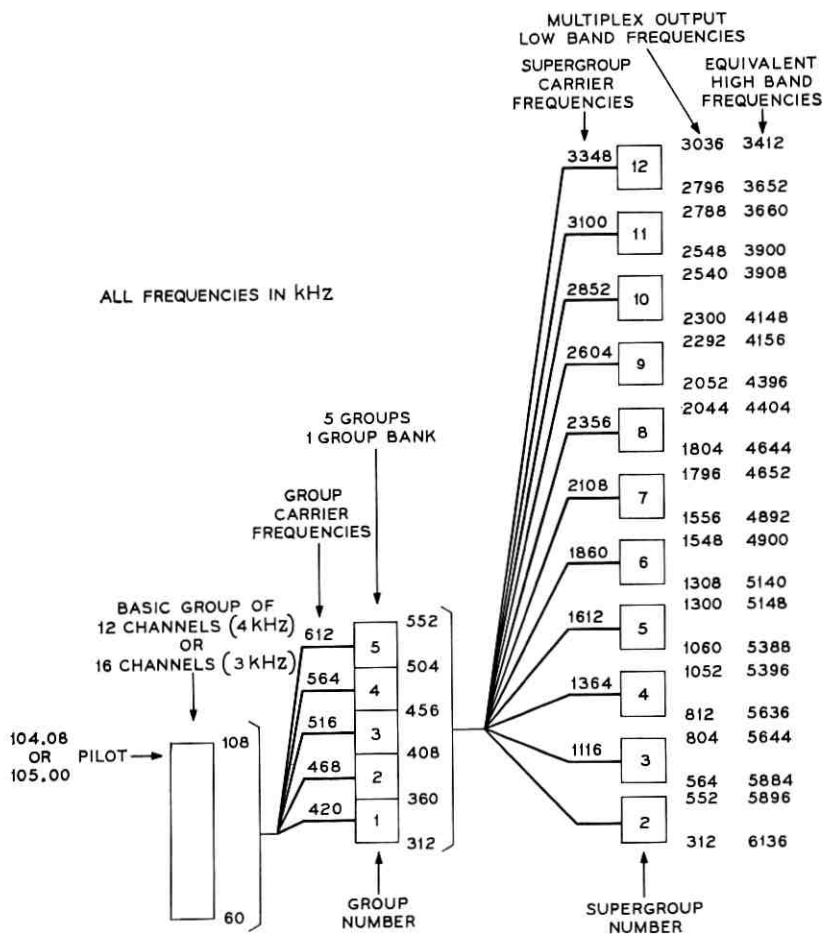


Fig. 2—Frequency allocation—SF Multiplex.

shore terminal. In the latter case, high-efficiency banks are generally used. They include sixteen 3-kHz-spaced channels in the 48-kHz group band that normally accommodates twelve 4-kHz-spaced channels. By careful design a voice bandwidth of about 2800 Hz is obtained. Although larger and more expensive than conventional channel banks, these prove to be economically more advantageous for use with submarine cable systems. Channel banks of this type, which are described in the literature,^{4,5} are manufactured by several telephone equipment suppliers.

2.2 Group and Supergroup Banks

The group and supergroup portion of the SF Multiplex is an extensively modified version of the new L Multiplex Terminal.⁶ Modifications were required to:

- (i) operate with 3-kHz-, as well as 4-kHz-spaced channels*,
- (ii) conform to CCITT supergroup frequency assignments,
- (iii) achieve the increased reliability required of a submarine cable terminal, and
- (iv) house the equipment in bays that are only 2.74 meters (9.0 feet) high.

By far the most difficult requirement was the first. Although not obvious, the reason is simple. All group and supergroup carrier frequencies (as well as a number of other tones generated by the carrier supply) are harmonics of 4-kHz. The power of these tones is generally quite high compared to that of a message circuit, and although normally subjected to high attenuation they nevertheless find their way into the transmission path (for example, via common-ground impedances, or because of inadequate modulator balance) with sufficient energy to potentially degrade performance. With 4-kHz operation, however, these carrier leaks, as they are often called, fall between channels and are of little concern.[†] On the other hand, many carrier leaks fall within 3-kHz-spaced channels and are of serious concern. It can be shown that for the SF Multiplex spectrum, the carriers for groups 4 and 5 of supergroup " n " fall as a 1050-Hz tone in channel 2 of groups 1 and 2 of supergroup $(n - 1)$, for $n = 4$ through 12. Similarly, the supergroup n carrier falls as a 1050-Hz tone in channel 14 of group 4, supergroup $(n + 2)$, for $n = 3$ through 9.

In group equipment, relatively minor modifications reduced the offending tones below the requirement of -65 dBm at zero transmission level (-65 dBm0) for inband frequencies, and -55 dBm0 otherwise. Modifications to the standard group bank design include the addition of shields to twisted pair wiring, some replacement of pairs by coaxial cable, changes in cable routing, and the addition of special, experimentally determined circuit ground arrangements.

The supergroup bank is another matter entirely. The need to hold

* Three- and four-kHz-spaced channel groups can not be mixed within a supergroup because the relatively slow cutoff of 4-kHz channel filters allows crosstalk between edge channels which are adjacent in the supergroup spectrum.

[†] Although this is true for voice service, the presence of spurious tones in broadband data services is of concern. Even here, however, requirements are not as severe as for voice channels.

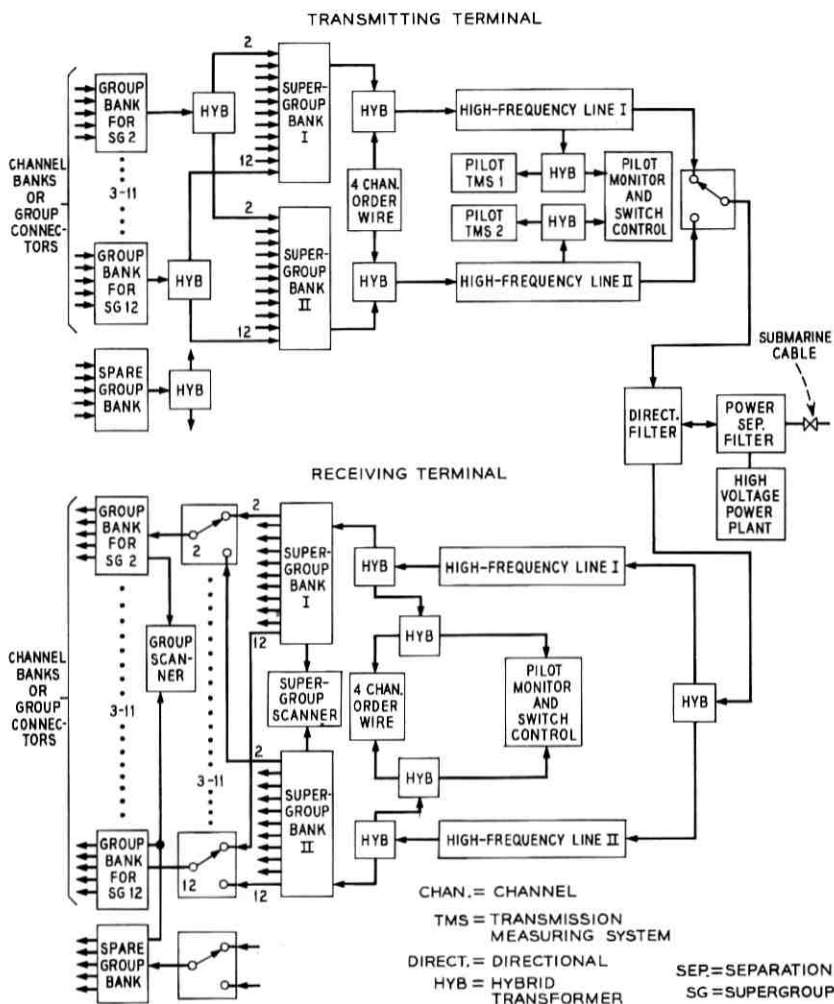


Fig. 3—Block diagram—SF Terminal transmission arrangement.

at least 135 dB isolation between the carrier leads and the lowest transmission level point in the transmitting supergroup bank dictated extensive mechanical and electrical redesign. Auxiliary filters were designed for each supergroup to provide additional loss in the transmission path at the carrier frequency, and a special carrier bandpass filter was added in each carrier lead to suppress spurious energy at transmission band frequencies. Use of CCITT frequency assignments forced the

design of three new supergroup bandpass filters, but in fact all ten supergroup filters were redesigned to meet more recent Bell System objectives on amplitude distortion in the supergroup bands.

The new physical layout of the supergroup bank was carefully designed to minimize length of critical leads, to separate high- and low-level cabling, to provide shielding between circuitry associated with different supergroups, and to preserve the coaxial structure through connectors linking the fixed portion of the bank and plug-in modules. The design makes use of a relatively small, flexible, double-shielded coaxial cable that is used extensively in other shore terminal equipment.

2.3 *Pilots*

A reference pilot tone is inserted into each group in the transmitting terminal to facilitate monitoring and adjustment of transmission at various points. It is also used to control the automatic gain regulation of the receiving group circuits. Conventional 104.08-kHz pilots are used with 4-kHz-spaced channel groups. This frequency is not compatible with 3-kHz operation, however. A frequency of 105 kHz was chosen for this case, since it has the same advantage, at least potentially, as the 104.08-kHz pilot; that is, it is but one channel spacing from the top (108 kHz) of the group band and so would not interfere with broadband services. Group reference pilots are inserted at a power of -20 dBm at zero transmission level.

2.4 *Carrier and Pilot Supply*

Like group and supergroup banks, the carrier and pilot supply portion of the SF multiplex is made up largely of modified equipment from the new L multiplex terminal.⁷ Primarily, modifications were necessary to provide:

- (i) duplicate equipment to achieve increased reliability;
- (ii) auxiliary group and supergroup carrier capacity for supplying the terminal order-wire facility, pilot monitoring and measuring equipment, and duplicate multiplex equipment;
- (iii) the capability of generating 105-kHz pilots required by 3-kHz channel banks;
- (iv) the capability of generating a 120-kHz channel carrier required by 3-kHz channel banks;
- (v) the capability of generating a 6448-kHz carrier required for the step of directional modulation; and
- (vi) the capability of generating a 620-kHz carrier used in the remote carrier-supply synchronization scheme.

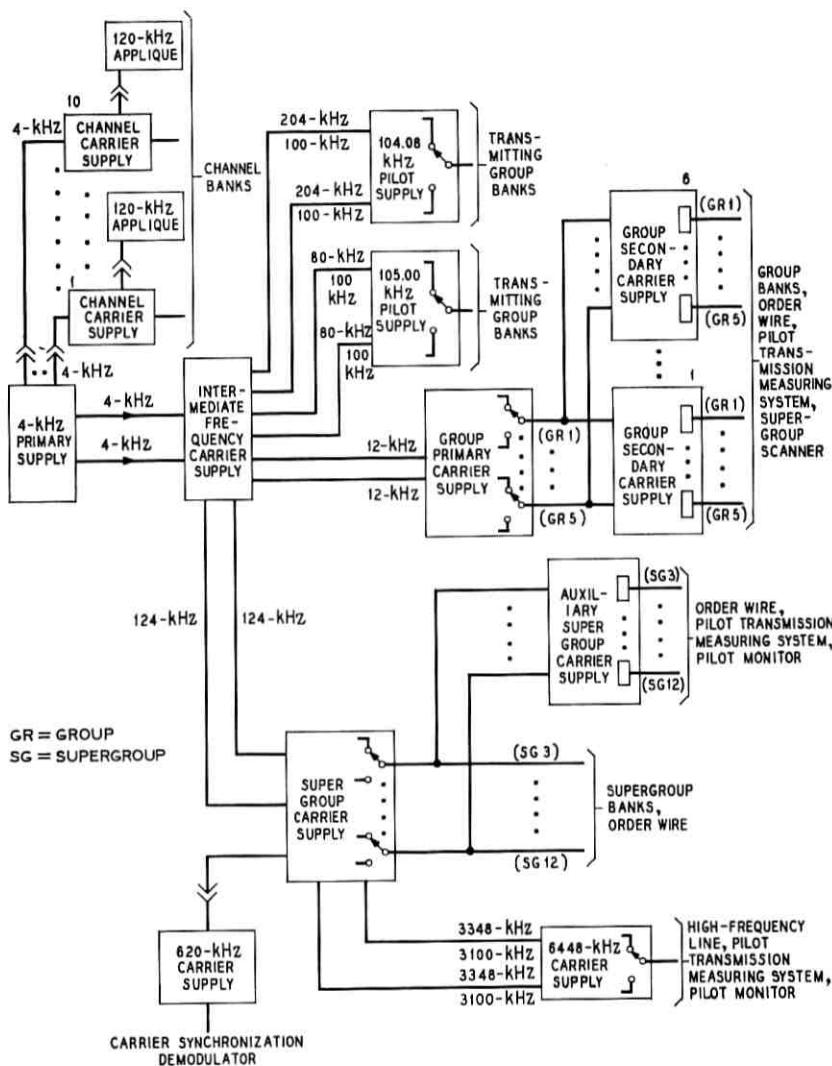


Fig. 4—Block diagram—carrier and pilot supply.

The carrier supply block diagram is shown in Fig. 4. Complete duplication of equipment is provided from the 4-kHz primary frequency supply to the distribution bus for the:

- (i) 104.08-kHz and 105-kHz pilot supplies,
- (ii) group primary carrier supply,

- (iii) supergroup carrier supply, and
- (iv) 6448-kHz carrier supply.

Trunks from a test tap on each of these distribution buses are brought to a carrier test panel where they can be connected by push-button selection to a measurement jack. Alarms are provided for each bus and there is automatic failure detection and switching between duplicate equipment. The group secondary and the auxiliary supergroup carrier-supply distribution buses include test point and alarm features.

The 105-kHz stabilized pilot is generated in a way similar to the 104.08 kHz. An 80-kHz signal is divided by 16 and the resulting 5 kHz is added to 100 kHz. The 80-kHz and 100-kHz signals are obtained from the standard intermediate-frequency supply. The output of the pilot supplies is stabilized so that it remains within 0.1 dB of the adjusted setting for variations in temperature between 0 and 50°C, changes in input power of ± 5 dB, and changes in supply voltage of ± 5 volts.

The 6448-kHz carrier frequency is obtained by adding the carriers associated with supergroups 11 (3100 kHz) and 12 (3348 kHz). The extra 120-kHz channel carrier required by 3-kHz channel banks is obtained directly from the output of the regular channel carrier supply by selective filtering.

2.4.1 Carrier Synchronization

At the heart of the carrier supply is the 4-kHz primary frequency generator.⁸ When the SF Terminal was planned, the standard primary supply, designed for domestic use, could be synchronized from an external source of either 64 or 308 kHz. Since neither frequency falls within the SF Transmission Band, frequency translation is necessary for synchronization of the far end primary supply over the undersea system. At the synchronizing terminal, an 88-kHz tone is inserted between two channels in the transmitting order-wire and translated to a low band line frequency of 1056 kHz. At the remote synchronized terminal, the tone undergoes two steps of demodulation* (to 308 kHz at an A station, or to 64 kHz at a B station). A non-standard 620-kHz carrier, which is required for the second step of demodulation in a B station, is obtained directly from the supergroup carrier supply harmonic generator by selective filtering.

* A change in frequency of any carrier used to demodulate the synchronizing tone introduces that change (magnitude and direction) in the tone itself. An even number of demodulation steps is required to insure that the error signal derived within the primary frequency supply (by comparing the received and locally generated tones) has the correct sense for stable operation.

III. HIGH-FREQUENCY LINE

3.1 Description

The high-frequency line, which is a completely new design, provides a transmission path between the multiplex and the ocean cable itself, via the power separation filter. It includes low-to-high-band modulation or high-to-low-band demodulation, as required, broadband equalization capability, and access to the transmission path for order-wire signals and test equipment. Figure 5 shows a considerably simplified block schematic of the high-frequency line. A number of amplifiers, attenuators, hybrid transformers, and filters are not shown. The role of the high-frequency line in system equalization is discussed separately in Section IV.

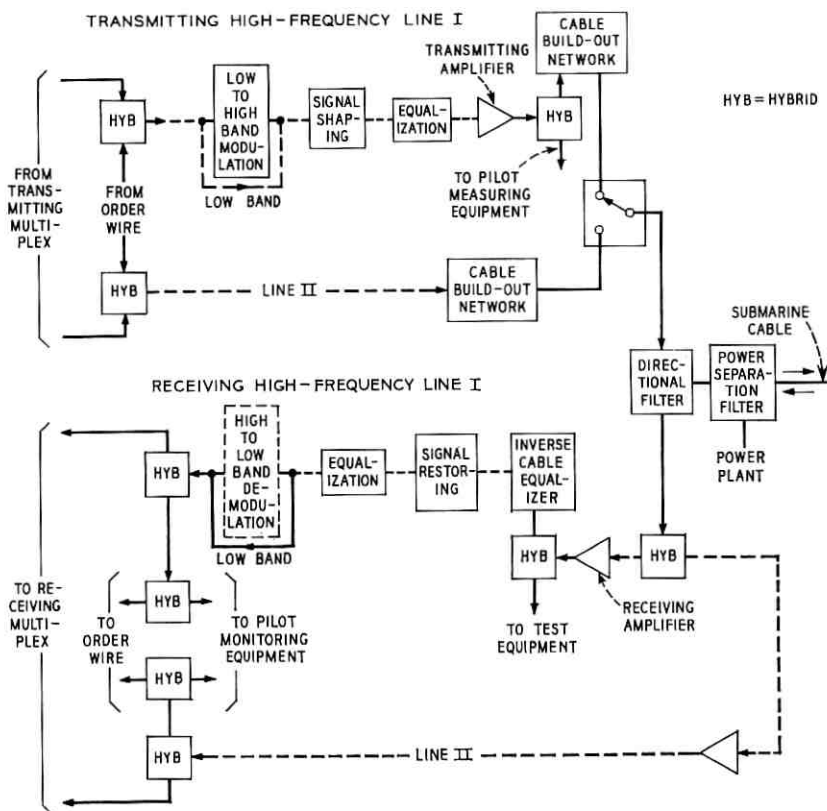


Fig. 5—Simplified schematic—high-frequency lines.

3.2 Design Considerations

An important design requirement is to keep the noise contribution of the high-frequency line negligible with respect to that of the undersea system. Potential sources of excessive high-frequency line noise are:

- (i) intermodulation noise from high level amplifiers,
- (ii) thermal noise from low level amplifiers, and
- (iii) intermodulation and thermal noise in the directional modulation circuitry.

In general, the last amplifier in the transmitting line will be operating at the highest level and the first amplifier in the receiving line at the lowest, because these levels depend on the maximum allowable length of the shore-end section of ocean cable.* For the SF System this length is equal to one repeater spacing. In that case transmission levels at the terminal-submarine cable interface are the same as for a repeater. Therefore, the power output capability of the transmitting amplifier and the thermal noise performance of the receiving amplifier must be at least comparable to those of a repeater to keep their noise contribution negligible relative to that of the undersea portion of even a short system. These requirements were most easily met by using the basic design of the input and output amplifiers developed for the basic repeater in the L4 System.⁹ The latest transistor and circuit design technology had achieved a low noise figure and a high-power linear output capability. Modification of the L4 design was necessary to remove the gain-frequency shape that is characteristic of analog transmission system line repeaters. Only minimal changes were necessary in the critical high-frequency cutoff region of the feedback loop (to beyond 100 MHz).

Typical noise performance of the two modified amplifier designs is shown in Fig. 6. The low thermal noise preamplifier has a maximum noise figure of about 6.0 dB and a flat 75-ohm insertion gain of 15.1 dB. The power amplifier has a flat 75-ohm insertion gain of 24.4 dB. The intermodulation noise portion of the curves of Fig. 6 are based on measurements using broadband random noise loading to simulate the multiplex signal. Correlation of intermodulation noise with amplifier output transmission level is based on a simulated rms signal power of +19.0 dBm0 per band, the maximum design load of the SF System.²

Figure 7(a) shows the low-to-high-band directional modulation arrangement. A conventional double-balanced diode ring modulator is used. Because of the relatively wide baseband and its closeness in

* The shore-end section is that cable between the terminal and the first repeater.

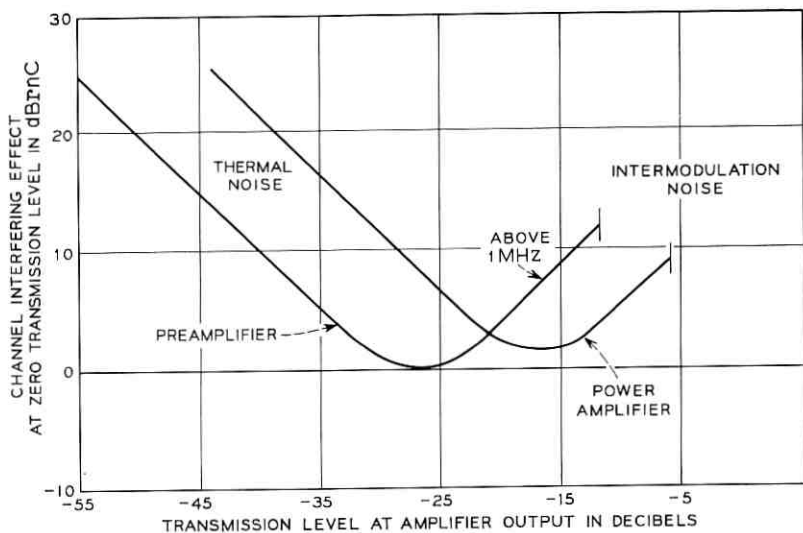


Fig. 6—Thermal and intermodulation noise performance of high-frequency line amplifiers.

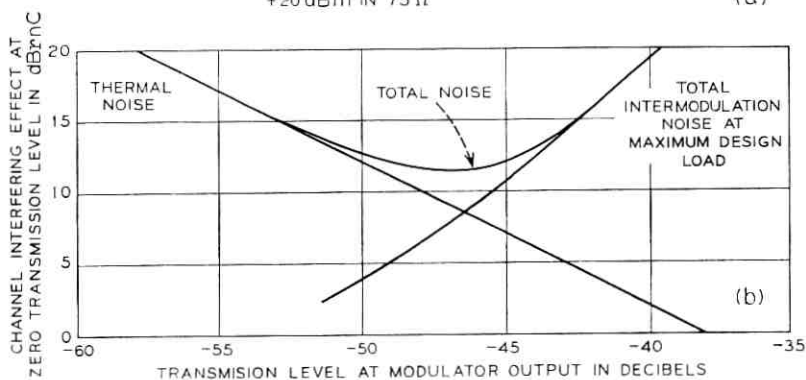
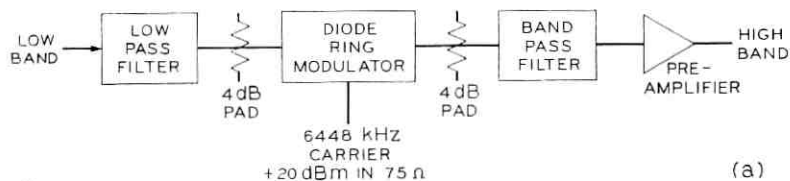


Fig. 7—Directional modulation scheme: (a) block schematic; (b) noise performance.

frequency to the wanted sideband (see Fig. 1), a great many intermodulation products fall into the transmission spectrum. Second- and third-order sideband as well as second-order baseband intermodulation are significant. The low noise amplifier described in Section 3.2 is an integral part of the frequency translation scheme because the operating transmission level is set to make the intermodulation noise of the modulator itself equal to the combined thermal noise of both modulator and amplifier. The noise performance is shown in Fig. 7(b). The intermodulation product distribution versus frequency is such that the noise is approximately equal across the band. Demodulation is less critical because of fewer in-band intermodulation products; in this step a noise contribution of less than 10 dBmC at zero level is achieved.

3.3 Power Separation Filter

The primary function of the power separation filter (PSF) is to combine transmission signals and dc power for application to the ocean cable. It is designed to permit safe and convenient disconnection of the combined transmission and power path from the submarine cable when power is turned down in order that

- (i) the cable fault localization test set (CFLTS), which will be described later, can be connected to the cable or
- (ii) the power feed path can be reconnected to an auxiliary dc test load located within the PSF equipment bay.

With the auxiliary load connected, the high-voltage facilities can be tested up to the last possible point prior to the submarine cable itself.

All these functions are accomplished without compromising safety. A PSF key interlock system is used in combination with the key interlock scheme of the power plant¹ to ensure that the plant is shut down before connections can be changed and that all stored energy is dissipated before access is allowed to the high voltage areas. Carbon protector blocks are used to limit voltages that can be induced on the transmission or ocean ground cables and to provide a short circuit to building ground if the ocean ground is disabled while the power plant is operating.

Since signals received from the submarine cable are quite weak, another important design feature of the PSF is protection against radio frequency interference. The high-voltage enclosure within the PSF bay, where the ocean cable terminates, is a copper box that is normally at the potential of the cable's outer conductor (return tape ground). As a shield, it is effectively an extension of the coaxial structure of the

cable. The two other grounds at the PSF are the familiar building ground and an ocean ground, which is used instead of building ground for the dc cable return current to avoid corrosion by electrolysis to nearby underground metallic structures. The ocean ground is brought into the terminal via armored ocean-type cable from a ground bed usually located near the terminal building. At transmission frequencies, there is no need to maintain isolation between these grounds. In fact, judicious capacitive interconnecting, which may vary from location to location, is used to minimize the pickup of spurious signals.

IV. TERMINAL ASPECTS OF SYSTEM EQUALIZATION

4.1 *Initial Lineup*

Once an SF System has been installed and its actual transmission characteristics established, its signal-to-noise performance is a function of the operating signal levels on the system, as determined by the transmitting terminal. When the system is initially aligned, equalizer-networks in the transmitting high-frequency line are adjusted in an attempt to obtain maximum and equal channel signal-to-noise ratios. Equalizer-networks in the receiving high-frequency line are then adjusted to provide the correct level at the input to the receiving multiplex.

The specific design approach that realizes these objectives is illustrated in Fig. 8(a). (The means of coping with equalization changes with time are shown in Fig. 8(b). The means will be discussed later.) Signal shaping utilizes both fixed and adjustable equalization. The fixed *signal shaping network* would be adequate for an ideal system, that is, basically one with no misalignment. The adjustable *special shapes equalizer* is used to obtain more nearly optimum signal shaping for each specific installation. Figure 9 depicts the kinds of shapes obtainable with the special shapes equalizer. After all other transmitting terminal equalizer-networks have been adjusted, the setting of the special shapes equalizer is determined experimentally, using noise loading techniques.*

The *system equalizers* are a series of networks designed for a particular installation after the actual system misalignment is established. Because the SF System is modulation limited,^{2,10} a simple yet effective rule for approximately minimizing the noise penalty associated with a given misalignment characteristic is to equalize half in the transmitting

* The special shapes equalizer is adjusted to obtain approximately equal total noise (thermal plus intermodulation) near the top, middle, and bottom of both transmission bands when the system is carrying its maximum design rms load of +19.0 dBmO per band, as simulated by gaussian noise.

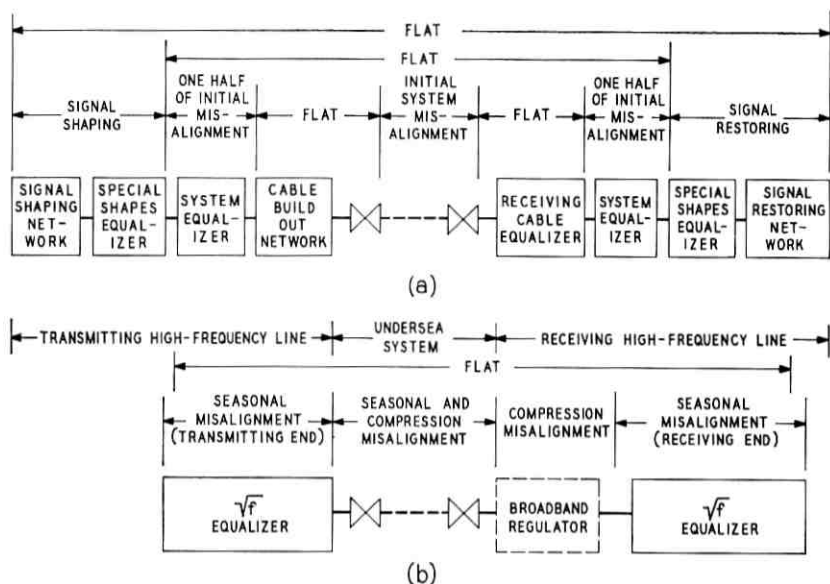


Fig. 8—SF Shore Terminal equalization plan: (a) initial lineup; (b) subsequent alignment.

terminal, and the other half in the receiving terminal. In practice, the system equalizers also compensate for transmission deviations in high-frequency line equipment.

The *cable buildout network* is adjusted with strapping options so that its loss plus the loss of the shore-end section of cable is equivalent to the loss of one full cable section (10 nm). This buildout is required because shore-end lengths vary from system to system. The *receiving cable equalizer* is made up of inverse cable-shape networks similarly adjusted for each system so that the sum of equalizer and shore-end section cable loss is approximately flat across the transmission band.

Signal restoring in the receiving high-frequency line, that is, *special shapes equalizer* plus *signal restoring network*, provides a shape complementary to that of signal shaping. Thus, the combined loss of signal shaping and restoring for a particular system is a constant.

4.2 Maintaining Alignment

Subsequent changes in undersea transmission with time must also be equalized by readjustments in the terminal. Relatively slow changes can occur due to seasonal variations in sea bottom temperature in the shallow water near shore. Permanent changes can occur, for example,

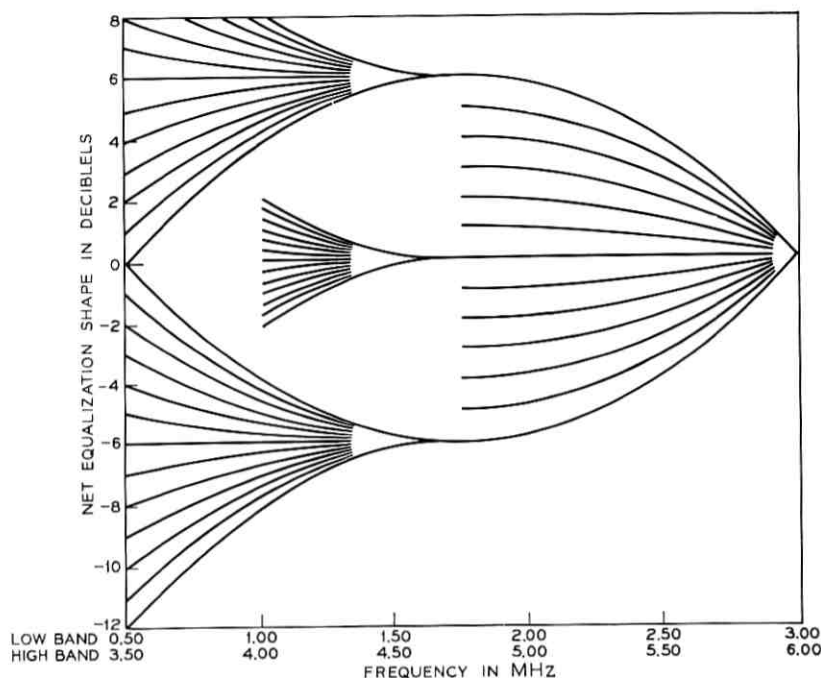


Fig. 9—Special shapes equalizer—some composite shapes.

due to aging of repeaters or cable, or as a result of undersea repairs. Gain compression in the undersea repeaters^{2,11} as a function of signal load results in rather rapid fluctuations in system gain.

The specific design approach for equalization of these changes with time is illustrated in Fig. 8(b). Since variations in undersea transmission because of seasonal temperature swings are, to a great extent, proportional to the square root of frequency, an adjustable $(f)^{\frac{1}{2}}$ equalizer* compensates for this effect. For example, when cable loss near shore increases, loss is removed from the transmitting $(f)^{\frac{1}{2}}$ equalizer to restore signal levels on the bulk of the system to their former value. At the same terminal, loss would be removed from the receiving $(f)^{\frac{1}{2}}$ equalizer to maintain the correct level at the input to the multiplex.

It is difficult to be specific about equalization of a permanent change in undersea transmission. In general, should such a change occur, it would be equalized in a manner similar to that used during the initial lineup.

* Equalizer loss = $K \pm n (f/6)^{1/2}$ dB, where K is a constant, f is frequency in MHz, and $n = 0, .5, 1, 1.5, 2.0 \dots 8$.

At present, it is expected that gain compression will be compensated for by a relatively high speed broadband *regulator* functioning as an expander, which will be located in the receiving high-frequency line. It will be controlled by a tone inserted in the distant terminal transmitting high frequency line. As a result of the dynamic behavior of the system gain the tone will be amplitude modulated. At the receiving terminal the variations in amplitude will be extracted by the regulator to provide an expansion characteristic complementary to that of system compression. This approach is feasible only when compression is approximately flat with frequency across the transmission band.

At this point, however, the design of the regulation scheme must still be considered experimental. An important parameter yet to be established is the speed at which the regulator must operate in order to eliminate transmission impairments resulting from compression. The answer certainly depends on the dynamic behavior of both the system compression characteristic and the signal load itself.

V. OPERATION AND MAINTENANCE

5.1 *General*

It was convenient, from the point of view of maintenance, to take full advantage of the duplicate supergroup and high-frequency line equipment by arranging two independent transmission paths, designated I and II (See Fig. 3). Hybrid transformers provide the required double feed. Duplicate paths meet again at hitless type switches, so that, in addition to fulfilling a reliability requirement, this arrangement permits a considerable amount of routine maintenance to be performed without even momentary service interruptions.

5.2 *Pilot Measuring and Monitoring Arrangements*

There are no special line pilots in the SF System. Terminal maintenance is based on use of group reference pilots. Built-in automatic and semi-automatic pilot measuring circuitry is part of the terminal. This includes:

- (i) pilot monitor and switch control equipment,
- (ii) the pilot transmission measuring system (transmitting scanner),
- (iii) receiving supergroup scanner, and
- (iv) receiving group scanner.

To avoid the difficulty of realizing very narrow bandpass (pilot pickoff) filters, which would otherwise be required to separate the pilots from the multiplex signal at SF line frequencies, all pilot measurements are

made at group baseband frequency (104.08 or 105 kHz). Thus, frequency translation is an integral part of the first three arrangements.

5.2.1 *Pilot Monitoring, Switch Control and Alarm Equipment*

In case of a failure in either duplicate transmitting or receiving high-frequency line, alarms are initiated and, if the failed path were the one in service, protection switching automatically substitutes the duplicate path. This is done by transmitting and receiving pilot monitor, switch control and alarm equipment. In the transmitting direction, two specific pilots are monitored continuously near the output of lines I and II* (See Fig. 3). In case of a failure, logical comparison of the results of the four pilot measurements automatically determines which of the two transmitting paths is more suitable for service. For maintenance purposes, a manual control allows either path to be locked in as the working path. A similar arrangement near the output of the two receiving high-frequency lines initiates alarms and controls the receiving switches.

5.2.2 *Pilot Transmission Measuring System (PIL TMS)*

The normal function of the pilot transmission measuring system is that of an automatic transmitting pilot scanner. In this operating mode, a PIL TMS is connected near the output of each transmitting high-frequency line. Each of the approximately 50 pilots is automatically measured in sequence and compared with a normal value (generally different for each pilot). The difference is displayed on a meter. A pilot deviation of more than 1.5 dB will initiate an alarm. The time required for a complete scan is about four and a half minutes.

In addition to the automatic pilot scanning capability, this measuring system can be used manually for a number of other pilot frequency measurements, as summarized below.

- (i) Direct measurement of deviation from normal pilot power at the following flat transmission level points:
 - (a) Transmitting channel bank or group connector output (2 frequencies),
 - (b) Transmitting group bank output (10 frequencies),
 - (c) Transmitting supergroup bank output (100 frequencies),
 - (d) Receiving high-frequency line output (100 frequencies), and
 - (e) Receiving supergroup output (10 frequencies).

* The group 1 pilot in supergroups 4 and 11.

- (ii) Measurement of absolute power at any low and high band pilot frequency within the range from -30 to -69 dBm.
- (iii) Measurement, on an out-of-service basis, of the conversion loss of group and supergroup banks at pilot frequencies.

Finally, a stable source of any of the pilot frequencies at 0 dBm can be obtained for general purpose terminal measurements.

5.2.3 Receiving Supergroup Scanner

In the scan mode, the group 1 pilot in each receiving supergroup is automatically measured in sequence by the supergroup scanner. Each pilot measurement is compared to the normal value and the difference is displayed. Detection of a pilot failure initiates an alarm. A complete scan of both receiving supergroup banks requires about one and a half minutes. The scan mode can be interrupted and manual measurements made by pushbutton and selector switch operation.

5.2.4 Receiving Group Scanner

Both pilot output power and receiving group amplifier gain are measured in sequence by the receiving group scanner. The duration of one complete scan cycle for the working and spare group banks is about four and a half minutes. In the scan mode this cycle is repeated continually. Since the receiving group amplifiers are regulated, the deviation from normal gain is a measure of the amount of regulation an amplifier is providing. Gain deviations up to ± 4.0 dB are allowed before an end-of-range alarm is initiated. The receiving group scanner can also be operated manually.

5.3 Order-Wire

A 4-channel order-wire facility is part of the SF Terminal. It allows communications over the submarine cable between shore stations without pre-empting commercial circuits for this purpose. Any of these circuits can be interconnected at voice frequency with similar inland facilities to allow, for example, communications between so-called gateway cities. To economize on the use of undersea transmission spectrum, these maintenance channels are transmitted between particular supergroups (See Fig. 1). Steps of "channel," "group," and "supergroup" modulation are used to place two 4-kHz channels in the 8-kHz spacing between supergroups 4 and 5, and two more between supergroups 5 and 6. It was necessary to add a special filter in transmitting channels 2 and 4 to increase band edge discrimination to avoid interference with adjacent

message channels. A modified A5 channel bank provides channel modulation. Group and supergroup functions are accomplished by a unique arrangement of standard and modified multiplex equipment, and some new apparatus designs. Carriers are obtained from the terminal carrier supply.

5.4 Undersea System Monitoring

To maintain near optimum system performance after initial lineup, it is necessary to monitor the behavior of the undersea transmission. One method is through use of supervisory tones. A supervisory oscillator, located in each undersea repeater,¹¹ continuously transmits a tone to a shore terminal. In a given SF System, the frequency associated with each repeater is unique. The tones, spaced nominally 100 Hz apart, lie in two 26-kHz wide bands: 528 to 554 kHz in the low band, and 5894 to 5920 kHz in the high band (Refer to Fig. 1). Adjacent repeaters have tones in opposite bands, so that the frequency sequence along the system is alternately high and low band. An A station receives only high band tones (from every other repeater) and a B station only low band tones. This arrangement is illustrated in Fig. 10(a). The means of launching these tones at the repeater are shown in Fig. 10(b). At the terminals these tones are measured with the supervisory tone test set,

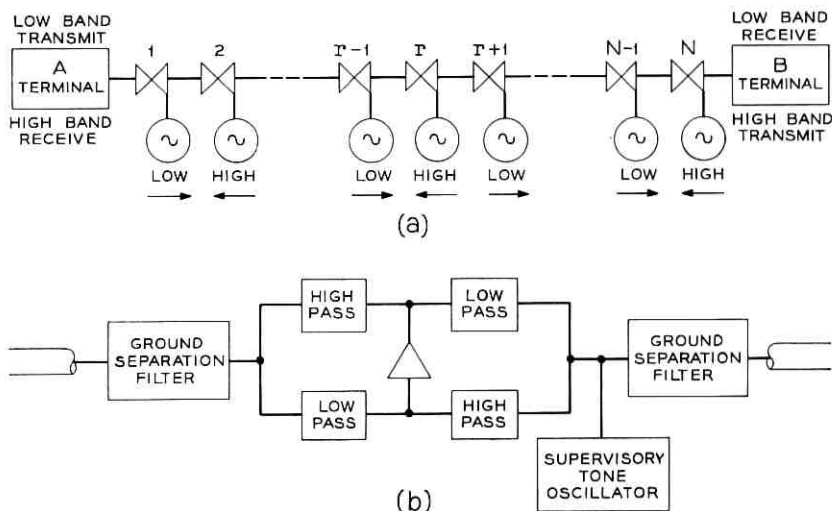


Fig. 10(a)—System layout with respect to supervisory tones.

Fig. 10(b)—Block diagram of repeater showing location of supervisory oscillator.

which will be discussed in some detail in the next section on fault location. Information about the distribution of transmission changes with distance along the cable can be inferred from these measurements.

A second monitoring method, which has been used in the past, is to measure the absolute power of the group reference pilots (at their line frequencies) simultaneously at each terminal physically as close to the ocean cable as practical. The difference at each frequency is approximately the loss of the undersea system at the time of measurement. Pilot measurements are most easily made with the pilot transmission measuring system.

One of the problems that system gain compression introduces is to make difficult the accurate measurement of undersea transmission changes due to causes other than compression. Data obtained from either of the above monitoring techniques will necessarily contain a certain, generally unknown, amount of compression effect. To precisely determine permanent changes from pilot measurements, it would be necessary to remove the system signal load. On the other hand, the amount a particular supervisory tone is affected by compression will, to a first approximation, be linearly proportional to the distance between the terminal and the repeater sending the tone. Thus tones from repeaters relatively close to the shore will be affected least. Fortunately, supervisory tone data from these repeaters are precisely what is needed to determine seasonal transmission changes because they are concentrated in cable relatively close to shore.

5.5 Data Handling

To facilitate recording and disseminating measurement data collected as part of the maintenance operation, a simple data handling scheme was assembled for the SF Terminal. It is composed of data translator, a teletypewriter equipped with tape punch and reader, and a data set. The output data of the pilot transmission measuring system and the supervisory tone test set are made available in binary coded decimal form. The data translator converts this into the ASCII* code acceptable to the teletypewriter, which produces printed page and punched paper tape output. Punched tape can be used to transmit data over the cable to the remote terminal. The *Teletype*[®] tape reader connects via a data set to an order-wire channel. Data can also be transmitted inland over a voice channel if a receiving terminal is suitably equipped.

For cable laying operations this setup is used with the Cable Laying Test Set.³

* American Standard Code for Information Interchange.

VI. FAULT LOCATING

6.1 *General*

If a fault occurs in the undersea system, it is essential that it be located accurately and quickly. The required accuracy is dependent on repair technology. In shallow water, the accuracy objective is about plus or minus one mile. The critical factor here is to determine whether repair is feasible without disturbing a repeater. In deep water, it is desirable to determine the location to within a repeater spacing at least. The time required to locate the fault should be short enough to avoid delaying the dispatch of a repair ship.

Three distinct methods are available for locating faults in the undersea portion of an SF System. The special test equipments designed to exploit each method are:

- (i) supervisory tone test set,
- (ii) repeater locating test set, and
- (iii) cable fault localization test set.

One distinction among these sets is that the first two can be used only when the system is powered, the last only when the system is not powered. At best, the first two can locate a fault to within a repeater section, because their measurements are specifically associated with individual repeaters. Other observations, however, such as a change in received noise power at one terminal, can provide further information that may allow a more specific fault location estimate to be made.

6.2 *Supervisory Tone Test Set*

The supervisory tone test set is a special purpose selective detector that can accurately measure the absolute power and frequency of individual tones in the supervisory bands. Frequency selectivity is about 25 Hz, and dynamic range is between -30 and -80 dBm. With external calibration the accuracy is ± 0.1 dB and at least ± 1 Hz. One of the difficulties normally encountered in tuning such a narrow-band detector is that of finding and holding the measured tone at the precise "bottom" of the selective filter. To solve this problem a phase-locked feedback loop was incorporated around the filter. In operation it is necessary only to tune to the neighborhood of a particular tone (about ± 5 Hz). The phase-lock will then automatically seek and maintain fine tuning.

Recall that in addition to providing fault locating information, supervisory tone measurements are used as a system maintenance

tool and, during cable laying operations, as an aid in determining transmission misalignment as a function of distance.^{2,3} Actually, supervisory tone measurements can only locate a fault unambiguously to within two adjacent repeaters and the intervening cable [See Figs. 10(a) and (b) and note that, relative to the bands in which they operate, high band oscillators are connected at the repeater inputs, whereas low band oscillators are connected at the outputs.] As an example, postulate a transmission fault that does not affect the ability to power the system. The A station would receive tones only from repeaters 2, 4, \dots , r , while the B station would receive only from repeaters $(r + 1)$, $(r + 3)$, \dots , $(N - 1)$. In this case the fault is probably in repeater $(r + 1)$ or the cable between r and $(r + 1)$. However, the fault could be in the ground separation filter or coupling of repeater r . This example is an oversimplification because it assumes loss of transmission in each direction beyond the fault. It is possible to obtain high resistance faults between the cable center conductor and sea water that have virtually no effect on transmission loss, but do generate considerable noise unless the dc voltage at the fault is adjusted to zero.* In this case, supervisory tone measurements are of little use. Fortunately, for faults of this type, the repeater locating test set becomes a very useful tool.

6.3 Repeater Locating Test Set

The repeater locating test set identifies repeaters by using their slight and normally undesirable nonlinearity. In response to a transmitted signal a return modulation product is obtained from each repeater. The mode of operation depends on whether operation is from an A or B station. At an A station, periodic bursts of a particular low band frequency are applied to the system. Bursts of second harmonic produced in each repeater fall in the high band, return to the A station, and are detected. The unique round trip delay to each repeater allows the returned signal from individual repeaters to be identified. At a B station, which transmits in the high band, a continuous tone as well as periodic bursts of another high band frequency are applied to the system. Bursts of the difference modulation product fall in the low band and return to the B station. All signals generated and detected by the test set fall in the standard group band of 60 to 108 kHz; terminal multiplex and high-frequency line equipment are used to translate these signals to and from line frequency.

The principal feature of the receiving portion of the set is a com-

* On a system with a power plant at each terminal, zero fault voltage is obtained if each plant can be adjusted to supply equal current.

mercial digital-memory oscilloscope. It can achieve a significant improvement in signal-to-noise ratio over conventional detection techniques through use of its computation and data storage capabilities. The results of a large number of repetitive measurements are automatically added and stored; the wanted signal adds systematically from sample to sample whereas the background noise adds on a random basis. When a measurement sequence is complete, the results are printed out in digital form. When the additional signal-to-noise advantage is not required, the detected signals can be directly displayed on a conventional oscilloscope. Two such displays are pictured in Fig. 11. The amplitude of a return pulse from any given repeater is a function of the nonlinearity of that repeater and both low and high band misalignments between it and shore. For example, the large decrease in pulse height between repeaters 11 and 12 in Fig. 11(a) is explained by the fact that about 900 meters of extra cable were added between these repeaters for purpose of equalization.

If a system fault occurs that does not affect the ability to maintain power, or if the system can be powered to the fault, then location to within a repeater spacing (two repeater ambiguity) is obtained by observing the presence or absence of return pulses, or, in some instances, amplitude changes from previous measurements.

One of the most difficult faults to locate is the relatively high im-

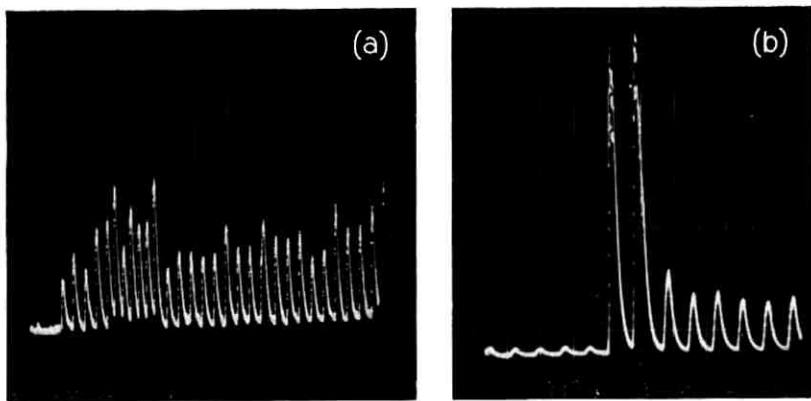


Fig. 11(a) and (b)—Repeater locating test set displays: (a) return pulses from first 30 repeaters out from the Jacksonville Beach, Florida, terminal; (b) a portion of the return pulses during a fault condition. Shore supplying normal current of 136 mA but ship receiving only 56 mA. Increase in modulation of current starved repeaters reveals location of fault to within a repeater spacing (two repeater ambiguity).

pedance shunt type, described in the previous section, which has little effect on transmission loss. A fault finding technique that proved very effective when such a trouble developed during installation of the new Florida-St. Thomas system was to adjust the power plants at each end of the system (in this case ship and shore) so that repeaters on one side of the fault were receiving substantially less current than those on the other. The increase in nonlinearity and decrease in gain of the current-starved repeaters was detected as shown in Fig. 11(b).

6.4 Cable Fault Localization Test Set

Historically, the most frequent cause of lost service time on submarine cable systems has been cable breaks resulting from fishing activity. It will usually be possible to power an SF System to the fault for most cable breaks and, using the repeater locating test set, locate the fault to within one repeater section. If it is not possible to power the system to the fault, the cable fault localization test set can be used.

Basically, this test set utilizes a complex artificial line that simulates the repeated cable at very low frequencies (≤ 1 Hz). Under fault conditions, the artificial cable is compared to the actual faulted cable in a bridge configuration. Provision is made for the insertion of a complex fault impedance (resistance and capacitance in parallel) at points along the artificial line until optimum balance conditions are obtained. The major difficulty in this method of fault location is that it is quite difficult to distinguish between the magnitude of the fault impedance and equivalent cable length; that is, similar bridge balances can be obtained at different points along the cable by varying the simulated fault impedance. This uncertainty can be reduced by testing at different low frequency currents and analyzing the resulting data with empirically-determined relationships that tend to separate fault impedance and cable parameters.

VII. PHYSICAL DESIGN OF TERMINAL EQUIPMENT

7.1 General

From the point of view of physical design, terminal equipment is separable into two categories:

- (i) duct type bay construction and
- (ii) cabinet construction.

The power separation filter and fault locating test sets are housed in 1.95-meter-high cabinets. This type of construction is illustrated in Fig.

12 which shows the cable fault localization test set. All operating controls and indicators are accessible from the front. Components are mounted in drawers that slide out individually to allow access for maintenance purposes.

The remaining terminal transmission equipment is mounted in 2.74-meter-high (9 feet), duct-type bays. Although not standard to

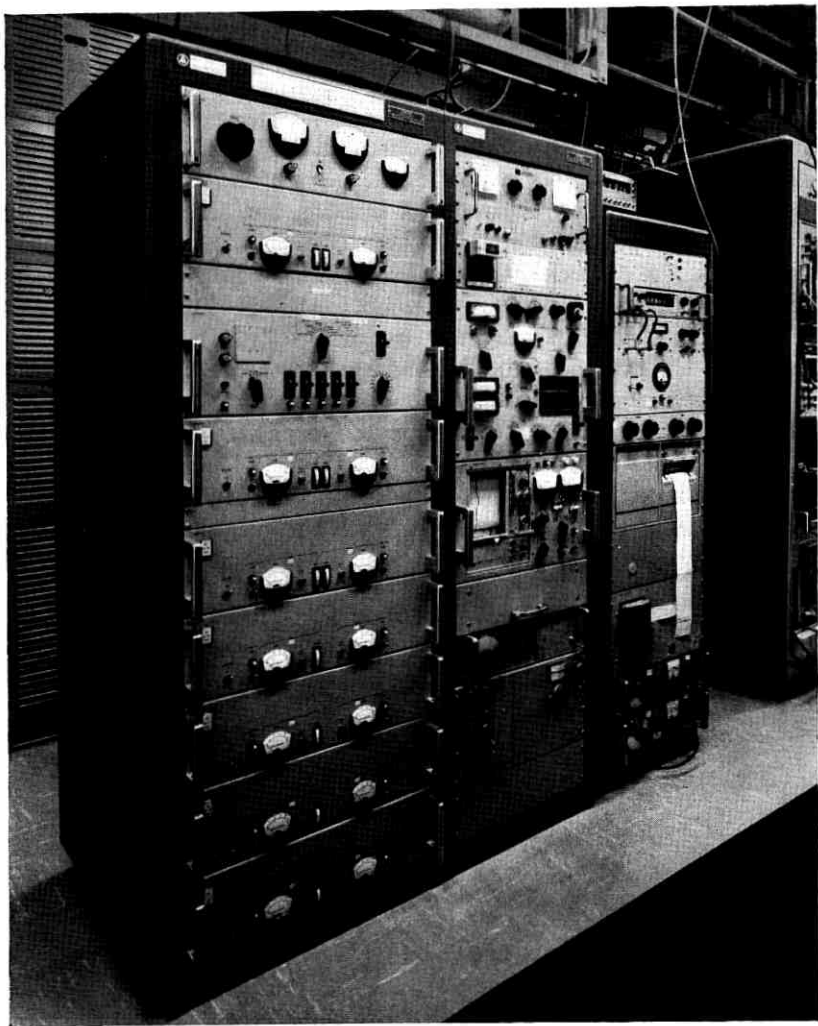


Fig. 12—Cable fault localization test set.

the Bell System, this height was chosen because it is more compatible with foreign equipment design practice and terminal buildings. The mechanical arrangement within the bays consists mainly of shelf-type units rigidly attached to the ducts, which contain the intershelf and interbay cabling. Apparatus of a passive nature are typically mounted on slide out units that rest on the shelves. Circuitry containing active devices is usually packaged in plug-in modules. In this way, ease of access is generally matched to the likelihood that maintenance will be required.

7.2 Bay Enclosures

The appearance of SF Terminal equipment has been a matter of more than usual concern, primarily because of the international character of submarine cable installations. It was recognized early in the design stage that harmonious appearance would be difficult to achieve because existing equipment of differing physical designs as well as equipment of new and modified designs would be included in the terminal. As a solution, a unique enclosure scheme was developed for use with duct-type bays to unify their appearance. A four bay lineup is pictured in Fig. 13. The front of the enclosure consists of two pairs of side-hinged doors for each bay in the lineup. When the doors are opened they can be slid into the area between ducts at the side of each bay. The upper and lower doors are separated by a rigid supplementary bay framework on which closed-door access panels are mounted. When the doors are in their normally closed position, they are flush with the panels, presenting a smooth uncluttered line. As the name implies, these closed-door access panels provide easy visual and physical access to fuses, alarm lamps, meters, keys, and test jacks required for routine operation of the system. The doors and access panels have a light gray textured vinyl finish. Surrounding this area is an interbay cabling trough at the top, a guard rail at the bottom, and end panels at the sides. The dark blue-gray textured finish of these items results in a picture frame effect.

VIII. ACKNOWLEDGMENTS

The design of the SF Terminal and fault locating equipment is the result of the efforts of many individuals within the Bell Telephone Laboratories. In addition to these individuals directly responsible for the overall designs, whose contributions are gratefully acknowledged here, mention should be made of the considerable design support provided by personnel of the network and multiplex areas.

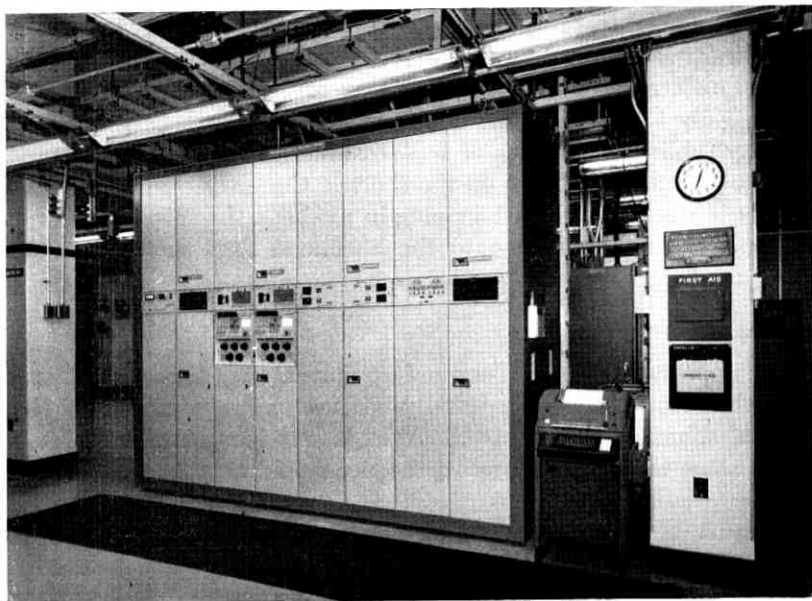


Fig. 13—Four bay lineup of transmission equipment showing the enclosure scheme.

REFERENCES

1. Calkin, E. T., and Golio, I., "Power Conversion," B.S.T.J., this issue, pp. 749-765.
2. Anderson, C. D., and Easton, R. L., "An Overview: Requirements and Performance," B.S.T.J., this issue, pp. 605-630.
3. Hirt, W. B., and Oldfather, D. O., "Transmission Tests, Computations and Equalization During Installation," B.S.T.J., this issue, pp. 783-798.
4. Law, H. B., Reynolds, J., and Simpson, W. G., "Channel Equipment Design for Economy of Bandwidth," *The Post Office Electrical Engineers Journal*, 53, part 2 (July 1960), pp. 112-117.
5. Tucker, R. S., "16-Channel Banks for Submarine Cable," *Bell Laboratories Record*, 38, No. 7 (July 1960), p. 248-252.
6. Graham, R. S., Adams, W. E., Powers, R. E., and Bies, F. R., "New Group and Supergroup Terminals for L Multiplex," B.S.T.J., 42, No. 2 (March 1963), pp. 223-278.
7. Albert, W. G., Evans, J. B., Jr., Ginty, J. J., and Harley, J. B., "Carrier Supplies for L-Type Multiplex," B.S.T.J., 42, No. 2 (March 1963), pp. 279-317.
8. Clark, O. P., Drazy, E. J., and Weller, D. C., "A Phase-Locked Primary Frequency Supply for L Multiplex," B.S.T.J., 42, No. 2 (March 1963), pp. 319-340.
9. Garrison, J. L., Labbe, L. P., and Rock, C. C., "Basic and Regulating Repeaters," B.S.T.J., 48, No. 4 (April 1969), pp. 841-889.
10. Members of the Technical Staff, Bell Telephone Laboratories, *Transmission Systems for Communications*, Murray Hill, New Jersey: Bell Laboratories, Inc., Fourth Edition, 1970, Chapters 13 and 14.
11. Buus, R. G., Kassig, J. J., and Yeisley, P. A., "Repeater and Equalizer Design," B.S.T.J., this issue, pp. 631-661.

Power Conversion

By E. T. CALKIN and I. GOLIOTO

(Manuscript received September 22, 1969)

This article describes the constant current, high voltage power facilities which power the SF Submarine Cable repeaters. The overall power system description includes discussions of the power connections to the cable, and a review of overall design requirements and objectives. The circuit description is covered in a general manner to demonstrate the basic concept. Particular emphasis is given to the description of several significant circuit innovations which include automatic load sharing by two constant current sources, 20 kHz inverter operation with output wave symmetry correction, solid-state alarm detectors, automatic turn-up and turn-down features, electrical noise suppression, redundant shutdown circuits and selective alarm cutoffs.

The physical design includes the extensive use of plug-in modules for ease of maintenance, convection cooling of the entire power supply including the test load, and a key interlock system which prevents access to high voltage components when the power supply is energized. A special high-voltage switch can transfer the cable from one power plant to another without interruption of service.

I. INTRODUCTION

At 7:15 A.M. on 8 August 1968, a technician pushed a button on a power supply in the Jacksonville Beach cable station. Within seconds another technician pushed a button on a power supply in the Magens Bay cable station. One minute later the first units of a new generation of submarine cable power supplies were in service powering the Florida to St. Thomas SF Submarine Cable System.¹

The direct current needed to power SF Submarine Cable repeaters² is carried to the repeaters by the coaxial cable's center conductor. The current originates in shore-based power supplies and is returned from the far cable end through sea (ocean) ground. Since the repeaters are connected in series, the current requirement is the same for all SF

Systems, but the end-to-end cable voltage drop is a function of system length.

The new submarine cable power supplies which went into service on 8 August 1968 are direct current sources designed to accommodate a wide range of voltage loads. These supplies are powered from duplicated -48 volt battery plants. The -48 volt battery-charging rectifiers are powered from commercial power lines. The interposition of batteries between the commercial power lines and the SF Submarine Cable power supply assures continuous power availability in the event of a commercial power failure. In addition, the provision of automatically or manually started engine or turbine driven alternators extends almost indefinitely the time period over which a commercial power failure may be tolerated.

Having provided for a reliable source of continuous input power, the SF Submarine Cable power supply designers have duplicated major power units within the power supply wherever practical and have specified highly reliable components for use where duplication is impractical. In doing this they have drawn heavily on the techniques used by the SB and SD Submarine Cable power supply designers.^{3,4} In addition, they have taken advantage of advances in the state of the art and incorporated many innovations which result in new ease of operation and maintenance in a smaller package.

II. DESIGN REQUIREMENTS

The power connections to an SF Submarine Cable System are shown in Fig. 1. Power supplies of opposite polarity are connected at each end to reduce, by half, the maximum voltage stress applied to the repeaters near to the cable ends. On systems short enough to require less than half the end-to-end cable voltage required for a maximum length system, the cable can be powered from one end only, or from both ends. If a short system is powered from both ends, then the power supply at either end has sufficient voltage reserve to power the entire cable. In addition, if a high impedance shunt fault develops in the cable in a short system powered from both ends, the system can be continued in service until repair is made. This is done by adjusting the load sharing between the power supplies until the voltage at the fault is zero.

The electrical design requirements placed on a power supply for one end of an SF Submarine Cable System are shown in Table I. The SF Submarine Cable power supply normally functions as a well-regulated current source. Under appropriate conditions, the supply shifts auto-

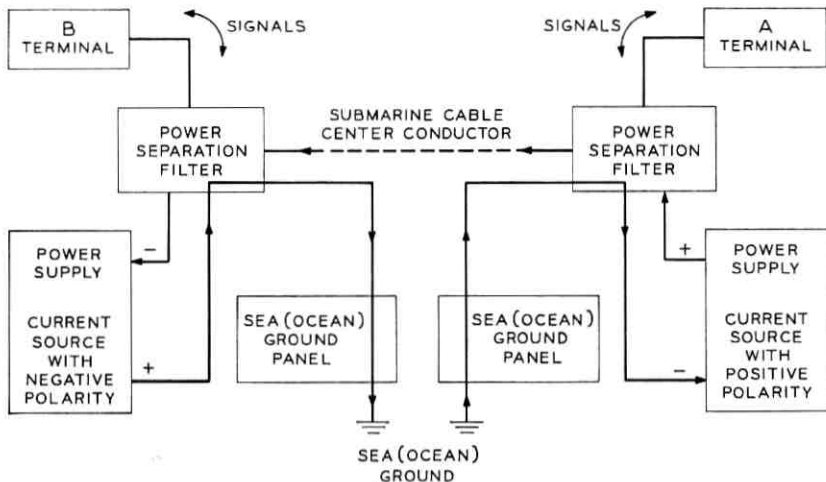


Fig. 1—Power connections to the SF Submarine Cable System.

matically to a voltage-limiting mode of operation to prevent excessive output voltages. The power supply is also required to have monitoring circuits which will warn of current or voltage abnormalities and will shut down either or both of its redundant current sources in the event of excessive voltage or current.

III. POWER SUPPLY ELECTRICAL SYSTEM

Figure 2 displays a complete power supply system as it would appear on one end of a cable system. Power continuity is achieved by using reliable, independent current sources which share the voltage load. The redundant current sources are the series connected inverter-rectifier 1 and inverter-rectifier 2. Either will automatically assume the entire voltage load should the other shutdown for any reason. Semiconductor diodes bypass the cable current between the output terminals of the shutdown inverter-rectifier current source.

Nonredundant paths achieve high reliability by use of highest quality, conservatively dc rated components.

Routine maintenance and repairs are facilitated by the inclusion of a second power plant in the system as a standby and by the provision of means of transferring the cable current load from the "in-service" to the standby plant without disrupting service. In Fig. 2, either power plant I or power plant II can be the standby plant. The load

TABLE I—SF SUBMARINE CABLE POWER SYSTEM ELECTRICAL REQUIREMENTS

Normal current	0.136 ampere
Nominal voltage	3500 volts
Maximum voltage (at normal current)	4500 volts
Maximum single frequency tone	-120 dBm at 500 kHz
(appearing at power separation filter transmission terminals)	-130 dBm at 6000 kHz
Normal accuracy of current control	±0.5 mA, absolute
Load current regulation:	
1000 volts change	±1.0%
one rectifier failure	-0.5%
Output resistance above 4500 volts	2500 ohms
External alarms, power plant output:	
Minor alarm—current	±2%
Minor alarm—voltage	±8%
Major alarm—current	±5%
Major alarm—voltage	±15%
External alarms, each rectifier output:	
Minor alarm—current	-5%
Minor alarm—voltage	200 to 500 volts
Major alarm—current	+5%
Major alarm—voltage	4975 volts
Protective shutdowns:	
Each rectifier	+5% current
	4975 volts (electronic)
	6500 volts (spark gap)
	6500 volts (spark gap)
Plant	

transfer provides the means of performing a "hot" transfer of cable load between the power plants, and also provides an adjustable test load to aid in servicing the standby power plant.

3.1 Power Separation Filter

The power separation filter is a separate bay in which the transmission signals and the power supply current are jointly applied to the undersea cable. Included in the power separation filter is a means of routing the current from the power supply to either the cable or an auxiliary load. This permits the testing of the power system up to the point of the cable connection before the cable is installed.

3.2 Power Plant Monitors

Each power plant contains a variety of circuits and devices which monitor the plant's output. These include meters, recorders, voltage and current alarms, overvoltage current droop circuits, and overvoltage shutdown circuits. These circuits monitor the combined output of the two inverter-rectifier current sources.

3.3 Inverter Rectifiers

The inverter-rectifier electrical system is illustrated in Fig. 3. The principle function of the inverter-rectifier is to produce well-regulated current at any voltage up to a maximum of 4500 volts. The conversion of the input battery power into a regulated current with a large voltage compliance is provided by the 20 kHz inverter, the regulator, and the high voltage rectifier. The inherent constant current characteristic of the saturable reactor is enhanced by a closed loop regulation system which reduces output current drift to the low level required of this power supply.

3.4 Load Sharing Between Inverter-Rectifiers

The requirement of low drift of output current results from the independent regulation of the two series-connected inverter-rectifiers and the large value of slope resistance connected across each inverter-rectifier's output. The slope resistance is selected at approximately the lowest value consistent with the "one rectifier failure" regulation requirement listed in Table I. The difference in output current drift

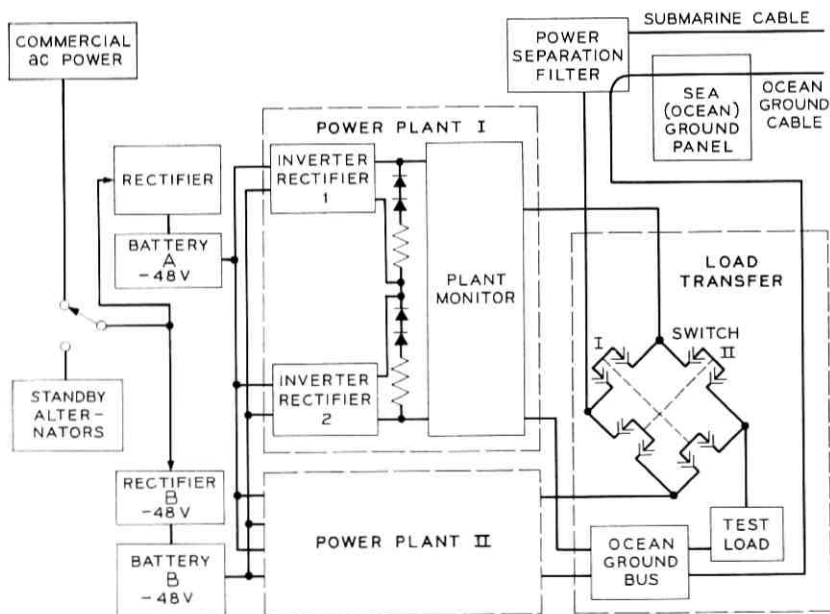


Fig. 2—SF Submarine Cable power system.

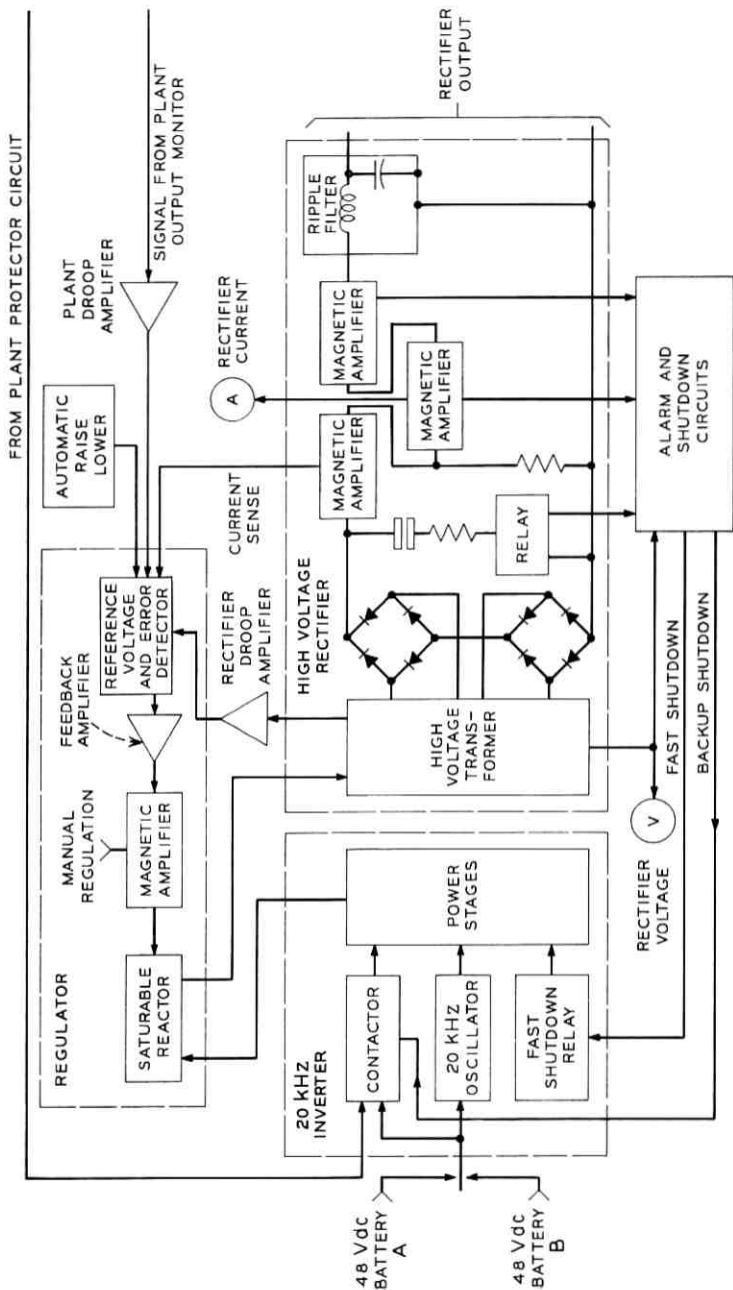


Fig. 3—SF Submarine Cable power supply inverter—rectifier system.

between the two inverter rectifiers must be less than 0.0215 percent if the drift in load sharing between inverter-rectifiers is to be less than fifty volts. (In the SD power supply, the equivalent figures were 0.025 percent current drift for a 75 volt load sharing drift—see pp. 1347–1348 of Ref. 4.) Both SD and SF have proven stable with respect to voltage sharing between rectifiers.

Normally the inverter-rectifiers are manually adjusted to equally share the load voltage. Equal load voltage sharing is essential neither to normal cable power nor to the automatic assumption of the full load voltage by one inverter-rectifier if the other should fail. Equal load sharing does, however, minimize the output voltage and current excursions which occur when one inverter-rectifier is shut down.

3.5 Load Sharing Between Cable Ends

Load sharing between power supplies on each end of a long cable is possible only because of the low output drift and the inclusion of a finite fixed slope which is introduced into each power supply by the slope resistances connected across their outputs. A lower slope resistance is used across the plant output because the "1000 volts change" regulation requirement listed on Table I is less severe than the "one rectifier failure" requirement. The lower slope resistance facilitates load sharing between power supplies. Load sharing between power supplies on a long system is essential since neither can power the complete system. Equal load sharing between power supplies on a short system powered from both ends is desirable from a monitoring point of view but not essential to system operation.

IV. POWER SUPPLY PHYSICAL SYSTEM

An SF Submarine Cable power supply is composed of either:

- (i) Two power plant bays and a load transfer bay as shown schematically in Fig. 2; or
- (ii) One power plant bay and a load transfer bay.

Figure 4 is a photograph of the two bays in the "one-power-plant" system. These bays are presently installed in the Jacksonville Beach cable station. A similar power supply with opposite polarity is installed on the other end of the system in the Magens Bay cable station.

A power plant bay is equipped with one plant monitor and two inverter-rectifier pullout units which house low voltage circuits. These are displayed in Fig. 5. Their associated high voltage components are mounted in the rear portion of the bay. The high voltage section is



Fig. 4—Single power plant SF Submarine Cable power supply.

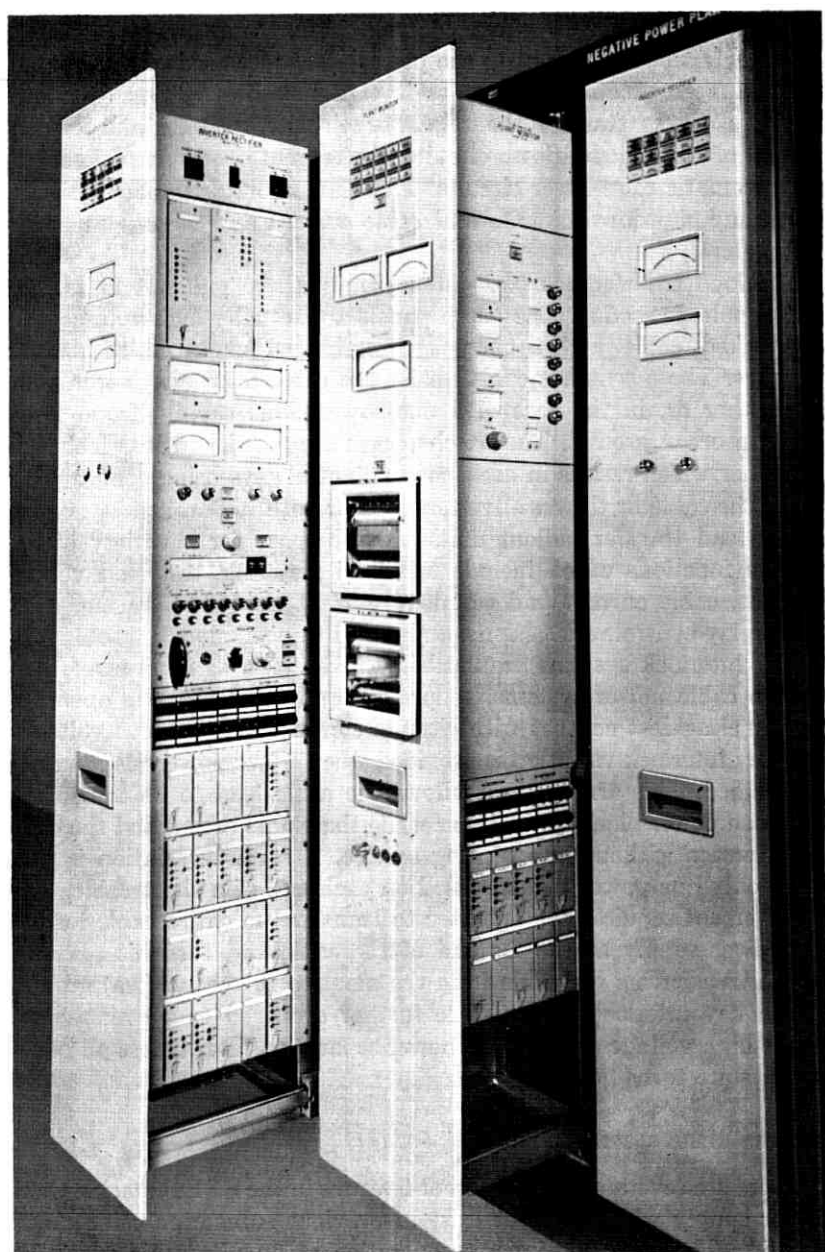


Fig. 5—SF Submarine Cable power plant pull-out units.

interlocked with power plant controls in a manner which permits access only when the input to the power plant is turned off.

The inverter-rectifier pullout unit contains the inverter (oscillator and two power stages), the regulator circuits, voltage and current meters, alarm and shutdown circuits, the current droop amplifiers, and the automatic raise-lower circuits. The plant monitor pullout unit has alarm and shutdown circuits and cable current and voltage recorders and meters.

The load transfer has two patch panels, three pull out units, and a high voltage section. The patch panels are actually high-voltage key-interlocked cable and connector assemblies that serve as manual switches. These switches can transfer the cable from one power plant to another or, in the case of a one-power-plant supply, from power plant to ocean ground. The patch panels are wired and interlocked in a manner which results in each power plant always being connected to either the cable or the resistive test load or both in parallel.

The load transfer pullout units provide space for a variety of low voltage functions which include test load metering, test load voltage controls, cable current and test load current meter calibrating, cable short relay controls and alarms, and alarm transmission.

Key interlock systems are used in this system, as in previous submarine cable power systems, to provide protection for the operating personnel against contact with circuits containing hazardous voltages. One key interlock variation is provided when a two-power-plant supply is specified. This arrangement allows the cable load to be transferred from one power plant to the other while the plants are on and the cable is in service without interrupting service. A second variation is used when only one power plant is used on each end of a short cable. This arrangement requires the operator to transfer the entire cable load to the power supply at the far end of the cable and turn his plant off before transferring the cable from the plant to the ocean (sea) ground.

The key interlock systems are further arranged to prevent access to the high voltage areas in the power separation filter unless all power supplies are turned off.

V. NEW ELECTRICAL FEATURES

There are several new or different features in the SF Submarine Cable power supply which did not exist in previous submarine cable power supplies. These include inverter operation at an ultrasonic frequency (20 kHz), an automatic output current raise-lower function on each

inverter-rectifier, automatic load sharing between power stages in an inverter-rectifier, output waveform symmetry correction in the inverter's oscillator-buffer circuit, solid-state alarm level detectors and selective remote alarm cutoff.

5.1 *Inverter Frequency and Power Supply Electrical Noise Suppression*

Direct voltage inversion to alternating voltage was introduced to submarine cable power supplies in the SD Submarine Cable power supply as a necessary adjunct to the use of a saturable reactor as a constant current source. A saturable reactor designed to work at 60 Hz would have been impractically large (see p. 1345 of Ref. 4). The state of the development of power semiconductors and broad experience then current in 400 Hz circuits led the SD Submarine Cable power supply designers to select an inverter frequency of 400 Hz even though they believed "higher frequencies would have been more suitable from a size standpoint" (see p. 1346 of Ref. 4).

The availability of improved power semiconductors and a dissatisfaction with the audible noise levels produced by the magnetic component cores of the SD Submarine Cable power supply led to the selection of 20 kHz as a nominal inverter operating frequency for the SF Submarine Cable power supply. The 20 kHz operating frequency and the reduced SF maximum output power requirement (approximately one-quarter of that required for SD) have enabled the SF physical designers to eliminate the separate inverter bay used in the SD Submarine Cable power supply by installing the SF inverter as one section of the inverter-rectifier pullout unit.

Operation of the SF inverter at 20 kHz has eliminated the audible noise problem but increased the electrical interference problem by moving the inverter's fundamental frequency of operation relatively closer to the lowest transmission frequency. Both the SD and SF inverters generate quasi-rectangular current and voltage waves. The 100 kHz lower frequency limit of the SD transmission band is the 250th harmonic of 400 Hz, but the 500 kHz lower frequency limit of the SF transmission band is only the 25th harmonic of 20 kHz. A further problem arises from the fact that the upper limit of the SF transmission is 6 MHz against the SD upper limit of 1.1 MHz. The net effect is to require better filtering and electrical noise suppression in the SF power systems than was necessary in previous submarine cable power systems.

Several new features are included in the SF power system's ripple and electrical noise suppression system. These include:

(i) The location of the power separation filter in a separate equipment bay which is installed at a distance from the power supply to reduce electrical noise which may result from direct electromagnetic radiation from the power supply;

(ii) The installation of new design, low pass filter networks in the output of each high voltage rectifier, in the power plant output and in the power separation filter;

(iii) The extensive use of shielded and coaxial leads throughout the power system; and

(iv) The extensive use of capacitors to by-pass longitudinal noise signals in the various input and output leads, other than high voltage leads, connecting to the power supply.

5.2 *Automatic Turn-up and Turn-down Circuits*

A major new operational feature introduced to submarine cable power supplies is the automatic raise-lower circuit installed in each inverter-rectifier. In the SD Submarine Cable systems, power is turned up by manually rotating the rectifier output current adjustment until the desired current level is achieved. On a long system, there are two "in-service" rectifiers on each end, or a total of four current adjustments which must be turned up.

The SF Cable system power may be turned up manually as is the SD Cable system, but SF is normally turned up automatically. In the automatic turn-up procedures, the power supplies on each end are manually preadjusted into their test loads. The adjustments are made with the automatic raise-lower circuits turned off. By doing the tedious and exacting initial adjustment on the test load when there is neither urgency nor danger of system damage, less pressure is placed on the operator. At the moment of the most tension, that is when the cable is being powered for the first time, the operators have nothing to do except push the cable current raise button and watch meter indications.

The automatic raise-lower function is achieved by means of a solid-state voltage ramp circuit which connects into the regulating control loop as illustrated in Fig. 3. The regulating control loop is a direct current amplifier system with negative feedback. The inverter-rectifier's output current is directly proportional to a reference voltage. The automatic raise-lower circuits achieve a smooth, automatic variation in the inverter-rectifier's output current by superimposing a voltage ramp on the reference voltage.

5.3 Load Sharing Between Inverter Power Stages

Another new feature is automatic load sharing between the two power stages within each SF power plant inverter. (This automatic sharing is not to be confused with the manually controlled load sharing between the virtually independent inverter-rectifiers.) The automatic load sharing is accomplished in a new circuit,⁵ the functioning of which depends on multiple gate windings on the power saturable reactor and corresponding multiple cores and primary windings on the high voltage transformer. The essential details of the circuit are shown in Fig. 6. The two gate windings are identical and the transformer's two primary windings are identical and wound on identical cores. A detailed description of the operation of this circuit is found in Ref. 5.

5.4 Inverter Switching Signal Symmetry

The inverter is composed of two power stages and an oscillator-buffer stage. The latter provides switching signals to the power stages. It is desirable that the switching signals to the power stages be symmetrical with respect to time, that is, the negative half-cycle which turns a power stage transistor off should be equal in length to the positive half-cycle which turns the same transistor on. The inherent oscillator output waveform was not as symmetrical as desired, due principally

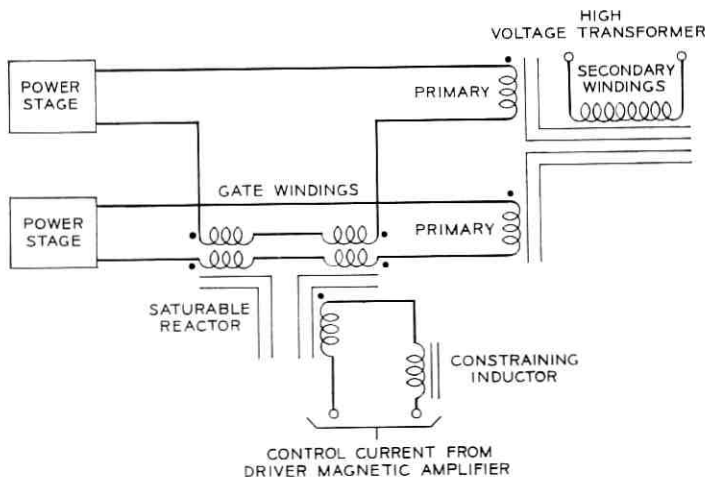


Fig. 6—Power stage forced load sharing circuit.

to differences in switching times between the two oscillator transistors. The symmetry is greatly improved by the circuit shown in Fig. 7.

This figure shows essential elements with the buffer stage, starting networks, and shaping networks omitted. A simplified description of the symmetry correction network circuit operation is as follows:

- (i) Equal voltages are assumed across C1 and C2.
- (ii) Assume that Q1 tends to conduct longer than Q2.
- (iii) When the on-time of Q1 equals the total on-time of Q2, then L1 saturates and a current flows in winding 1-2 of T1 in such a direction as to turn Q1 off, thereby correcting the tendency of Q1 to be on longer than Q2.

An expanded description of the circuit operation is given in Ref. 6.

5.5 Alarm-Level Detectors

The current alarm-level detectors are semiconductor circuits which can be adjusted to detect over or under currents to within a few tenths of a percent of the desired alarm level. A single detector circuit can detect both over and under current levels from a single signal current. The semiconductor level detector circuits permit easier adjustment than that provided by the meter type relay used in earlier voltage and current alarms. Furthermore, shutdowns can be tested and circuits calibrated while the power supply is connected to and powering the cable.

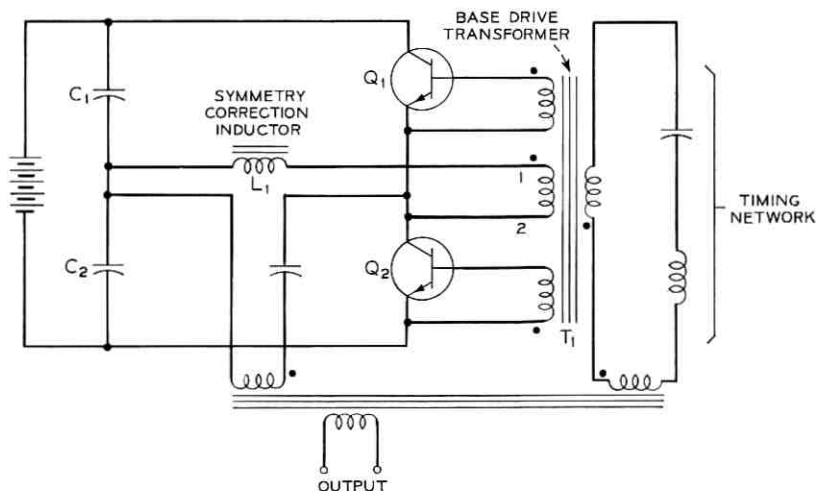


Fig. 7—Simplified SF oscillator circuit with a symmetry correction network.

5.6 *Selective Alarm Cutoff*

Still another new alarm feature is the selective cutoff of remote alarm transmission. There is only one alarm cutoff push button, but operation of this push button will cutoff the remote transmission of those alarms which exist at the instant the alarm cutoff push button is operated. If a new alarm condition appears, it is remotely transmitted and it is necessary to again operate the alarm cutoff push button to silence the remote alarm.

VI. NEW PHYSICAL FEATURES

The SF Submarine Cable power supply is the first submarine cable power supply to make extensive use of plug-in modules. The majority of the components in the power supply other than high voltage items, are mounted on 54 plug-in modules. The number of different types of plug-in modules is minimized by putting only the common features of similar circuits on a plug-in module. Hence, the 54 plug-in modules represent only 17 different types.

Two advantages of this commonality result. First, component density is increased, permitting smaller overall power supply dimensions than would otherwise be possible. Second, the small number of different types permits the storage of at least one spare of each type in the power supply. The presence of these spares facilitates maintenance procedures and repairs.

6.1 *Precision Calibrator and Test Load Location*

The separate equipment bay that housed the test load circuits and the precision current meter calibration circuit (see p. 1360 of Ref. 4) in the SD Submarine Cable power supply, are not needed in the SF Submarine Cable power supply. Both the calibration circuit and the test load circuits are installed in the SF load transfer bay.

The SF precision current meter calibration circuit is similar (except for shunt resistances) to that used in the SD Submarine Cable power supply.

The SF test load maximum power dissipation is approximately one-quarter of that required for the SD test load. Hence, not only is the SF test load located in the load transfer bay, but it is cooled by natural convection rather than by the fan driven air circulation.

6.2 *Natural Convection Cooling*

It has been possible to dispense with cooling fans in the SF Submarine Cable power plant despite the location of the inverter's power stage and

oscillator modules in the inverter-rectifier pullout unit. The lower power rating of the SF power supply (relative to the SD power supply) and the physical design of the equipment (to take advantage of natural convection) make this possible. The principal heat producers, the inverter's oscillator and power stage plug in modules, are located at the top of the inverter-rectifier pullout unit. An opening in the front panel, just below the inverter modules, allows air to come in and flow through the inverter modules before exiting through the top of the bay. This arrangement keeps the rest of the equipment at or near room temperature.

VII. SHIPBOARD POWER SUPPLY

The shipboard power supply (see pp. 1364-65 of Ref. 4) formerly designed and used for SD Cable laying and repair operations, now has been modified to provide power for SF Submarine Cable systems. The modification consisted of the installation of magnetic amplifier and meter options.

VIII. SUMMARY

The SF Cable power supply contains many design innovations. Inverter operation at 20 kHz and convection cooling result in silent operation in a smaller plant. These features introduced several challenging problems in the area of electromagnetic noise suppression and magnetic apparatus design.

The sophisticated use of semiconductors in combination with relays permits stability, small size and reliability in the alarm and protection systems. In addition, automatic turn-up and turn-down circuits and modular physical design contribute to new ease of operation and maintenance.

IX. ACKNOWLEDGMENTS

The SF Submarine Cable power supply design is the result of the team effort of over twenty circuit, physical, and apparatus designers. We particularly acknowledge the guiding and co-ordinating roles of the following persons: circuit design—Mr. J. D. Bishop; physical design—Mr. S. Mottel; and magnetic component design—Messrs. T. G. Blanchard and B. E. Stevens.

REFERENCES

1. Anderson, C. D., and Easton, R. L., "An Overview: Requirements and Performance," *B.S.T.J.*, this issue, pp. 605-630.
2. Buus, R. G., Kassig, J. J., and Yeisley, P. A., "Repeater and Equalizer Design," *B.S.T.J.*, this issue, pp. 631-661.
3. Meszaros, G. W., and Spencer, H. H., "Power Feed Equipment for the North Atlantic Link," *B.S.T.J.*, 36, No. 1 (January 1957), pp. 139-162.
4. Bishop, J. D., and Mottel, S., "Cable Power Facility," *B.S.T.J.*, 43, No. 4, Part 1 (July 1964), pp. 1339-1366.
5. Kadri, F. V., "A Reliable High-Powered Constant Current Supply," *IEEE Transactions on Magnetics*, *MAG-3*, No. 1 (March 1967), pp. 22-28.
6. Ianniello, J. W., "Symmetry Correction for a Free-Running Transistor Inverter," *IEEE Transactions on Industry and General Applications*, *IGA-5*, No. 1 (January/February 1969), pp. 68-73.

Submarine Cable Route Engineering

By R. C. ALLEN, D. H. ORT, and L. M. SCHINDEL

(Manuscript received May 12, 1969)

The objective of route engineering in submarine cable work is to select a path for the proposed cable which is both economical and highly reliable from the standpoint of system integrity. The engineering effort begins with selection of the terminal sites and is carried through to a complete route layout. Echo sounding is the principal technique used for deep water route investigations. In shallow water where extraordinary protection measures such as plowing-in of cable may be required, special survey techniques are used to evaluate a route. These latter techniques include side-scan sonar, subbottom profiling, and the application of a Bell Telephone Laboratories developed survey sled.

I. INTRODUCTION

The need for a submarine cable link develops as a result of anticipated or existing demands for communication services. Several factors enter into a general route plan. Among these are: (i) system fit—that is, how the submarine cable system will be integrated into existing and planned networks, (ii) economics, in a first-cost sense, where the rule is that the shortest cable distance between two points is the most economical, and (iii) system reliability, where communications integrity is a prime factor. Economics also plays a part in system reliability in that repair costs above normal system maintenance costs must be considered.

Route engineering of the system begins almost as soon as the communications demand is anticipated. System fit locates the end points, or terminal stations, of the cable system within somewhat restricted geographical limits. Once the general locations of the end points have been established, it remains to apply the remaining guide lines to the overall system. Although some cable routes may be short (several hundred kilometers), major routes may link points separated by 3000 to 6000 kilometers. In the longer routes a great circle route may be the

shortest and most economical, but may contain certain environmental hazards which dictate that a devious path be followed. The problem is to select a route for the proposed cable that is economical and yet such that system integrity is assured. In the North Atlantic Ocean, historically, telephone system integrity has suffered most in the relatively shallow (less than 600 meter) depths in the ocean. Here fishing trawlers have caused extensive damage. Deep water portions of cable routes have been relatively trouble free.

The submarine cable system may, for route engineering applications, be considered composed of several parts: (i) the terminal station, (ii) the continental margin (0 to 600 meters), and (iii) deep water portions.

II. TERMINAL SITE SELECTION

The terminal sites of the submarine cable system should be located as near to the shore as possible because the necessary land connection between the terminal building and the sea is subject to the same hazards as any other land cable (excavation by other workmen, conflict with other structures, and so on). The failure of this relatively short portion of the system would, of course, disable an entire transoceanic system. Deep trenching, conduit encased in concrete, warning signs, and in some cases full ownership of the right-of-way are means used to minimize the possibility of failure.

One of the unique requirements of a modern submarine cable system is for a current-carrying ground (called the ocean ground) to be located adjacent to the terminal station. It is part of the return path for the dc current (136 milliamperes in the case of SF Cable) that is carried on the center conductor of the cable to energize the submerged repeaters. This ocean ground, which may be made up of several buried silicon-iron rods, must be located where a uniformly low resistance of about one ohm maximum can be maintained the year around. The ground must be close enough to the station that the series resistance of the connecting cable is correspondingly low. Adequate separation from other underground metallic structures, including the submarine cable itself, must be provided such that the other structures are not adversely affected by the cable system ground current. If the terminal station is sufficiently close to the sea, a satisfactory ground can usually be obtained in a salt water environment adjacent to the landing beach; otherwise, earth resistivity measurements must be made to find a suitable location of the ground and determine the number of silicon-iron rods required.

The nature of the landing beach and adjacent ocean bottom are

factors which influence the ultimate location of the terminal building. The beach itself should be open, unobstructed, and located near access roads such that heavy construction equipment—bulldozers, trucks, and power shovels—can be used during installation of the cable, and later for normal maintenance work.

Ideally, the sea bottom approaching the beach should be firm, gradually sloping, and free of boulders or outcropping rock. The slope of the bottom should be such that the cable ship placing the shore end of the cable is able to anchor and still have adequate distance from the land and adjacent shoal areas for maneuvering.

The bottom near the shore is usually examined by divers. If necessary, certain measures such as underwater blasting, split cast iron sleeves, and concrete bags are used to provide a suitable protected path for the cable. Severe erosion may occur in the surf zone, and often cable is excavated or jetted into the beach to a depth of about six feet.

Inland of the terminal building, the communications system is brought into the existing land network. Where possible the submarine cable system is linked to a hardened (blast resistant) coaxial cable route. Redundancy is often provided by a radio relay system.

III. DEEP WATER ROUTE CONSIDERATIONS

The goal of route engineering for a submarine cable system is to determine the best path for the cable from the combined standpoints of first cost and future maintenance. Certain ocean survey operations are required to define engineering parameters.

In the past 30 years, with the advent of electronic depth sounders and other types of exploratory equipment, a great deal has been learned about the configuration and environment of the ocean bottom. Existing knowledge is by no means complete, however, and in certain instances is even subject to error.

Many nations publish navigational and oceanographic charts and exchange information pertinent to them. Charts originated by the U. S. Navy Oceanographic Office (NOO) and Environmental Science Services Administration (ESSA) are among those of greatest value. Bottom contour (BC) charts are available which are to a scale of about 20 kilometers to an inch. A dense concentration of sounding lines is considered to be anything closer than eight kilometers apart.

A great deal of information on the nature of the sea bottom and the movement of water masses may be obtained from published reports of oceanographic institutions such as Scripps Institute of Oceanography

(California), Lamont Geological Observatory (New York), or Woods Hole Oceanographic Institution (Massachusetts). Their reports may include photographs of the bottom and analyses of the nature and depth of bottom sediments. Frequently, oceanographic specialists in particular fields of interest are consulted or retained for studies of specific problems.

Of great value also are the trouble histories of existing submarine cables, particularly the older (19th century) cables. Correlation of their histories with presently available oceanographic information can lead to predictions of the performance of proposed cables.*

It is beyond the scope of this article to present a detailed discussion of the physiography of the sea bottom. The list of references¹⁻⁷ will lead the interested reader to the vast literature on the subject. It is enough to say that, just as on land, the sea bottom is divided into physiographic regions — mountain ranges, plateaus, plains, and valleys. The sections shown in Fig. 1 illustrate some of the factors to be considered in route engineering. The major divisions are *continental margin* (including continental shelf and slope), *ocean basin*, and *mid-ocean ridge*; each presents unique environmental considerations.

The principal tool for investigating a proposed route is the echo sounder. This is a hull-mounted, 12-kHz unit, and when used in conjunction with a precision recorder such as the precision depth recorder (PDR), the sounding system provides a depth record which can be read to ± 2 meters (Fig. 2).

The shortest route between terminal sites is initially considered. This route is altered as necessary on the basis of existing information. Thus, gross trouble spots are eliminated. A ship is chartered and one or more echo sounder lines are then made on and parallel to the initial route. During this survey, an area of particular interest, such as the mid-ocean ridge or continental slope, may be examined in detail by means of a closely spaced grid pattern. A resultant, "best possible" deep water route is selected.

IV. SHALLOW WATER ROUTE CONSIDERATIONS

As stated earlier, in the North Atlantic Ocean deep ocean portions of cable routes have been relatively trouble free. The most hazardous portions of routes for cable lie in depths of a few meters to about 600 meters where fishing trawler activity is greatest. The high incidence of cable break and cable damage in these depths led to the development

* For classical studies of this type the reader is referred to various papers in the literature by B. C. Heezen and M. Ewing on the "1929 Grand Banks Turbidity Current." This current was particularly destructive to several existing telegraph cables. See Refs. 1 through 7.

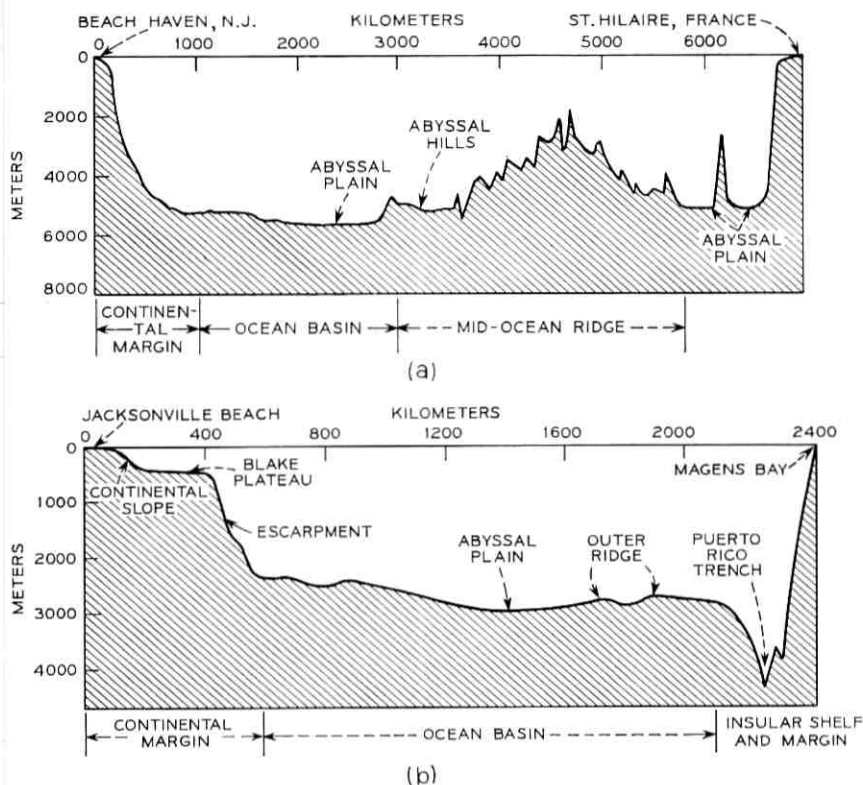


Fig. 1—Profiles of ocean bottom: (a) TAT-4 cable, and (b) Florida-St. Thomas SF Cable.

of the ocean cable plow described by R. E. Mueser and H. A. Baxter.⁸ Certain route engineering factors and related survey techniques are associated with use of the plow.

4.1 Plowing Route Survey

The survey system used to define a plowing route must provide as much detailed information as possible about the ocean bottom in order to reduce to a minimum the uncertainties associated with the burying operation. In particular, the survey system used must supply information on such matters as:

- (i) the presence and extent of obstacles such as boulders, rock outcrops, and shipwrecks;
- (ii) the thickness and extent of sedimentary material;

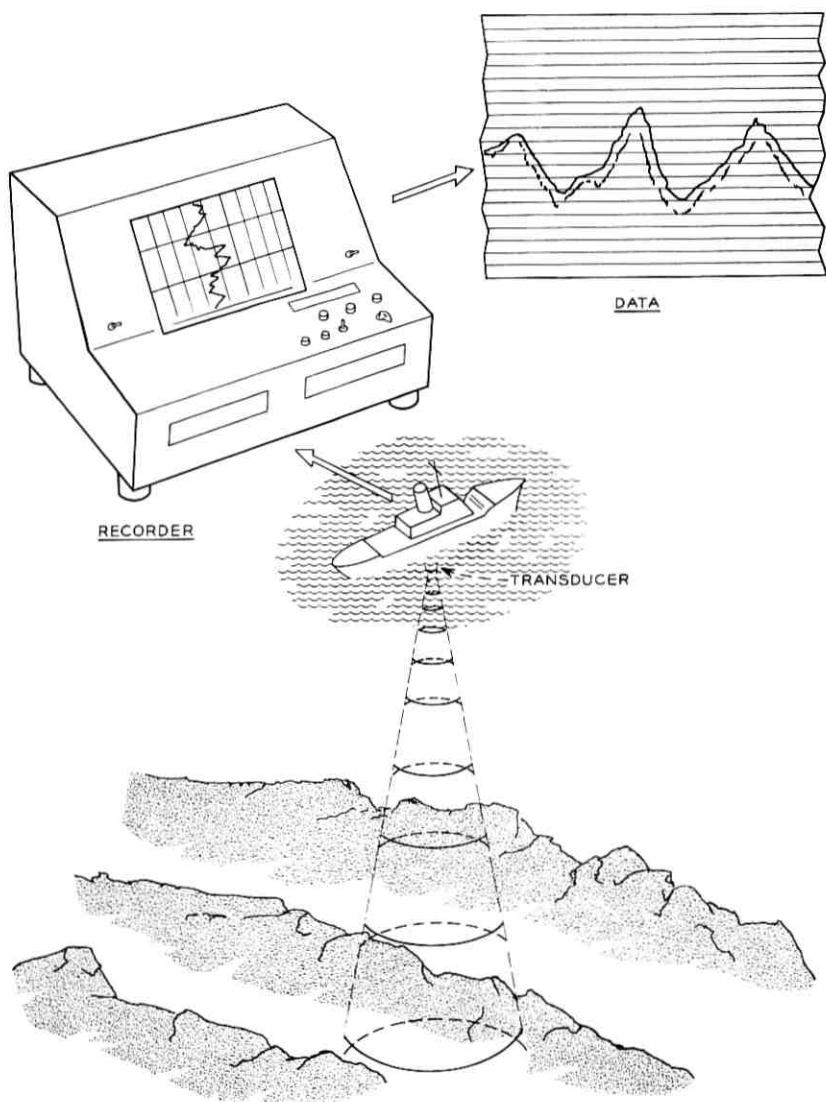


Fig. 2—Echo sounder system.

- (iii) the engineering properties of bottom materials; and
- (iv) the overall plowability of the bottom.

The survey technique used to attain these objectives includes three phases:

- (i) Phase I —reconnaissance survey,
- (ii) Phase II —sled survey, and
- (iii) Phase III—trials with ocean cable plow.

The three phases are carried out in a specific time framework. Phases I and II must be completed well in advance of cable installation to permit forwarding of specifications of cable types and lengths to the factory. This lead time is on the order of 12 to 18 months.

Phase I begins with a review of information pertinent to the area of interest and is initiated as soon as the general route plan is formulated. This literature search often is sufficient to describe in broad terms the conditions to be encountered in any area. The scope and type of subsequent work is also determined. The field portion of Phase I typically involves going to sea with a suitable vessel and sufficient equipment to obtain the necessary information for plowing operations.

Phase II, a sled survey, consists of towing a survey sled (equipped with a plow share) along the route selected by Phase I operations. This operation provides a means of evaluating the plowability of the tentative route selected by the reconnaissance survey.

Phase III, towing the ocean cable plow over the selected route to establish a final plowing route, theoretically can be carried out anytime between the completion of the analysis of Phase II data and the actual burying. In practice, it is generally carried out immediately preceding the actual burying operation as part of the burying system sea trials and crew training period.

Proper navigational control is required in each of the survey phases to tie the observed data into spacial coordinates. During Phase I activities, standard Loran or Decca navigation systems are used where available. Positional repeatability to approximately ± 1 kilometer ($\pm \frac{1}{2}$ mile) is easily achieved with these systems. For Phase II and III work, a precision navigation system such as Decca HI-FIX with a repeatability of ± 50 meters (150 feet) is used. The precision network is referenced as accurately as possible to fixed geodetic positions nearest to the shore station.

4.1.1 Phase I Reconnaissance Survey

Because the reconnaissance survey covers a large area, general practice is to lay out proposed survey tracks in a grid network in order to ade-

quately cover the region. The survey equipment used in Phase I includes:

(i) Echo sounder (Fig. 2)—This equipment provides a continuous plot of the depth of the ocean along the ship's track.

(ii) Seismic profiler (Fig. 3)—The seismic profiler provides information concerning the constituency of the ocean subbottom, (for example, the material from near the water-ocean bottom interface to a few hund-

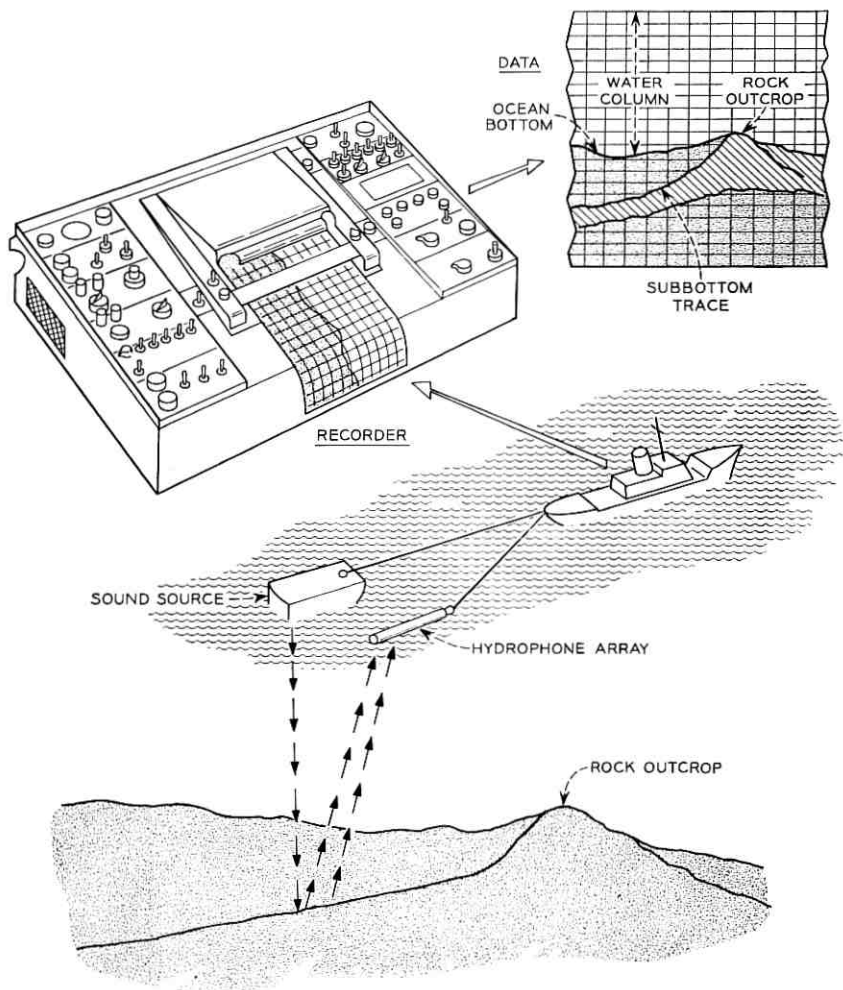


Fig. 3—Seismic profiler system.

red meters beneath the interface). Thus, information concerning rock outcrops and depth to bedrock can be obtained from these data. The equipment consists of a towed sound source, a towed hydrophone array, and a shipboard-mounted recorder.

(iii) Side-scanning sonar (Fig. 4)—A side-scanning sonar unit (150 to 250 kHz, depending upon the unit used) provides data concerning materials and objects on the ocean bottom determined from acoustic reflections. The system includes a towed underwater vehicle which contains several banks of acoustic transducers used both as transmitters and receivers. A recorder aboard ship displays the data received by the towed vehicle. This unit provides an effective tool for locating surface rocks, boulders, or shipwrecks which might lie in the survey area.

(iv) Bottom sampling (Fig. 5)—This technique consists of dropping a long (on the order of three meters), weighted pipe or a "grab" type sampler into the ocean bottom to obtain samples of the sediment near the surface. The samples can be examined for such characteristics as shear strength, water content, grain size, and so on, and allow a definition of the engineering properties of the materials.

(v) Underwater photography (Fig. 6)—A 35-mm camera system consisting of steel frame, camera, flash unit, triggering mechanism, and supporting cable is used to take still pictures of the ocean bottoms at locations selected from the seismic-profiler, echo sounder, and side-scan data.

Upon completion of Phase I field work and a suitable program of data analysis, a composite, three-dimensional description of the area of interest is formulated. The description includes the vertical and horizontal distribution of materials, the engineering properties of these materials, and detailed photographic data for discrete areas within the larger surveyed area. From this information, a route is selected which is optimum for plowing vehicle operations.

4.1.2 Phase II Sled Survey

The second phase survey technique consists of towing a Bell Laboratories designed survey sled over the route selected from the Phase I survey data in order to evaluate the plowability of the bottom. The sled (Fig. 7) is about 4 meters long by 2.5 meters wide by 2 meters high and weighs approximately 2700 kilograms in air. A 5-centimeter-thick Bell System "C" plow share is mounted to the aft end of the sled by means of a pivot bolt and a shear pin. The share is provided with two sets of mounting holes which permit the depth of cut of the share to be adjusted

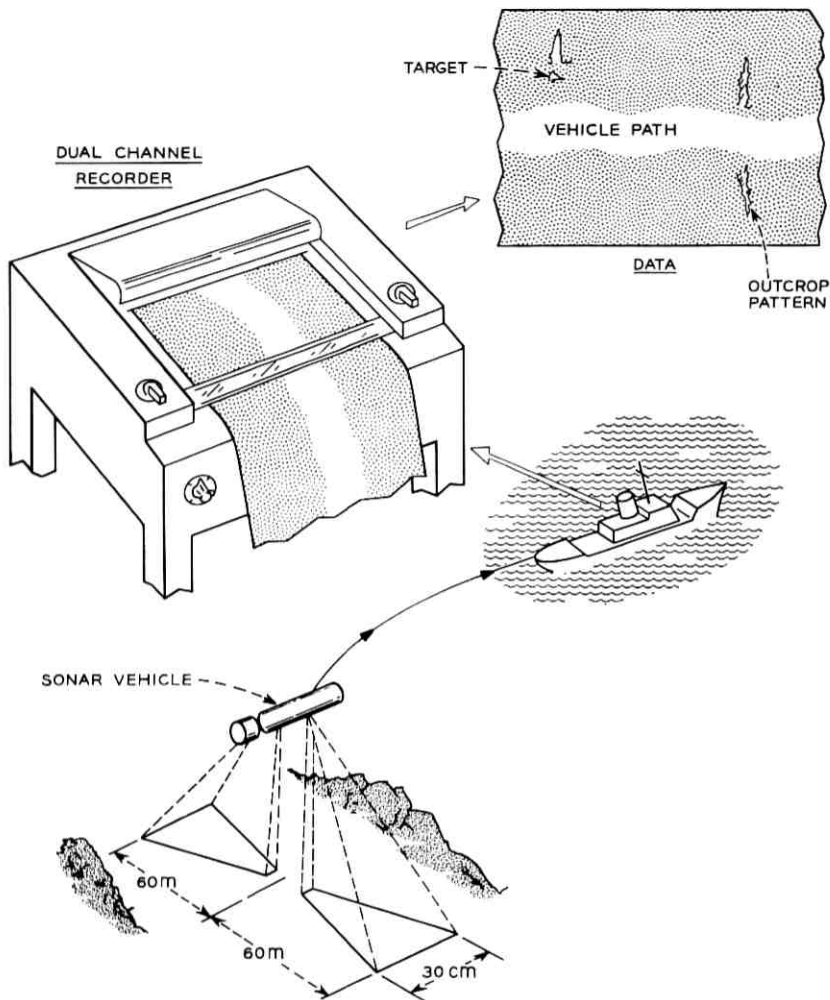


Fig. 4—Side scanning sonar system.

to either 53 or 33 centimeters disturbed depth. The shear pin is included in the design to permit the share to retract should excessively hard bottom be encountered during plowing. If this protective pin does shear, the sled is recovered to replace the broken pin.

An additional safety feature is included in that the towing pad eye on the front of the sled will detach from the sled when towing tensions exceed 20,000 kilograms. This detachment results in transfer of the

attached tow chain to the upper (lifting) pad eye, causing the sled to tumble free of the obstacle. Should this happen, the sled is recovered and refitted with slings in preparation for the next launch.

An underwater television camera is provided on the sled to permit visual examination of the ocean bottom over which the sled is towed. The camera is attached to a pan-and-tilt mechanism, thus permitting viewing of the ocean bottom ahead of the sled as well as most of the sled proper. Two mercury vapor lights are used for the television camera, one light mounted forward and one aft near the plow share. A self-contained 70-mm camera and strobe light are also included for taking photographs of the ocean bottom. Shipboard instrumentation includes towing tension and vehicle depth readouts, a television monitor and tape

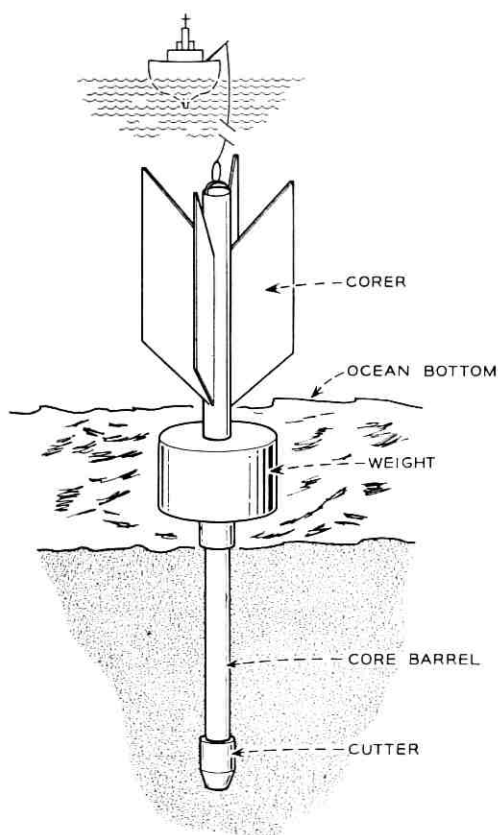


Fig. 5—Bottom sampling.

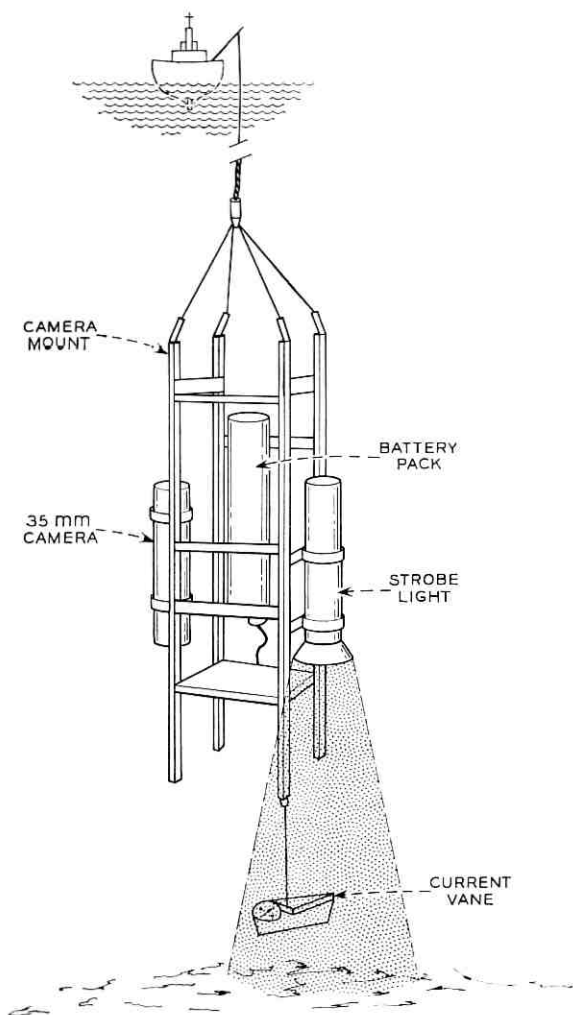


Fig. 6—Camera system (35 mm).

recorder, pan-and-tilt controls, and associated power supplies and supporting electronics.

The sled is connected to the ship by means of a tow wire (2.22 cm diameter and 40,000 kg nominal breaking strength) and a control cable (2.54 cm diameter with 24 pairs of 18-gauge conductors and one 728A coaxial). The tow wire is handled over the bow of the ship; the control

cable (which has floats attached to achieve slightly positive buoyancy in water) is handled from the stern. Launching and recovering of the sled are accomplished by using a yard and stay technique.

4.1.3 Phase III Plow Survey

For final route verification, the Phase III survey includes the towing of the ocean cable plow along the proposed burying route (as specified by the Phase I and II surveys) without actually burying any cable. This procedure ensures that the route selected has no hidden hazards and that the ocean cable plow is indeed capable of plowing in the selected areas. Also, this period of time (usually immediately prior to the actual cable burying operation) provides a trial period for both the equipment and personnel.

V. FINAL ROUTE LAYOUT

When all of the survey information has been obtained and plotted, it is carefully analyzed, section by section, and the most feasible final route is laid out. It is at this stage that the long, straight courses of

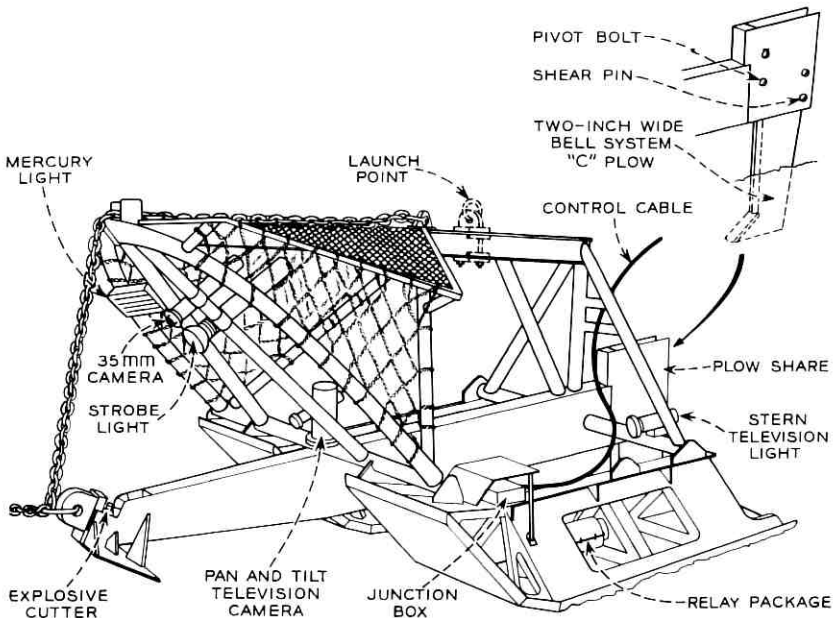


Fig. 7—Ocean bottom survey vehicle.

the preliminary route are broken into a number of shorter ones between angle points known as alter course points. The total route length is determined, first by scaling the distances between alter course points, and, more precisely, by Mercator sailing computations which are usually made independently by more than one person (to insure that the total length of the cable route is accurately determined).

A profile of the ocean bottom along the finally selected route is drawn based on the soundings obtained during the surveys. The profile is usually drawn to a vertical/horizontal scale distortion of 40/1 to make the relief stand out clearly and reduce the total length of the profile chart.

Since slack cannot be laid (nor is it needed) with the ocean cable plow, the length of cable required where plowing is done is virtually the same as the route length. Elsewhere the bottom profile is analyzed to determine: (i) the amount of cable slack which must be payed out so that the cable will rest on the bottom; and (ii) the maximum speed of the cable ship, section by section, when the cable is laid. Both of these quantities are functions of the slope of the sea bottom. The ideal final condition is to have the cable lie uniformly on the sea bottom at zero tension, hence, without suspensions. Insufficient slack or too fast laying speed would prevent attaining this condition; excessive slack or too slow speed would be wasteful of cable or would require greater ship time than necessary.

A armorless cable is specified for the sections to be plowed in and for laying in water depths greater than 600 or 800 meters. Double-armored cable is usually called for between the beach and the start of plowing where chafing and surf action may occur. Single-armored cable is used where lighter protection is adequate, such as on land and in the less exposed sections in shallow water. Electromagnetic shielding is provided on all cable on and adjacent to land.

The total length of cable of all types that must be manufactured for the project, exclusive of spare for future maintenance, is the sum of:

- (i) the total route length between terminals,
- (ii) the total amount of slack to be laid based on the bottom profile,

and

(iii) an amount, based on judgement and experience, to cover contingencies arising from deviations from course and other unpredictable occurrences during the laying of the cable.

The cable route engineering program discussed here leads to a complete cable system plan which includes designation of landing sites,

overall route configuration, specification of cable types and lengths and, as required, a program of cable protection including cable burying.

If the program is carried out carefully, it can make a vital contribution to the successful placement of the cable system over a route which is economical in first cost, and may be expected to yield a long, low-maintenance service life.

REFERENCES

1. Elmendorf, C. H., and Heezen, B. C., "Oceanographic Information for Engineering Submarine Cable Systems," *B.S.T.J.* 36, No. 5 (September 1957), pp. 1047-1093.
2. Guilcher, A., *Coastal and Submarine Morphology*, Methuen and Company, Ltd., 1964.
3. Heezen, B. C., Tharp, M., and Ewing, M., "The Floors of the Ocean: I The North Atlantic," *Geological Soc. of Amer. Special Paper* 65, April 11, 1959, pp. 1-107.
4. Hill, N. M., (editor), *The Sea*, Vol. 1-3, London: Interscience, 1962.
5. Shepard, F. P., *Submarine Geology*, (second edition), New York: Harper and Row, 1963.
6. Sverdrup, H. U., Johnson, M. W., and Fleming, R. H., *The Oceans*, (fourth edition), New York: Prentice-Hall, 1952.
7. Turekian, K. K., *Oceans*, Englewood Cliffs, New Jersey: Prentice-Hall, 1968.
8. "Burying Ocean Cable," *Bell Laboratories Record*, 45, No. 6 (June 1967), pp. 187-191.

Transmission Tests, Computations and Equalization During Installation

By W. B. HIRT and D. O. OLDFATHER

(Manuscript received July 30, 1969)

The SF Submarine Cable System relies on equalization of accumulated misalignment at regular intervals along the length of the system to achieve performance objectives. Transmission tests of the undersea system during installation are required to implement this equalization. This article describes the tests that are conducted and the computations used to select equalizer settings just before an equalizer is overboarded.

I. INTRODUCTION

A continuous test program forms a part of the SF Submarine Cable System¹ installation procedures from the time material arrives at the dock for loading aboard ship through the final line-up before service. The tests and computations we describe are necessary to verify satisfactory performance of cable, repeaters, and equalizers; to obtain data for determining the optimum ocean block equalizer settings; and to obtain data for system line-up and for system administration and fault location.

The major undersea portions of the SF System are laid using the ship, equipment and techniques developed for the SD Submarine Cable System.²⁻⁵ Transmission testing coincides with these cable-laying activities, although the transmission test equipment installed on the Cable Ship Long Lines (shown in Fig. 1) is predominantly new. A parallel transmission effort is also performed during the burial of SF System cable,⁶ but with the use of considerably more primitive equipment aboard ship.

II. GENERAL

The SF Submarine Cable System has a maximum design length of 4000 nm which causes the total transmission loss at the highest system

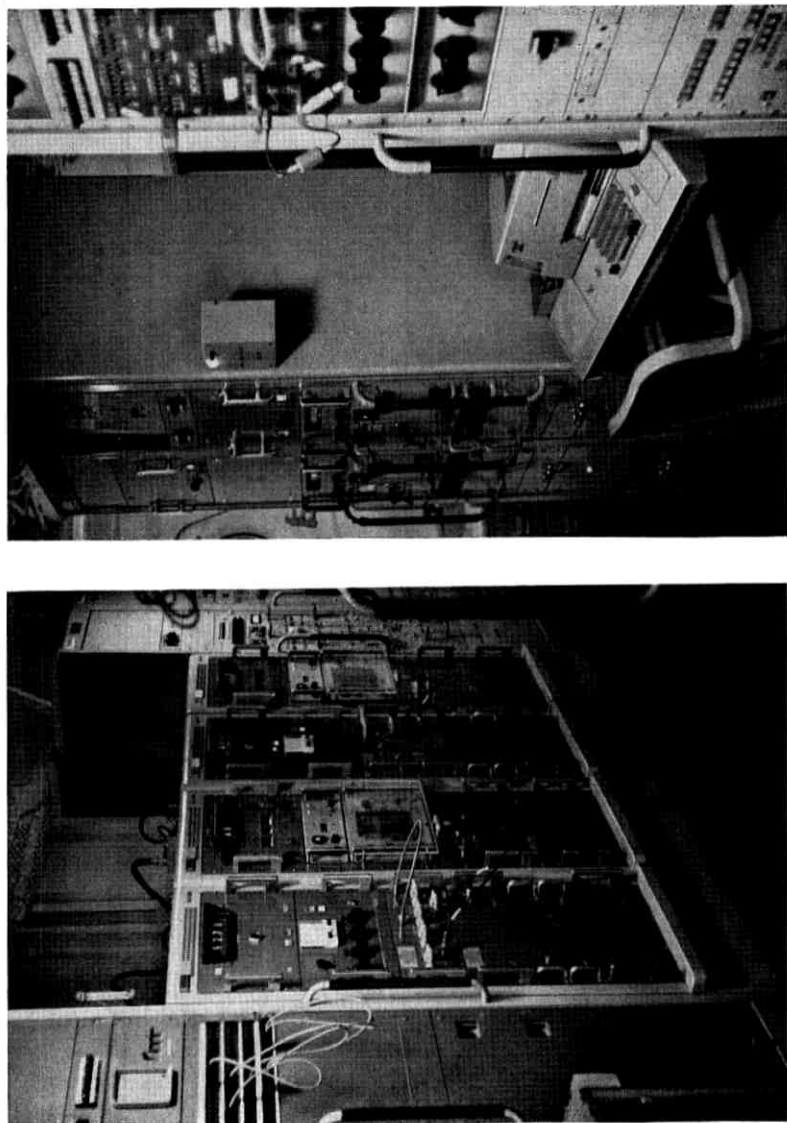


Fig. 1—Long Lines transmission test room.

frequency to reach 16,000 dB. This loss must be matched by the gain of some 400 repeaters. The repeater gain is shaped to match the cable loss at a temperature of 3°C and a pressure corresponding to a depth of 2000 fathoms. Since the installed cable experiences the actual temperature and pressure of the ocean floor, the length of each cable section is cut at the factory to obtain the desired loss at the highest system frequency. This step of equalization, based on oceanographic data on temperature and pressure for each cable section along the route, adjusts the top frequency loss for known departures from design temperature and pressure. All other equalization of the undersea system takes place in the adjustable ocean block equalizers situated after every 20 repeaters. This equalization corrects for production deviations in cable loss and repeater gain and uncertainties in predicted ocean-bottom temperature and depth as well as any cable attenuation changes due to handling. Ideally, complete information on which to base the equalization is available only after the complete system has been installed on the sea bottom. However, in order to adjust an equalizer, it must be physically accessible. The optimum adjustment for an equalizer must therefore be made before the equalizer goes overboard. Selection of the best setting for a particular equalizer must therefore be based on measurements made between the shore and that equalizer before this portion of the system has completely reached equilibrium at the sea bottom. Hence, detailed and continuous transmission measurements and computations are necessary during the actual laying of the system.

An ideal, no-misalignment system would have equal transmission levels at the output of every repeater at a given frequency. Since deviations will accumulate, however, the ocean block equalizers are used to periodically reduce the misalignment to minimize deviations from the average level. Repeaters operating at a high level contribute above-average modulation noise, while low-level repeaters cause increased thermal noise. The SF System is modulation-limited, so the undesirable modulation noise of high-level repeaters comes into effect before gross overload is reached. Figure 2 shows how misalignment accumulated and was periodically reduced at 5000 kHz on the Jacksonville-St. Thomas system, while Fig. 3 shows the improvement obtained across the transmitting and receiving bands at the end of block four.

Although the aim of transmission tests during laying is to determine the deviations of cable and repeaters when fully stabilized at sea-bottom temperature and pressure, the measurements must be made between

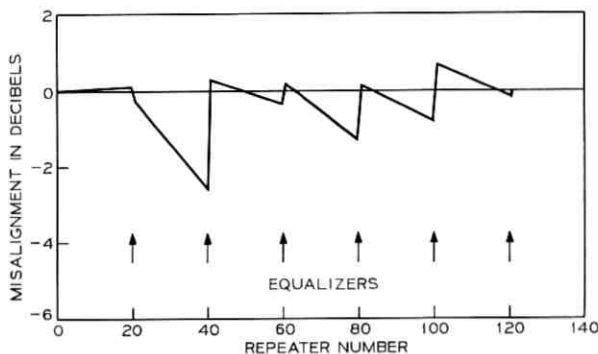


Fig. 2—Typical misalignment versus repeater number plot (data apply to Jacksonville-St. Thomas SF System at 5000 kHz).

the shore and the block-end equalizer to go overboard next. The cable in that portion trailing down from the ship to the ocean floor is warmer than it will be when it reaches the ocean floor. As the cable cools, its loss decreases. On the other hand, as the cable is subjected to increased pressure its loss increases. As a result of these two effects, the transmission changes continuously as the laying proceeds. For this reason, measurements are made at hourly intervals, and the data are plotted versus miles laid. At a prescribed time before the equalizer must go overboard, the transmission data are extrapolated to the end of the

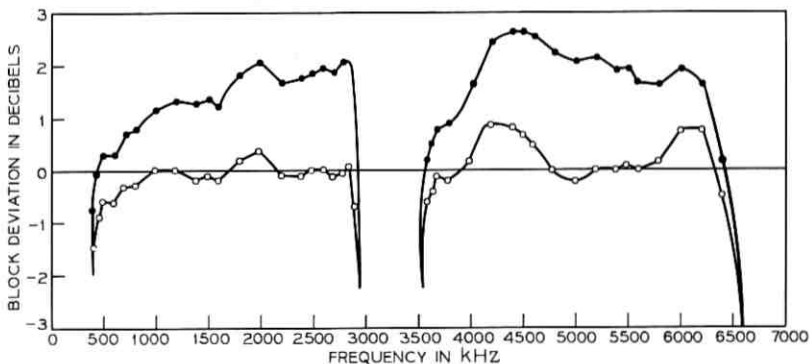


Fig. 3—Typical block deviation before and after equalizer adjustment (data apply to block four (repeater 81) of Jacksonville-St. Thomas SF System). Dots show data before equalizer adjustment and circles show data after equalizer adjustment.

block and compared with desired values. This gives the accumulated deviation corresponding to having the entire portion stabilized at sea-bottom conditions. The switchable shapes available in the equalizer are then examined to determine the optimum settings to minimize the deviation in each band. After the equalizer has been switched, measurements are made to check that the desired result has been achieved, the equalizer is prepared for overboarding, and measurements from the shore through the next block-end equalizer are begun.

III. LOADING

The submarine cable, repeaters, and ocean block equalizers which form the undersea system are tested individually at their place of manufacture and are shipped separately to be loaded aboard the cable ship. The repeaters and equalizers are brought aboard and placed in the repeater stacks in proper sequence for laying. The cable sections are coiled in the various cable tanks according to the loading plan, which is devised so that when the cable and repeaters are spliced together the system can be payed out continuously without snarls or knots. During the loading of the cable it is given the following tests: insulation resistance between center and outer conductors, dc loop resistance of center and outer conductors in series, pulse echo observation to verify length and check for impedance irregularities, and a continuous monitoring of the dc cable loop resistance while the section is being coiled in the tank. The ends of each cable section are brought to the repeater stack and placed at the appropriate end of the repeater or equalizer to which they are later spliced to form an assembled shipload.

IV. ASSEMBLED SHIPLOAD TESTS

Since this is the first time the individual parts actually form a transmission system and since the shipping, loading, and splicing could cause faults, the assembled shipload is tested to ensure that the system is free from faults prior to laying.

4.1 *Transmission Tests*

The connections necessary to power the assembled shipload and to utilize the shipboard test room equipment are shown in Fig. 4. From this figure it can be seen that for these tests the shipboard transmission equipment must transmit and receive in both bands, with

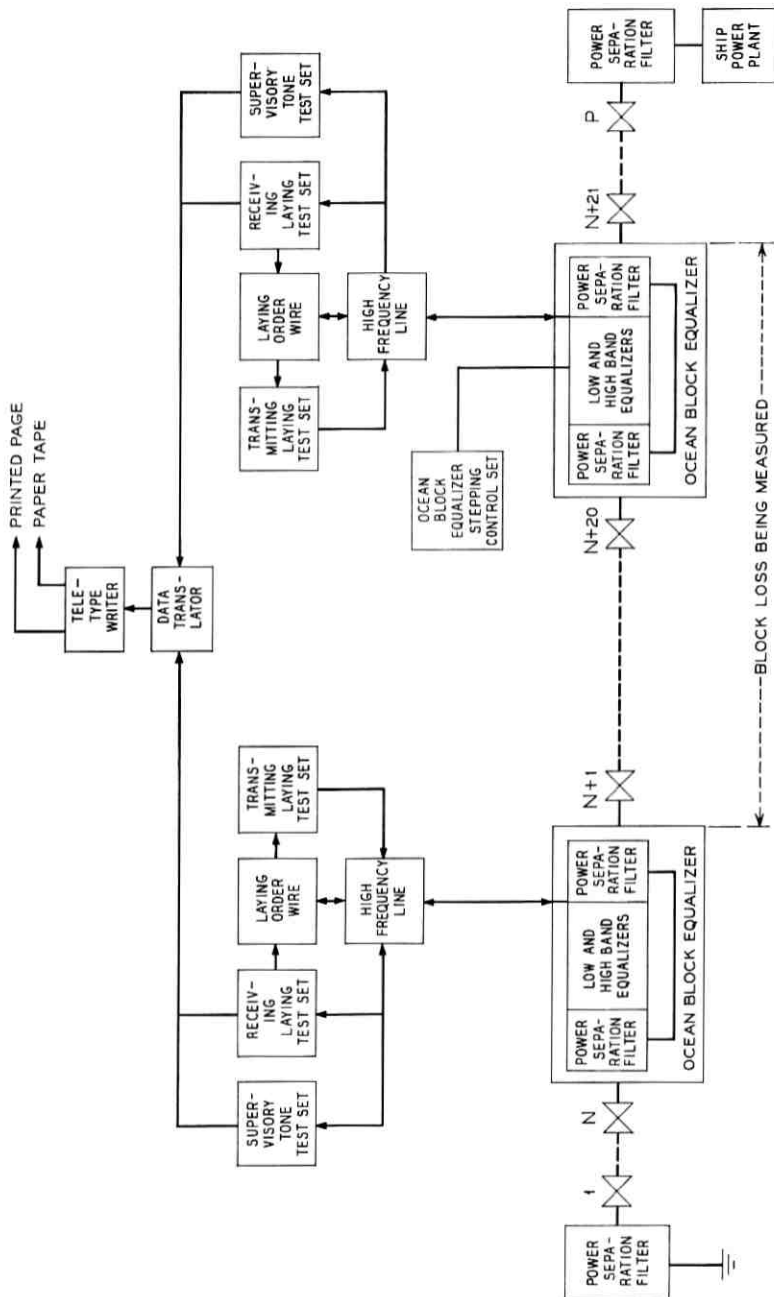


Fig. 4—Assembled shipload transmission tests.

the equipment at one end performing the function the shore equipment will perform when tests are made between ship and shore. The length of each assembled shipload is short enough to allow power to be supplied by a power supply at one end only, with the other end grounded. The connection block diagram therefore shows that each end is placed in a power separation filter, which connects the shipboard dc power plant to the center conductor at one end, grounds the center conductor at the other end, and provides 75-ohm coaxial connections for transmission tests. Before dc power is applied to the system, the insulation resistance between inner and outer conductors is measured with the far-end open circuited to ensure that there are no faults to ground which would prevent the proper application of power and cause the power plant to shut down. A measurement of the dc resistance through the center conductor path of all the cable, repeaters, and equalizers ensures that the system is properly connected.

Transmission tests through the entire assembled shipload are not feasible because, in general, the high shipboard temperature relative to that of the sea-bottom causes the cable loss to be very high. Therefore, signals are sent only through one block; that is, between two adjacent ocean block equalizers or between an equalizer and the adjoining cable end. Testing one block at a time has the added advantage that it is easier to localize any trouble that may be indicated by the test.

Refer again to Fig. 4. Transmission tests are made at a fixed set of frequencies in the low band and the high band. The paperwork, laying test sets, and computer programs are prearranged for these frequencies. For the assembled shipload tests 0-dBm test tones are generated by the test set at one end in, say, the high band and are received by the set at the other end of the block. For tests in the low band the test set at each end is reconnected to interchange transmitting and receiving functions. The laying test sets measure with an accuracy of 0.1 dB and go from one frequency to the next automatically. The frequency and received power are printed by the teletypewriter, which also prepares a punched paper tape of the data. The transmission tests at the 26 standard frequencies in the low band and 24 frequencies in the high band are completed with a permanent typed copy and a paper tape prepared in about 5 minutes per block. The power of the test tones applied to the repeaters is controlled by attenuators in the high-frequency lines. Different transmitting and receiving attenuator settings are required depending on whether the transmission connection is at an equalizer or a power separation filter, due to the different paths the test tones take between the nearest repeater and the test equipment.

4.2 Computations

The raw transmission test data become useful only if one can interpret them and decide whether or not the system is operating properly. The factory measurements on actual repeaters, equalizers, and cable sections in each block are compiled and used as a standard with which all test results are compared. The factory measurements on these items and the place in the system to which they are assigned are the inputs to a computer program which processes the data and prints out the compiled information, called the system data book. The system data book, prepared at Bell Telephone Laboratories before the loading or laying begins, is available in printed form and is also stored in the shipboard computer. Data at two conditions, the predicted sea-bottom temperature and pressure and at atmospheric pressure and 10°C are provided. Assembled shipload test data are compared with the system data book block loss (Fig. 4) at atmospheric pressure and 10°C, since these conditions require smaller corrections to match the shipboard temperature and pressure than sea-bottom conditions do. To determine the block loss from the measured data, the losses and gains encountered by the test tones at each frequency in the transmitting and receiving high-frequency lines, the power separation filters, all patch cords and test trunks, and the path through the equalizer taken by the tone must be known. These gains and losses are measured, at the 50 standard frequencies and for all the various attenuator settings needed, prior to the tests on the system.

The connection to each equalizer used for transmission tests is a bridge across the main transmission path between the power separation filter in the equalizer and the high- and low-band equalizing networks (Fig. 4). The test connection and the orientation of each equalizer in the system put the equalizing networks in both bands between the shore station and the ship test connection during laying. The test equipment has a 75-ohm impedance which is bridged across the 59.4-ohm system impedance. A bridging loss of 3.91 dB used in the calculations allows for the mismatch loss and an additional factor to convert the test measurements to the power which would represent the power at the bridging point if the bridge connection were not there, since this lead is removed (open-circuited) after tests are completed.

After the computation of the measured block loss from the measured data, the cable temperature coefficient must be used to correct the data to 10°C for direct comparison with the system data book. This requires temperature data from various points in the ship's cable tanks, obtained using thermocouple sensors and a multichannel readout

panel. This temperature detection is the least accurate part of the test, so the difference between the temperature-corrected measured block loss and the system data book value often has a characteristic proportional to the square root of frequency, caused by the temperature measurement inaccuracy. This disparity becomes recognizable when plotted and can be subtracted out to give the real transmission deviation. This process amounts to using the transmission data as a measure of average block temperature.

All computations described here for assembled shipload tests and also all computations during laying are performed by the ship and shore personnel using specially prepared forms and procedures. The same computations are also performed using the shipboard computer. The shipboard computer installation is shown in Fig. 5. (The teletypewriter shown in Fig. 5 for use with the computer is separate from the teletypewriter shown in Figs. 4 and 6 for use with the transmission test equipment.) A shipboard computer rather than a land based computer was chosen for several reasons:

(i) Certain computations are done as the ship steams to the cable grounds; land based computers demand the maintenance of a data link between ship and shore.

(ii) A shipboard computer is available as a real time, on-line machine with a system of utility programs for general computations.

(iii) A shipboard computer allows laying a system from any shore station.

The repeater supervisory tones⁷ are also measured during assembled shipload tests. Odd-numbered repeaters send tones in one direction and even-numbered repeaters in the other. The difference between the power of two tones gives the loss between the two repeaters at the frequency of the tone. Simple computation refers the absolute power and frequency of each tone to the repeater output for comparison with the system data book values to ensure that each tone is working

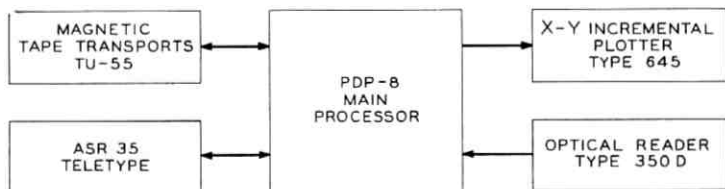


Fig. 5—Shipboard computer installation.

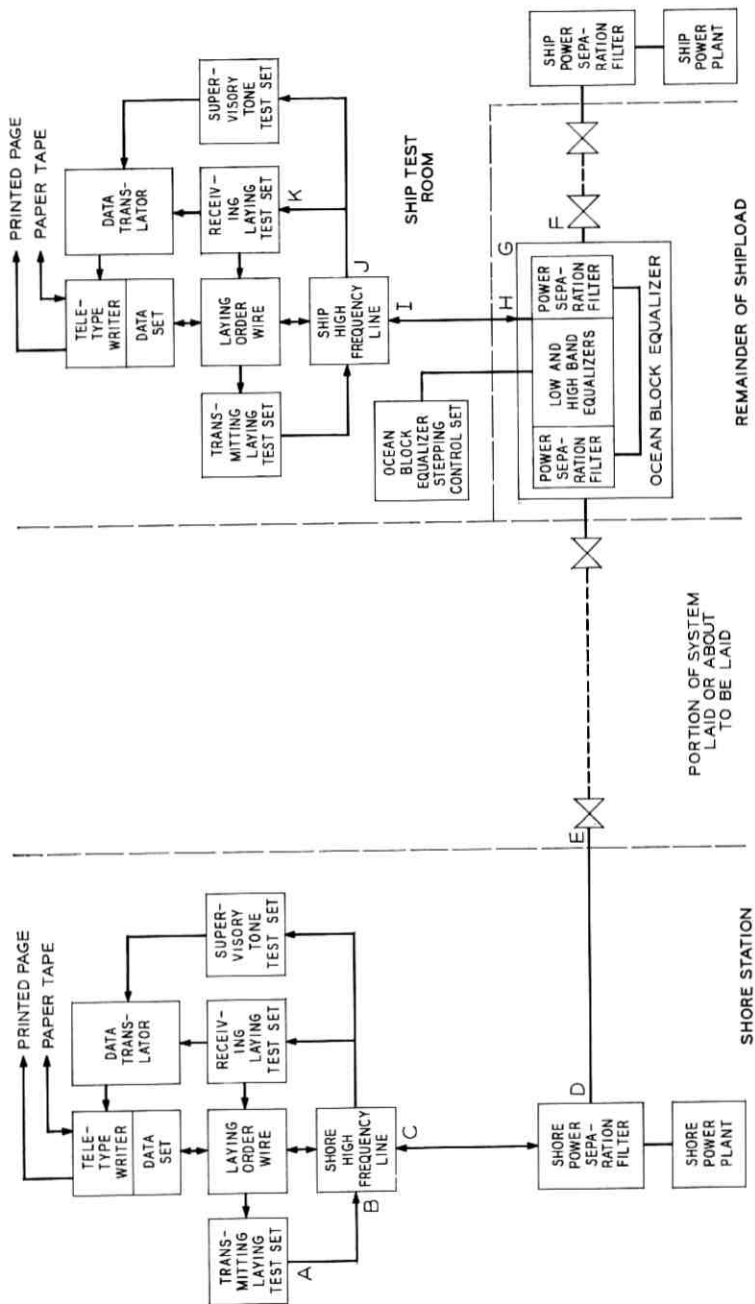


Fig. 6—Transmission tests during laying.

properly. The loss between two repeaters can also be manipulated for direct comparison with the system data book information.

4.3 *Ocean Block Equalizer Adjustment Tests*

The switchable networks in each equalizer⁷ are used to modify the transmission characteristic to correct for misalignments detected during laying. It is important to verify that the networks are all on a known reference setting prior to beginning the lay and also to test the equipment used to switch the networks. This is done during assembled shipload tests by making three transmission runs on each block: the first run with the equalizer networks as shipped from the factory, the second with each of the seven networks in each band switched to the complement of the desired reference setting, and the third with the networks all set to the reference setting. This testing exercises all the switching relays and stepping equipment. Addition and subtraction of the three runs shows whether or not the equalizer was shipped with the correct reference settings and ensures that the equalizer has the reference settings for the lay. Tables of the expected transmission change between the reference setting and each of the 127 other possible combinations in each band, for each of the 50 standard frequencies, are used to check the measurements.

V. TESTS DURING LAYING

The actual laying of a shipload can begin under a variety of circumstances. For example, the first 92 nm at the Florida end of the Jacksonville-St. Thomas SF System were installed by the c. c. g. s. JOHN CABOT and part of this load was buried for protection.⁸ Transmission tests were performed on this part of the system, but no equalizers were involved. The first shipload carried by the c. s. LONG LINES was installed starting at the end of the cable left by the CABOT.

5.1 *Transmission Tests*

After bringing the previously laid cable end aboard ship, tests are made from the recovered end to the shore station to verify the integrity of the previously laid system, followed by the splicing of the recovered end to the outboard end of the assembled shipload. At this point the shore, ship, and undersea parts of the system appear as in Fig. 6. Schematically, Figs. 4 and 6 are similar, but the physical separation of ship and shore by hundreds or thousands of miles represented in Fig. 6 makes the performance of transmission tests quite a different

problem from the assembled shipload tests. The main communication link is the laying order wire, which provides twelve 4-kHz channels for voice, teletype data, and laying test set automatic control. After splicing, dc power is applied from the ship's power plant, and from the shore if the length requires double-ended feed.⁸ The first transmission tests are made from the shore through the shipboard equalizer which will be laid first.

Transmission tests at the standard frequencies in both bands are made hourly during laying. At both ship and shore, attenuators in the high-frequency lines are adjusted before the first test through each block so that the tone power at the output of the first repeater in the transmitting direction is -10 dBm at the top frequency of both bands. The laying test set on the ship generates 0 dBm at each frequency in the ship-to-shore band, and when the shore test set has measured one tone it automatically goes to the next frequency and signals the set on the ship to do likewise via the laying order wire. For this case the shore teletypewriter prints the measured frequency and power and punches a paper tape. Since both shore and ship have two complete test sets, transmission tests in both bands can be made simultaneously. As the measurements are being made, only the receiving station for each band obtains a printed page and a paper tape of the data, so the data are exchanged between ends by running the tapes through the teletypewriter tape readers which transmit via the data sets and laying order wire.

Other tests during laying include: (i) continuous monitoring of one frequency in each band when no transmission tests are being made, with automatic alarms for a ± 1 dB variation, (ii) tests of the received power and frequency of the supervisory tones from the repeaters, (iii) measurements of thermal noise, (iv) transmission tests at more closely spaced frequency intervals, (v) transmission tests through the next ocean block immediately before starting to lay the block to re-verify its integrity, and (vi) monitoring of the power plant current and voltage.

5.2 *Setting an Equalizer*

The general discussion at the beginning of this article explains that the data from each transmission run must be plotted versus cable mileage, for each of the 50 frequencies, since the measured power continuously changes as the system is laid. The test tones are transmitted at a constant power, but as the cable leaves the ship's tanks and sinks to the ocean floor it cools and the pressure on it increases. The net

effect is a decrease in loss with time which is linear (in decibels) if the ship's laying speed, ocean bottom temperature, and depth are constant. Fig. 7 gives a typical example, showing that the absolute value of the received power increases as the laying proceeds.

About three hours before the predicted time at which the equalizer should go overboard, the received power versus cable mileage data (Fig. 7) are used to predict what the measured power would be at the end of the block, that is, if it were possible to continue the transmission measurements until all the cable up to the equalizer had stabilized at ocean-bottom conditions. Extrapolating the power versus mileage data to a point near the actual equalizer mileage* provides this information. Direct comparison of the extrapolated block-end power with the transmission objective gives the block deviation. The transmission objective for each block is the expected test set received power computed assuming that the misalignment in the system is a desired amount called the misalignment objective. The misalignment objective allows the system to be laid and equalized with a desired misalignment, to take care of factors such as annual temperature variations, so that the time-averaged misalignment is minimized. The transmission objective is found by tracing the path of the 0-dBm transmitted test tone through all the known gains and losses up to the receiving test set. Referring to Fig. 6, the transmission objective for the shore-to-ship band, in minus dBm (decibel referred to 1 milliwatt), is: 0-dBm transmitted power + patch cord loss AB + shore transmitting loss BC + shore PSF loss CD + shore section loss DE - misalignment objective EF - cable loss FG - equalizer PSF loss GH + equalizer bridging loss H + test trunk loss HI - ship receiving gain IJ + patch cord loss JK.

The block deviation is the cumulative error in the system up through the equalizer about to be laid and represents the correction which should be applied by switching the equalizer's networks from the reference settings to their final settings. The block deviation definition for both directions of transmission specifies that excess gain produces a positive deviation. The 128 possible switchable combinations in each band are tabulated and plotted versus frequency as added loss relative to the reference setting, so that a direct comparison of the block deviation gain with the new equalizer setting's added loss can be made. The comparison can be made manually by superimposing a plot of deviation and plots of the available equalizer settings and choosing the setting

* The extrapolation mileage used is a function of the ship's speed, shipboard cable temperature, ocean temperature and depth, and the thermal time constant of the cable.

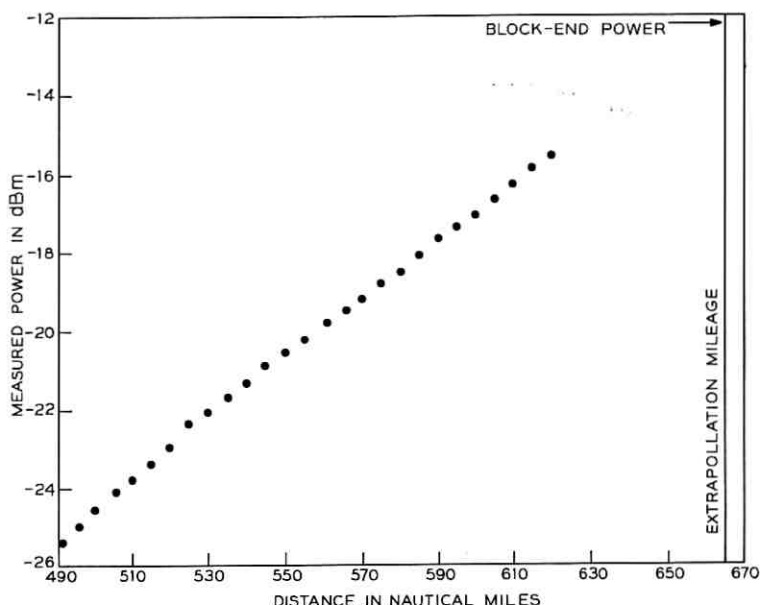


Fig. 7—Typical transmission test received power versus mileage plot (data apply to block four and 5000 kHz on Jacksonville-St. Thomas SF System).

which most closely matches the deviation over the desired frequency range. A separate decision must be made for each band. Fig. 3 shows the effect of setting the equalizer of block four of the Florida-St. Thomas SF system.

While it can be done manually, the shipboard computer operator performs the calculations which result in choosing the optimum equalizer setting by using programs stored in the computer. There are two classes of programs. Those that are designed to compute specific items such as transmission objectives, and a general set of utility routines which can be used to manipulate data under the control of the operator. Both classes of programs reside on one of the magnetic tape units and are called into the machine for use at the option of the operator. One of the other tape units is used for the logging of measured data. Operationally, a punched paper tape containing received power versus frequency and cable mileage is prepared using the test sets, data translator, and associated teletypewriter. This tape is read by the computer with the optical reader using a utility routine and logged onto magnetic tape. When approximately 20 frequency runs have been logged in

this way the extrapolation routine is called into core. The data are extrapolated to find the extrapolated block-end power, defined above, and the results of this extrapolation are both printed and logged on the data tape. Next, a special program is used to compute the block deviation, using the transmission objective and extrapolated block-end power. Finally, the equalization program is used to choose the optimum equalizer setting for each band. The equalization program minimizes the frequency weighted sum of squares of the difference between the block deviation and the available equalizer switchable shapes. A weighting function that gives equal weight to the deviations at the standard frequencies which are in-band but away from the band edges and that gives zero weight to the deviations at the other standard frequencies has been found to give satisfactory results.

Two of the factors which are considered when making the equalization decision are (i) the misalignment of the last repeater in the block being laid and (ii) the misalignment of the first repeater in the next block. Figure 1 shows typical misalignments of first and last repeaters, with the equalizer causing the sharp changes between these adjacent repeaters and an assumed linear change over the remainder of the repeaters in a block. The misalignment of the last repeater in the block being laid cannot be changed by switching the equalizer. This misalignment is computed from the block deviation and the system data book value of the loss between the outputs of the repeaters on either side of the equalizer for the equalizer reference setting. The misalignment of the first repeater in the next block equals \pm the block deviation for the equalizer reference setting and can be computed for any other equalizer setting by using the equalizer added loss due to switching. The question of which sign (\pm) becomes evident upon considering which band is involved, and the facts that (i) gain always gives a positive block deviation and (ii) during cable laying the output of the repeater nearest shore is always taken as the point of zero misalignment for both directions of transmission. Hence, during cable laying a system gain in both bands will give a positive misalignment in the shore-to-ship band and a negative misalignment in the ship-to-shore band.

After the equalization decision has been made and about two hours before the equalizer must go overboard, a transmission test is made, the equalizer stepping control equipment is used to switch the 14 networks, followed immediately by another transmission test. These two transmission tests are compared to verify that the transmission change does represent the desired new equalizer setting. The trans-

mission and switch control leads are then transferred to the next equalizer in the system, and the two leads on the equalizer to be laid are sealed. Testing is then begun from the shore through the next equalizer.

REFERENCES

1. Anderson, C. D., and Easton, R. L., "An Overview: Requirements and Performance," B.S.T.J., this issue, pp. 605-630.
2. Ehrbar, R. D., "A Cable Laying Facility," B.S.T.J., 43, No. 4 (July 1964), pp. 1367-1372.
3. Grismore, O. D., "Cable and Repeater Handling System," B.S.T.J., 43, No. 4 (July 1964), pp. 1373-1394.
4. Gretter, R. W., "Cable Payout System," B.S.T.J., 43, No. 4 (July 1964), pp. 1395-1434.
5. Altenburg, C. J., Butler, J. H., McSweeney, R. J., and Sutton, L. E., "Design and Powering of Cable Ship Long Lines," B.S.T.J., 43, No. 4 (July 1964), pp. 1435-1459.
6. Mueser, R. E., and Baxter, H. A., unpublished work.
7. Buus, R. G., Kassig, J. J., and Yeisley, P. A., "Repeater and Equalizer Design," B.S.T.J., this issue, pp. 631-661.
8. Calkin, E. T., and Golioto, I., "Power Conversion," B.S.T.J., this issue, pp. 749-765.

Power-Frequency Characteristics of the TRAPATT Diode Mode of High Efficiency Power Generation in Germanium and Silicon Avalanche Diodes

By D. L. SCHARFETTER

(Manuscript received October 23, 1969)

We calculate in this paper the output power obtainable, versus frequency, for TRAPATT diodes. This high efficiency mode of operation is analyzed by means of a simplified model for both germanium and silicon avalanche diodes. The model evolved from a study of detailed computer simulations of experimental diode-circuit systems. The simplified analysis assumes: (i) The avalanche zone transit, plus the recovery time to the swept-out state, occurs in a half period of the TRAPATT frequency. (ii) The ratio of IMPATT frequency to TRAPATT frequency is 3:1. (iii) The diode area is chosen to provide 10 ohms negative resistance, a reasonable value for microwave circuits. The calculated electrical characteristics agree well with experimental observations. Consideration of circuit and thermal limitations results in a design for maximum power output for a millimeter wave silicon oscillator. Power output in excess of 1 watt CW, with an efficiency of 40 percent, is predicted at a frequency of 50 GHz.

1. INTRODUCTION

High efficiency oscillations in silicon avalanche diodes were first reported in April 1967 by Prager, Chang, and Weisbrod.¹ The new high efficiency mode was shown experimentally to be capable of converting dc to microwave power with efficiency greatly exceeding that predicted for the classical IMPATT mode.²⁻⁴ For example pulsed bias operation with 60 percent efficiency at about 750 MHz was reported.⁵ Normal IMPATT operation for this structure would be at about 4 GHz with perhaps 10 percent efficiency.

The first theoretical calculation⁶ of the high efficiency mode was obtained by computer simulation of carefully characterized experimental results⁷ on germanium structures. Pulsed bias operation, with output power of 7.5 watts at 2.6 GHz with efficiency of 40 percent was reported.⁸ The mode of oscillation was named TRAPATT, for TRapped Plasma Avalanche Triggered Transit, which describes the new physical mechanisms discovered by the computer simulations.

Characterization of experimental circuits⁹ which support TRAPATT oscillations and additional more detailed simulations¹⁰ of the diode-circuit system further verify the avalanche-triggered-trapped plasma description of the high efficiency mode.

In this paper a simplified model, previously described,^{11,12} for the physical processes occurring in TRAPATT operation is employed in order to analyze a wide range of structures and operating conditions. Detailed terminal voltage waveforms are calculated, using appropriate ionization rates, velocities, and other materials parameters, for germanium and silicon structures. The avalanche-zone-transit model, first described in Ref. 11, is presented in detail, and in a form suited for the calculation of the detailed spacial variation of carriers and electric field. Analysis of the diode recovery from the "trapped-plasma" state follows the treatment presented in Ref. 12. Solutions are obtained by applying a terminal current waveform with sharp wavefronts appropriately chosen to represent the effects of the multiresonant circuit to the diode terminals.^{9,10} A power generating cycle commences with the application of a large terminal current to a completely swept-out diode structure in which the only carriers available for impact ionization result from thermal generation via defect centers. A rapidly moving avalanche zone results which sweeps through the depletion region from the p-n junction to the heavily doped substrate-base interface. A dense electron-hole plasma is left behind the zone and constitutes a high conduction current essentially zero voltage state which lasts for approximately a half period. Recovery to the essentially zero conduction current-high voltage state is also controlled by terminal current waveform. The avalanche zone transit occurs in a time considerably less than the transit time of a carrier moving at saturated velocity. The sweep-out recovery period cannot be less than the transit time.

The TRAPATT phenomena described above cannot exist as a small signal effect. Instead it results as the natural (and ultimate) large signal limit of the chain of events which occur when a sufficiently large drive is applied to a swept-out avalanche diode. Obviously, a wide range of characteristics result over the range from small signal to large signal,

and between single resonance and multiresonance circuits. In this paper, we calculate only results in the range of the TRAPATT large-signal limit. Furthermore, to facilitate comparisons between germanium and silicon semiconductors, structure designs, bias levels, and operating frequencies, we present detailed results for only a particular terminal current waveform. This waveform, however, would be expected to permit the highest power output at a given frequency, and analysis of the output spectrum agrees with experimental circuit characteristics. This waveform is a perfect square wave of terminal current chosen so that the avalanche zone transient and the subsequent sweep-out of the generated electron-hole plasma occurs precisely over the positive half period. The negative half period is one of high reverse voltage and low saturation current.

The next two sections (II and III) are mathematical and develop the expressions* which are evaluated for a wide range of parameters in Section IV. Section II develops in detail the mathematical characteristics of the avalanche zone transient, which occurs when the step in current is applied to a completely swept-out diode. Section III is a development of the mathematical expressions which describe the recovery period back to the completely swept-out state. Section IV contains: (i) A summary of the expressions developed in Sections II and III, required for the complete analysis. (Readers not interested in mathematical detail may omit Sections II and III.) (ii) A discussion of the structure design which results as a requirement imposed by the particular current waveform. (iii) Results of efficiency, input power density, and diode impedance versus operating frequency for two typical structure designs. (iv) For particular optimum structures, results of depletion layer thickness, epitaxial impurity concentration, diode junction area, average voltage, average current, and output power versus operating frequency. Section V presents a discussion of these results, and their comparison with experimental information.

II. SIMPLIFIED MODEL FOR ZONE TRANSIENT

Selected frames of a computer simulation,¹³ shown in Figs. 1-3 illustrate the zone transit of TRAPATT operation. Shown are solutions for hole and electron concentrations and electric field profiles at successive instants of time. A zone transit begins, as shown in Fig. 1, when a large current charges a previously swept-out diode structure until impact avalanche generation of carriers begins. For circuit conditions,

* The results of the analysis are identical to those presented in Ref. 12, although the derivation follows from a different point of view.

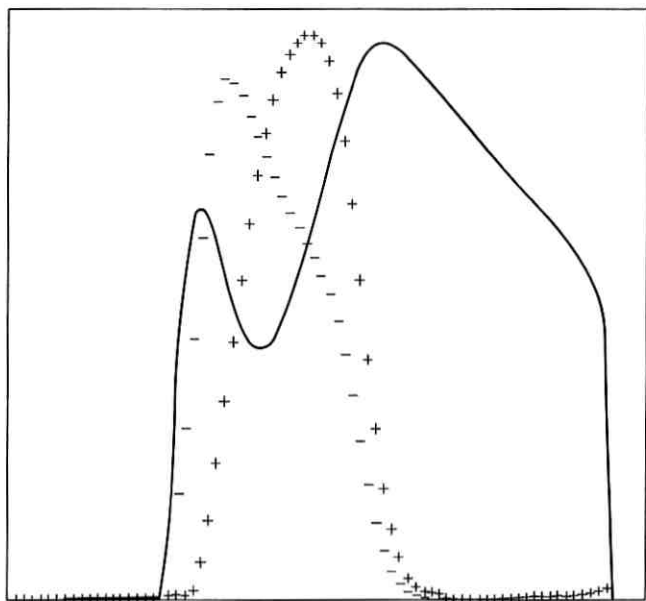


Fig. 1—Start of zone transient. Hole concentration (+), electron concentration (-), electric field (solid) versus distance.

which maintain this large current, the avalanche zone sweeps through the depletion region, as shown in Fig. 2. A dense electron-hole plasma is left, as shown in Fig. 3, after the completion of the zone transit. The computer simulations utilize a model for the semiconductor diode,⁵ which is quite general and has been quite successful in simulating large-signal operation of the various modes of power generation in avalanche diodes. Simulated solutions are obtained for the evolution in time of the diode variables (hole and electron concentrations and electric fields) and circuit variables (terminal current and voltage). The main disadvantage of the program is the cost, which generally prohibits a large-scale parameter investigation of the diode-circuit system.

In this section the expressions are derived which form the basis for a large-scale parameter study for the zone transient. The treatment is necessarily simplified, and it is assumed that holes and electrons have identical properties, except for sign, and in particular the ionization rate α and carrier velocity v are identical functions of the electric field \mathcal{E} . The one-dimensional continuity, transport, and space-charge equations are simplified to

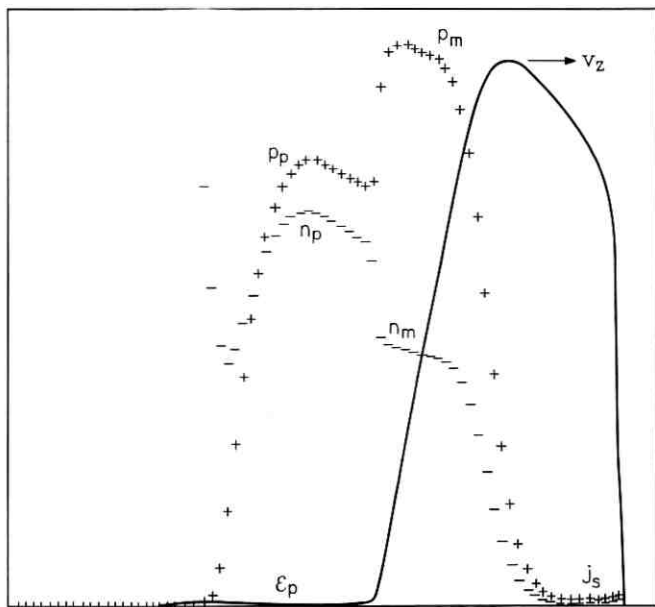


Fig. 2—Zone transient period. Indicated are multiplied carriers p_m and n_m , plasma concentrations p_p and n_p , and plasma field ϵ_p , resulting from initial concentration j_s and zone velocity v_z .

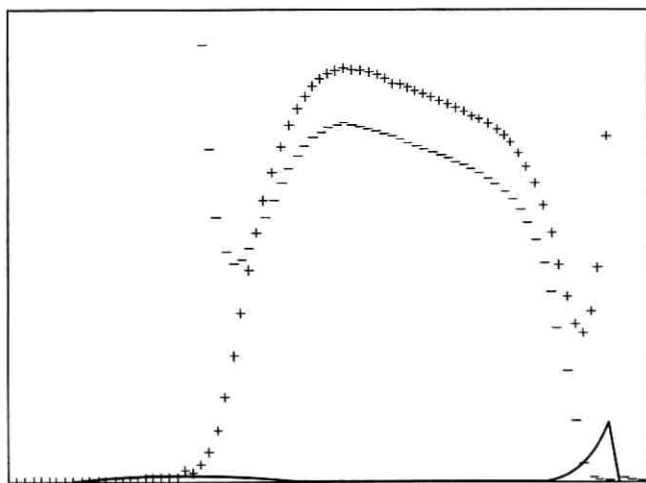


Fig. 3—Termination of zone transient.

$$q \frac{\partial p}{\partial t} = g_T + \alpha(j_p + j_n) - \frac{\partial j_p}{\partial x}, \quad (1)$$

$$q \frac{\partial n}{\partial t} = g_T + \alpha(j_p + j_n) + \frac{\partial j_n}{\partial x}, \quad (2)$$

$$j_p = qvp, \quad (3)$$

$$j_n = qvn, \quad (4)$$

$$\frac{\partial \mathcal{E}}{\partial x} = \frac{q}{\epsilon} (p - n + \text{Dop}), \quad (5)$$

in which diffusion transport has been neglected, g_T represents thermal generation via defects where significant, and Dop is the net ionized impurity concentration.

A reduction in the number of required equations can be obtained by differentiating the space-charge equation and substituting from the expression which results when the two continuity equations are added:

$$\epsilon \frac{\partial}{\partial x} \left(v \frac{\partial \mathcal{E}}{\partial x} \right) = q \frac{\partial}{\partial x} \left(\frac{j_p - j_n}{q} + v \cdot \text{Dop} \right), \quad (6)$$

$$\frac{\partial}{\partial t} \left(\frac{j_p + j_n}{v} \right) = 2g_T + 2\alpha(j_p + j_n) - \frac{\partial}{\partial x} (j_p - j_n) \quad (7)$$

or

$$\epsilon \frac{\partial}{\partial x} \left(v \frac{\partial \mathcal{E}}{\partial x} \right) = q \frac{\partial}{\partial x} (v \cdot \text{Dop}) + 2g_T + 2\alpha j_c - \frac{\partial}{\partial t} \left(\frac{j_c}{v} \right) \quad (8)$$

where j_c is the conduction current ($j_c = j_p + j_n$) and

$$j_c = J_T - \epsilon \frac{\partial \mathcal{E}}{\partial t}, \quad (9)$$

where J_T is terminal current density.

A single equation in \mathcal{E} results by substituting equation (9) into (8), but is of no particular advantage. The solution for \mathcal{E} and j_c is instead obtained from the following two equations, where doping gradients have been neglected:

$$\begin{aligned} \frac{\partial j_c}{\partial t} &= 2v(g_T + \alpha j_c) + \frac{j_c v'}{\epsilon v} (J_T - j_c) \\ &\quad - \epsilon v^2 \frac{\partial^2 \mathcal{E}}{\partial x^2} + v v' \frac{\partial \mathcal{E}}{\partial x} \left(q \text{Dop} - \epsilon \frac{\partial \mathcal{E}}{\partial x} \right), \end{aligned} \quad (10)$$

$$\epsilon \frac{\partial \mathcal{E}}{\partial t} = J_T - j_c. \quad (11)$$

Equation (10) is obtained from equation (8) by expressing $\partial v/\partial x$ in terms of an incremental mobility $v' = dv/d\varepsilon$, that is, $\partial v/\partial x = v'(\partial\varepsilon/\partial x)$, and $(\partial v/\partial t) = v'(\partial\varepsilon/\partial t)$.

2.1 Coordinate Transformation

The zone transient period of a TRAPATT oscillation is distinctively characterized by the rapid movement of an avalanche zone from the p-n junction to the base-substrate interface,^{11,12} in which computer simulations^{6,10,13} indicate that all quantities at a time $t_0 + \Delta t$ have profiles related to those existing at time t_0 , by the relationship

$$\varepsilon(x, t_0 + \Delta t) = \varepsilon(x - v_z \Delta t, t_0),$$

where the electric field is used for illustration. Incorporation of this "experimentally" observed phenomena into the mathematical model results in the elimination of all partial derivatives with respect to time, and replacing them with an ordinary special derivative. That is by making a change of variable $z = x - v_z \Delta t$,

$$\frac{\partial}{\partial t} \rightarrow -v_z \frac{d}{dz} \quad (12)$$

where v_z is the velocity of the zone and is assumed constant. Incorporation of transformation (12) into equations (10) and (11) yields

$$\frac{v_z}{v_s} \frac{dj_c}{dz} = \epsilon v_s \frac{d^2 \varepsilon}{dz^2} - 2v_s (g_T + \alpha j_c) \quad (13)$$

and

$$-\epsilon v_s \frac{d\varepsilon}{dz} = J_T - j_c, \quad (14)$$

for the case where the velocity is saturated at value v_s . The velocity of the zone is readily obtained^{11,12} from equation (14), when evaluated in the region in front of the avalanche, where $j_c \ll J_T$ and $d\varepsilon/dz = -(q/\epsilon)N_A$: $v_z = J_T/qN_A$. If thermal generation is neglected, once the avalanche zone is formed, equations (13) and (14) can be separated and integrated to yield the detailed structure of the avalanche zone for any $\alpha(\varepsilon)$ function, and most interestingly predicts the magnitude of multiplication of conduction current through the zone—independent of the exact form of the $\alpha(\varepsilon)$ function: Substitution into equation (13) of the derivative of equation (14) with respect to z yields

$$2\alpha j_c = \left(\frac{v_s}{v_z} - \frac{v_z}{v_s} \right) \frac{dj_c}{dz}. \quad (15)$$

Further substitution of dz from equation (14) yields

$$2\alpha j_c = \left(\frac{v_s}{v_z} - \frac{v_z}{v_s} \right) \frac{dj_c}{d\varepsilon} \left(\frac{j_c - J_T}{\epsilon v_z} \right),$$

which can be rearranged as

$$2\epsilon v_z \alpha(\varepsilon) d\varepsilon = \left(\frac{v_s}{v_z} - \frac{v_z}{v_s} \right) \left(\frac{j_c - J_T}{j_c} \right) dj_c. \quad (16)$$

With the condition that the conduction current in front of the zone has constant value j_s (as a result of thermal generation during the previous clear-out sequence), the relationship between value of electric field in the zone and conduction current is given by

$$\epsilon \int_0^{\varepsilon(i_c)} \alpha(\varepsilon) d\varepsilon = \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) \left[J_T \ln \left(\frac{j_c}{j_s} \right) - j_c + j_s \right] / (2v_z). \quad (17)$$

From equation (15) we see that v_z must be greater than v_s in order to multiply up conduction current from the value j_s . The total current multiplication j_{cm} across the entire zone is obtained by evaluating the upper limit of the left side of equation (17) also at zero field, in which case the left side equals zero and j_{cm} is given by

$$j_{cm} = j_s e^{(i_{cm} - i_s / J_T)}, \quad (18)$$

which shows that the effect of an increase in j_s , is a decrease in j_{cm} .¹¹

2.2 Individual Hole and Electron Concentrations

The individual hole and electron concentrations can be obtained from the value of j_c . The expressions are derived for two regions. One is anywhere within the zone where $v = v_s$, and the other is to the left of the zone where v may be less than v_s .

For $v = v_s$:

$$j_c = qv_s(p + n),$$

which can be solved simultaneously with

$$\frac{q}{\epsilon} (p - n - N_A) = \frac{j_c - J_T}{\epsilon v_z},$$

obtained by equating equations (5) and (14), as

$$p = \frac{j_c}{2q} \left(\frac{1}{v_s} + \frac{1}{v_z} \right); \quad (19)$$

$$n = \frac{j_c}{2q} \left(\frac{1}{v_s} - \frac{1}{v_z} \right). \quad (20)$$

Therefore if the detailed variation of j_c is determined, for a particular $\alpha(\mathcal{E})$ function, from equation (17), equations (19) and (20) yield the detailed profiles of carriers through the zone. In particular the values of holes and electrons, see Fig. 2, when $j_c = j_{cm}$ are simply

$$p_m = \frac{j_{cm}}{2q} \left(\frac{1}{v_s} + \frac{1}{v_z} \right), \quad (21)$$

$$n_m = \frac{j_{cm}}{2q} \left(\frac{1}{v_s} - \frac{1}{v_z} \right). \quad (22)$$

2.3 For $v \leq v_s$:

In this region impact ionization can be neglected and we return to the individual hole and electron continuity equations (1) and (2) written as

$$-v_z \frac{dp}{dz} = -\frac{d}{dz}(pw), \quad (23)$$

$$-v_z \frac{dn}{dz} = \frac{d}{dz}(nw), \quad (24)$$

where v is the carrier drift velocity. Inspection of the above equations reveals that any solution must satisfy the following constraints:

$$c_p = p(v_z - v), \quad (25)$$

$$c_n = n(v_z + v), \quad (26)$$

where c_p and c_n are constants. The validity of equations (25) and (26) can be checked by differentiation.

The value of c_p and c_n can be obtained by substitution of p_m and n_m from equations (21) and (22), and reveals that $c_p = c_n = c$:

$$c = \frac{j_{cm}}{2q} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right). \quad (27)$$

The values of p and n for any $v \leq v_s$ are given by

$$p = \frac{j_{cm}}{2q} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) / (v_z - v(\mathcal{E})), \quad (28)$$

$$n = \frac{j_{cm}}{2q} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) / (v_z + v(\mathcal{E})). \quad (29)$$

2.4 Trapped Plasma Solution

The profiles of holes, electrons and field in the transition region over which the carriers change from values p_m and n_m to their values in the trapped plasma, p_p and n_p , can be obtained from the simultaneous solution of equations (28) and (29) and equation (14) for any arbitrary $v(\mathcal{E})$ relation. However the solution in the trapped plasma is independent of the exact form of $v(\mathcal{E})$, and the plasma concentration of holes and electrons is even independent of the value of low field mobility.

In the trapped plasma the values of holes, electrons, and field (p_p , n_p , and \mathcal{E}_p) are related by

$$p_p - n_p = N_A, \quad (30)$$

$$J_T = qv_p(p_p + n_p), \quad (31)$$

$$v_p = \mu \mathcal{E}_p. \quad (32)$$

The solution for v_p is obtained from equations (28), (29) and (30):

$$v_p = \frac{j_{cm}}{2qN_A} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) \left\{ \left[1 + \left(\frac{2qN_A v_z}{j_{cm} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right)} \right)^2 \right]^{\frac{1}{2}} - 1 \right\}. \quad (33)$$

Once v_p is found, the values of p_p and n_p are given by

$$p_p = \frac{N_A}{2} \left(\frac{v_z}{v_p} + 1 \right), \quad (34)$$

$$n_p = \frac{N_A}{2} \left(\frac{v_z}{v_p} - 1 \right). \quad (35)$$

2.5 Discussion of Zone Transient

A simplified model for the zone transient which results from the transformation, equation (12), has been presented. In the numerical analysis the calculation begins at the point where the electric field is a maximum and $j_e = J_T$. Solutions are then obtained by the simultaneous solution of equations (16) and (14) for j_e and \mathcal{E} as functions of position to the left and to the right of this point. The hole and electron profiles are then obtained from equations (19) and (20). The value of plasma velocity v_p is obtained from a solution of equation (33), and the plasma density from equations (34) and (35).

It is of great significance to note that the solution in the trapped plasma is independent of the exact form of the $\alpha(\mathcal{E})$ and $v(\mathcal{E})$ relations and is determined uniquely by the value of j_e , N_A , and J_T .

A simplification for the trapped plasma solution results for the condi-

tion $v_p \ll v_s$. Equations (33), (34) and (35):

$$\frac{v_p}{v_s} = \frac{J_T}{j_{cm} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right)}, \quad (36)$$

$$p_p = \frac{N_A}{2} \left(\frac{j_{cm}}{J_T} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) + 1 \right), \quad (37)$$

$$n_p = \frac{N_A}{2} \left(\frac{j_{cm}}{J_T} \left(\frac{v_z}{v_s} - \frac{v_s}{v_z} \right) - 1 \right). \quad (38)$$

In the next section we present an analysis of the recovery period of the diode from the trapped plasma state to the state in which the diode is completely free (except for thermal generation) of mobile carriers.

III. RECOVERY TRANSIENT

In this section the transient solution is calculated for a plasma filled n^+p^+ diode after the application of constant reverse bias current J_R , which is a good approximation to the recovery portion of a TRAPATT oscillation. The analysis also applies to the devices treated by Benda and Spence for the reverse recovery in power rectifiers,¹⁴ and by Moll and Hamilton for step recovery diodes.¹⁵ However, the recovery from the large density plasma to a swept out state of large voltage, which occurs in TRAPATT operation, differs substantially from the earlier treatment^{14,15} which applies to low density plasmas and slow removal rates.

It is found that the removal rate of the electron-hole plasma and build-up of the electric field profile is quite accurately described by a simple model¹² which includes the effects of displacement current and carrier drift at constant mobility and saturated velocity. For a given p-region width the total recovery time can be as small as the transit time of the slower carrier if the plasma density is less than a critical value and the reverse current equals a critical value. For current greater than a critical value avalanche occurs before complete carrier removal. The first part of this section pertains to a more general system in which holes and electrons are not required to have equal characteristics; while the second part is for the special case of equal properties and couples the zone transient analysis with the recovery analysis.

3.1 Part 1

Figures 4a and 4b are frames from a computer simulation run to test the analysis to follow and will serve for notation.

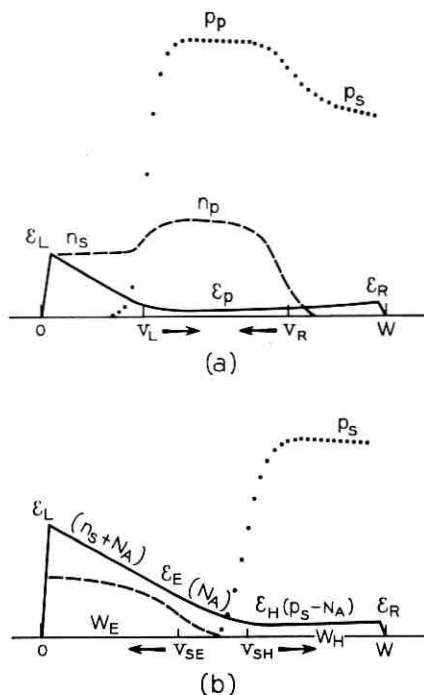


Fig. 4—Recovery transient. Hole concentration, electron concentration, and electric field versus distance. (a) Plasma velocity period and (b) saturated velocity period.

At $t = 0$ the diode has a uniform plasma p_p and n_p and the center region is characterized by:

$$p_p = N_A + n_p,$$

$$J_R = q\epsilon_p(\mu_p p_p + \mu_n n_p),$$

where ϵ_p is the field in the plasma μ_p and μ_n the hole and electron mobilities, and J_R the reverse or recovery current.

The solution for carrier density p_s , n_s and velocities v_L and v_R^* is obtained by an integration of the continuity equations over the transition region from the plasma to the regions where the field has built up to that required for saturated velocity:

$$q \frac{\partial}{\partial t} \int p \, dx = - \int dj_p$$

* v_L and v_R are the hole and electron velocities, respectively, in the plasma.

and

$$q \frac{\partial}{\partial t} \int n dx = \int dj_n$$

or

$$(p_p - p_s)v_R = v_{SH}p_s - \mu_n p_p \mathcal{E}_p,$$

$$(n_p - n_s)v_L = v_{SE}n_s - \mu_n p_p \mathcal{E}_p,$$

$$n_p v_R = \mu_n \mathcal{E}_p n_p,$$

$$p_p v_L = \mu_p \mathcal{E}_p p_p.$$

From the relations above:

$$v_R = \mu_n \mathcal{E}_p, \quad (39)$$

$$v_L = \mu_p \mathcal{E}_p, \quad (40)$$

$$p_s = p_p \frac{(v_R + v_L)}{(v_R + v_{SH})}, \quad (41)$$

$$n_s = n_p \frac{(v_L + v_R)}{(v_L + v_{SE})}, \quad (42)$$

where v_{SH} and v_{SE} are the saturated hole and electron velocities. The solution as shown in Fig. 4a, with values for v_L , v_R , p_s , n_s as given above (and verified by computer simulations), is valid until a time t_0 , when the plane moving at velocity v_L reaches the plane moving at velocity v_R :

$$t_0 = W/(v_R + v_L). \quad (43)$$

For $t \leq t_0$ the terminal voltage is given by

$$V = \frac{q}{\epsilon} \frac{t^2}{2} [v_R^2 p_s + v_L^2 n_s + N_A (v_L^2 - v_R^2)]. \quad (44)$$

For the remainder of the clear-out cycle refer to Fig. 4b. For times greater than t_0 three regions exist:

Region 1 extends from $x = 0$ to the plane moving at velocity v_{SE} . The field here has slope given by

$$\frac{\partial \mathcal{E}}{\partial x} = -\frac{q}{\epsilon} (N_A + n_s), \quad (45)$$

and the field \mathcal{E}_L has a magnitude given by

$$\epsilon \mathcal{E}_L / t = J_R - q v_{SE} n_s = q v_L (n_s + N_A). \quad (46)$$

Region 3 extending from the plane moving at velocity v_{SH} to $x = W$. The field here has slope given by

$$\frac{\partial \mathcal{E}}{\partial x} = \frac{q}{\epsilon} (p_s - N_A), \quad (47)$$

and the field \mathcal{E}_R has a magnitude given by

$$\epsilon \mathcal{E}_R / t = J_R - qv_{SH}p_s = qv_R(p_s - N_A). \quad (48)$$

Region 2, the remainder, has electric field with slope

$$\frac{\partial \mathcal{E}}{\partial x} = -\frac{q}{\epsilon} N_A. \quad (49)$$

The terminal voltage V for $t \geq t_0$ is obtained, with reference to Fig. 4b as follows.

$$\begin{aligned} V &= (\mathcal{E}_L + \mathcal{E}_E)/2 \cdot W_E, \\ &+ (\mathcal{E}_R + \mathcal{E}_H)/2 \cdot W_H, \\ &+ (\mathcal{E}_E + \mathcal{E}_H)/2 \cdot (W - W_E - W_H), \end{aligned}$$

or rearranged as

$$V = \frac{1}{2}[\mathcal{E}_L W_E + \mathcal{E}_R W_H + \mathcal{E}_E(W - W_H) + \mathcal{E}_H(W - W_E)], \quad (50)$$

where

$$W_E = v_L t_0 - v_{SE}(t - t_0),$$

$$W_H = v_R t_0 - v_{SH}(t - t_0)$$

\mathcal{E}_E and \mathcal{E}_H are given by equations (46) and (48) and

$$\mathcal{E}_E = \mathcal{E}_L - \frac{q}{\epsilon} (N_A + n_s) W_E,$$

$$\mathcal{E}_H = \mathcal{E}_R - \frac{q}{\epsilon} (p_s - N_A) W_H.$$

Equation (50) is valid until W_E or W_H equals zero, the time required to completely remove electrons or holes. The times required to completely clear-out electrons and holes are

$$\Delta t_E = \frac{W}{v_{SE}} \frac{n_p}{n_s}, \quad (51)$$

$$\Delta t_H = \frac{W}{v_{SH}} \frac{p_p}{p_s}, \quad (52)$$

where p_s and n_s are given by equations (41) and (42).

3.2 Minimum Recovery Time

If the recovery current J_R is chosen so that $v_R = v_{SE}$ and $v_L = v_{SH}$, that is,

$$J_R = q(v_{SH}p_p + v_{SE}n_p) \equiv J_c, \quad (53)$$

the clear-out will proceed at its most rapid rate. Any further increase in J_R will not speed up the recovery but will result in a $\partial\mathcal{E}/\partial t$ in the plasma which will result in an unnecessarily large value of \mathcal{E}_L or \mathcal{E}_R near the end of the recovery period and premature avalanche before clear-out. For $J_R = J_c$:

$$p_s = p_p,$$

$$n_s = n_p.$$

A constraint on p_p is obtained by requiring the field \mathcal{E}_L (obtained from equation (46) with $J_R = J_c$) to be less than the value obtained there in steady-state breakdown \mathcal{E}_0 , that is,

$$p_p < \epsilon\mathcal{E}_0/qW, \quad (54)$$

for $t = \Delta t_H$. Since $p_p > N_A$, equation (54) indicates that the diode must have a punch-through voltage less than the breakdown voltage. The computer simulation shown in Fig. 4 is for silicon with equal properties assumed for holes and electrons at 300°K, $W = 10$ microns, $p_p = 1.52 \times 10^{15}$, $n_p = 0.514 \times 10^{15}$, $N_A = 1.0 \times 10^{15}$, and $J_R = 1.5 \times 10^3$ A/cm².

3.3 Part II

For the remainder of this section it is assumed that holes and electrons have equal properties and therefore only one plasma velocity v_p replaces v_R and v_L , and one saturated velocity v_s replaces v_{SE} and v_{SH} in the previous equations. Furthermore, the plasma velocity obtained during the zone-transit period while current J_T flows is not the same as the plasma velocity characteristic of the recovery period while current J_R flows; unless $J_T = J_R$. Therefore the following notation is used:

$$v_{PT} = \text{velocity of plasma while current } J_T \text{ flows,}$$

$$v_{PR} = \text{velocity of plasma while current } J_R \text{ flows.}$$

A factor F is defined by reference to Fig. 5. The breakdown voltage of the diode is V_0 , the field at the n⁺p junction at breakdown is \mathcal{E}_0 , and

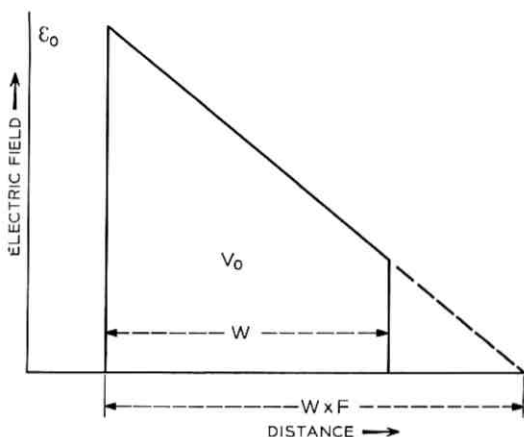


Fig. 5—Definition of punch-through factor F . Electric field versus distance at breakdown.

the width of the depletion region is smaller than the value $\epsilon\epsilon_0/qN_A$ by a factor $F > 1$. That is F is a measure of the degree of punch-through. (In the numerical calculation ϵ_0 and W are obtained by an iterative process for any given value of N_A and F .)

It is the purpose of this section to derive design requirements for TRAPATT operation by coupling the zone-transit analysis and the recovery analysis. It will be shown that the value of recovery current J_R which will result in complete removal of the plasma just as the diode voltage (and field profile) are the breakdown values V_0 and ϵ_0 is,

$$J_R = qN_A v_s (2F - 1);$$

and for the condition that the plasma be extracted in one transit time the following constraint must be imposed on the zone transit:

$$v_z/v_{PT} = 2F - 1.$$

The above constraint yields a unique value for J_T , for any given structure. This value is uniquely determined by material (Ge or Si), N_A , F , and J_S . For an additional constraint, for example that we minimize the sum of zone-transit time and recovery time or that this time period have a fixed relationship to the IMPATT transit time, a unique value of F can be determined. By such constraints then a unique (except for weak dependence on J_S) power-versus-frequency curve can be obtained for a given material. The derivation of the expressions required follows.

At the end of the zone transit the diode has plasma density given by equations (34) and (35). The clear-out densities p_s and n_s are

$$p_s = N_A \left(\frac{v_z}{v_{PT}} + 1 \right) \frac{v_{PR}}{v_{PR} + v_s}, \quad (55)$$

$$n_s = N_A \left(\frac{v_z}{v_{PT}} - 1 \right) \frac{v_{PR}}{v_{PR} + v_s}. \quad (56)$$

Utilizing equations (55) and (52) the electric field ε_L , obtained from equation (46), just as the carriers are cleared out has value

$$\varepsilon_L(\Delta t) = qv_{PR}(n_s + N_A) \left(\frac{Wp_p}{\epsilon v_s p_s} \right),$$

and when set equal to ε_0 (from Fig. 5 $\varepsilon_0 = (q/\epsilon)N_AFW$) yields

$$(n_s + N_A) \frac{v_{PR}P_p}{v_s P_s} = N_A F.$$

Further substitution from equations (41) and (56) yields

$$2F = 1 + \frac{v_z}{v_s} \left(\frac{v_{PR}}{v_{PT}} \right)$$

or utilizing equations (41) and (42)

$$J_R = qN_A v_s (2F - 1). \quad (57)$$

The equation above results solely from the requirement that right when the last remnants of the plasma are leaving the depletion region, the diode is at breakdown voltage V_0 . We have the interesting result that the degree of punch-through (F) *uniquely* determines—an upper limit on the reverse recovery current, independent of the zone transit current J_T , ionization rate, mobility or saturation current.

The critical current for removal of the plasma in one transit time is

$$J_c = qv_s(p_p + n_p).$$

For p_p and n_p given by equations (34) and (35),

$$J_c = qv_s N_A v_z / v_{PT},$$

and for the condition $J_R = J_c$,

$$\frac{v_z}{v_{PT}} = 2F - 1. \quad (58)$$

Equation (58) imposes a constraint on the zone transit process so that

the resulting plasma can be removed in one transit time. From the zone-transit analysis, in particular equation (33), a relationship was derived relating v_z , v_s , j_{cm} , N_A and v_{PT} . Therefore, combining equations (33) and (58) yields a relationship between F , N_A , v_s , J_T and j_s , which must be satisfied if the plasma is to be extracted in one transit time. However, the condition for maximum frequency is one which minimizes the sum of three characteristic times: the charging time t_1 , the zone transit time t_2 and the recovery time (compare Fig. 6).

The power output of an avalanche diode is limited by realizable circuit impedance or thermal limitations. Both have the effect of limiting the area of the diode. For a given upper limit on circuit impedance the best design is one which maximizes the product of power output times frequency squared. This figure of merit (F.M.) will have a maximum value determined by a judicious choice for J_T , J_R , the duration of time they are applied, and the duration of time the breakdown voltage-saturation current state persists. Such optimum conditions are not sought in the present analysis, inasmuch as after they were found it would probably be quite difficult to design a realistic circuit which could provide such a waveform. Instead, optimum values of F and current amplitude are found such that the F.M. is a maximum for perfect square wave current drive. (This is experimentally observed¹⁰ where the square wave is made up of a limited number of odd harmonics; the circuit is an open circuit at all even harmonics,

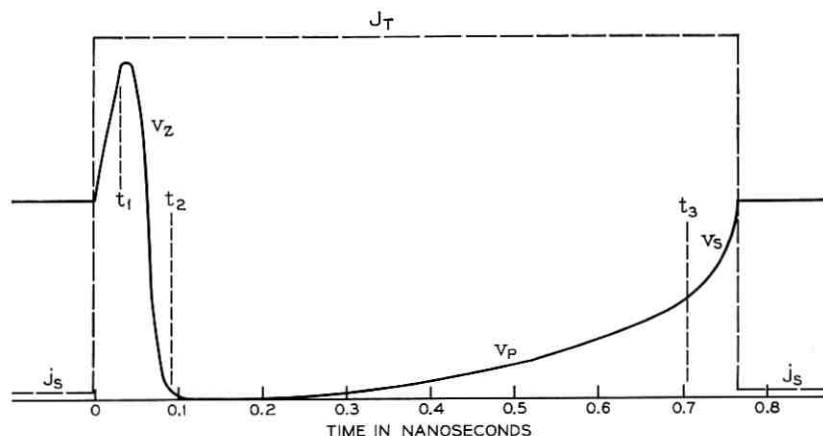


Fig. 6—Example of calculated voltage (solid) and assumed diode current (dashed) versus time. Zone transit (v_z), plasma velocity recovery (v_p), and saturated velocity (v_s), periods indicated.

the fundamental has impedance as calculated herein, and the third harmonic corresponds to the IMPATT frequency.) The next section reports results of a calculation for such a power-frequency characteristic.

IV. POWER OUTPUT AND STRUCTURE DESIGN

The application of terminal current J_T to a swept-out diode results in the zone transients shown in Figs. 1-3. In Section II we presented a considerably simplified analysis for the zone transient, as shown for example in Fig. 2. In Section III we analyzed the recovery transient, as shown in Figs. 4. A recovery current $J_R = qN_A v_s(2F - 1)$, equation (57), is required to completely remove the plasma in as short a time as possible, such that the diode voltage and field profile have just attained the breakdown values. The structure factor F and breakdown values are defined by Fig. 5. In the first part of this section results are presented on a per unit area basis. The voltage waveform, as shown for example in Fig. 6, is calculated for various values of N_A , ranging from 10^{15} cm $^{-3}$ to 10^{17} cm $^{-3}$, and values of F equal to 1.5 and 2.0. The current density J_T is determined by equation (57). The resulting waveform is Fourier analyzed, the frequency determined, and results presented in Figs. 7-9. In the second part, results are presented for an optimally chosen structure: (i) The value of F is found for each value of N_A such that the ratio of IMPATT frequency (π transit angle) to TRAPATT frequency is 3 : 1. (ii) The diode area is chosen to provide 10 ohms negative resistance, a reasonable value for microwave circuits. These results are presented in Figs. 10-15.

4.1 Part I

As shown in Section II the zone transient solution is logarithmically related to the thermally generated current j_s . In the calculations the current j_s is given by

$$j_s = qn_i W / \tau. \quad (59)$$

The intrinsic carrier concentration is evaluated at 300°C. for silicon and 100°C. for germanium, and the lifetime τ is 100 nsec. Effects of variations of j_s on device design are considered elsewhere.¹² The saturated velocity v_s is 8.9×10^6 cm/sec for silicon and 6.5×10^6 cm/sec for germanium. The ionization rate $\alpha = ae^{-b/E}$ has a and b 1.6×10^6 cm $^{-1}$, 1.9×10^6 V/cm for silicon and 3.5×10^5 cm $^{-1}$, 7.5×10^5 V/cm for germanium.¹⁶

The waveform of diode voltage for a particular structure is shown

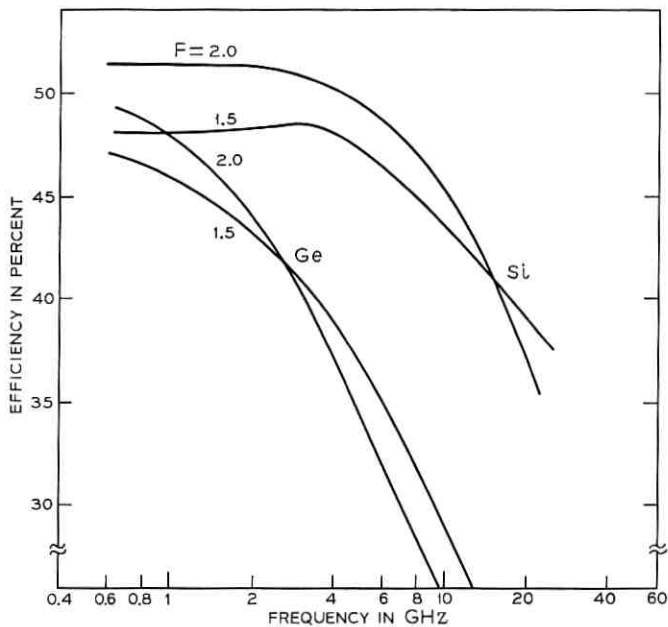


Fig. 7—Efficiency versus frequency, $F = 1.5$ and 2.0 .

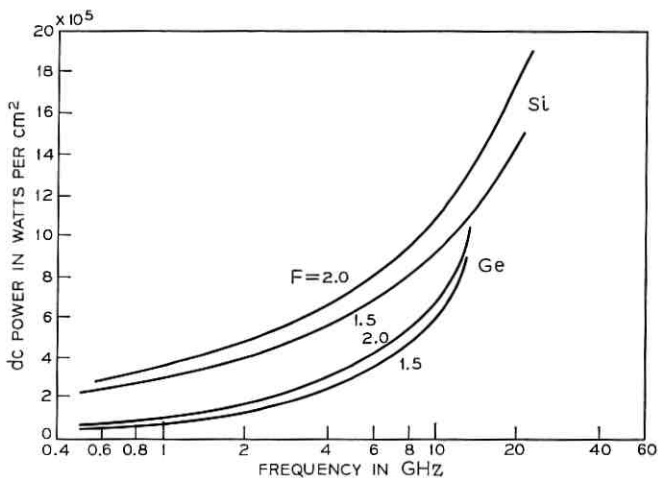


Fig. 8—Input (average) power density versus frequency, $F = 1.5$ and 2.0 .

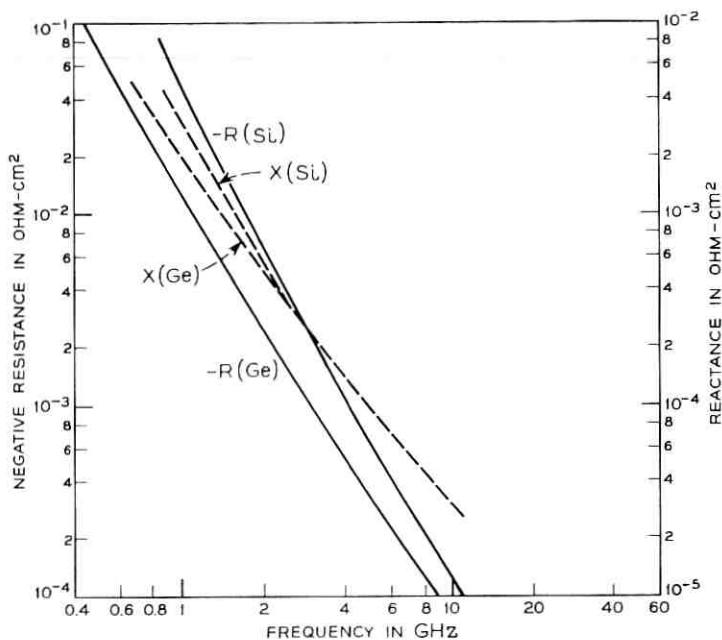


Fig. 9—Diode impedance versus frequency, $F = 1.5$.

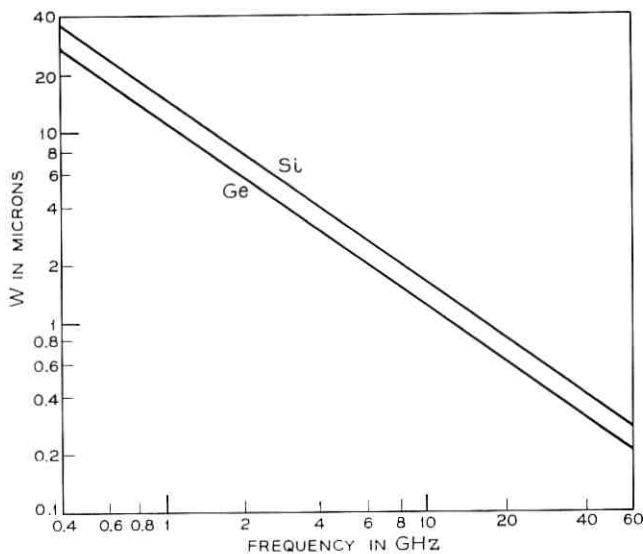


Fig. 10—Depletion layer thickness W versus frequency.

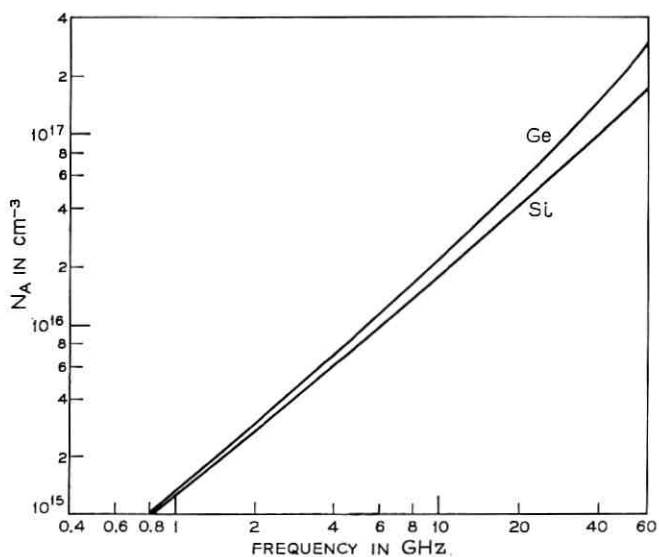
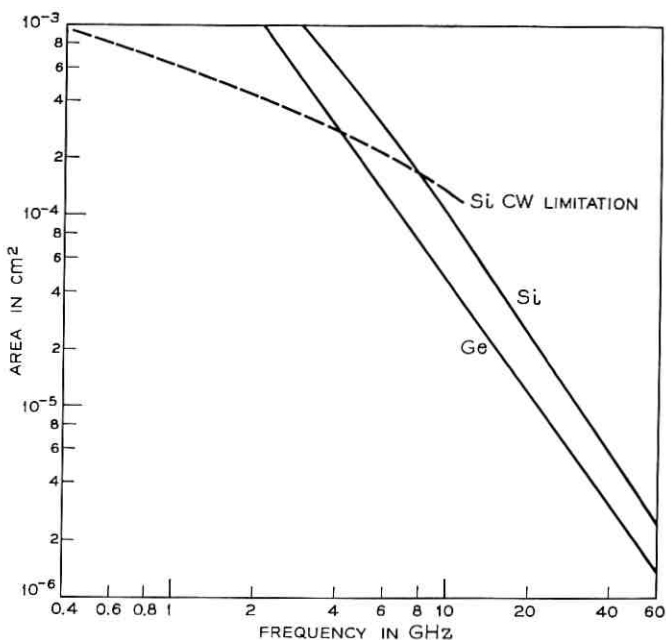
Fig. 11—Impurity concentration N_A versus frequency.

Fig. 12—Diode area versus frequency.

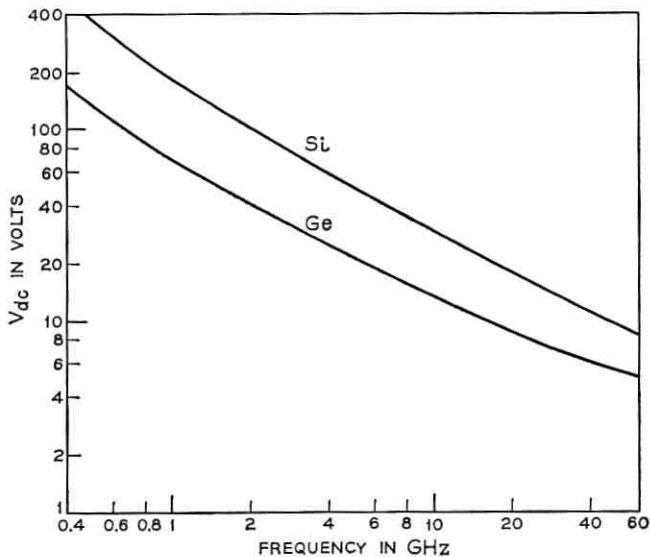


Fig. 13—Average voltage versus frequency.

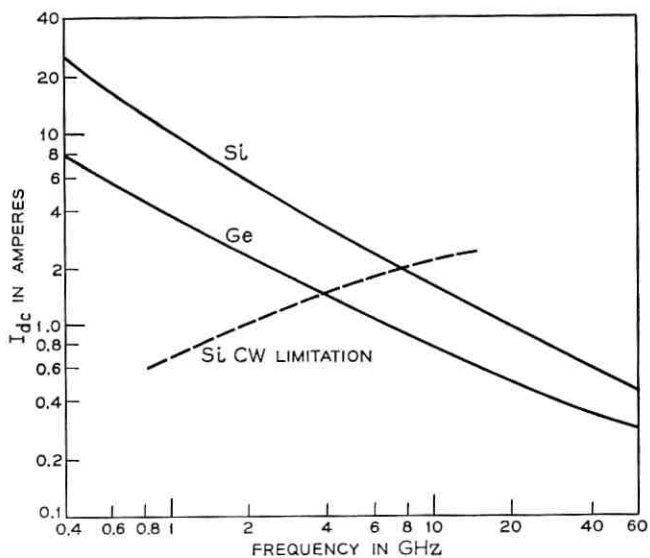


Fig. 14—Average current versus frequency.

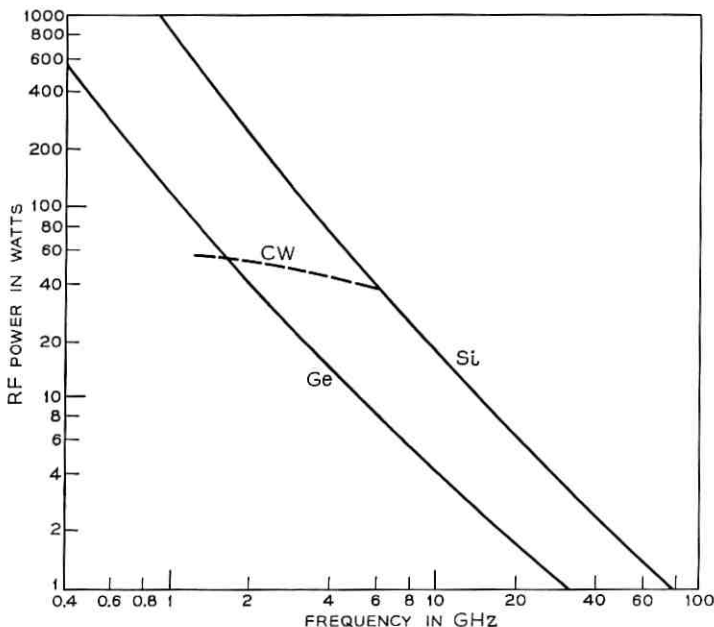


Fig. 15—Output power versus frequency.

in Fig. 6. The time t_1 is a depletion layer capacitance charging time, required to overdrive the diode to the point at which the avalanche zone transient begins. The zone transient lasts until time t_2 . The recovery transient characterized by the plasma velocity v_p lasts until time t_3 . The remainder of the recovery time, to t_4 , is characterized by carriers moving at saturated velocity. After complete sweep-out of carriers, the diode voltage equals the breakdown value and the diode current equals the saturation value j_s , for a half period of duration t_4 .

The voltage waveforms are calculated according to the simplified models presented in Sections II and III. The voltage for times between t_1 and t_2 , the zone transient, are calculated by "sliding" the field profile, shown for example in Fig. 2, from the point where the maximum field occurs at the N^+P junction, corresponding to time t_1 , to the right at velocity v_s . The time t_2 corresponds to the time when the location of the right-most point of the plasma region would enter the substrate. The voltage is the area under the field curve, within the confines of the depletion region. The voltage thus calculated is only an approximation at the beginning of this transient. However, the value of trapped

plasma left in the depletion region at time t_2 is independent of the actual avalanching conditions, and this plasma constitutes the initial conditions for the recovery voltage waveforms from time t_2 until t_4 . The voltage waveform in the neighborhood of time t_1 is determined by complicated space charge effects shown in Fig. 1. It is felt that detail occurring over such a short period of time would not significantly affect the overall power output and efficiency calculation, but this is one of the major limitations of this simplified analysis.

Results for two values of punch-through factor F are shown in Figs. 7 and 8. The frequency is calculated (for particular values of N_A) and results shown versus frequency. Figure 7 is a plot of calculated power conversion efficiency versus frequency; Fig. 8, dc power density versus output frequency. The diode impedance, as calculated by Fourier analysis of the waveforms, is shown in Fig. 9 for $F = 1.5$.

4.2 Part II

Results shown in Figs. 7-9 are on a per unit area basis, and no optimization for the F factor has been attempted. The area of an avalanche oscillator is determined by circuit impedance limitations and thermal dissipation limitations for CW operation. Results of the calculation are presented in Figs. 10-15 for an area chosen such that 10 ohms negative resistance at the TRAPATT frequency is obtained. In addition, the F factor was chosen by an iterative process for optimization of the diode structure. Initially optimization was sought for the criterion of maximization of the products of output power times operating frequency squared. Results clearly indicated that the maximum was for values of F only slightly larger than 1.0. For such cases the recovery time is very nearly one transit time, the minimum value. The total half period of the oscillation would be only slightly larger than two transit times, and therefore a ratio of IMPATT-to-TRAPATT frequency of 2 : 1 could be obtained. Instead results for a frequency ratio of 3 : 1 are presented. This will be somewhat less than the optimum, but more realistic, as this mode of operation is more easily obtained experimentally.^{9,10}

Device depletion layer thickness W , epitaxial impurity concentration N_A , and device area A , calculated for the 3 : 1 mode, are shown in Figs. 10-12, versus output frequency. Operating characteristics are presented in Figs. 13-15. Plotted versus frequency are average voltage, average current, and output power. As discussed previously, device area was chosen to provide 10 ohms negative resistance at the operating (TRAPATT) frequency. No other circuit limitations on area are in-

cluded, such as might be imposed by IMPATT considerations. Such considerations will be discussed later. The output power calculated for a silicon device, if the area is chosen for CW operation, is shown as a dashed line in Fig. 15. The choice of maximum device area permitting CW operation was obtained from silicon IMPATT diode thermal resistance and burnout analysis collected over the past several years at this laboratory.¹⁷ Based on this analysis, circuit impedance limitations determine device area for operating frequencies above about 10 GHz. In addition, device area may have to be reduced to provide sufficient negative resistance to overcome losses at the IMPATT frequency. This effect has not been considered in the present analysis.

V. DISCUSSION

We have presented calculated operating characteristics of the TRAPATT mode of high efficiency power generation in germanium and silicon avalanche diodes. The calculated results were obtained by analysis of solutions of a simplified model for the avalanche zone transit and reverse recovery periods which characterize the TRAPATT mode. The calculated solutions agree well with solutions obtained by detailed computer simulation studies, except for detail missing at the beginning of the zone transit, as the avalanche zone is forming.

Characteristics were presented, on a per unit area basis, for diodes with two degrees of punch-through, of the electric field at breakdown, into the terminating substrate. The silicon structures are more efficient than the germanium but require considerably more input power. For both materials input power requirements increase with F , the degree of punch-through, at a particular frequency.

A particular design was obtained by choosing F such that the TRAPATT frequency would be one-third the IMPATT (π transit angle) frequency of the structure. The calculated epitaxial impurity concentration and depletion layer width were presented. The calculated results are in excellent agreement with our experimental Ge structure ($f = 2.4$ GHz, $W = 4.5$ μm , $N_a = 4.2 \times 10^{15}$ cm^{-3}).^{8,10} Calculated results of power output versus frequency were presented for diode areas chosen to yield 10 ohms negative resistance at the TRAPATT frequency, and for silicon diodes with areas determined by thermal limitations. Power output in excess of 1 watt CW, with efficiency of 40 percent is predicted at a frequency of 50 GHz.

VI. ACKNOWLEDGMENTS

The work presented here evolved from computer simulations utilizing a detailed semiconductor program developed by the author and H. K. Gummel, and from the careful experimental characterization of TRAPATT waveforms by R. L. Johnston and precise circuit characterization by W. J. Evans. The author gratefully acknowledges many discussions with those named above and D. J. Bartelink, B. C. DeLoach, Jr., and R. M. Ryder, in which the understanding of the TRAPATT modes developed. The analytical treatments of the avalanche zone and reverse recovery transients, in part or in total, has been carefully and independently derived by D. J. Bartelink and B. C. DeLoach, Jr.

REFERENCES

1. Prager, H. J., Chang, K. K. N., and Weisbrod, S., "High Power, High Efficiency Silicon Avalanche Diodes at Ultrahigh Frequencies," *Proc. IEEE*, *55*, No. 4 (April 1967), pp. 586-587.
2. Read, W. T., "A Proposed High-Frequency Negative-Resistance Diode," *B.S.T.J.*, *37*, No. 2 (March 1958), pp. 401-466.
3. DeLoach, Jr., B. C., and Johnston, R. L., "Avalanche Transit-Time Microwave Oscillators and Amplifiers," *IEEE Trans. on Electron Devices*, *ED-13*, No. 1 (January 1966), pp. 181-186.
4. Scharfetter, D. L., and Gummel, H. K., "Large-Signal Analysis of A Silicon-Read Diode Oscillator," *IEEE Trans. on Electron Devices*, *ED-16*, No. 1 (January 1969), pp. 64-77.
5. Prager, H. J., Chang, K. K. N., and Weisbrod, S., "Anomalous Silicon Avalanche Diodes for Microwave Generation," *Proc. Cornell Conf. on High Frequency Generation and Amplification*, August 29-31, Ithaca, N. Y., pp. 266-280.
6. Scharfetter, D. L., Bartelink, D. J., Gummel, H. K., and Johnston, R. L., "Computer Simulation of Low Frequency High Efficiency Oscillation in Germanium" *IEEE Solid State Device Research Conference*, June 17-19, 1968.
7. Johnston, R. L., and Scharfetter, D. L., "Low Frequency High Efficiency Oscillations in Germanium Impatt Diodes," *IEEE Solid State Device Research Conference*, June 17-19, 1968; *IEEE Trans. El. Dev.*, *ED-16*, No. 11 (November 1969), pp. 905-911.
8. Johnston, R. L., Scharfetter, D. L., and Bartelink, D. J., "High-Efficiency Oscillations in Ge Avalanche Diodes Below the Transit-Time Frequency," *Proc. of IEEE*, *56*, No. 9 (September 1968), pp. 1611-1613.
9. Evans, W. J., "Circuits for High Efficiency Avalanche Diode Oscillators," 1969 *IEEE G-MTT International Microwave Symposium*, May 5-7, 1969, Dallas, Texas, *IEEE, G MTT-17*, No. 12 (December 1969), pp. 1060-1067.
10. Evans, W. J., and Scharfetter, D. L., "Characterization of Avalanche Diode Trapatt Oscillators," *IEEE, GED-17*, No. 5 (May 1970).
11. Bartelink, D. J., and Scharfetter, D. L., "Avalanche Shock Fronts in p-n Junctions," *Appl. Phys. Lett.*, *14*, No. 10 (May 15, 1969), pp. 320-323.
12. DeLoach, Jr., B. C., and Scharfetter, D. L., "Device Physics of Trapatt Oscillators," *IEEE Trans. El. Dev.*, *ED-17*, No. 1 (January 1970), pp. 9-21.
13. Scharfetter, D. L., "High Efficiency Operation of Impatt Diodes," *Int. Electron Devices Meeting*, October 23-25, 1968, Washington, D. C.

14. Benda, H., and Spenke, E., "Reverse Recovery in Silicon Power Rectifiers," Proc. of IEEE, 55, No. 8 (August 1967), pp. 1331-1354.
15. Moll, J. L., and Hamilton, S. A., "Physical Modeling of the Step Recovery Diode for Pulse and Harmonics Generation Circuits," Proc. of IEEE, 57, No. 7 (July 1969), pp. 1250-1259.
16. Dunn, C. N., and Decker, D. R., unpublished work.
17. Ciccolella, D. F., unpublished work.

An Integral Charge Control Model of Bipolar Transistors

By H. K. GUMMEL and H. C. POON

(Manuscript received December 17, 1969)

We present in this paper a compact model of bipolar transistors, suitable for network analysis computer programs. Through the use of a new charge control relation linking junction voltages, collector current, and base charge, the model includes high injection effects. The performance substantially exceeds that of existing models of comparable complexity. For low bias and with some additional idealization, the model reduces to the conventional Ebers-Moll model.

I. INTRODUCTION

Since its formulation in 1954, the Ebers-Moll model¹ has been the major large-signal model for bipolar transistors. It is based directly on device physics and covers all operating regimes, that is, active, saturated, and cut-off operation. But various approximations limit the accuracy of the model. In the original paper the frequency dependence was described in terms of frequency-dependent current generators. For time-domain analysis, this required use of Laplace transforms.

In 1957 Beaufoy and Sparkes² analyzed the bipolar transistor from a charge control point of view. The charge control model² or the equivalent^{3,4} charge control form of the Ebers-Moll model, is directly useful for transient analysis. It is this model, in various forms, that is presently used in all the major general purpose network analysis programs. More elaborate models⁵⁻⁷ are useful for detailed device studies, but they have not found widespread use in analysis programs due to their complexity.

As device technology evolved over the years making possible devices of reproducible characteristics, and as deeper understanding was gained in device-theoretical studies, many new effects were identified that are not represented by the basic Ebers-Moll model. Among these are a finite, collector-current-dependent output conductance due to basewidth modulation⁸ (Early effect), space-charge-layer generation and recomb-

nation⁹ (Sah-Noyce-Shockley effect), conductivity modulation in the base¹⁰ (Webster effect) and in the collector¹¹ (Kirk effect), and emitter crowding.^{12,13}

Some of these effects have been included into the Ebers-Moll model in various ways. The common approach has been to specify some parameters of the model as functions of bias and to describe this bias dependence in tabular form or through parametric expressions. Thus, in the NET-1 program, for example,¹⁴ the common-emitter current gain is given by a series expansion in the emitter-base voltage. In the CIRCUS program,¹⁵ forward and reverse current gains and forward and reverse transit times are specified as functions of collector current in tabular form. Such "curve-fitting" modeling tends to require large numbers of parameters or table entries for an accurate description. Also, frequently the parameters are not easily interpretable in terms of the device structure and thus can be obtained only *a posteriori*, from detailed measurements, but cannot be predicted.

The present paper makes use of a new, general charge-control relation¹⁶ which links junction voltages, collector-current, and base charge. This new charge-control relation, used in conjunction with conventional charge-control theory, allows many of the effects not contained in the basic Ebers-Moll model to be incorporated in an integral, physical way, in compact form and with good parameter economy.

In the conventional treatment of the transport equation, as used in the derivation of the Ebers-Moll model, attention is focused on recombination in the base (which in modern transistors is generally *not* the dominant source of base current). Through the use of simplified "boundary conditions" relating carrier concentrations at the edges of the base to the junction voltages, and through neglect of the variation of the electric field in the base with bias, the voltage dependence of emitter and collector currents contains errors, except near zero bias.

The new charge control relation arises from the treatment of the transport equation for the carriers that pass between emitter and collector. Use is made of the fact that recombination has only a very small effect on the junction-voltage dependence of the current passing from emitter to collector (later called the dominant current component). Hence for this dependence, but of course not for the base current, recombination is neglected. A direct, closed-form solution of the transport equation from inside the emitter to inside the collector is then possible. An important feature of the result, quoted in equation (10), is a denominator containing the base charge, that is the charge of all those carriers that communicate with the base terminal.

With the concentration in the base of the carriers carrying the dominant current component known, recombination in the base could be computed, and the small effect that this recombination has on the dominant current component could be applied as a correction.¹⁷ Actually, however, for typical transistors this correction is negligible. The base current is modeled by empirical relations since detailed recombination parameters are not available.

The general features of the new model are presented in Section II. Details of a particular implementation using 21 model parameters are given in Section III.* Table I contains a list of the model parameters. Appendix A describes the treatment of base push-out. Discussion of some aspects of the model, and a set of computed characteristics of a sample transistor are given in Section IV. A compact summary of the model is given in Appendix B. A discussion of the significance of the model parameters and of default values is contained in Appendix C. This paper considers mainly the dc and low-frequency aspects of the model. Use of the model for ac or transient analysis will be presented elsewhere.

II. GENERAL CONSIDERATIONS

As a point of departure we will consider the Ebers-Moll equations.¹ They may be written in the form

$$\begin{pmatrix} I_e \\ I_c \end{pmatrix} = [T] \begin{pmatrix} \exp(qV_{eb}/kT) - 1 \\ \exp(qV_{cb}/kT) - 1 \end{pmatrix} \quad (1)$$

where T is a symmetric matrix of coefficients that are constant, that is, bias independent. At the time when the Ebers-Moll model was developed, attainable base widths were large by today's standards, and in order that useful current gains could be obtained lifetime in the device had to be long. Reverse saturation currents were used as indicators of lifetime. These circumstances are reflected in the notation used in Ref. 1 for the elements of T :

$$T = \begin{pmatrix} \frac{I_{e0}}{1 - \alpha_n \alpha_i} & -\frac{\alpha_i I_{c0}}{1 - \alpha_n \alpha_i} \\ -\frac{\alpha_n I_{e0}}{1 - \alpha_n \alpha_i} & \frac{I_{c0}}{1 - \alpha_n \alpha_i} \end{pmatrix} \quad (2)$$

where I_{e0} and I_{c0} are emitter and collector reverse saturation currents

* A simplified implementation of model, and a comparison with measurements, have been presented in the Digest of Technical Papers, 1970 IEEE International Solid State Circuits Conference, Paper 7.1.

TABLE I—MODEL PARAMETERS

Group 1: Knee parameters and transit times

- I_k Knee current*
 V_k Abs. value of knee voltage, in unit of kT/q
 τ_f Forward tau (forward transit time)
 r_t Tau ratio (ratio of reverse to forward transit time)

Group 2: Base Current

- i_1 Ideal base current coefficient
 i_2 Nonideal base current coefficient
 n_e Forward base current emission coefficient
 i_3 Reverse base current coefficient
 n_c Reverse base current emission coefficient

Group 3: Emitter Capacitance

- $P_e \begin{cases} v_{oe} & \text{Abs. value of emitter offset voltage, in units of } kT/q \\ m_e & \text{Emitter grading coefficient} \\ a_{e1} & \text{Emitter zero bias capacitance coefficient} \\ a_{e2} & \text{Emitter peak capacitance coefficient} \end{cases}$

Group 4: Collector Capacitance

- $P_c \begin{cases} v_{oc} & \text{Abs. value of collector offset voltage, in units of } kT/q \\ m_c & \text{Collector grading coefficient} \\ a_{c1} & \text{Collector zero bias capacitance coefficient} \\ a_{c2} & \text{Collector peak capacitance coefficient} \end{cases}$

Group 5: Base Push-out

- v_{rp} Abs. value of base push-out reference voltage, in units of kT/q
 r_w Effective base width ratio
 r_p Base push-out transition coefficient
 n_p Base push-out exponent

Auxiliary Quantities

$$\begin{aligned}
 e_k &= \exp(-v_k) \\
 e_{ke} &= \exp(-v_k/n_e) \\
 e_{kc} &= \exp(-v_k/n_c)
 \end{aligned}$$

* Quantity is negative for npn transistor.

and α_n and α_i are forward and reverse common base current gains. A simpler and more appropriate but fully equivalent notation is

$$T = \begin{pmatrix} (1 + 1/\beta_f)I_s & -I_s \\ -I_s & (1 + 1/\beta_r)I_s \end{pmatrix}. \quad (3)$$

Here β_f and β_r are forward and reverse common emitter current gain and I_s is the "intercept" current, that is, the current obtained when on a semilog plot of I_c vs. V_{cb} the collector current is extrapolated to $V_{cb} = 0$. The notation of equation (3) is considered more appropriate since the intercept current I_s is nearly independent of current gains. The matrix elements T_{12} and T_{21} in equation (2) show an apparent dependence on lifetime through the forward and reverse current gains.

Actually, this dependence is nearly cancelled by that of I_{e0} and I_{c0} . To a very good approximation the intercept current I_s depends only on the total number of impurities in the base.¹⁸

Equation (1) with matrix T given by equation (3) suggests the following interpretation: The emitter and collector current have a common, dominant component

$$I_{cc} = -I_s[\exp(qV_{cb}/kT) - \exp(qV_{cb}/kT)]. \quad (4)$$

In addition the emitter and collector currents have each a separate component I_{be} and I_{bc} , which is proportional to the reciprocal forward and reverse common emitter current gain:

$$I_{be} = (I_s/\beta_f)[\exp(qV_{cb}/kT) - 1], \quad (5)$$

$$I_{bc} = (I_s/\beta_r)[\exp(qV_{cb}/kT) - 1]. \quad (6)$$

The terminal currents are then given by

$$I_e = -I_{cc} + I_{be}, \quad (7)$$

$$I_c = I_{cc} + I_{bc}, \quad (8)$$

$$I_b = -I_{be} - I_{bc}. \quad (9)$$

So far, we have made no changes in the content of the Ebers-Moll equations; we have only brought them into a form that will facilitate the development of the new model. We shall retain equation (7) through (9)* but we will replace equations (4) through (6) by relations giving an improved representation of the physical processes in the transistor.

The separation of emitter and collector currents into the dominant I_{cc} component and the base current components in equations (7) and (8) allows us to give different voltage dependence to the individual components. For example, at low injection levels, collector current and emitter-base voltage are related through the "ideal" diode law; that is, the collector current is proportional to $\exp(qV_{cb}/nkT)$ where the "emission coefficient" n is very close to unity. The base current at low forward bias, on the other hand, is typically a "nonideal" current,¹⁹ that is it has an emission coefficient n with values typically between 1.5 and 2. This nonideal current results from space charge recombination⁹ or surface recombination, or the presence of both effects. At higher forward emitter-base voltages the base current is dominated by an ideal component.¹⁹

In principle, it is possible to compute the base current as a function

* These equations are not independent. Equation (7) will be dropped later.

of V_{eb} (for given V_{cb}) for one-dimensional structures,^{17,20-22} provided that the doping profile and the recombination parameters, for example the concentration as a function of distance of the important species of recombination centers, as well as their energy levels and capture cross sections, are known. However, in practice, the recombination properties are not known to the detail required for such calculations. Even the assumption of a constant concentration of one species of recombination centers is a gross oversimplification. In real transistors the lack of lattice perfection in heavily doped regions causes enhanced recombination, and interfaces between substrates and epitaxial layers provide local regions of high recombination.

Nevertheless, detailed studies have confirmed that the base current can be described by a sum of terms exponential in voltage with emission coefficients of the magnitude indicated above. Equations (5) and (6) are thus replaced by more general functions of V_{eb} and V_{cb} that are characterized by pre-exponential coefficients and emission coefficients which become model parameters and which represent the overall recombination properties and influence the dependence of forward and reverse current gain on bias. In the next section we will use a specific set of base current components.

The new charge control relation¹⁰, derived from basic physical considerations, permits the dominant current components I_{cc} to be obtained without use of low-injection approximations:

$$I_{cc} = -I_s Q_{b0} \frac{\exp(qV_{eb}/kT) - \exp(qV_{cb}/kT)}{Q_b}. \quad (10)$$

Here Q_b is the "base charge", that is, the charge of all carriers of the type that communicate with the base terminal: electrons in a pnp transistor and holes in an npn transistor. The base charge is a function of bias; Q_{b0} is the zero-bias value. The bias dependence of Q_b is the subject of the conventional charge control theory¹² which approximates the excess (or stored) base charge $Q_b - Q_{b0}$ as the sum of capacitive contributions and products of collector current components and transit times. The essential feature of the model presented here is the simultaneous use of conventional charge-control theory and relation (10). As shown before,¹⁰ relation (10) is a good approximation for one-dimensional transistors for a wide range of bias conditions, including those leading to high-injection and base push-out effects.

It is of interest to note that the Ebers-Moll equations embody *superposition*, that is that the collector current can be expressed as the sum of a function of the emitter voltage and a function of the collector

voltage. For real transistors, violations of the superposition principle are easily observed. Consider, for example, the Early effect⁸ that is, the dependence of the low-frequency output conductance on bias. As shown schematically in Fig. 1, a region of bias exists in which the collector current varies approximately linearly with collector-emitter voltage for fixed base current, in such a way that the straight-line sections, when extrapolated, intersect (approximately) at a negative voltage which we shall call the "Early voltage", V_A . For superposition to be valid, the lines would have to be nearly parallel to each other. The base charge Q_b in the denominator of equation (10) through its dependence on collector voltage via the collector capacitance disables superposition and provides a realistic description of the output conductance.

Another point of interest concerns high-injection effects in the base region. The "ideal" voltage dependence of I_{cc} on V_{cb} is caused primarily by the dependence of the minority carrier concentration in the base near the emitter as $\exp(qV_{cb}/kT)$. This dependence holds, however, only as long as the minority carrier concentration is small compared to the doping concentration. If the minority carrier (the word minority starts to lose its literal meaning here) concentration is large compared to the doping concentration, then it varies as $\exp(qV_{cb}/2kT)$, and so does, approximately, I_{cc} except for additional complications due to

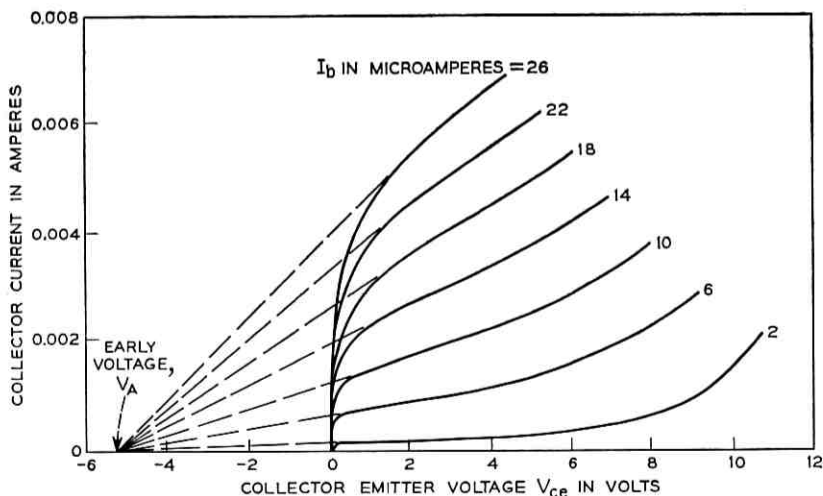


Fig. 1—Definition of Early voltage.

base push-out. The intersection of the $n = 1$ and $n = 2$ asymptotes to I_{cc} represents an important characteristic feature of the transistor. We shall call it the "knee point" and use its coordinates, V_k and I_k as model parameters and as a basis for normalizations.*

Consolidating the development up to this point, the model, exclusive of parasitic effects, is described by

$$I_c = -I_s Q_{bo} \frac{\exp(qV_{eb}/kT) - \exp(qV_{cb}/kT)}{Q_b} + I_{bc}, \quad (11)$$

$$I_b = -I_{be} - I_{bc}. \quad (12)$$

As discussed above, the base current components I_{be} and I_{bc} depend strongly on the recombination properties of the structure and are in practice not readily calculated from first principles. By contrast, the base charge as a function of bias depends primarily on the doping profile and is nearly independent of recombination properties. Hence, given the doping profile, Q_b as a function of V_{eb} and V_{cb} can be computed by existing techniques. However, for such a calculation considerable computer resources (memory and time) are required. For network analysis programs it is preferred to approximate Q_b by simple algebraic or algorithmic (implicit functions) representations which, depending on complexity, can give reasonable accuracy. One such representation is given in the next section. Special features are use of a modified representation of junction capacitance²³ which avoids the problem of an infinite capacitance when the junction voltage equals the built-in voltage, and use of a 4-parameter representation of base push-out.

III. SPECIFIC MODEL IMPLEMENTATION

In this section we present a specific implementation of the general model, equations (11) and (12). The bias dependence of base charge and base current will be modeled. The polarity assumed is that of a pnp transistor.

The dominant current component I_{cc} may be separated into an emitter and a collector component, or a forward and reverse component:

$$I_{cc} = -I_s Q_{bo} \frac{\exp(qV_{eb}/kT) - 1}{Q_b} + I_s Q_{bo} \frac{\exp(qV_{cb}/kT) - 1}{Q_b} \equiv -I_f + I_r. \quad (13)$$

* The exact definition of these parameters is given in the following Section.

We then model the excess base charge as consisting of emitter and collector capacitive contributions Q_e and Q_c , and of forward and reverse current-controlled contributions*

$$Q_b = Q_{b_0} + Q_e + Q_c - \tau_f B I_f - \tau_r I_r. \quad (14)$$

Here τ_f and τ_r are forward and reverse transit times. The coefficient B has been included to model the increase of the transit time when base push-out occurs; it has a value of unity in the absence of base push-out.

For modeling these parameters it is convenient to normalize all charges in equation (14) with respect to the zero bias charge Q_{b_0} and to denote the normalized charges by lowercase symbols. Also, we replace I_f and I_r according to equation (13). Then

$$\underbrace{1 + q_e + q_c + \frac{1}{q_b} \frac{I_s}{-Q_{b_0}} \{ \tau_f B [\exp(qV_{cb}/kT) - 1] + \tau_r [\exp(qV_{cb}/kT) - 1] \}}_{q_1} + \frac{1}{q_b} q_2. \quad (15)$$

Multiplication of equation (15) by q_b removes q_b from the denominator of the last term on the right side of equation (15) and gives rise to a quadratic equation in q_b . Its solution gives q_b explicitly in terms of the junction voltages, except for a possible q_b -dependence of B :

$$q_b = \frac{q_1}{2} + \left[\left(\frac{q_1}{2} \right)^2 + q_2 \right]^{\frac{1}{2}}. \quad (16)$$

The term q_1 represents the sum of the zero-bias charge and the charge associated with the junction capacitances; q_2 represents the excess base charge, or the current-dependent charge associated with diffusion capacitances. The latter charge contains a dependence on the base push-out effect through the parameter B which is treated in Appendix A.

For high forward bias the charge q_2 is the dominant component of the base charge q_b . Except for the base push-out term B , it is characterized by four parameters: I_s , Q_{b_0} , τ_f , and τ_r . It will be convenient to normalize these parameters. For this we define the knee voltage V_k as the emitter voltage for which q_2 equals unity (for zero collector voltage and with neglect of terms small compared to the exponential of the emitter voltage):

* The minus sign arises because the base charge contains electrons neutralizing the positive charges $\tau_f B I_f$ and $\tau_r I_r$. Q_b , Q_{b_0} , Q_e , and Q_c are all negative quantities for positive V_{cb} and V_{cb} .

$$V_k = \frac{kT}{q} \ln \frac{-Q_{b0}}{I_k \tau_f}. \quad (17)$$

The low-injection-extrapolated collector current for $V_{eb} = V_k$ is

$$I_k = \frac{-Q_{b0}}{\tau_f}. \quad (18)$$

It will be convenient to normalize all quantities having dimensions of current with respect to I_k , and to express voltages by their difference from V_k , in units of kT/q . We shall use lower-case symbols for normalized quantities. Thus, we define

$$v_k = \frac{qV_k}{kT}, \quad (19)$$

$$v_e = \frac{qV_{eb}}{kT} - v_k, \quad (20)$$

$$v_c = \frac{qV_{cb}}{kT} - v_k, \quad (21)$$

$$e_k = \exp(-v_k). \quad (22)$$

With these normalizations, equations (11) and (12) become

$$I_c = I_k i_c = I_k \left(-\frac{e^{v_e} - e^{v_c}}{q_b} + i_{bc} \right), \quad (23)$$

$$I_b = I_k i_b = I_k (-i_{be} - i_{bc}). \quad (24)$$

For the base charge we obtain

$$q_2 = B(e^{v_e} - e_k) + (\tau_r/\tau_f)(e^{v_c} - e_k), \quad (25)$$

$$q_b = q_1/2 + [(q_1/2)^2 + q_2]^{\frac{1}{2}}, \quad (26)$$

$$Q_b = -I_k \tau_f q_b. \quad (27)$$

With these normalizations, we can use the set $I_k, v_k, \tau_f, r_i \equiv \tau_f/\tau_r$ as the four model parameters describing q_2 for the case $B = 1$. These four parameters constitute Group 1 of the model parameters listed in Table 1.

The charge contribution from the emitter and collector junction capacitances will be considered next. The conventional representation of junction capacitance is through an expression containing three parameters:

$$C = \frac{\text{const}}{(V - V_{\text{built-in}})^m}. \quad (28)$$

The parameters are $V_{\text{built-in}}$ (which for silicon is typically $\approx 0.7V$), the grading coefficient m , and the constant in the numerator which can be related to the zero-bias capacitance. This expression causes difficulties when the junction voltage V approaches the built-in voltage and C goes to infinity. In a real transistor, of course, a finite amount of charge is stored for all bias conditions, and the derivatives of charge with respect to junction voltages are finite. As shown in detail in Ref. 23, equation (28) can be modified so as to be free of singularities by the introduction of a fourth parameter which relates to the forward-bias capacitance inferred from measurements of transit time vs. emitter current. For compactness of notation, we express the four parameters as elements p_1, p_2, p_3 , and p_4 of a four-dimensional vector \mathbf{P} (see Table I and Appendix C). If we now define a function

$$f(v, \mathbf{P}) = p_3 \left\{ \frac{1}{(1 + p_4)^{p_3}} + \frac{\frac{v}{p_1} - 1}{\left[\left(\frac{v}{p_1} - 1 \right)^2 + p_4 \right]^{p_3}} \right\}, \quad (29)$$

then the normalized emitter and collector charges are given by:

$$q_e = f\left(\frac{qV_{eb}}{kT}, \mathbf{P}_e\right), \quad (30)$$

$$q_c = f\left(\frac{qV_{cb}}{kT}, \mathbf{P}_c\right), \quad (31)$$

where \mathbf{P}_e and \mathbf{P}_c are the four-parameter vectors describing the emitter and collector junctions, respectively. These two vectors constitute Groups 3 and 4 of the model parameters listed in Table I. As discussed in Appendix C, some of the parameter values can be estimated or approximated in terms of other model parameters. In any case, these parameters are readily amenable to numerical evaluation from the device structure.

As mentioned in the previous section, the recombination in transistors is best handled through a description of the base current as a sum of exponentials in the junction voltages. Pertinent model parameters are pre-exponential factors and emission coefficients. For typical transistors the forward base current is adequately described by two components, one ideal ($n = 1$) and the other nonideal ($n = n_e$). For the reverse base current a single nonideal ($n = n_c$) component is adequate. We define

$$e_k = \exp(-v_k), \quad (32)$$

$$e_{k_e} = \exp(-v_k/n_e), \quad (33)$$

$$e_{kc} = \exp(-v_k/n_c). \quad (34)$$

Then we let

$$i_{be} = i_1(e^{v_e} - e_k) + i_2(e^{v_e/n_e} - e_{ke}), \quad (35)$$

$$i_{bc} = i_3(e^{v_c/n_c} - e_{kc}). \quad (36)$$

The quantities i_1 , i_2 , n_e , i_3 , n_c are the five parameters which characterize the recombination behavior of the transistor. They are listed as Group 2 in Table I.

The last set (Group 5) of the model parameters listed in Table I describes the base push-out effect. For its description four parameters are required. The details of the base push-out modeling are given in Appendix A.

We thus have a total of 21 parameters which describe the transistor model. For several of these parameters default values can be used as discussed in Appendix C.

The following features are a consequence of the normalizations used:

- (i) I_k is proportional to the emitter area. All other parameters are, to first order, independent of area. Area scaling (neglecting complications caused by lateral effects such as emitter crowding) is achieved simply by changing the value of I_k . This feature is particularly convenient for integrated circuit work, where transistors on a given slice differ nominally only in their lateral dimensions.
- (ii) For pnp transistors all model parameters have positive numerical values. For npn transistors two changes are required: a) I_k must be made a negative quantity; b) the Boltzmann voltage kT/q must be given a negative value (or the Boltzmann voltage is given the sign of I_k for pnp's and npn's). When this is done, the polarity of terminal currents and voltages is in agreement with standard practice (currents positive if flowing *into* the device).
- (iii) The offset voltages V_{oe} and V_{oc} used in modeling the capacitive charges (see Appendix C) are approximately proportional to the absolute temperature. Hence use of constant, that is, temperature independent, values for the normalized quantities v_{oe} and v_{oc} implements automatically the temperature dependence of the offset voltages. The normalized knee voltage v_k is not, to first order, temperature independent, but varies with temperature approximately as

$$v_k(T) = v_k(T_o) + \frac{qV_o}{kT_o} \left(\frac{T_o}{T} - 1 \right) \quad (37)$$

where T_0 is a reference temperature (for example, room temperature) and V_g is the band-gap voltage (1.12 V for silicon).

IV. DISCUSSIONS AND EXAMPLES

This Section will illustrate the performance of the new model. The following features are of interest:

- (i) For low bias so that Q_b is nearly equal to Q_{b0} (or $q_b \approx 1$) and with the choice $n_e = n_c = 1$, the model reduces to the Ebers-Moll model.
- (ii) Superposition which is operative in the Ebers-Moll equations, is disabled through the base-charge denominator Q_b in equation (11). The dependence of Q_b on collector voltage produces a finite output conductance, that is, the Early effect.
- (iii) The rapid increase of Q_b when base push-out occurs causes fall-off of current gain and frequency response.
- (iv) If we define an effective emission coefficient n by

$$\frac{1}{n} \equiv \frac{kT/q}{I_c} \frac{dI_c}{dV_{eb}} \Big|_{V_{cb}=\text{constant}}, \quad (38)$$

then it is seen that for the model, n varies from approximately unity at low currents to approximately two at high currents (and larger values when base push-out occurs). The shift to a value of two represents high injection effects. The $n = 1$ and $n = 2$ asymptotes intersect approximately at $I_c = I_k$ and $V_{eb} = V_k$ (see Fig. 2).

If we define an emitter capacitance C_e by

$$C_e = \frac{dQ_e}{dV_{eb}} \Big|_{V_{cb}=\text{constant}}, \quad (39)$$

then the effective emission coefficient at low current values, where the current contribution to Q_b , equation (14), is negligible, is given by*

$$n = 1 + \frac{(kT/q)C_e}{Q_{b0}}. \quad (40)$$

Thus, small deviations from the ideal exponential law are caused by the emitter capacitance. These deviations are present even at

* If the transistor is used in a common emitter configuration it may be useful to define the effective emission coefficient as in equation (39), but with V_{ec} instead of V_{cb} held constant. For this case, shown in Fig. 2, the emitter capacitance C_e in equation (40) is to be replaced by the sum of emitter and collector capacitances.

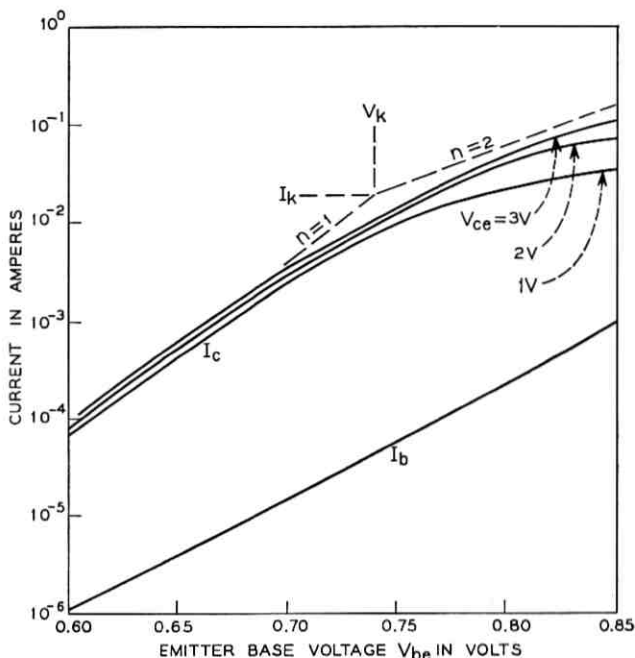


Fig. 2—Collector and base currents for three values of collector-emitter voltage versus base-emitter voltage. Also shown are "knee" point (V_K , I_K), and slopes corresponding to values of 1 and 2 for the emission coefficient n .

low forward currents[†]. Additional increases of n at intermediate currents are caused by base resistance (not considered in this paper).

- (v) For currents low enough such that base widening effects are negligible, the emitter-collector delay time τ_d for common emitter operation is given by

$$\begin{aligned} \tau_d &= \left. \frac{dQ_b}{dI_c} \right|_{V_{ce}=\text{const}}, \\ &= \frac{\tau_f + \frac{kT}{q} \frac{(C_e + C_c)}{I_c}}{1 - \frac{kT}{q} \frac{(C_e + C_c)}{Q_b}}. \end{aligned} \quad (41)$$

[†] In principle, equation (40) could be used to obtain the emitter capacitance from a dc semilog plot of I_c vs. V_{be} . However, very accurate temperature control would be required.

At low current values, and for $C_c \ll C_e$, the denominator of equation (41) is approximately $1/n$, equation (40). Then the value of the emitter capacitance C_e may be obtained from the slope $n(kT/q)C_e$ of a plot of delay time τ_d versus reciprocal collector current (see Fig. 6). It is this forward-bias emitter capacitance that may be used to set the parameter a_{e2} (see equation 1 of Appendix C).

As a demonstration of the proposed model, various characteristics for a double diffused silicon npn transistor at room temperature ($kT/q = 0.02585$ V) were computed with the parameter values listed in Table II. The results are shown in Figs. 2 through 7. We treat the intrinsic transistor model, without inclusion of parasitics, such as base resistance and inactive base region. Figure 2 is a semilog plot of collector and base current vs. emitter-base current for $V_{ce} = 1, 2$ and 3 V. Note the transition from $n = 1$ to $n = 2$ near $I_k = 19$ mA and $V_k = 0.74$ volts. Figure 3 shows common-emitter low-frequency current gain β vs. collector current for various collector voltages. The frequency dependence is conveniently characterized by the low-frequency approximation to the unity-gain frequency,²⁴ f_L . (For high-current-gain transistors in the active region, f_L is synonymous with the conventional cut-off frequency f_T .) Figure 4 shows f_L vs. collector current for various collector voltages.

TABLE II—MODEL PARAMETER VALUES

$I_k = -1.875 \times 10^{-2}$ amps.
$v_k = 28.7$
$\tau_f = 3.2 \times 10^{-10}$ secs.
$r_t = 10.0$
$i_1 = 2.35 \times 10^{-4}$
$i_2 = 2.19 \times 10^{-3}$
$n_e = 1.5$
$i_3 = 7.65 \times 10^{-3}$
$n_c = 1.5$
$v_{oe} = 27.1$
$m_e = 0.24$
$a_{e1} = 0.337$
$a_{e2} = 1.03 \times 10^{-2}$
$v_{oc} = 27.1$
$m_c = 0.1265$
$a_{c1} = 0.187$
$a_{c2} = 7.17 \times 10^{-3}$
$v_{rp} = 18.2$
$r_w = 10.0$
$r_p = 4.55$
$n_p = 3.0$

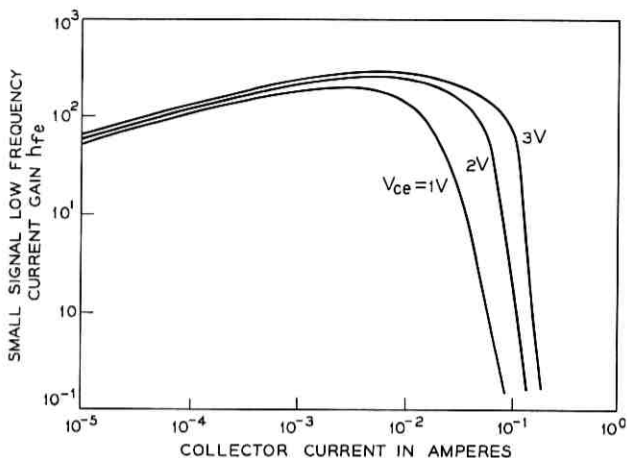


Fig. 3—Small-signal low-frequency current gain h_{fe} versus collector current for three values of collector-emitter voltage.

Figure 5 presents a family of I_c vs. V_{ce} characteristics, with I_b as parameter. Figure 6 gives results of simulated τ_d vs. $1/I_c$ measurements of the model. The same information is displayed as f_L contour plots in Fig. 7.

V. CONCLUSION

With the aid of a new charge control relation, a compact bipolar transistor model, suitable for network analysis computer programs, has been developed. This model is an "integral model" in that parameters important for the ac response also shape the dc characteristics. The model includes high-injection effects; for low bias and idealized emission coefficients it is equivalent to the Ebers-Moll model. Figures 2-7 exhibit many features of real transistors that conventional models do not contain, or can give only by curve-fitting with the use of many parameters. These figures were obtained for an implementation of the model using 21 parameters (Table I, Appendix B). For several of these parameters, values can be estimated *a priori*. Such values may be used as default values. Sometimes users may be willing to make some sacrifice in the accuracy of the model in return for having to specify only few key parameters. It appears feasible to require that only I_k (proportional to emitter area), the Early voltage V_A , maximum β (at some collector voltage, for example, $V_{ce} = 5$ V), maximum f_L (at $V_{ce} = 5$ V) and the collector current at which maximum f_L occurs, need be specified and

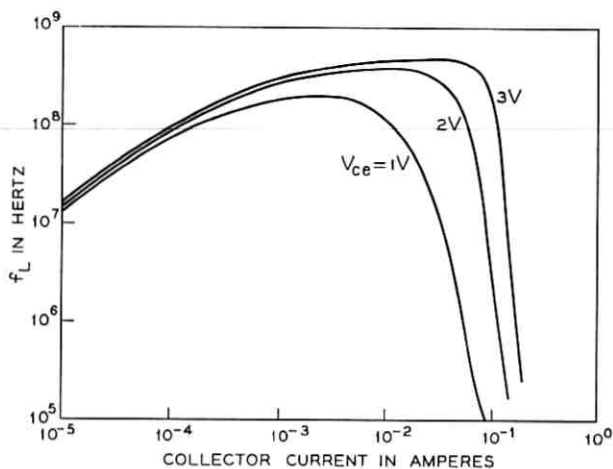


Fig. 4—Low-frequency approximation to unity-gain frequency f_L versus collector current for three values of collector-emitter voltage.

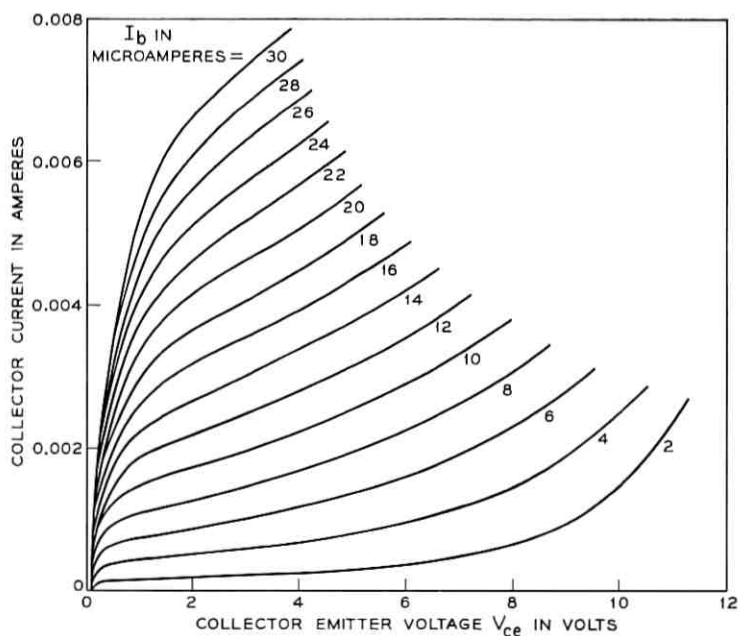


Fig. 5—Collector current versus collector-emitter voltage for various values of base current.

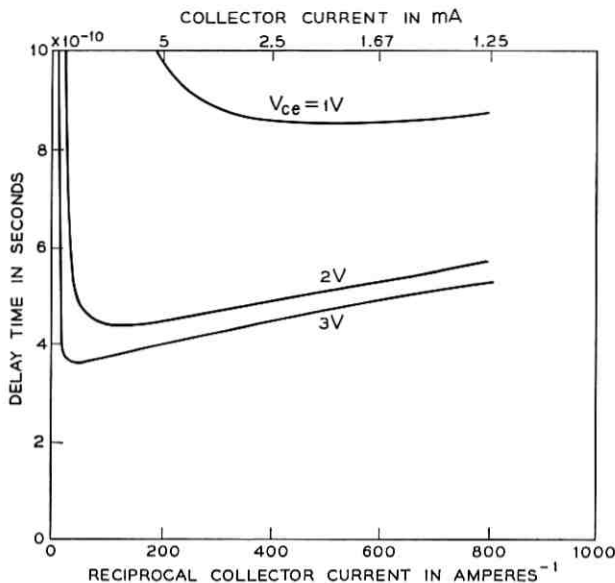


Fig. 6—Emitter-collector delay time versus reciprocal collector current for three values of emitter-collector voltage.

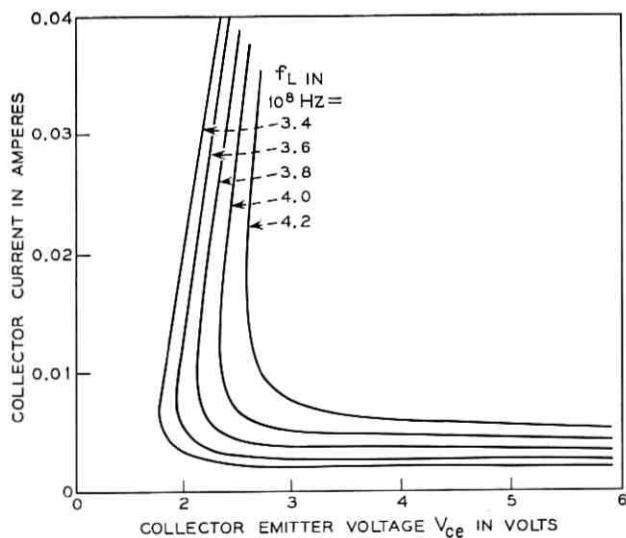


Fig. 7—Contours of constant f_L in collector-emitter voltage, collector-current plane.

that the full set of model parameters is automatically computed with the use of default values.

Not included in this paper, but to be included in careful modeling, are parasitics such as base resistance, inactive base region, collector resistance, and so on. The base resistance may be modeled as consisting of a constant, exterior resistance incurred in the inactive base and an interior resistance. To reflect conductivity modulation of the base, the value of the interior resistance may be made inversely proportional to the base charge Q_b . It appears feasible to extend the intrinsic model, for example to include carrier generation by impact ionization.

VI. ACKNOWLEDGMENT

The authors are indebted to F. M. Smits for encouragement and for substantial improvements in the presentation of the material.

APPENDIX A

Base Push-Out Effect

The approach towards modeling the base push-out effect was guided by results obtained in a detailed analysis of this effect.²⁵ The discussion is based on pnp polarity (I_c is a negative quantity). Assuming constant resistivity ρ in the collector region adjacent to the base (epitaxial region), the base push-out effect starts approximately at a collector current value

$$I_1 = \frac{A_e(V_{oc} - V_{cb})}{\rho W_c} \quad (42)$$

where A_e is the emitter area and W_c is the width of the lightly doped collector region. Let W_{eff} be the effective width of the base. For $-I_c < I_1$, the effective base width is equal to the metallurgical base width, W_b ,

$$W_{eff} = W_b. \quad (43)$$

For $-I_c > I_1$, the effective base width is approximately given by

$$W_{eff} = W_b + W_c \left(1 - \frac{I_1}{I_c}\right). \quad (44)$$

Equations (43) and (44) can be written as

$$W_{eff} = W_b + \frac{W_c}{2} \frac{[(I_c + I_1)^2]^{1/2} - (I_c + I_1)}{I_c}. \quad (45)$$

Equation (45) gives much too abrupt a transition for W_{eff} from W_b to

$(W_b + W_c)$, and becomes numerically unstable when I_c approaches zero. Hence, we modify equation (45) with an additional parameter I_2 , such that

$$W_{\text{eff}} = W_b + \frac{W_c}{4} \frac{\{[(I_c + I_1)^2 + I_2^2]^{\frac{1}{2}} - (I_c + I_1)\}^2}{I_c^2 + I_2^2}. \quad (46)$$

The low-current forward transit time τ_f is to be modified by the square of the ratio of the effective base width to metallurgical base width, to give the total base transit time τ_{fb} :

$$\tau_{fb} = \tau_f \left(\frac{W_{\text{eff}}}{W_b} \right)^2 \equiv \tau_f B \quad (47)$$

with

$$B = \left\{ 1 + \frac{W_c}{4W_b} \frac{\{[(I_c + I_1)^2 + I_2^2]^{\frac{1}{2}} - (I_c + I_1)\}^2}{I_c^2 + I_2^2} \right\}^2. \quad (48)$$

The quantity B may be expressed in terms of model parameters and the normalized collector current $i_c = I_c/I_K$

$$B = \left\{ 1 + \frac{r_w}{4} \frac{\{[(i_c + i_1)^2 + r_p]^{\frac{1}{2}} - (i_c + i_1)\}^2}{i_c^2 + r_p} \right\}^2 \quad (49)$$

with

$$r_w = \frac{W_c}{W_b}, \quad r_p = \left(\frac{I_2}{I_K} \right)^2,$$

and

$$i_1 = \frac{(V_{oc} - V_{cB})A_e}{\rho W_c I_K} = \frac{v_{oc} - v_k - v_c}{v_{rp}}. \quad (50)$$

So far, equation (49) models effects in a one-dimensional transistor. Emitter crowding and carrier storage in the inactive base cause the transit time at high currents to increase more strongly than given by equation (49). For a first-order modeling of emitter crowding we replace the exponent 2 outside the square brackets by an adjustable model parameter n_p ("push-out exponent"):

$$B = \left\{ 1 + r_w \frac{\{[(i_c + i_1)^2 + r_p]^{\frac{1}{2}} - (i_c + i_1)\}^{n_p}}{4(i_c^2 + r_p)} \right\}^2. \quad (51)$$

The quantities r_w , r_p , v_{rp} , n_p and v_{oc} are model parameters (groups 5 and 4).

APPENDIX B

A Summary of the Model

Terminal quantities are: collector current I_c , base current I_b , emitter-base voltage V_{eb} , collector base voltage V_{cb} , and base charge Q_b . Define the following normalized quantities: $i_c = I_c/I_k$, $i_b = I_b/I_k$, $v_e = (qV_{eb}/kT) - v_k$, $v_c = (qV_{cb}/kT) - v_k$, $q_b = Q_b/(I_k\tau_f)$, and define the function

$$f(v, P) = p_3 \left\{ \frac{1}{(1 + p_4)^{p_2}} + \frac{(v/p_1 - 1)}{[(v/p_1 - 1)^2 + p_4]^{p_2}} \right\}.$$

Then the following relations hold between the normalized quantities

$$i_{be} = i_1(e^{v_e} - e_k) + i_2(e^{v_e/n_e} - e_{ke}), \quad (52)$$

$$i_{bc} = i_3(e^{v_c/n_c} - e_{kc}), \quad (53)$$

$$i_b = -i_{be} - i_{bc}, \quad (54)$$

$$i_4 = i_c + (v_{oc} - v_c - v_k)/v_{\tau p}, \quad (55)$$

$$B = \left\{ 1 + r_w \frac{\{[i_4^2 + r_p]^{\frac{1}{2}} - i_4\}^{n_p}}{4(i_c^2 + r_p)} \right\}, \quad (56)$$

$$q_1 = 1 + f(v_e + v_k, P_e) + f(v_c + v_k, P_c), \quad (57)$$

$$q_2 = B(e^{v_e} - e_k) + r_t(e^{v_c} - e_k), \quad (58)$$

$$q_b = q_1/2 + [(q_1/2)^2 + q_2]^{\frac{1}{2}}, \quad (59)$$

$$i_c = -\frac{(e^{v_e} - e^{v_c})}{q_b} + i_{bc}. \quad (60)$$

APPENDIX C

Discussion of Model Parameters

The knee parameters I_k and v_k , and τ_f are related to the conventional intercept current I_s (saturation current), the base charge at zero bias Q_{bo} , and the nominal value of the low frequency approximation to the unity-current-gain frequency f_L , by

$$I_k = \frac{-Q_{bo}}{\tau_f}, \quad (61)$$

$$v_k = \ln \left(\frac{-Q_{bo}}{\tau_f I_s} \right), \quad (62)$$

and

$$\tau_f = \frac{1}{2\pi f_L} \quad (63)$$

In terms of structural parameters τ_f is given approximately by

$$\tau_f = \frac{w_b^2}{2\eta D} \quad (64)$$

where w_b is the base width and where η represents the drift effect in the base.²⁶ η is unity for uniform base doping and has typical values between two and ten for diffused-base transistors.

The zero-bias base charge is approximately given by

$$Q_{b0} = -A_e q N_b \quad (65)$$

where A_e is the emitter area, q the electronic charge and N_b the number of impurities per unit area in the base. Typical values for N_b are a few times 10^{12} per cm^2 (lower values would cause premature punch through and high values cause low injection efficiency and/or low f_L). The intercept current is given by

$$I_s = A_e \frac{q n_i^2 D}{N_b} = -\frac{A_e^2 q^2 n_i^2 D}{Q_{b0}} \quad (66)$$

where n_i is the intrinsic carrier concentration and D is the effective diffusivity of carriers in the base. From equations (62) and (64) through (66), one obtains

$$v_k = 2 \ln \frac{N_b/w_b}{n_i} + \ln 2\eta \quad (67)$$

The parameters i_1 , i_2 , i_3 determine the current gain of the transistor. The variation of forward gain with current at low currents is governed by n_e and the inverse gain variation by n_c .

The group 3 parameters model the emitter junction capacitance. The offset voltage V_{oe} is approximately the conventional "built-in" voltage, which has a typical value near 0.7 volts for silicon at room temperature, or $v_{oe} = V_{oe}/(kT/q) = 27$. The grading coefficient m_e depends on the type of doping transition: it is one-fourth and one-sixth for ideal step and linearly graded junctions respectively. Typical values for emitter junctions are in the neighborhood of 0.2. Parameter a_{e1} is related to the zero-bias capacitance C_{oe} by²³

$$a_{e1} = \frac{C_{oe} V_{oe}}{Q_{b0}(1 - 2m_e)} \quad (68)$$

If one assumes the emitter junction to be a step junction and the base doping to be uniform, and if one expresses base doping and base width in terms of the number of impurities per cm^2 in the base, N_b , and in terms of f_L , then an approximate formula for a_{e1} may be derived to be

$$a_{e1} = \frac{1}{(1 - 2m_e)} \left(\frac{\pi \epsilon^2 V_{e0}^2}{4q^2 D} \right)^{\frac{1}{2}} \frac{(f_L)^{\frac{1}{2}}}{(N_b)^{\frac{1}{2}}}, \quad (69)$$

$$= A \frac{(f_L/10^9 \text{ Hz})^{\frac{1}{2}}}{(N_b/5 \times 10^{12} \text{ cm}^{-2})^{\frac{1}{2}}}, \quad (70)$$

where ϵ is the permittivity and D the diffusivity of electrons (holes) in an npn (pnp) transistor. For a silicon npn transistor, the numerical value for A based on equation (69) is 0.147. For an actual double diffused transistor of $f_L = 400$ MHz, the value of A obtained from parameter fitting was found to be 0.202.

The last parameter a_{e2} in this group is related to the forward bias capacitance, C_{ef} , deduced from the slope of delay time versus reciprocal emitter current by²³

$$a_{e2} = \left[\frac{(1 - 2m_e)rC_{ef}}{C_{oe}} \right]^{-m_e} \quad (71)$$

where r is a numerical coefficient approximately equal to unity, the exact value depending on the doping profile. Typical values for a_{e2} range from 10^{-2} to 10^{-3} . If emitter capacitance effects are not of importance to the users, and if the user does not want to specify group 3 parameter values, the following default values are suggested:

$$v_{oe} = 27, \quad \text{silicon};$$

$$m_e = 0.2;$$

$$a_{e1} = 0.2 \frac{(f_L/10^9 \text{ Hz})^{\frac{1}{2}}}{(N_b/5 \times 10^{12} \text{ cm}^{-2})^{\frac{1}{2}}};$$

$$a_{e2} = 3 \times 10^{-3}.$$

The group 4 parameters model the collector junction capacitance. The parameters v_{oc} , m_c , a_{c1} and a_{c2} have similar meanings as their counterparts in the emitter junction capacitance. Typical default values (for silicon transistors) for v_{oc} , m_c , and a_{c2} are

$$v_{oc} = 27, \quad \text{silicon};$$

$$m_c = 0.15;$$

$$a_{c2} = 10^{-3}.$$

We shall relate a_{c1} to the output characteristics of the transistor. The Early voltage as defined in Section II is of magnitude comparable to that of the punch-through voltage V_T , defined as that voltage for which the charge associated with collector capacitance, Q_{cc} , equals minus Q_{bo} . We denote the coefficient relating the Early voltage and the punch-through voltage by r_A

$$V_T = r_A V_A. \quad (72)$$

Then a_{c1} is given in terms of the Early voltage by

$$a_{c1} = \frac{1}{\left(\frac{r_A V_A - V_{oc}}{V_{oc}}\right)^{1-2m_c} - 1}. \quad (73)$$

The exact value of r_A depends on details of the doping profile and on the region in the I_c vs V_{cc} domain from which the Early voltage is extrapolated; a typical value of r_A is 1.7. Thus, given the Early voltage and the other group 4 parameters, equation (73) is convenient for estimating a_{c1} . It should be noted that a_{c1} refers to the collector capacitance of the intrinsic transistor. The terminal collector capacitance will be dominated by that of the inactive base region.

The group 5 parameters model the base push-out effects. $V_{rp} = (kT/q)v_{rp}$ is the resistive voltage drop across the collector, caused by a current of magnitude I_K . The ratio of the width of the collector epitaxial region to the width of the metallurgical base is designated r_w . The parameter r_p determines the steepness of the variation of the forward delay time as a function of a collector current in the current range where base push-out is incipient. The base push-out exponent n_p determines the fall-off of f_L for high currents. For $n_p = 2$, f_L has a tendency to level off after it has decreased from its maximum value by a factor of $(1 + r_w)^2$. For $n_p > 2$, the decrease continues beyond this level. In the example given in Section IV of this paper, the following values for v_{rp} , r_w , r_p and n_p are used.

$$v_{rp} = 18.2,$$

$$r_w = 10.0,$$

$$r_p = 4.55,$$

$$n_p = 3.0.$$

REFERENCES

1. Ebers, J. J., and Moll, J. L., "Large-Signal Behaviour of Junction Transistors," Proc. IRE, *42*, No. 12 (December 1954), pp. 1761-1772.
2. Beaufoy, R., and Sparkes, J. J., "The Junction Transistor as a Charge-Controlled Device," ATE J. (London), *13*, No. 4 (October 1957), pp. 310-324.
3. Hamilton, D. J., Lindholm, F. A., and Narud, J. A., "Comparison of Large Signal Models for Junction Transistors," Proc. IEEE, *52*, No. 3 (March 1964), pp. 239-248.
4. Koehler, D., "The Charge-Control Concept in the Form of Equivalent Circuits, Representing a Link Between the Classic Large Signal Diode and Transistor Models," B.S.T.J., *46*, No. 3 (March 1967), pp. 523-576.
5. Linvill, J. G., "Lumped Models of Transistors and Diodes," Proc. IRE, *46*, No. 6 (June 1958), pp. 1141-1152.
6. Murphy, B. T., "Diode and Transistor Self-Analogues for Circuit Analysis," B.S.T.J., *47*, No. 4 (April 1968), pp. 487-502.
7. Ohtsuki, T., and Kani, K., "A Unified Modeling Scheme for Semiconductor Devices with Applications of State Variable Analysis," IEEE Transactions on Circuit Theory, *17*, No. 1 (February 1970), pp. 26-32.
8. Early, J. M., "Effects of Space Charge Layer Widening in Junction Transistors," Proc. IRE, *40*, No. 11 (November 1952), pp. 1401-1406.
9. Sah, C. T., Noyce, R. N., and Shockley, W., "Carrier Generation and Recombination in P-N Junctions and P-N Junction Characteristics," Proc. IRE, *45*, No. 9 (September 1957), pp. 1228-1243.
10. Webster, W. M., "On the Variation of Junction Transistor Current Gain Amplification Factor With Emitter Current," Proc. IRE, *42*, No. 6 (June 1954), pp. 914-920.
11. Kirk, Jr., C. T., "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," IRE Trans. on Electron Devices, *ED-9*, No. 2 (March 1962), pp. 164-174.
12. Pritchard, R. L., and Coffey, W. N., "Small Signal Parameters of Grown-Junction Transistors at High Frequencies," 1954 IRE Convention Record, Pt. 3, pp. 89-98.
13. Fletcher, N. H., "Self-Bias Cutoff Effect in Power Transistors," Proc. IRE, *43*, No. 11 (November 1955), p. 1669.
14. Malmberg, A. F., and Cornwell, F. L., "NET-1 Network Analysis Program," Los Alamos Scientific Laboratory of the University of California, TID-4500 (19th Ed.), April 30, 1963.
15. Milliman, L. D., Massena, W. A., and Dickhaut, R. H., "CIRCUS, A Digital Computer Program for Transient Analysis of Electronic Circuits," Harry Diamond Lab. User's Guide, HDL 346-1, Program Manual HDL 346-2, January 1967.
16. Gummel, H. K., "A Charge Control Relation for Bipolar Transistors," B.S.T.J., *49*, No. 1 (January 1970), pp. 115-120.
17. Gummel, H. K., "A Self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations," IEEE Trans. on Ed, *11*, No. 10 (October 1964), pp. 455-465.
18. Gummel, H. K., "Measurement of the Number of Impurities in the Base Layer of a Transistor," Proc. IRE, *49*, No. 4 (April 1961), p. 834.
19. Iwersen, J. E., Bray, A. R., and Kleimack, J. J., "Low Current Alpha in Silicon Transistors," IRE Trans. on ED, *9*, No. 6 (November 1962), p. 474-478.
20. Gwyn, C. N., Scharfetter, D. L., and Wirth, J. L., "The Analysis of Radiation Effects in Semiconductor Junction Devices," IEEE Trans. Nuclear Science, *NS-14*, No. 6 (December 1967), pp. 153-169.
21. Scharfetter, D. L., and Gummel, H. K., "Large-Signal Analysis of a Silicon Read Diode Oscillator," IEEE Trans., *ED-16*, No. 1 (January 1969), pp. 64-77.
22. Caughey, D. M., "Simulation of UHF Transistor Small-Signal Behaviour to 10 GHz for Circuit Modeling," Conference on Computerized Electronics, Cornell University, Ithaca, N. Y., August 1969.

23. Poon, H. C., and Gummel, H. K., "Modeling of Emitter Capacitance," Proc. of IEEE Letter, *57*, No. 12 (December 1969), pp. 2181-2182.
24. Gummel, H. K., "On the Definition of the Cutoff Frequency f_T ," Proc. IEEE, *57*, No. 12 (December 1969), p. 2159.
25. Poon, H. C., Gummel, H. K., and Scharfetter, D. L., "High Injection in Epitaxial Transistors," IEEE Trans., *ED-16*, No. 5 (May 1969), pp. 455-457.
26. Moll, J. L., and Ross, I. M., "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity," Proc. IRE, *44*, No. 1 (January 1956), pp. 72-78.

The Potential Due to a Charged Metallic Strip on a Semiconductor Surface

By E. WASSERSTROM[†] and J. McKENNA

(Manuscript received December 30, 1969)

We solve numerically the problem of finding the potential and electric field around a negatively charged metallic contact on the surface of an n -type semiconductor. The semiconductor, which has permittivity ϵ_1 , fills the half-space $y < 0$. The contact is an infinitely long strip of width $2a$, defined by $y = 0, 0 \leq x \leq 2a, -\infty < z < \infty$. The region $y > 0$ is vacuum with permittivity ϵ_0 . In suitable dimensionless coordinates the potential ϕ satisfies Laplace's equation in $y > 0$ and the equation $\nabla^2 \phi = e^\phi - 1$ in $y < 0$. On the boundary $y = 0, \phi = \phi_0 < 0, 0 \leq x \leq 2a$, and the usual electromagnetic boundary conditions at the remainder of the interface. Finite difference schemes are used to solve the resulting boundary value problem.

In most practical cases $|\phi_0| \gg 1$ and $\eta = \epsilon_0/\epsilon_1 \ll 1$. We examine in considerable detail the limiting case $\eta = 0$, first for the less practical situation where $|\phi_0| \ll 1$ and then for $|\phi_0| \gg 1$. In case the $|\phi_0| \ll 1$ we show that our numerical solution agrees well with the exact analytical solution of a linearized version of the problem. For $|\phi_0| \gg 1$, we give plots of the equipotential curves, curves of equal charge density, and curves of constant electric field amplitude. These results also yield expressions for the capacitance of both a strip and a circular electrode. The modifications of these results when $\eta > 0$ are also given in some detail. Finally, we discuss the numerical calculations at some length.

I. INTRODUCTION AND FORMULATION OF THE PROBLEM

In the study of a number of solid-state devices, it is important to know the electrostatic potential in the neighborhood of a metal-semiconductor contact (Schottky diode) and in a metal insulator-semiconductor structure (MOS capacitor).

Motivated by this interest, we consider in this paper the following

[†] On leave from the Technion-Israel Institute of Technology, Haifa, Israel, when this work was conducted.

problem. An infinitely long metallic strip of width $2a^*$ and zero thickness occupies the region $0 \leq x^* \leq 2a^*$, $y^* = 0$, as shown in Fig. 1. The region $y^* > 0$ is filled with air and the region $y^* < 0$ is filled with an n -type semiconductor. The metallic strip is charged to a negative potential, $\phi_0^* < 0$, and we wish to calculate the potential, the electric field and the electric charge density distribution in the semiconductor under the assumption of no current flow. As will be shown later, this solution also determines the potential, field and charge density around a circular metallic contact.

For the doping levels normally encountered in such devices, say 10^{20} - 10^{22} m^{-3} the electrostatic potential ϕ^* in the semiconductor is governed by the Poisson equation,

$$\nabla^{*2}\phi^* = -\rho^*/\epsilon_1, \quad (1)$$

where ∇^{*2} is the Laplacian, ϵ_1 is the permittivity of the semiconductor, and the net volume charge density ρ^* is given by¹

$$\rho^* = qN_d[1 - \exp(q\phi^*/kT)], \quad (2)$$

where $-q$ is the charge of an electron, N_d is the donor number density, k is Boltzmann's constant, and T is the absolute temperature. In writing down equation (2) we have neglected the contact potential Ψ between the metal strip and the semiconductor,² since in many cases of practical interest $|\phi_0| \gg |\Psi|$. Here and in the following, starred quantities have rationalized MKS dimensions; unstarred quantities, except for a few obvious physical parameters, are dimensionless.

As in Ref. (1) we introduce the dimensionless length and potential

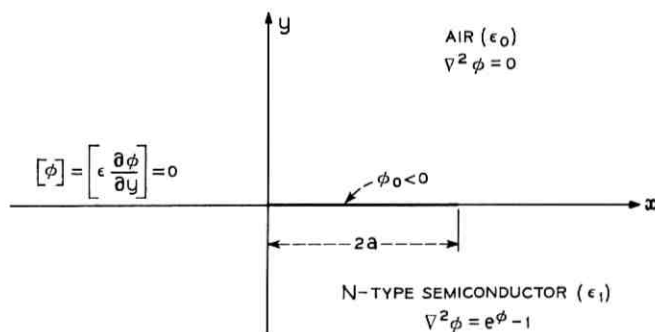


Fig. 1—A diagram of the geometry of the problem P1 showing the conducting strip at the air-semiconductor interface and the coordinate system used. The dimensionless parameters shown are defined in equation (3). The symbol [] is defined in equation (8).

by the relations

$$\begin{aligned}x &= x^*/\lambda_D, & y &= y^*/\lambda_D, & a &= a^*/\lambda_D, \\ \phi &= q\phi^*/kT, & \rho &= \rho^*/qN_d,\end{aligned}\quad (3)$$

where the Debye length is given by³

$$\lambda_D = (\epsilon_1 kT/q^2 N_d)^{\frac{1}{2}}. \quad (4)$$

Typical values for a lightly doped semiconductor device are $N_d = 10^{21} \text{ m}^{-3}$, $\epsilon_0/\epsilon_1 = .0625$ for germanium and $\epsilon_0/\epsilon_1 = .078$ for silicon, and $a^* = 10^{-4} \text{ m}$, where ϵ_0 is the permittivity of free space. Then, for example, for silicon at $T = 300^\circ\text{K}$, $\lambda_D = 1.35 \times 10^{-7} \text{ m}$, $q/kT = 38.8 \text{ volt}^{-1}$, and $a = 740$.

In terms of these dimensionless quantities the boundary value problem to be solved for the potential ϕ can be summarized mathematically as follows:

(i) In the air, $y > 0$ (see Fig. 1), the potential satisfies Laplace's equation

$$\nabla^2 \phi = 0. \quad (5)$$

(ii) In the semiconductor, $y < 0$, the potential satisfies Poisson-Boltzman's equation

$$\nabla^2 \phi = e^\phi - 1. \quad (6)$$

(iii) On the plate, $y = 0$ and $0 \leq x \leq 2a$, the potential is a given (negative) constant

$$\phi = \phi_0 < 0. \quad (7)$$

(iv) At the interface, $y = 0$, $|x - a| > a$, the potential and the normal component of the electric displacement vector are continuous⁴

$$[\phi(x, 0)] = \phi(x, 0+) - \phi(x, 0-) = 0, \quad (8a)$$

$$\left[\epsilon \frac{\partial \phi}{\partial y}(x, 0) \right] = \eta \frac{\partial \phi}{\partial y}(x, 0+) - \frac{\partial \phi}{\partial y}(x, 0-) = 0, \quad (8b)$$

where

$$\eta = \epsilon_0/\epsilon_1. \quad (8c)$$

(v) At infinity

$$\lim_{r \rightarrow \infty} (\phi) = 0, \quad (9)$$

where $r = (x^2 + y^2)^{\frac{1}{2}}$.

We refer to the boundary value problem defined by conditions (i)-(v) as P1. We have been unable to solve this problem analytically, and instead have studied its solution numerically.

Qualitatively, for large $|\phi_0|$ the potential can be described easily. Within the semiconductor, the negative charge on the conducting strip repels the mobile electrons. This action produces a layer around the strip, called the depletion layer, from which almost all the mobile electrons have been expelled. The positive donor ions left behind make this a region of uniform positive volume charge density. Far from the plate the semiconductor is almost neutral, and these two regions are connected by a transition layer. When $\eta = 0$ this transition layer is sharp and well defined and is several Debye lengths thick. However, when $\eta \neq 0$, this transition region becomes broader and diffuse near the semiconductor-air interface. In the air on the other hand, the potential is essentially that due to the dipole formed by the negative charge on the conducting strip and the positive charge due to the donor ions in the depletion layer.

In Section II, we consider the special case $\eta = 0$. For bias voltages ϕ_0 which are typically encountered in semiconductor devices, the thickness of the depletion layer is large compared to a Debye length but small compared to the width of the strip, $2a$. We will show that we then only need to consider a strip completely embedded in an n -type semiconductor. This will reduce the solution of P1 to the solution of P2

$$(i) \text{ In the semiconductor } \nabla^2 \phi = e^{\phi} - 1. \quad (6)$$

$$(ii) \text{ On the plate, } y = 0, 0 \leq x \leq 2a, \phi = \phi_0 < 0. \quad (10)$$

$$(iii) \text{ At infinity } \lim_{r \rightarrow \infty} (\phi) = 0. \quad (11)$$

Solutions of P2 are obtained by the method of finite differences. For $|\phi_0| \ll 1$, equation (6) can be linearized to

$$\nabla^2 \phi = \phi. \quad (12)$$

The linearized version of P2 in the limit $a = \infty$, has been solved analytically by Lewis⁵ and for small ϕ_0 his results agree excellently with our numerical solution in the region $x < a$. This provides a good check on our numerical methods. Note, however, that in our formulation we neglected the contact potential and therefore the problem is not physically meaningful for this limit. For $|\phi_0| \gg 1$ numerical calculations of both the electric field and the potential are presented in considerable detail. Finally, the capacitance per unit length of the strip and the capacitance of a circular electrode are presented with particular emphasis on edge effects.

In Section III we briefly discuss what modifications of the results of Section II must be made when $\eta \neq 0$. It is shown that the chief effect of positive η in the semiconductor is a smearing out of the transition region near the air-semiconductor interface.

In Section IV we give some of the details of the methods of numerical analysis used.

II. THE CASE $\eta = 0$

In most practical applications $\epsilon_0 \ll \epsilon_1$, as was pointed out in Section I, and it is therefore of interest to treat first the simpler, limiting problem with $\eta = 0$. When $\eta = 0$ condition (8b) becomes

$$\frac{\partial \phi}{\partial y}(x, 0-) = 0, \quad (13)$$

and the solution for the potential in the semiconductor is decoupled from the solution for the potential in the air. In fact, the solution of P1 in the semiconductor is now identical with the solution of P2 because of symmetry.

There are three characteristic lengths in the problem of the finite width strip: the half width of the strip, the Debye length, and the thickness of the depletion layer. If the half width of the strip is very large compared to both the Debye length and the thickness of the depletion layer, the solution below the plate and sufficiently far from the edges must be the one dimensional solution (independent of x). Thus, for a "sufficiently wide" finite strip, there are really only two characteristic lengths for the solution in the semiconductor: the Debye length and the depletion layer thickness.

We determine more precisely what "sufficiently wide" means. It is well known that if one considers the one dimensional problem of an n -type semiconductor filling the region $y < 0$, with the plane $y = 0$ held at a large negative potential $\phi_0 < 0$, then the thickness of the depletion layer, R , is accurately given by²

$$R = |2\phi_0|^{1/2}. \quad (14)$$

[The accuracy of this approximation is also discussed in Ref. (1).] The Debye length is equal to unity in our nondimensional coordinates. Thus, if in addition to $\eta = 0$, we have

$$a \gg 1, \quad (15a)$$

and

$$a \gg R, \quad (15b)$$

then a disappears from the problem in the semiconductor as a characteristic length. Reverse bias voltages of the order of $\phi_0^* \sim -5$ to -50 volts, ($\phi_0 \sim -200$ to -2000) are typical, and this corresponds to $R \sim 19.7$ to 63.0 at room temperature. Thus in most practical situations condition (15) is satisfied, since $a \gtrsim 740$, and we can expect the solutions to be independent of a also near the edges of the strip.

If in addition to equations (13) and (15a) we have

$$|\phi_0| \ll 1, \quad (16)$$

then P2 can be linearized, for equation (5) can be approximated by

$$\nabla^2 \phi = \phi, \quad (17)$$

since $|\phi| \leq |\phi_0|$ everywhere. This linear problem with $a = \infty$ has been solved analytically by Lewis,⁵ and in polar coordinates (see Fig. 2) the solution is given by

$$\begin{aligned} \phi/\phi_0 = \frac{1}{2} \exp(r \sin \theta) & \left\{ 1 - \operatorname{erf} \left[r^{\frac{1}{2}} \left(\cos \frac{\theta}{2} + \sin \frac{\theta}{2} \right) \right] \right\} \\ & + \frac{1}{2} \exp(-r \sin \theta) \left\{ 1 + \operatorname{erf} \left[r^{\frac{1}{2}} \left(\cos \frac{\theta}{2} - \sin \frac{\theta}{2} \right) \right] \right\}, \end{aligned} \quad (18)$$

where $\operatorname{erf}(z)$ is the error function.⁶ We have solved the (nonlinear) P2 by a finite difference method to be described in Section IV. The numerically calculated equipotentials are shown in Fig. 3, and in Fig. 4 we compare the numerical solution of P2 (crosses) with the analytical solution (18) of the linearized problem (continuous line).

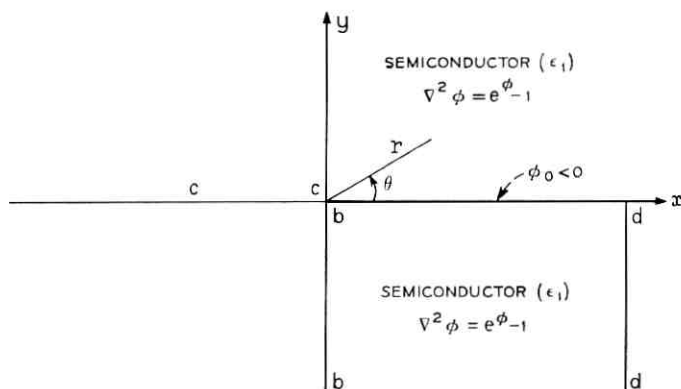


Fig. 2—A diagram of the geometry of the problem P2 showing the coordinate system used.

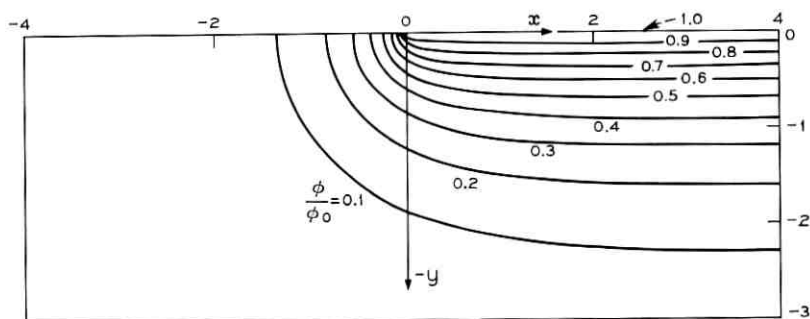


Fig. 3—Equipotential curves for the case $\eta = 0$, $\phi_0 = -0.01$, and $a = 4$.

These figures correspond to $\phi_0 = -0.01$ and $a = 4$, using the mesh size $\Delta x = \Delta y = .1$. The three curves in Fig. 4 are the potential along the lines $b - b(x \leq 0, y = 0)$, $c - c(x = 0, y \leq 0)$ and $d - d(x = 4, y \leq 0)$. These three lines are shown in Fig. 2. It is seen that the two solutions agree very well.

In most practical cases, however, ϕ_0 is very large (typically $200 < |\phi_0| < 2000$), so we shall concentrate on the large potential problem from now on. When ϕ_0 is not small, the problem cannot be linearized. We have been unable to find approximate analytic solutions, and so we have had to solve the problem numerically.

From the results of Ref. (1), we should expect that if distances are normalized with respect to the depletion layer thickness, then the potential when normalized with respect to the plate potential should be essentially independent of ϕ_0 as $|\phi_0| \rightarrow \infty$. We introduce this normalization here:

$$\bar{x} = x/R, \quad \bar{y} = y/R, \quad \bar{a} = a/R, \quad \bar{\phi} = \phi/\phi_0, \quad \bar{\rho} = \rho, \quad (19)$$

where R is the depletion layer thickness given in equation (14). In terms of our new variables, the basic equation (6) for $\bar{\phi}(\bar{x}, \bar{y})$ reads

$$\nabla^2 \bar{\phi} \equiv \left(\frac{\partial^2}{\partial \bar{x}^2} + \frac{\partial^2}{\partial \bar{y}^2} \right) \bar{\phi} = 2(1 - e^{\bar{\phi}}) = 2\bar{\rho}. \quad (20)$$

We have solved P2 in the normalized (tilde) variables by the method of finite differences. A detailed description of the method will be given in Section IV. Calculations for $\phi_0 = -100$ and -500 using the mesh sizes $\Delta \bar{x} = \Delta \bar{y} = .05$ and for $\phi_0 = -2500$ using the mesh sizes $\Delta \bar{x} = \Delta \bar{y} = .025$ for $\bar{a} = 3$ and 4, have been carried out. In Fig. 5 we show the equipotential ($\bar{\phi}$) lines and in Fig. 6 the lines of constant charge $\bar{\rho} (= \rho)$

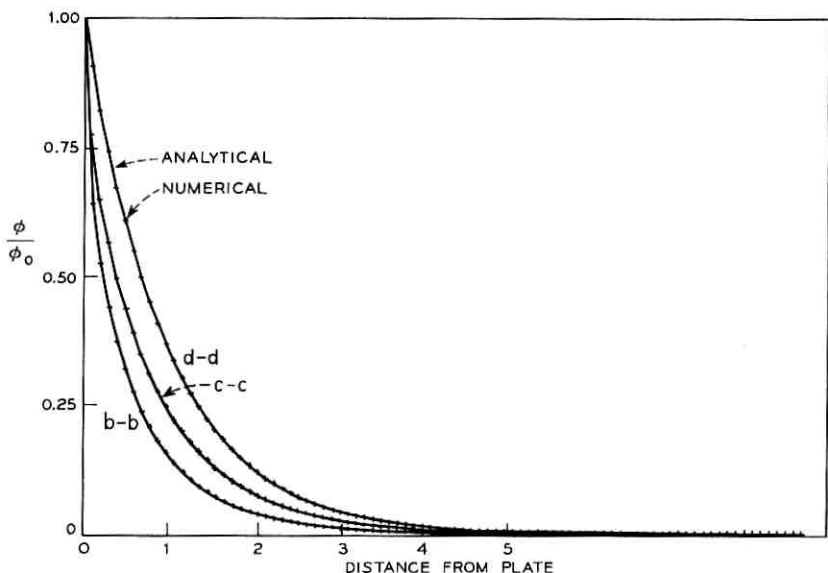


Fig. 4—A comparison of the analytical solution of the linearized problem (continuous line) with the numerical solution of the nonlinear problem P2 (crosses) for the case $\eta = 0$, $\phi_0 = -.01$. The three curves are the potential along $b - b$ ($x \leq 0, y = 0$), $c - c$ ($x = 0, y \leq 0$) and $d - d$ ($x = 4, y \leq 0$).

for the case $\phi_0 = -500$. These curves show clearly the depletion layer and the transition layer. The curves for $\phi_0 = -100$ and -2500 (which we do not show) are essentially the same; they differ only in that for $\phi_0 = -100$ the transition layer is thicker while for $\phi_0 = -2500$ it is sharper. It might be pointed out that in the tilde coordinates the thickness of the transition layer becomes vanishingly small as $|\phi_0| \rightarrow \infty$,

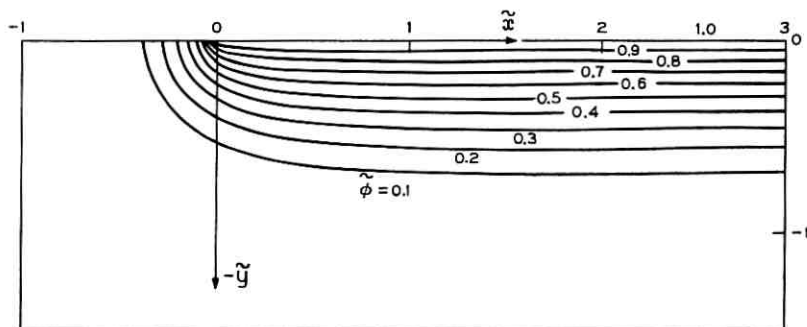


Fig. 5—Equipotential curves for the case $\eta = 0$, $\phi_0 = -500$, $d = 3$.

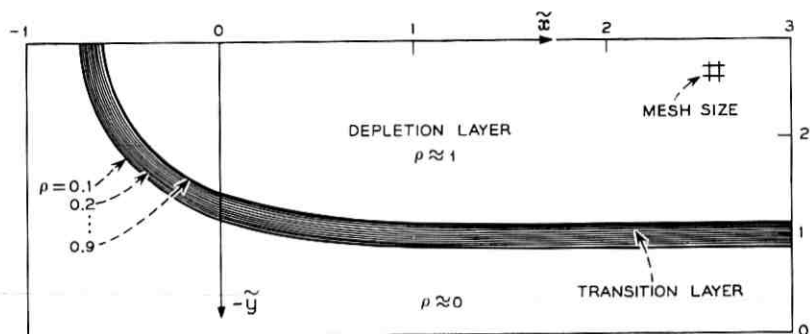


Fig. 6—Constant charge density curves for the case $\eta = 0$, $\phi_0 = -500$ and $a = 3$.

but in the true dimensional coordinates, the thickness is $\sim 4\lambda_D$, independent of ϕ_0 as $|\phi_0| \rightarrow \infty$.

Figures 5 and 6 also show that for $x \gtrsim 2$, the equipotential and equi-charge curves are essentially parallel to the x axis. Thus $a \gtrsim 2R$ is a sufficient condition for replacing the strip of finite width by a semi-infinite strip. Furthermore, since $\partial\bar{\phi}/\partial x = \partial^2\bar{\phi}/\partial x^2 = 0$ in the region $x \gtrsim 2$, the solution is one dimensional here. In Ref. (1) it was shown that an excellent approximate solution of the one dimensional problem is the "zeroth-order matching" solution

$$\bar{\phi} = \begin{cases} (1 - |\bar{y}|)^2, & -1 \leq \bar{y} \leq 0 \\ 0, & \bar{y} \leq -1. \end{cases} \quad (21)$$

From equation (21) we also see that for $x \geq 2$

$$\begin{aligned} \bar{E}_x &= -\frac{\partial\bar{\phi}}{\partial x} = 0, \\ \bar{E}_y &= -\frac{\partial\bar{\phi}}{\partial y} = \begin{cases} -2(1 - |\bar{y}|), & -1 \leq \bar{y} \leq 0, \\ 0, & \bar{y} \leq -1. \end{cases} \end{aligned} \quad (22)$$

In Fig. 7 we plot $\bar{\phi}$ along the lines $b - b(x \leq 0, \bar{y} = 0)$ $c - c(x = 0, \bar{y} \leq 0)$ and $d - d(x = 3, \bar{y} \leq 0)$ for the case $\phi_0 = -500$ and $a = 3$. We superimpose on this a plot of $\bar{\phi}(3, \bar{y})$ as given by (21). The agreement on $d - d$ between $\bar{\phi}$ calculated by the finite difference method and $\bar{\phi}$ given by (21) is excellent. This will be the case even for larger mesh sizes since the truncation error due to approximating the Laplacian by a five point difference scheme is proportional to the fourth derivatives of $\bar{\phi}$, which should be small since $\bar{\phi}$ is essentially parabolic along $d - d$.

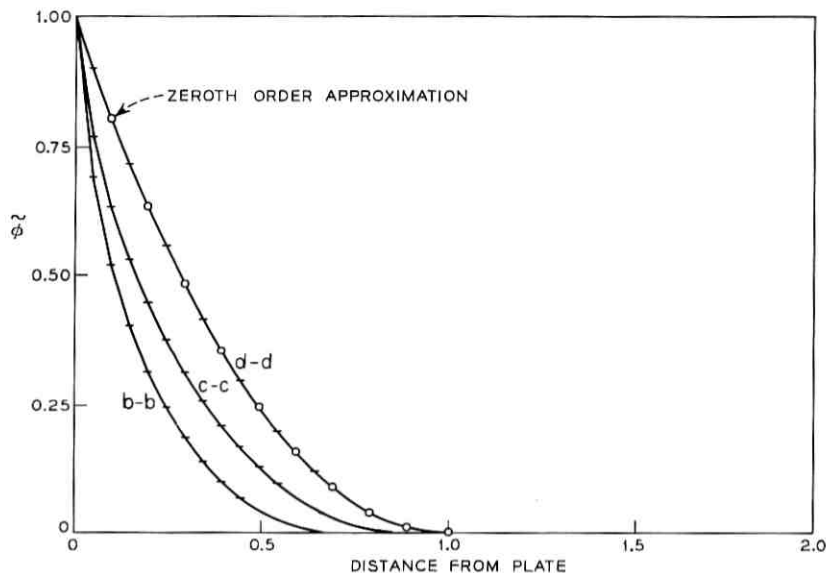


Fig. 7—Graphs of ϕ along the lines $b - b$ ($x \leq 0, \bar{y} = 0$), $c - c$ ($x = 0, \bar{y} \leq 0$) and $d - d$ ($x = 3, \bar{y} \leq 0$) for the case $\eta = 0, \phi_0 = -500$, and $a = 3$. Superimposed on the $d - d$ curve is a plot of $\phi(3, \bar{y})$ (circles) given by equation (21).

In Fig. 8 we plot $|\tilde{E}_x|$ along $b - b$ and $|\tilde{E}_y|$ along $c - c$ and $d - d$ for the same case, and we superimpose on this a plot of $|\tilde{E}_x(3, \bar{y})|$ as given by equation (22).

Because of the singularities of the electric field (and all higher derivatives of the potential) near the edge $x = \bar{y} = 0$, one cannot expect to obtain a uniformly valid numerical solution there. Nevertheless, comparison of the numerical and analytical solutions for the small potential case (Fig. 4) shows that even at the nearest mesh points to the plate edge, the error in the numerical computation of the potential is not large. The error in calculating the electric field is naturally greater. In order to decrease the truncation error (the difference between the exact solution of the differential equation and the solution of the difference equations), one can decrease the mesh size uniformly over the computational field. This can become quite expensive and is not necessary. A more efficient scheme is to refine the mesh size only in a small region around the plate edge and to find the numerical solution inside this region using at the boundary the values obtained from the coarser grid. We have done this for a region of size 1×1 around the plate edge with a mesh size of $\Delta x = \Delta y = .0125$ in the case $\phi_0 = -500$ and $\tilde{a} = 3$.

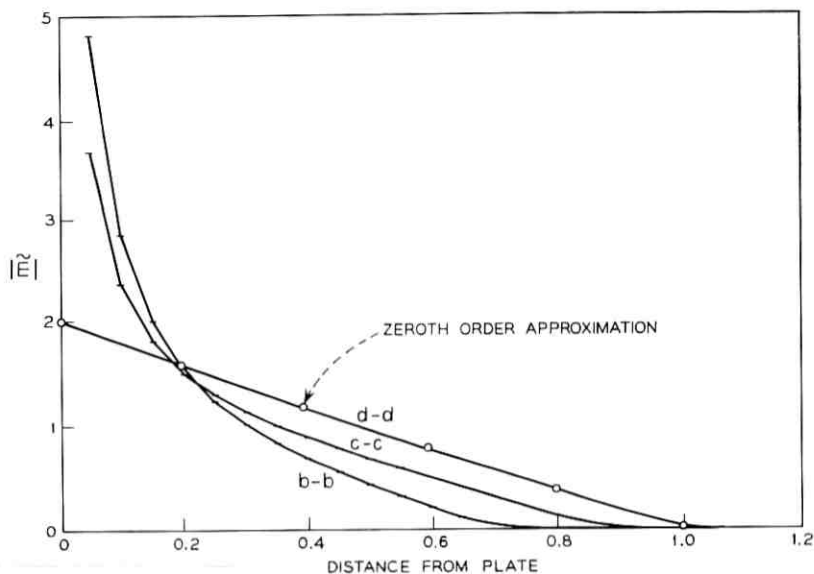


Fig. 8—A plot of $|\tilde{E}_z|$ along $b - b$ ($\tilde{x} \leq 0, \tilde{y} = 0$) and $|\tilde{E}_y|$ along $c - c$ ($\tilde{x} = 0, \tilde{y} \leq 0$) and $d - d$ ($\tilde{x} = 3, \tilde{y} \leq 0$) for the case $\eta = 0, \phi_0 = -500$, and $a = 3$. Superimposed on the $d - d$ curve is a plot (circles) of $|\tilde{E}_y(3, \tilde{y})|$ given by equation (22).

For this refined solution we plot $\tilde{\phi}$ along $b - b$ and $c - c$ in Fig. 9, and in Fig. 10 we plot $|\tilde{E}_z|$ along $b - b$ and $|\tilde{E}_y|$ along $c - c$. The squares are points obtained from the solution using the coarser mesh ($\Delta\tilde{x} = \Delta\tilde{y} = .05$), and we see that several grid points away from the plates, the two solutions agree nicely.

In addition to equipotential curves, curves of constant field amplitude can also be plotted. We define

$$\tilde{\psi}(x, \tilde{y}) = \left[\left(\frac{\partial \tilde{\phi}}{\partial \tilde{x}} \right)^2 + \left(\frac{\partial \tilde{\phi}}{\partial \tilde{y}} \right)^2 \right]^{1/2} = \|\tilde{\mathbf{E}}\|. \quad (23)$$

From equation (23) we see that $\max_{\tilde{y}} \tilde{\psi}(x, \tilde{y}) = \tilde{\psi}(x, 0) = 2$ for $x \geq 2$. Near the plate edge, however, there are much higher fields. In Fig. 11 we draw the curves of constant $\tilde{\psi}$, again for the case $\phi_0 = -500$ and $\tilde{a} = 3$. The plate edge region can be defined as the region where $\tilde{\psi}(x, \tilde{y}) \geq 2$. In Fig. 12 we show in more detail the curves of constant $\tilde{\psi}$ in the plate edge region. The data is taken from the calculation of the potential using a refined mesh around the plate edge discussed above.

An important application of the numerical solution is the computation

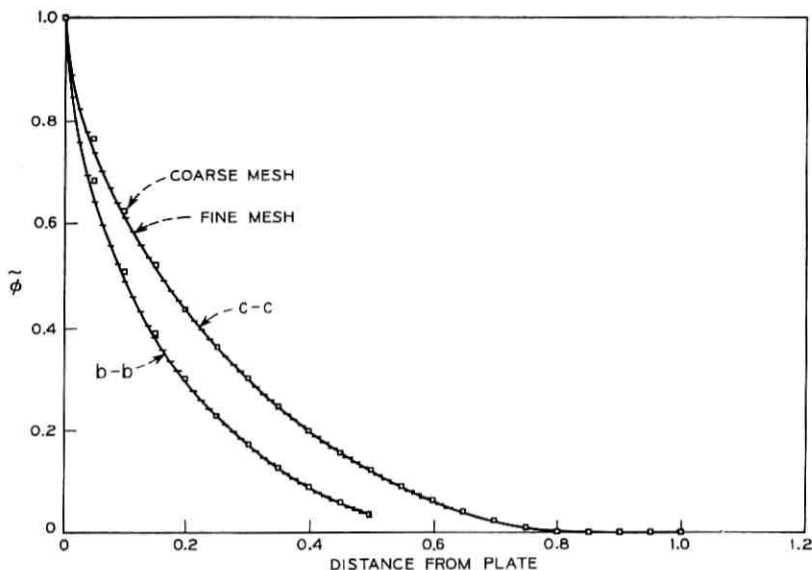


Fig. 9—Plots of $\bar{\phi}$ along $b-b$ ($\bar{x} \leq 0, \bar{y} = 0$) and $c-c$ ($\bar{x} = 0, \bar{y} \leq 0$) for the case $\eta = 0, \phi_0 = -500$, and $a = 3$. The solutions were obtained with a mesh size of $\Delta\bar{x} = \Delta\bar{y} = .0125$ (continuous line) and $\Delta\bar{x} = \Delta\bar{y} = .05$ (squares).

of the capacitance per unit length of the strip, C^* , where

$$C^* = \frac{\partial Q^*}{\partial \phi_0^*}, \quad (24)$$

and Q^* is the total (dimensional) charge per unit length on the strip. To calculate Q^* we note that it is just equal and opposite to the total net charge per unit length in the semiconductor:

$$Q^* = - \int_{-\infty}^0 \int_{-\infty}^{\infty} \rho^*(x^*, y^*) dx^* dy^*. \quad (25)$$

We introduce the nondimensional (tilde) coordinates normalized to the depletion layer thickness and write

$$Q^* = -qN_a\lambda_D^2R^2 \int_{-\infty}^0 \int_{-\infty}^{\infty} \bar{\rho} d\bar{x} d\bar{y} = 2\epsilon_1\phi_0^* \int_{-\infty}^0 \int_{-\infty}^{\infty} \bar{\rho} d\bar{x} d\bar{y}. \quad (26)$$

This last integral, which is almost independent of ϕ_0^* for large $|\phi_0^*|$, was computed numerically by evaluating the sum $\sum \sum \bar{\rho} \Delta\bar{x} \Delta\bar{y}$ and we write the result in the form

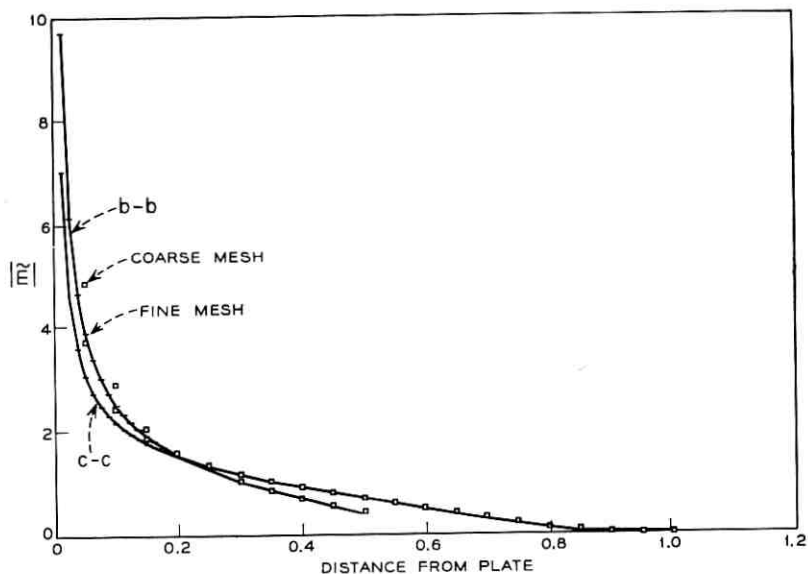


Fig. 10—Plots of $|\vec{E}_x|$ along $b - b$ ($\tilde{x} < 0, \tilde{y} = 0$) and $|\vec{E}_y|$ along $c - c$ ($\tilde{x} = 0, \tilde{y} \geq 0$) for the case $\eta = 0, \phi_0 = -500$, and $\tilde{a} = 3$. The solutions were obtained with a mesh size of $\Delta\tilde{x} = \Delta\tilde{y} = .0125$ (continuous line) and $\Delta\tilde{x} = \Delta\tilde{y} = .05$ (squares).

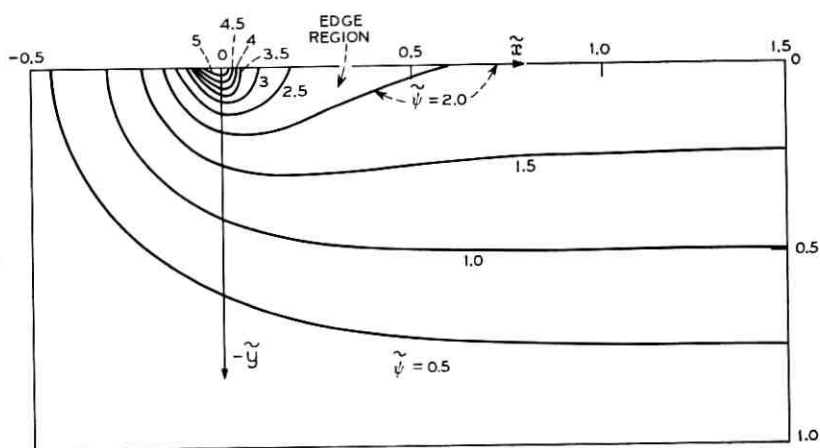


Fig. 11—Curves of constant field amplitude for the case $\eta = 0, \phi_0 = -500$, and $\tilde{a} = 3$.

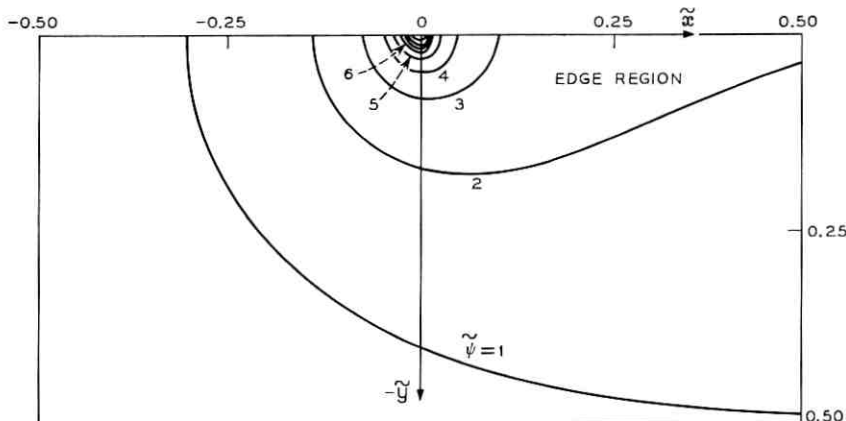


Fig. 12—A detailed plot of the constant field curves near the edge of the plate for the case $\eta = 0$, $\phi_0 = -500$ and $\bar{a} = 3$.

$$\iint \bar{\rho} \, d\bar{x} \, d\bar{y} = 2\bar{a} + 2D_0. \quad (27)$$

The first term on the right of equation (27) is the value we would obtain if the effects of the edges and the transition layer were ignored, and D_0 is the correction for these effects. For $\bar{a} \gg 1$ and $|\phi_0| \gg 1$, D_0 should be essentially independent of \bar{a} and ϕ_0 , that is, of a^* and ϕ_0^* . For our basic computation, $\bar{a} = 3$, $\phi_0 = -500$ and $\Delta\bar{x} = \Delta\bar{y} = .05$, we obtained

$$D_0 = 0.354. \quad (28)$$

This value of D_0 was very insensitive to increasing \bar{a} to 4, and it changed only slightly when $|\phi_0|$ was increased to 2500 (with $\Delta\bar{x} = \Delta\bar{y} = .025$). When we insert equation (27) into (26), expressing \bar{a} in dimensional coordinates, we obtain

$$Q^* = 4\epsilon_1\phi_0^* \left(\frac{a^*}{R^*} + D_0 \right), \quad (29)$$

where

$$R^* = \lambda_D R = (-2\epsilon_1\phi_0^*/qN_d)^{\frac{1}{2}} \quad (30)$$

is the dimensional depletion layer thickness. From equations (24), (29) and (30) we get

$$C^* = \frac{2a^*\epsilon_1}{R^*} \left(1 + 2D_0 \frac{R^*}{a^*} \right). \quad (31)$$

It is clear that our calculations also give an accurate approximation to the potential around a conducting disc of radius r_0^* , situated at the air-semiconductor interface, and charged to a large negative potential, as long as

$$r_0 \gg R, \quad (32)$$

where

$$r_0 = r_0^*/\lambda_D. \quad (33)$$

In all previous expressions for the potential and fields we need only interpret x^* as the radial coordinate r^* .

We can now also calculate the capacitance of a circular disc. In this case we have

$$Q^* = -qN_d\lambda_D^3 R^3 \int_{-\infty}^0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\rho} \, dx \, d\bar{y} \, dz, \quad (34)$$

and

$$\int_{-\infty}^0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{\rho} \, dx \, d\bar{y} \, dz = \pi\bar{r}_0^2 + 2\pi\bar{r}_0 D_0, \quad (35)$$

where

$$\bar{r}_0 = r_0/R, \quad (36)$$

and of course we still have $D_0 = 0.354$. If we substitute equations (35) and (36) into (34) and differentiate with respect to ϕ_0^* , we get

$$C^* = \frac{\epsilon_1 \pi r_0^{*3}}{R^*} \left(1 + 4D_0 \frac{R^*}{r_0^*} \right). \quad (37)$$

Finally, the potential in the air, $\bar{y} > 0$ is related to the potential on the interface $\bar{\phi}(\bar{x}, 0)$ by Green's formula⁷

$$\bar{\phi}(\bar{x}, \bar{y}) = \frac{1}{\pi} \int_{-\infty}^{\infty} \bar{y} [(x - \xi)^2 + \bar{y}^2]^{-1/2} \bar{\phi}(\xi, 0) \, d\xi, \quad (38)$$

where implicit use has been made of boundary condition (8a). Using the values of $\bar{\phi}(\xi, 0)$ obtained from the finite difference solutions in $\bar{y} < 0$, the integral in (38) has been evaluated numerically to give the potential in $\bar{y} > 0$.

III. THE CASE $\eta \neq 0$

When $\eta = \epsilon_0/\epsilon_1$ is not zero, the problem of solving P1 is complicated by the fact that the solutions for the potential in the air and in the

semiconductor become coupled and cannot be found separately. Nevertheless, the effects of this coupling can be taken into account in the semiconductor, without actually solving the problem in the air, by using Green's formula (38). We did this and found the solution in the semiconductor by an iterative finite difference scheme. Using this solution on the boundary, the integral in (38) was evaluated numerically to obtain the solution in the air. The details of this numerical scheme are given in Section IV. We continue to concentrate on the case of large potential ($|\phi_0| \gg 1$) and a wide conducting strip ($a \gg R$), so the normalized (tilde) variables introduced in equation (19) are retained.

Except in a wedge shaped region at the interface at each edge of the strip (see Fig. 13), the solution in the semiconductor is very insensitive to changes in η and is little different from the solution for $\eta = 0$. The main new feature of the solution in the semiconductor is the appearance of a "shoulder" in the depletion layer near the surface, and a smearing out of the transition layer there. In fact, the larger η is, the less sharp the transition layer in this region is. This effect increases the charge in the wedges and thus also increases the capacitance of the plate. The solution in the air is, however, more sensitive to changes in η .

We illustrate these qualitative remarks by a number of graphs. The case illustrated in all these graphs is $\phi_0 = -500$, $\bar{a} = 3$, and the mesh size used was $\Delta\tilde{x} = \Delta\tilde{y} = .05$. In Fig. 14 we show the equipotential ($\tilde{\phi}$) curves and in Fig. 13 we show the curves of constant charge ($\tilde{\rho}$) for $\eta = .1$. These two figures illustrate the previous remarks. In Fig. 15 we give graphs of $\tilde{\phi}$ along the lines $b - b(\tilde{x} \leq 0, \tilde{y} = 0)$ and $c - c(\tilde{x} = 0, \tilde{y} \leq 0)$ for $\eta = 0, .05$, and $.1$. In Fig. 16 we give a vector plot of the field around the plate for $\eta = .1$.

Finally, we have numerically evaluated the integral (27) in order to

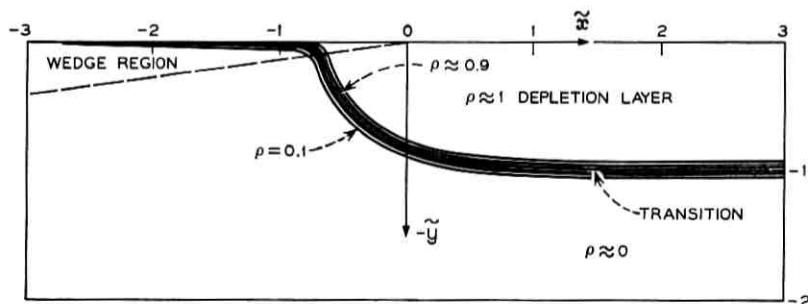


Fig. 13—Constant charge density curves for the case $\eta = .1$, $\phi_0 = -500$, and $\bar{a} = 3$.

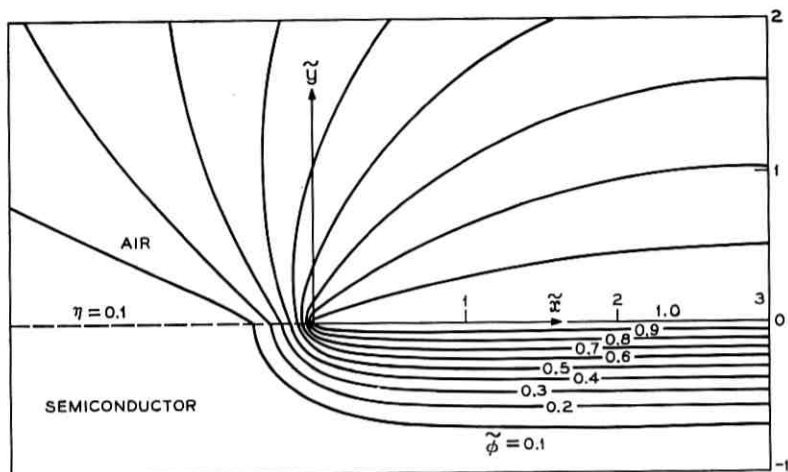


Fig. 14—Equipotential curves for the case $\eta = .1$, $\phi_0 = -500$, and $a = 3$.

calculate the capacitance for $\eta \neq 0$. This was done for $\phi_0 = -500$, $\tilde{a} = 3$, for $\eta = 0, .05$, and $.1$ with $\Delta x = \Delta y = .05$, and for $\eta = 0, .05, .1, .15$ and $.2$ with $\Delta x = \Delta y = .1$. For this range of values of η , it was found that (27) and the expressions (31) and (37) for the capacitance remain valid if $D_0 = .354$ is replaced by

$$D(\eta) = .354 + .43 \eta. \quad (39)$$

We believe the number .43 in equation (39) is correct to about 10 percent.

If we had approximated the depletion layer by a rectangle with sides $2a^*$ and R^* completed at each end with a quarter circle of radius R^* , we would have obtained $D(\eta) \equiv \pi/4$. With this value of $D(\eta)$, equation (37) yields the approximation of Goodman⁸ and of Sze and Gibbons.⁹ By a completely different technique Copeland¹⁰ has independently obtained the values $D(.078) = .375$ and $D(.0625) = .365$ which agree rather well with our results.

IV. DETAILS OF THE NUMERICAL METHODS

In this section we discuss some of the details of the various finite difference schemes used in solving numerically the boundary value problems P1 and P2. In each of the three cases studied, $\eta = 0$ and $|\phi_0| \ll 1$, and $\eta \neq 0$, $|\phi_0| \gg 1$, the basic nonlinear partial differential equation (6) or (20) is replaced by a system of finite difference equations,

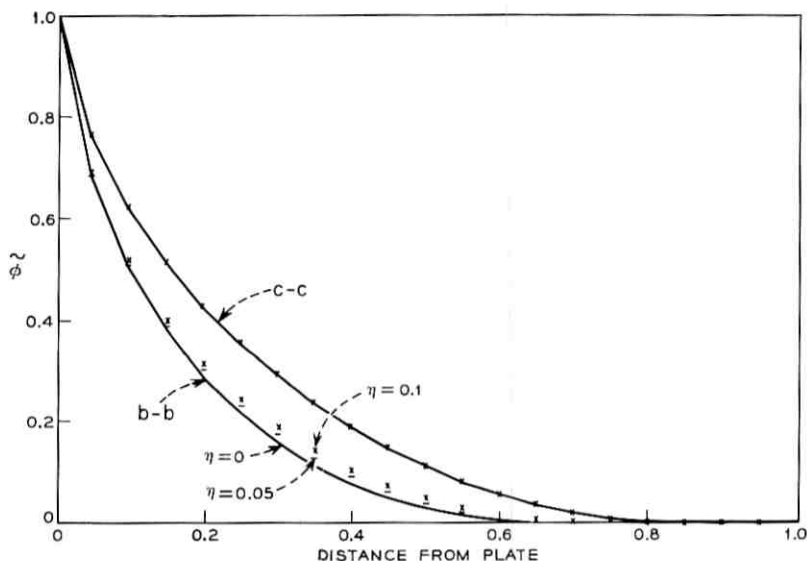


Fig. 15—Graphs of $\bar{\phi}$ along the lines $b - b$ ($x \leq 0, \bar{y} = 0$) and $c - c$ ($x = 0, \bar{y} \leq 0$) for the case $\phi_0 = -500$ and $\bar{a} = 3$ for $\eta = 0, .05$, and $.1$.

but the boundary conditions, or the method of solving the finite difference equations differ from case to case.

In each case the infinite plane is replaced by a finite rectangle, as shown in Fig. 17, where advantage is taken of the symmetry of the problem around $x = a$. The rectangle is subdivided into a square mesh with mesh spacing h , so the rectangle has length Mh and depth Nh . In the small potential case, we define

$$\phi_{i,j} = \phi(a - (i - 1)h, -(j - 1)h), \quad (40a)$$

while in the large potential case

$$\phi_{i,j} = \bar{\phi}(\bar{a} - (i - 1)h, -(j - 1)h), \quad (40b)$$

where ($1 \leq i \leq M + 1, 1 \leq j \leq N + 1$). In either case, at each interior mesh point, $\nabla^2 \phi_{i,j}$ is approximated by the five point formula⁶

$$\nabla_h^2 \phi_{i,j} = (\phi_{i+1,j} + \phi_{i-1,j} + \phi_{i,j+1} + \phi_{i,j-1} - 4\phi_{i,j})/h^2. \quad (41)$$

At the boundary points ($i = 1, 2 \leq j \leq N$) we make use of the basic symmetry to write

$$\nabla_h^2 \phi_{1,j} = (2\phi_{2,j} + \phi_{1,j+1} + \phi_{1,j-1} - 4\phi_{1,j})/h^2, \quad (42)$$

and similarly in the case $\eta = 0$ we write

$$\nabla_h^2 \phi_{i,1} = (\phi_{i+1,1} + \phi_{i-1,1} + 2\phi_{i,2} - 4\phi_{i,1})/h^2 \quad (43)$$

for $(M_1 + 1 < i \leq M)$, where $M_1 h = a$ for $|\phi_0| \ll 1$ and $M_1 h = \bar{a}$ for $|\phi_0| \gg 1$.

When $\eta = 0$, $|\phi_0| \ll 1$, equation (6) is replaced by the difference equations

$$\nabla_h^2 \phi_{i,j} = \exp(-|\phi_{i,j}|) - 1 \quad (44)$$

for $(1 \leq i \leq M, 2 \leq j \leq N)$ and $(M_1 + 1 < i \leq M, j = 1)$. Since $\phi \leq 0$ everywhere, the replacement of $\exp \phi$ by $\exp -|\phi|$ changes nothing analytically, but numerically it eliminates certain instabilities in some iteration schemes. Equation (43) must be supplemented by further equations obtained from the boundary conditions. The condition $\phi = \phi_0$ on the strip yields

$$\phi_{i,1} = \phi_0, \quad (1 \leq i \leq M_1 + 1). \quad (45)$$

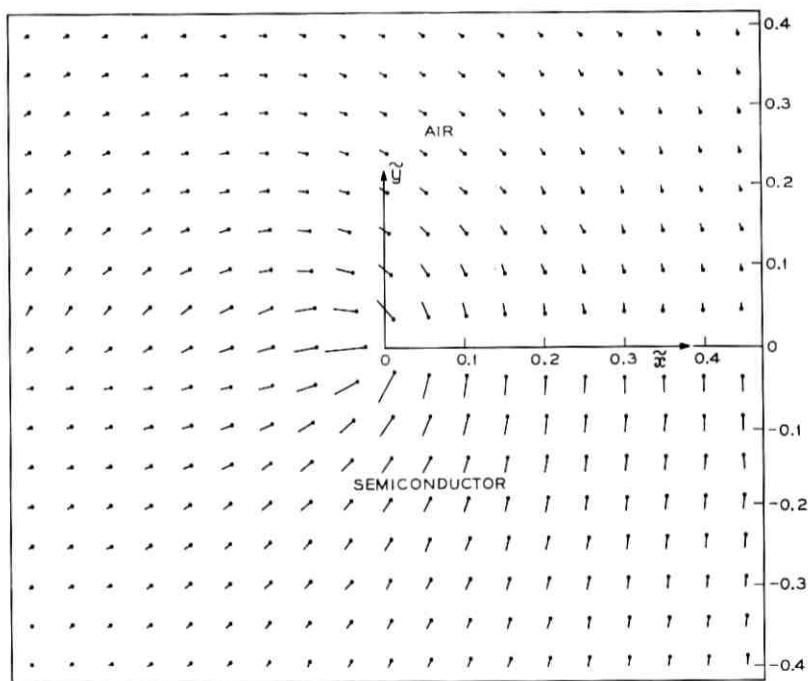


Fig. 16—A vector plot of the field around the edge of the plate for the case $\eta = .1$, $\phi_0 = -500$, and $a = 3$. Each line segment has the direction of the field and the length of the segment is proportional to the magnitude of the field.

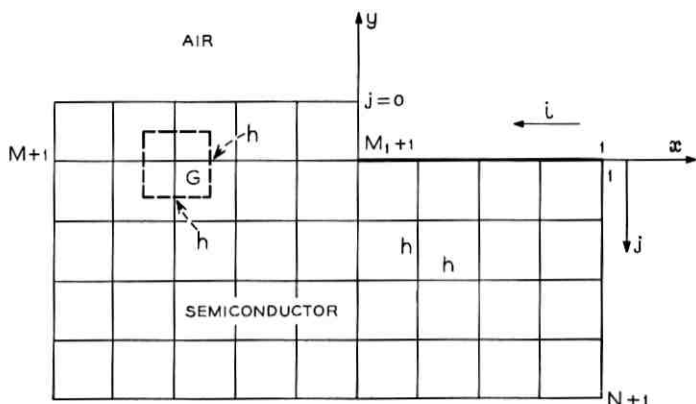


Fig. 17—A schematic drawing of the square grid used in the numerical calculations.

On the remaining boundaries condition (11) is approximated by requiring the vanishing of the normal derivatives, which decay faster than the potential itself. This yields the finite difference equations

$$\phi_{M+1,j} = \phi_{M,j}, \quad (1 \leq j \leq N), \quad (46)$$

$$\phi_{i,N+1} = \phi_{i,N}, \quad (1 \leq i \leq M). \quad (47)$$

Equations (44), (45), (46) and (47) form a system of $MN + M + N$ equations for the values of $\phi_{i,j}$ at $MN + M + N$ mesh points ($\phi_{M+1,N+1}$ does not appear in these equations).

This set of transcendental equations has a unique solution¹¹ which was determined numerically by a point successive over-relaxation iterative method.^{12,13} Briefly described, the method is as follows: an initial guess, $\phi_{i,j}^{(0)}$, is made for $\phi_{i,j}$. At each interior point subsequent iterations $\phi_{i,j}^{(n)}$ are determined by the equations

$$\bar{\phi}_{i,j}^{(n+1)} = \frac{1}{4} \{ \phi_{i+1,j}^{(n)} + \phi_{i-1,j}^{(n+1)} + \phi_{i,j+1}^{(n)} + \phi_{i,j-1}^{(n+1)} - h^2 \exp(-|\phi_{i,j}^{(n)}|) + h^2 \}, \quad (48)$$

$$\phi_{i,j}^{(n+1)} = \omega \bar{\phi}_{i,j}^{(n+1)} + (1 - \omega) \phi_{i,j}^{(n)}. \quad (49)$$

The over-relaxation parameter ω is given by¹³

$$\omega = 1 + \sigma^2 / [1 + (1 - \sigma^2)^{\frac{1}{2}}]^2 \quad (50a)$$

$$\sigma = \frac{1}{2} \{ \cos(\pi/N) + \cos(\pi/M) \}. \quad (50b)$$

This value of ω is optimal for the Dirichlet problem for Laplace's

equation. Equation (48) must be modified in an obvious way at the boundary points where (44) is satisfied, and the $\phi_{i,j}$ at the remaining boundary points must be eliminated with the aid of equations (45) through (47). The iterations are repeated until either

$$\delta_{\infty} = \{ \max_{i,j} | \phi_{i,j}^{(n+1)} - \phi_{i,j}^{(n)} | \} / | \phi_0 | \quad (51a)$$

or

$$\delta_2 = \{ \sum_i \sum_j | \phi_{i,j}^{(n+1)} - \phi_{i,j}^{(n)} |^2 \}^{1/2} / \{ | \phi_0 | (MN + M + N)^{1/2} \} \quad (51b)$$

is suitably small. It is important to note that the nonlinear term, $\exp(\phi_{i,j})$, appears on the right hand side of equation (48). For this reason the method is referred to as an explicit scheme. Typically in the small potential case we chose $M_1 = 40$, $M = 100$, $N = 92$, $h = .1$, and $\phi_{i,j}^{(0)} = 0$, and after 50 iterations we obtained $\delta_{\infty} < 10^{-3}$ and $\delta_2 < 10^{-4}$.

The iteration scheme just outlined is only conditionally stable, and it can be shown that an approximate condition for its convergence when applied to a system of equations of the form

$$\nabla_k^2 \phi_{i,j} = f(\phi_{i,j}) \quad (52)$$

is

$$\omega \left(1 + \frac{h^2 C}{4} \right) < 2, \quad (53a)$$

where

$$C = \max f'(\phi). \quad (53b)$$

When $| \phi_0 | \ll 1$, condition (53a) does not impose a severe restriction, since the region over which the solution varies significantly is small. Then h , M and N can be chosen so that M and N are not too large and at the same time h and ω satisfy equation (52) for economically acceptable ω . When $| \phi_0 | \gg 1$, however, the depletion layer is so large that M and N must be chosen excessively large in order for (52) to be satisfied.

In order to avoid this difficulty, we have employed an implicit scheme instead of an explicit one. For convenience we first introduce rescaled (tilde) coordinates in (19). For $\eta = 0$, the resulting finite difference equations are just those given in equations (41) through (47) except that (44) is replaced by

$$\nabla_k^2 \phi_{i,j} = 2 \{ 1 - \exp(-| \phi_0 \phi_{i,j} |) \}. \quad (54)$$

The explicit iteration scheme defined by equations (48) and (49) is replaced by the following implicit scheme. The quantity $\bar{\phi}_{i,j}^{(n+1)}$ is now the root of the transcendental equation

$$\begin{aligned} \bar{\phi}_{i,j}^{(n+1)} - \frac{h^2}{2} \exp \{-|\phi_0 \bar{\phi}_{i,j}^{(n+1)}|\} \\ = \frac{1}{4} \{\phi_{i+1,j}^{(n)} + \phi_{i-1,j}^{(n+1)} + \phi_{i,j+1}^{(n)} + \phi_{i,j-1}^{(n+1)} - 2h^2\}. \end{aligned} \quad (55)$$

The solution of this equation, which can be found by the Newton-Raphson method, is then substituted into equation (49) to obtain $\phi_{i,j}^{(n+1)}$. Equation (55) must be modified in an obvious way at the boundary points where (43) is satisfied, and the $\phi_{i,j}$ at the remaining boundary points are eliminated as before. When $0 < \omega < 2$, the iteration scheme just defined can be shown to be unconditionally stable,¹² that is, it converges for any mesh size h and any $\phi_{i,j}^{(0)}$. However, in practice we found $\omega \sim 1.5$ to provide the most rapid convergence. The rapidity of convergence is not very sensitive to small changes in ω around $\omega = 1.5$. Typically in the large potential case for $\eta = 0$ we chose $M_1 = 60$, $M = 100$, $N = 40$, and $h = .05$, and after ~ 80 iterations we obtained $\delta_\infty < 10^{-4}$ and $\delta_2 < 1.6 \times 10^{-5}$.

When $\eta \neq 0$, this iteration scheme must be modified to take into account the coupling of the potential in the semiconductor to the potential in the air. If we attempt to solve directly for the potential in both the air and in the semiconductor by the method of finite differences, we encounter difficulties. The potential in the air decays very slowly at infinity, so that if we use a reasonable number of mesh points, boundary conditions such as (46) and (47) introduce fairly large errors into the calculation. In order to circumvent this problem without using an inordinately large number of mesh points we proceed as follows. We replace boundary condition (8b) by the equivalent condition that⁴

$$-\int_{\partial G} \epsilon \frac{\partial \phi^*}{\partial n^*} dl^* = \int_G \int \rho^* dx^* dy^*, \quad (56)$$

where G is a small square shown in Fig. 17, of side h centered at the boundary point ($M_1 + 1 < i < M$, $j = 1$). The integral on the left of equation (56) is a line integral around the boundary of G , and $\partial/\partial n^*$ denotes the normal derivative. The finite difference approximation to (56) yields the equation in tilde coordinates

$$\begin{aligned} \phi_{i,2} + \eta \phi_{i,0} + \frac{(1+\eta)}{2} (\phi_{i+1,1} + \phi_{i-1,1}) - 2(1+\eta)\phi_{i,1} \\ = h^2 \{1 - \exp(-|\phi_0 \phi_{i,1}|\}\}, \end{aligned} \quad (57)$$

where $\phi_{i,0}$ is the potential at the mesh point in the air ($y = h$). Equation (57) now replaces (43) at the air-semiconductor interface points, and yields the iteration equations

$$\begin{aligned} \bar{\phi}_{i,1}^{(n+1)} - \frac{h^2 \exp(-|\phi_0 \bar{\phi}_{i,1}^{(n+1)}|)}{2(1+\eta)} = \frac{1}{4} \{ \phi_{i-1,1}^{(n+1)} + \phi_{i+1,1}^{(n)} \} \\ + \frac{1}{2(1+\eta)} \{ \eta \phi_{i,0}^{(n)} + \phi_{i,2}^{(n)} - h^2 \}, \quad (M_1 + 1 < i < M). \end{aligned} \quad (58)$$

Once the $(n+1)$ iterates $\phi_{i,i}^{(n+1)}$ have been determined for all points within and on the boundary of the semiconductor, the $\phi_{i,0}^{(n+1)}$, ($M_1 + 1 < i \leq M$), are calculated from the finite difference version of Green's formula (38) using the values $\phi_{i,1}^{(n+1)}$ already known. We did not prove that this modified iteration scheme converges, but it works well in practice. In this case also, the empirically determined value $\omega = 1.5$ seems optimal.

Finally, we make several brief comments about the difference between the true solutions of P1 or P2 and the numerical solutions, that is, the truncation errors. It is well known¹³ that when the true solution is smooth enough (specifically when the fourth derivatives are bounded) the truncation error is $O(h^2)$. However, in our problem, near the plate's edge $x = y = 0$, even the first derivative is not bounded, and the above estimate fails. Wasow¹⁴ considered a problem similar to ours with smooth boundaries and piecewise analytic boundary values and found that the truncation error vanished when $h \rightarrow 0$.

It can be shown¹⁵ that near the corner $x = y = 0$, the singularity in the field for arbitrary η is

$$\nabla \phi = O(r^{-\frac{1}{2}}) \quad (59)$$

for both the small and large potential cases (in the appropriate coordinates) where r is the distance from the corner. Bramble, and others,¹⁶ investigated the truncation (or discretization) error of such problems, and if we combine equation (59) with their Theorem 3.1, we find that near the corner the truncation error is

$$|\phi_{i,i} - \phi_{\text{exact}}(x, y)| = O(h^{\frac{1}{2}}), \quad (60)$$

and it vanishes as $h \rightarrow 0$.

The truncation error is not uniform, and because of the slow convergence near the corner, it is worthwhile to refine the mesh size in a small region around it (as was described in Section II). This was done for the small potential case for $\eta = 0$ and $\phi_0 = -.01$, for a refined mesh size of $h = .0125$ and in Fig. 18 we compare the analytic solution

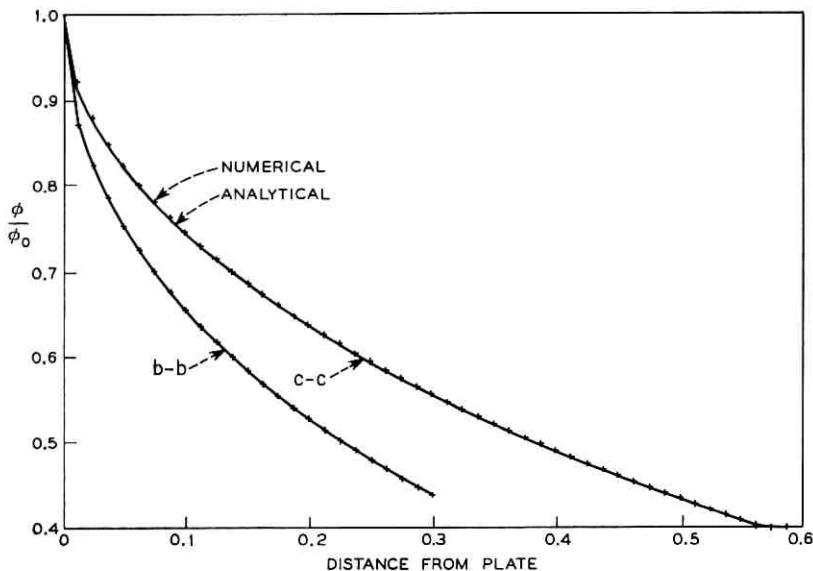


Fig. 18—A comparison of the analytical solution of the linearized problem (continuous line) with the numerical solution of the nonlinear problem P2 (crosses) for the case $\eta = 0$, $\phi_0 = .01$ and for a mesh size near the plate edge of $\Delta x = \Delta y = .0125$.

(continuous line) with the numerical one (crosses) along the lines $b - b$ and $c - c$.

All the calculations described in this paper were performed on a GE 635 digital computer. The graphical output was obtained with the aid of a routine for calculating level curves written by G. S. Deem, L. K. Russell, and N. J. Zabusky.¹⁷

V. ACKNOWLEDGMENTS

The authors are indebted to H. K. Gummel, J. A. Lewis, E. G. Nicolian, R. J. Strain and S. Sze for many helpful conversations during the work on this problem. They also wish to acknowledge the expert assistance of Miss C. G. Miller in efficiently translating into computer programs the many numerical schemes proposed.

REFERENCES

1. Lewis, J. A., McKenna, J., and Wasserstrom, E., "The Field of Negative Point, Line or Plane Charges in an n -Type Semiconductor," unpublished work.
2. Sze, S., *Physics of Semiconductor Devices*, New York: John Wiley and Sons, Inc., 1969, pp. 368, 370.

3. Landau, L. D., and Lifshitz, E. M., *Statistical Physics*, London: Pergamon Press, 1968, p. 231.
4. Panofsky, W. K. H., and Phillips, M., *Classical Electricity and Magnetism*, Cambridge, Mass.: Addison-Wesley Publishing Co., 1955, pp. 26, 27.
5. Lewis, J. A., "The Flat Plate Problem for a Semiconductor," to be published in the September 1970 issue of the B.S.T.J.
6. Abramowitz, M., and Stegun, I. A., *Handbook of Mathematical Functions*, Washington: National Bureau of Standards, 1964, pp. 297, 885.
7. Titchmarsh, E. C., *Theory of Fourier Integrals*, Oxford: Oxford University Press, 1948, p. 125.
8. Goodman, A. M., "Metal-Semiconductor Barrier Height Measurement by the Differential Capacitance Method-Ore Carrier System," *J. Appl. Phys.*, *34*, No. 2 (February 1963), p. 329.
9. Sze, S. M., and Gibbons, G., "Effect of Junction Curvature on Breakdown Voltage in Semiconductors," *Solid-State Electronics*, *9*, No. 9 (September 1966), p. 831.
10. Copeland, J. A., "Diode Edge Effect on Doping-Profile Measurement," *IEEE Trans. Electron Devices*, *ED-17*, No. 5 (May 1970).
11. Bers, L., "On Mildly Nonlinear Partial Difference Equations of Elliptic Type," *J. Research Natl. Bur. Standards*, *51*, No. 11 (November 1953), p. 229.
12. Ortega, J. M., and Rockoff, M. L., "Nonlinear Difference Equations and Gauss-Seidel Type Iterative Methods," *J. SIAM Numer. Anal.*, *3*, No. 9 (September 1966), p. 497.
13. Varga, R. S., *Matrix Iterative Analysis*, Englewood Cliffs, N. J.: Prentice-Hall, 1962, pp. 59, 165, 203.
14. Wasow, W. R., "The Accuracy of Difference Approximations to Plane Dirichlet Problems with Piecewise Analytic Boundary Values," *Quart. Appl. Math.*, *15*, No. 4 (April 1957), p. 53.
15. Lewis, J. A., and Wasserstrom, E., "The Field Singularity at the Edge of an Electrode on a Semiconductor Surface," to be published in the July-August 1970 issue of the B.S.T.J.
16. Bramble, J. H., Hubbard, B. E., and Zlamal, M., "Discrete Analogues of the Dirichlet Problem with Isolated Singularities," *SIAM J. Numer. Anal.*, *5*, No. 3 (March 1968), p. 1.
17. Deem, G. S., Russell, L. K., and Zabusky, N. J., unpublished work.

Television Transmission of Holograms Using a Narrow-Band Video Signal

By J. E. BERRANG

(Manuscript received December 18, 1969)

It has been experimentally verified that holograms can be transmitted via television using the synthesized carrier-frequency method. A vidicon camera system was used to transmit three on-axis holograms to the receiver where the video signal was used to amplitude modulate a 15 MHz carrier. This modulated carrier was applied to the control grid of a cathode-ray tube, and the synthesized carrier-frequency hologram was then photographed from the tube screen. Several of the experimental problem areas, such as mechanical stability, camera tube storage effects, and display-recording process optimization, are considered in detail.

I. INTRODUCTION

Enloe, Murphy, and Rubinstein transmitted off-axis or carrier-frequency holograms by means of television.¹ Although they were successful, they were severely restricted by the limited resolution of the television camera tube. This fundamental difficulty has its roots in the fact that a carrier-frequency hologram has spatial frequencies at least four times higher than the highest spatial frequency contained in the subject. (This fact is elaborated on in Ref. 2.)

Burckhardt and Doherty³ were later able to produce carrier-frequency holograms using an on-axis reference beam. Figure 1 shows the arrangement they used. A Mach-Zehnder interferometer was used to align the subject and reference beams, and a means was provided for shifting the phase of the reference beam. A transmission grating was placed in front of the photographic plate used to record the hologram. A series of four exposures was made on a single photographic plate. Between exposures, the transmission grating was translated laterally by one-fourth of its spatial period, and the reference beam was shifted in phase by 90°; the resulting synthesis at the photographic plate was a carrier-frequency hologram.

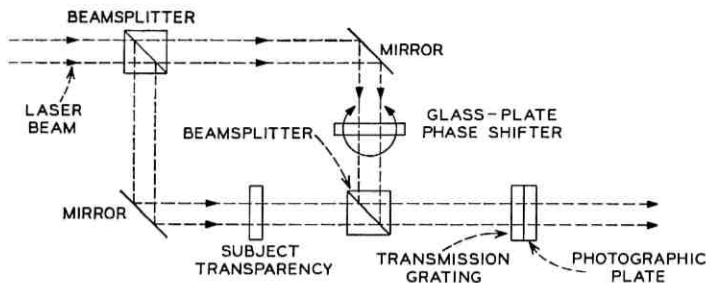


Fig. 1—Experimental arrangement used by Burckhardt and Doherty to synthesize carrier-frequency holograms from multiple on-axis holograms.

Burckhardt and Enloe² suggested that this method is applicable to the transmission of holograms over a television system and showed that its use would lower the resolution requirements on the television camera by a factor of four. They also indicated that the minimum number of on-axis holograms required to synthesize a carrier-frequency hologram is three, with corresponding phase shifts of 120° being used. This paper reports on the implementation of their proposal. Figure 2 shows the basic technique.

A hologram of a subject transparency is formed with an on-axis reference beam at the photosensitive surface of a vidicon camera tube and is transmitted as a baseband signal to a receiver where it is multiplied with an electrical grating signal and applied to the control grid of a cathode-ray tube. The resulting display is captured photographically.

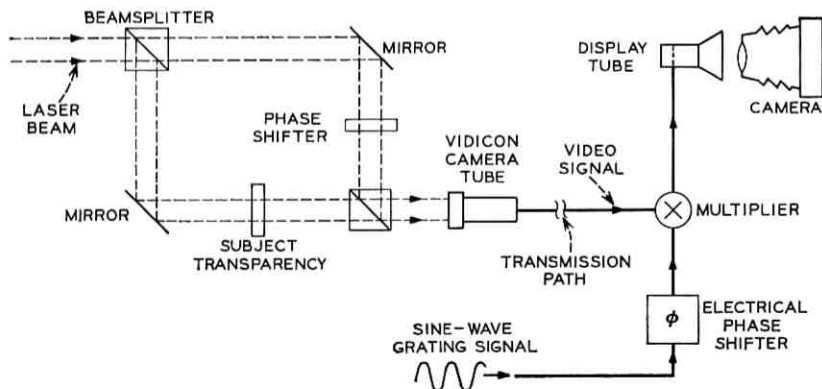


Fig. 2—Hologram transmission via television, with reduced resolution requirements on the camera tube.

Three exposures are recorded on a single transparency with the reference beam and electrical grating signal each shifted in phase by 120° between exposures. The result is a synthesized carrier-frequency hologram which, when illuminated with coherent light, reconstructs an image of the subject.

II. DESCRIPTION OF THE EXPERIMENT

2.1 *General*

Figure 3 is a diagram of the system that was used for producing and recording synthesized carrier-frequency holograms with a television system. The light source used was a Spectra-Physics, model 124, 15 milliwatt, helium-neon laser operating at 6328 Angstroms; an attenuator was used to reduce the light intensity to a level suitable for application to the vidicon camera tube. An on-axis hologram of a two-dimensional transparency was formed directly on the photosensitive surface of the vidicon camera tube, no camera lens being used. A rotatable glass plate was used in the reference beam path to provide phase shift.

The television camera system used was a RAM (Canoga Electronics) V1000 closed-circuit television system, and the cathode-ray display tube was a Westinghouse WX-30176P-4 with dynamic focusing. The dynamic-focusing system provided a component of focus voltage which was a function of the distance from the scanning spot to the center of the display screen. This system was designed by M. E. Lukacs.⁴

The output of a 30 MHz crystal oscillator was divided by two in a single binary count-down stage to give the 15 MHz signal which was used to provide the electrical equivalent of the optical transmission grating used by Burckhardt and Doherty. This electrical grating signal was then shifted in phase by passing it through appropriate lengths of coaxial cable to provide three phases with 120° separation between them. The desired phase shift was selected by means of a manual switch. This grating signal was then amplitude modulated with the holographic video information from the television camera, amplified in a tuned amplifier, and applied to the control grid of the cathode-ray tube.

The 15 MHz grating frequency was divided by 1024 in a ten-stage binary count-down circuit giving the system horizontal line rate and at the same time insuring that the grating remained fixed with respect to the horizontal dimension of the cathode-ray tube display.

The vertical scanning rate was 60 Hz, synchronized to the ac power line; consequently, any 60 or 120 Hz modulation of the horizontal scanning due to power-frequency hum was fixed on the cathode-ray tube

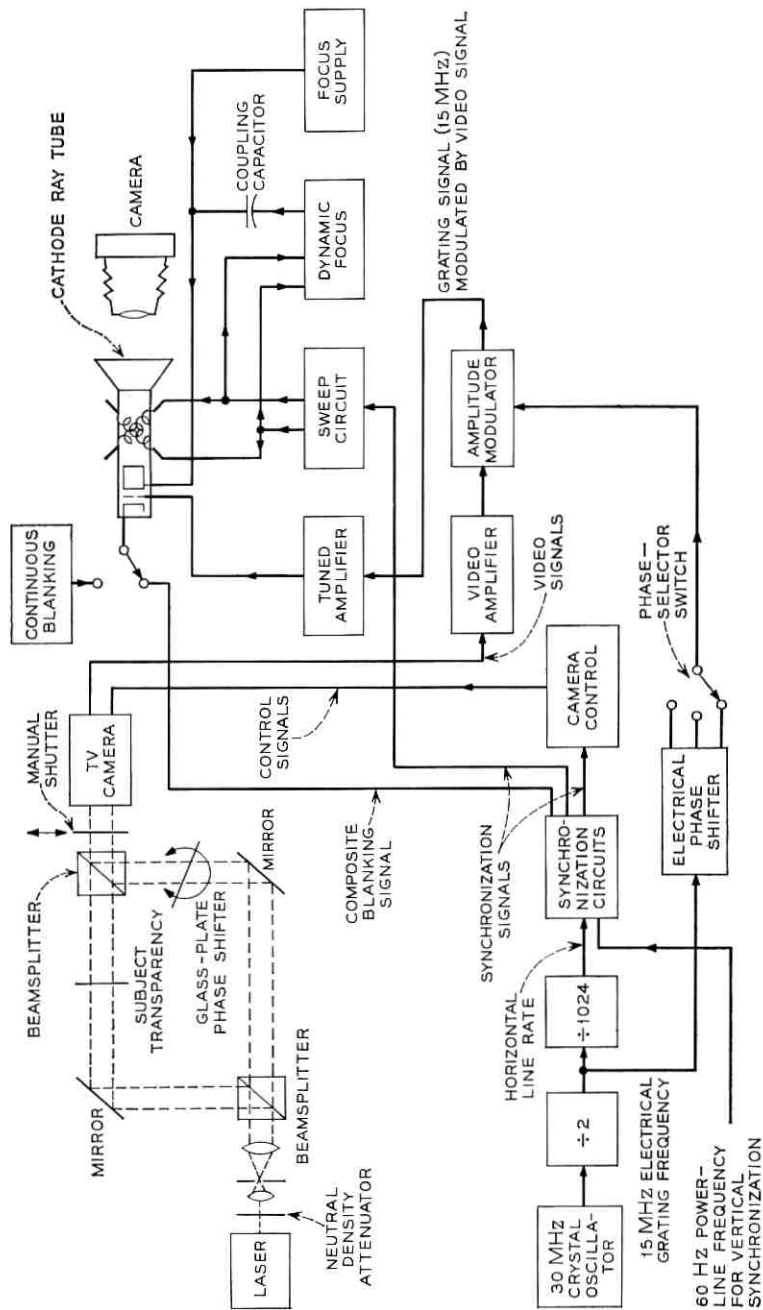


Fig. 3—Experimental arrangement used to transmit holograms over a television system.

screen. This stabilization of hum patterns was essential because of the extreme fineness of the spatial grating produced during each exposure.

The system horizontal and vertical scanning rates were not related. This scanning format allowed for removal of the scanning-line structure by means of photographic time exposures. The photographic recording system consisted of a tripod-mounted camera.

2.2 *Technique*

As part of the experimental procedure, care was taken to prevent reflections of laser light from causing unwanted interference patterns at the vidicon photosensitive surface. The front surface of the vidicon face plate was given an antireflecting coating as were both sides of the glass plate used for phase-shifting the reference beam, and light-absorbing screens were provided where they were deemed necessary.

Any differential change in the dimensions of the two legs of the interferometer during the formation of a hologram would cause the phase of the reference beam to shift with respect to that of the object beam. In this experiment, where three phases of reference beam were used, it was necessary to maintain the relative path lengths constant to about $1/30$ of a wavelength of light. The interferometer, as well as the laser and television camera, were mounted on a heavy iron table which was floated on partially inflated innertubes to isolate it from building vibrations. The table was completely covered with a plexiglass shield to keep out air currents, and remote controls were provided for all adjustments which had to be made to the apparatus on the table during the production of a hologram.

If the vidicon camera tube were illuminated with one of the three phase patterns for an extended time, that pattern would remain stored on the vidicon photosensitive surface interfering with the proper addition and cancellation of the three phases of reference beam during the making of a hologram. To obviate this difficulty, the vidicon was mechanically shuttered between exposures and when not in use.

The image aspect ratio at the camera had to be faithfully reproduced at the cathode-ray tube screen. That is to say, the horizontal and vertical size relationship of an image at the cathode-ray tube was required to be the same as it was at the camera tube. Otherwise, an additional cylindrical lens would have to be used during reconstruction to overcome the resulting astigmatism.

A switch was provided to permit operation of the cathode-ray tube in either of two blanking modes: (i) retrace blanking, and (ii) continuous blanking. To prevent displacement of the recording camera during the

multiple-exposure process, the camera shutter was left open for the full time (about two minutes), and the interexposure shuttering was done by blanking the cathode-ray tube. Exposure times were relatively long, and this shuttering could be performed manually.

An experimentally developed two-step photographic recording process was used. Three exposures of 25 seconds each at $f/8$ were made on a single Polaroid 55 PN film. The light output from the cathode-ray tube during this part of the recording process was about 0.15 foot-lamberts. To make the best use of the available resolution, the image size was adjusted to fill the Polaroid film. When reconstructions were made directly from the Polaroid transparencies, certain degradations (apparently due to phase distortion) were observed. Therefore, a second photographic step was introduced in which the synthesized hologram was transferred to a Kodak 649F glass plate. Because of the high resolution of the 649F emulsion and to save processing time, several (two to four) holograms were usually placed on a single plate at a proportionately reduced size. The Polaroid transparencies were placed between two glass plates and illuminated from behind with a uniform incandescent light source. The luminance of the transparencies when thus illuminated was adjusted to be about 650 foot-lamberts. The 649F plate was then exposed at $f/22$ for 3 minutes and 20 seconds.

Care was taken during each photographic step to insure that exposure was restricted to the linear portion of the film's characteristic curve. It was experimentally determined that following the exposure data given above and the manufacturers' recommendations for processing gave good results. If clipping or compression of part of the visual information were experienced due to operation in the nonlinear regions of the transfer characteristics of either of the films involved or of the cathode-ray tube, the operating conditions outlined by Burekhardt and Enloe² were not achieved and poor reconstructions resulted.

2.3 Reconstruction

The holograms were reconstructed using the system shown in Fig. 4. A hologram was illuminated with a collimated beam of laser light and the resulting output Fourier transformed with a lens, L . A spatial-frequency filter was used to discard one image and the direct beam, leaving the other image to be viewed or photographed. Because the reconstructed image was small, an additional lens was used to magnify it. Figure 5 shows examples of reconstructed images.

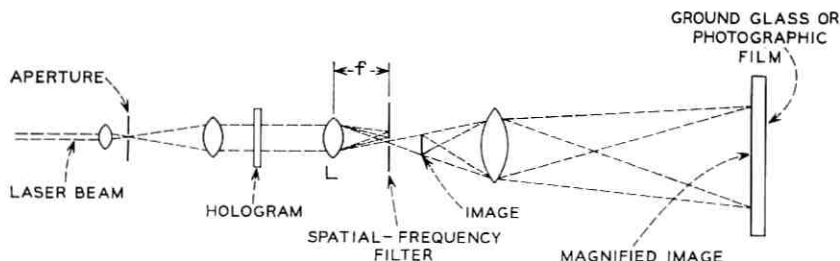


Fig. 4—Reconstruction of synthesized carrier-frequency hologram.

2.4 Remarks

The modulator and amplifier circuits (Fig. 3) contain broadly-tuned stages designed to have a bandwidth of 1 MHz. The finest horizontal detail that can be recognized in the reconstructions (Fig. 5) is that represented by the girl's teeth, or 45 cycles per picture width. The received holograms occupied 80 percent of the scanned width at the cathode-ray tube, and the scan-to-retrace ratio was two. The horizontal line rate was $15 \times 10^6/1024$ lines per second. The highest spatial frequency that is apparent in the reconstructions is, therefore, equivalent to an electrical frequency of

$$45 \frac{\text{cycles}}{\text{picture width}} \times 1 \frac{\text{picture width}}{0.8 \text{ scanned width}} \times 1.5 \frac{\text{scanned width}}{\text{line}} \\ \times \frac{15 \times 10^6 \text{ lines}}{1024 \text{ second}} = 1.2 \text{ MHz.}$$

Wide-band circuits could of course be used, but because of the low signal-to-noise ratio of the narrow-band reconstructions, no experiments were made with wider bandwidth signals.

Several problem areas (that is, display tube resolution, relative motion between camera and display tube, and film resolution) associated with the display and recording of the hologram could be eased by reducing the frequency of the electrical grating signal. The frequency of this signal could easily be reduced by a factor of two, even if a wider band video signal were used.

The two-step photographic process described above could be reduced to a one-step process by using a glass plate having a higher speed than the 649F plate. The Kodak Super Panchro-Press plate is four orders of

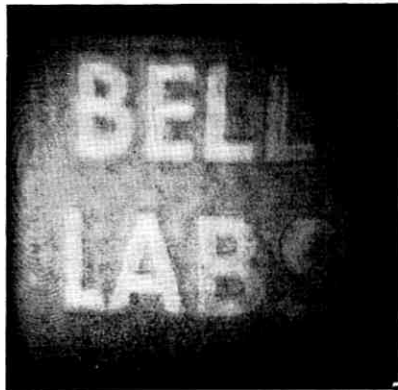


Fig. 5—Original and reconstructions of "girl in checkered shirt," and reconstruction of "BELL LABS."

magnitude faster than the 649F and has sufficient resolution but is somewhat more granular. It would be a good plate to try.

III. SUMMARY

It has been experimentally verified that holograms can be transmitted via television using the synthesized carrier-frequency method discussed by Burckhardt and Enloe.² A vidicon camera system was used to transmit three on-axis holograms to the receiver where the video signal was used to amplitude modulate a 15 MHz carrier. This modulated carrier was applied to the control grid of a cathode-ray tube, and the synthesized carrier-frequency hologram was then photographed from the tube screen. Reconstructions were obtained which contain resolution compatible with the video bandwidth used but which have low signal-to-noise ratios. Several of the experimental problem areas, such as mechanical stability, camera tube storage effects, and display-recording process optimization, were considered in detail, and suggestions were made for lessening some of these difficulties.

IV. ACKNOWLEDGMENTS

The author wishes to express his gratitude to D. W. Doughty for antireflecting coatings on the vidicon and the glass-plate phase shifter, to M. E. Lukacs for the design and construction of the dynamic-focusing system, and to C. B. Burckhardt, L. H. Enloe, and A. B. Larsen for many invaluable discussions.

REFERENCES

1. Enloe, L. H., Murphy, J. A., and Rubinstein, C. B., "Hologram Transmission via Television," *B.S.T.J.*, 45, No. 2 (February 1966), pp. 335-339.
2. Burckhardt, C. B., and Enloe, L. H., "Television Transmission of Holograms with Reduced Resolution Requirements on the Camera Tube," *B.S.T.J.*, 48, No. 5 (May-June 1969), pp. 1529-1535.
3. Burckhardt, C. B., and Doherty, E. T., "Formation of Carrier-Frequency Holograms with an On-Axis Reference Beam," *Appl. Opt.*, 7, No. 6 (June 1968), pp. 1191-1192.
4. Lukacs, M. E., unpublished work.

Burst Distance and Multiple-Burst Correction*

By J. D. BRIDWELL and J. K. WOLF

(Manuscript received December 22, 1969)

This paper is concerned with burst error, burst erasure and combined burst-error and burst-erasure correction. Part I introduces the concept of burst distance and subsequently develops burst-correcting properties of a code relative to its burst distance. Part II discusses product codes for multiple-burst correction (MBC). The MBC properties of a product of two codes are derived from the properties of the original codes. The correction of spot errors is generalized to multiple-spot correction. Theorems are presented which strengthen the single-burst correcting (SBC) properties of some codes. A class of codes which corrects single, triple and quadruple bursts and 5 single errors is developed, and a decoding procedure is given. Finally, a code from the new class of MBC codes is compared with three other MBC codes.

I. INTRODUCTION

It is a property of many burst-noise channels that bursts occur not singly, but in bursts of bursts or in random multiple bursts.¹ For this reason, single burst-correcting (SBC) codes do not give good error-control performance on such channels. It is therefore desirable to have codes which correct multiple bursts within a given block. Some multiple-burst correcting (MBC) codes have been known for some time. However, until recently the complexity of the decoding process has not been comparable to that of correcting single bursts. Naturally, we do not expect the decoding process to be as simple for MBC codes as for SBC codes. We would expect the ideal complexity of a double-BC code to be in the same ratio as a SBC code that a double-error-correcting complexity is to a single-error-correcting complexity.

* This work was partially supported by the U. S. Air Force Office of Scientific Research under Contract AF 49 (638)-1600. This material was taken from a dissertation submitted to the Faculty of the Polytechnic Institute of Brooklyn in partial fulfillment of the requirements for the Ph.D. degree in Electrical Engineering.

The Reed-Solomon² codes are useful for multiple-burst correction as well as for single-burst correction. Consider a t -error-correcting Reed-Solomon code over $GF(2^m)$ transmitted in binary. No burst of $(\lceil t/l \rceil - 1)m + 1^*$ can corrupt more than t/l successive symbols of a code word. Therefore, all patterns of l bursts each of length $(\lceil t/l \rceil - 1)m + 1$, $l = 1, \dots, t$ are correctible. The disadvantage of these codes is that the decoding process is more difficult than we would like. Operations must be performed over $GF(2^m)$, which introduces equipment complexity that such operations over $GF(2)$ do not require. MacWilliams³ has had some success in converting Reed-Solomon codes over $GF(2^3)$ into binary codes. In general, however, the operations that must be performed to decode a t -error-correcting Reed-Solomon code are nonbinary operations.

Interleaved codes are suitable for MBC as well as SBC.^{4,5} A t -error-correcting code interleaved b times corrects all patterns of $\lceil t/l \rceil$ bursts each of length lb for each integer $l = 1, \dots, t$. Binary interleaved codes have implementation advantages over nonbinary Reed-Solomon codes. However, interleaved codes reduce the single problem of correcting t bursts of b to b separate problems of correcting t errors. Correcting t errors becomes a less and less trivial problem as t becomes large.

Other multiple-burst-correcting codes have been proposed by Wolf,⁶ Stone,⁷ and by Kasahara and Kasahara.⁸

Bahl and Chien⁹ show that the 3-dimensional product codes with simple even parity check subcodes are double-burst-correcting. They also demonstrate a simple decoding procedure. Bahl and Chien generalized their results, showing that $m + 1$ -dimensional product codes with simple even parity check subcodes correct m bursts within a block. A 3-dimensional Bahl and Chien code of block length $n_1 n_2 n_3$ is generated by

$$g(x) = \frac{(x^{n_1 n_2} + 1)(x^{n_1 n_3} + 1)(x^{n_2 n_3} + 1)}{(x^{n_1} + 1)(x^{n_2} + 1)(x^{n_3} + 1)}$$

where n_1 , n_2 and n_3 are pairwise relatively prime.

This paper generalizes the use of product codes for multiple-burst correction and presents a class of MBC product codes.

II. BURST DISTANCE

2.1 Introduction

In this part, we introduce a distance measure for bursts and relate the burst error, burst erasure and combined burst-error and burst-

* $\lceil x \rceil$ represents the largest integer within x .

erasure-correcting properties of a code to its burst distance, we also give a simple example of the synthesis of a code with given burst distance.

2.2 A Distance Measure for Bursts

A burst of length b is a set of b consecutive symbols the first of which is nonzero. A pattern of m bursts of length b is measured in an analogous manner. Let the first burst of b be b consecutive symbols beginning with a nonzero symbol. Let the first nonzero symbol following that burst begin the second burst of b , and so forth. This measure can be taken cyclically only when closed-loop burst patterns are allowed. If closed-loop patterns are allowed, the first burst of b is allowed to begin at any nonzero symbol among the first b symbols. The true burst measure will be defined then as the minimum measure thus obtained. For example, consider the pattern:

$$\begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{array}$$

Beginning at the second position, a total of 3 bursts of 5 are measured. Beginning at position 4, however, only 2 bursts of 5 are measured. Since there are only 2 nonzero symbols among the first 5 symbols, these two cases are sufficient to define the measure as 2 bursts of 5. The minimum number of bursts of b in a nonzero code word is denoted d_b , the minimum burst- b distance.

2.3 Correction Capabilities of a Code With Burst Distance d_b

We next relate the correction capability of a code to its minimum burst- b distance. Theorem 1.0 is a generalization of the relation of the number t of errors correctible by a code to d_1 , its minimum Hamming distance¹:

$$t = \left\lfloor \frac{d_1 - 1}{2} \right\rfloor.$$

Theorem 1: A linear code with minimum burst- b distance d_b corrects all patterns of $\lfloor (d_b - 1)/2 \rfloor$ bursts of b .

Proof: The sum of two patterns each of $\lfloor (d_b - 1)/2 \rfloor$ or fewer bursts of b cannot be a code word, since all code words have at least d_b bursts of b . Therefore, all patterns of $\lfloor (d_b - 1)/2 \rfloor$ or fewer bursts of b are correctible.

Theorem 2 is a generalization of the erasure correction capability e of a code with minimum Hamming distance d_1 : $e = d_1 - 1$.²

Theorem 2: A code with minimum burst- b distance d_b corrects all patterns of $d_b - 1$ erasure bursts of b .

Proof: Suppose a code word C_1 is transmitted and that $d_b - 1$ or fewer erasure bursts of b occur. Then filling in the erased symbols with all possible combinations from the signaling alphabet guarantees that at least one code word, C_1 , will result. Suppose that another code word, say C_2 , also results. Then C_1 and C_2 differ by only $d_b - 1$ or fewer bursts of b , meaning that another code word, $C_1 + C_2$ has burst distance $d_b - 1$ or less, contrary to hypothesis. Then all patterns of $d_b - 1$ or fewer erasure bursts of b are correctible.

Theorem 3 combines single-burst-erasure-correcting (SBXC) with SBC. Its principal importance is its application to a cyclic code with r check digits, which is $r - SBXC$.²

Theorem 3: A linear code that is $r - SBXC$ corrects any erasure burst of length $e > 0$ directly preceded by a burst of b_1 and directly followed by a burst of b_2 provided that b_1 and b_2 are arbitrary but fixed integers such that $b_1 + b_2 + e \leq r$.

Proof: Erasing the b_1 symbols preceding the erasure burst and the b_2 symbols following the erasure burst produces an erasure burst of r or less, which is correctible.

The final theorem 4, states the combined MBC and MBXC ability of a code with minimum burst- b distance d_b .

Theorem 4: A linear code with minimum burst- b distance d_b corrects all patterns of m_1 bursts of b and m_2 erasure bursts of length $e_i \neq 0$, each directly preceded by a burst of b_{1i} and followed by a burst of b_{2i} ($i = 1; \dots, m_2$) if $2m_1 + m_2 < d_b$ and if b_{1i} and b_{2i} are arbitrary but fixed integers such that $b_{1i} + e_i + b_{2i} \leq b$.

Proof: The first step is to make each error-erasure burst a pure erasure burst by erasing the b_{1i} symbols preceding and the b_{2i} symbols following the i th erasure burst e_i , $i = 1, \dots, m_2$. The result is m_1 pure error bursts and m_2 pure erasure bursts, each of length b or less. Assume that the code word C_1 was transmitted. Fill in all erasures with all possible combinations of symbols. At least one filled-in result C'_1 is m_1 bursts of b from a code word, thus C'_1 can be corrected to C_1 . Assume there are two code words C_1 and C_2 each m_1 or fewer bursts of b from C'_1 . Then C_1 and C_2 differ from each other by $m_1 + m_2$ or fewer bursts of b . Then some code word $C_1 + C_2$ has burst distance $2m_1 + m_2$, which is impossible. Next suppose that some other filled-in sequence C_2

is within m_1 bursts of b of a code word C_3 . Then C_1 and C_3 differ by $m_1 + m_1 + m_2$ or fewer bursts of b , which is impossible. Therefore there is one and only one code word that will result from the decoding procedure defined.

2.4 An Example of a Code with Minimum Burst- b Distance d_b

The correction capabilities of a code with minimum burst- b distance d_b have been presented. No mention has been made of how to construct a code with minimum burst- b distance d_b , however. An example of such a construction is interleaving b times a code with minimum Hamming distance $d_1 = d_b$. The interleaved code has minimum burst- b distance d_b , while maintaining the minimum distance, also d_b , of the original code. This example illustrates that the minimum burst- b distance of a code need be no greater than the minimum Hamming distance. Product codes are investigated in Section III for MBC properties, further illustrating the usefulness of the theory developed in this section.

III. PRODUCT CODES

3.1 Introduction

In this part, we show the burst distance of a product code to be a function of the parameters of the subcodes. Such parameters are the burst distance, Hamming distance, number of check symbols and block length. Elspas' spot-error correction is generalized to multiple-spot correction.⁷ We introduce a class of product codes which corrects single, triple and quadruple bursts and five single errors, and present a simple decoding algorithm which allows decoding by subcodes. Finally, the performance of codes from this class is compared with other known MBC codes.

3.2 Product Codes

A two-dimensional product code^{2,7} is a two-dimensional array as indicated in Figure 1. Each row is a code word from a systematic block code with block length n_1 , number of information symbols k_1 , number of check symbols $n_1 - k_1 = r_1$, and minimum Hamming distance $d_{1,1}$. This subcode* is an (n_1, k_1) code. The column subcode, also a systematic block code, is an (n_2, k_2) code with minimum Hamming distance $d_{1,2}$.

* The use of the terminology "subcode" for the row code or column code is at variance with another use of this term. In this paper, a subcode will always refer to either the row code or the column code of a product code.

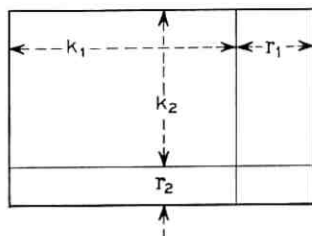


Fig. 1—A two-dimensional product code.

This two dimensional definition of a product code can readily be extended to a multidimensional product code. For example, a three-dimensional product code is formed by making each row in Figure 1 a code word from a two-dimensional product code. The column code is just a one-dimensional code. The generalization to more than three dimensions is easily made.

All product codes considered here are two-dimensional unless otherwise noted. Moreover, all subcodes are assumed to be linear, hence all product codes considered are linear. An (n, k) code that is the product of an (n_1, k_1) row code and an (n_2, k_2) column code has block length $n = n_1 n_2$, number of information symbols $k = k_1 k_2$ and minimum Hamming distance $d_1 = d_{1,1} d_{1,2}$.¹⁰ For notation, the product code (n, k) is given by $(n, k) = (n_1, k_1) \times (n_2, k_2)$. The transmission rate $R = k/n$ of the product codes is the product of the rates of the subcodes :

$$R = \frac{k}{n} = \frac{k_1 k_2}{n_1 n_2} = R_1 R_2 .$$

It is therefore clear that a product code with high rate requires subcodes with even higher rates. A moderate-rate product code with reasonably powerful subcodes might be readily achieved, while a high-rate product code with very powerful subcodes might be difficult. This is an important observation, since the properties of the product code depend on the properties of the subcodes. We investigate this at length later in Section 3.3.

The technique of iteration of codes was introduced by Elias.¹⁰ Elias proposed a coding system for use on the binary symmetric channel that produces an arbitrarily small error probability at a nonzero transmission rate. It is the only known block coding scheme for the binary symmetric channel without feedback with this property. Moreover, Elias' decoding strategy is straightforward and simple. The decoding scheme does not, however, correct all patterns of $\lfloor (d_1 - 1)/2 \rfloor$ errors.

It succeeds by correcting many patterns of more than $\lfloor (d_1 - 1)/2 \rfloor$ errors.

A general problem of decoding product codes by subcodes is that errors guaranteed* correctible by the minimum distance of the product code are not necessarily correctible by the technique of subcode decoding. Reddy¹² gave an algorithm that guarantees correction of $\lfloor (d_1 - 1)/2 \rfloor$ errors by subcode decoding if at least one of the subcodes is majority decodable.¹³ No less stringent general condition has yet been found.

Burton and Weldon¹⁴ showed that under certain conditions, product codes of cyclic subcodes are themselves cyclic. Burton and Weldon termed such a product code a cyclic product code. Abramson investigated cyclic product codes, introducing an interesting interleaving argument,¹⁵ which is summarized below.

Let the $n_1 n_2$ digits in the array of Figure 2 be represented by a poly-

		COLUMN INDEX c						
		0	1	2	3	4	5	6
ROW INDEX r	0	0	15	30	10	25	5	20
	1	21	1	16	31	11	26	6
	2	7	22	2	17	32	12	27
	3	28	8	23	3	18	33	13
	4	14	29	9	24	4	19	34

Fig. 2—An example of the mapping.

$$r = i \bmod (n_2); c = i \bmod (n_1).$$

nomial $f(x)$ of degree $n_1 n_2 - 1$ or less, that is,

$$f(x) = \sum_{i=0}^{n_1 n_2 - 1} f_i x^i.$$

If n_1 and n_2 are relatively prime, if the row and column sub-codes are cyclic, and if for

$$0 \leq i \leq n_1 n_2 - 1$$

we define

$$\text{and } \left. \begin{aligned} r &= i \bmod n_2 \\ c &= i \bmod n_1 \end{aligned} \right\}, \tag{1}$$

* Slepian¹¹ introduced the product code terminology. He showed that the generator matrix for the iteration of two codes is combinatorially equivalent to the tensor product of the individual generator matrices.

and map the digit in row r and column c of the array into f_i of $f(x)$, then the product code is cyclic. An example of this mapping for which $n_1 = 7$ and $n_2 = 5$ is shown in Fig. 2. The integer 17 corresponding to column index 3 and row index 2 is the 18th transmitted symbol or the term $f_{17}x^{17}$.

The generator polynomial of the product code is given by

$$g(x) = \frac{g_1(x^{n_2})g_2(x^{n_1})}{g.c.d. \{g_1(x^{n_2})g_2(x^{n_1})\}}$$

where $g_1(x)$ generates the row subcode and $g_2(x)$ generates the column subcode. Abramson points out that the mapping corresponds to interleaving a code word from $g_1(x)$ on every n_2 th digit of $f(x)$ and interleaving a code word from $g_2(x)$ on every n_1 th digit of $f(x)$. The interleaving argument has merit in providing simple proof of Burton's and Weldon's result and in providing a simple form for $g(x)$ of the cyclic product code. Another interesting consequence of the interleaving argument is that it suggests burst correction. Burst-correcting codes are trivially constructed by interleaving random-error-correcting codes. So why not form burst-correcting codes by using a rather peculiar interleaving?

Elsas investigated the burst-correcting properties of product codes.⁷ He showed that many error patterns could be corrected by detecting errors through column decoding, then erasing the columns with detected errors and using the row subcode to fill in the erased symbols. An interesting application is spot correction. A spot error is a two-dimensional pattern of errors within a product code array. Such errors might occur in a communication system in which the digital channel is considered to be a storage medium—a magnetic tape, for example.

3.3 Multiple-Burst Correction with Product Codes

We assume a two-dimensional (n, k) product code with minimum burst- b distance d_b . The row subcode is an (n_1, k_1) code with minimum burst- b distance $d_{b,1}$. The (n_2, k_2) column subcode has minimum burst- b distance $d_{b,2}$. For example, $d_{1,1}$ is the Hamming distance of the row subcode, while d_1 is simply the Hamming distance of the product code.

Several theorems will be proven regarding the minimum burst distance of product codes. Two different orders of transmitting the n_1n_2 digits will be considered. If the transmission is row-by-row, it will be called row transmission. If the transmission is in accord with the mapping (1), it will be called cyclic transmission. In some cases, one method of

transmission will allow stronger application of the theorem than the other. Any differences resulting from the transmission method will be discussed following the theorem.

Theorem 4 relates the minimum burst- b distance of a product code to the minimum burst- b distance of one of its subcodes and the minimum Hamming distance of the other subcode. We assume the row code to have minimum burst- b distance $d_{b,1}$ and the column code to have minimum Hamming distance $d_{1,2}$.

Theorem 4: $d_b \geq d_{b,1}d_{1,2}$.

Proof: We must show that every nonzero word in the product code has at least $d_{b,1}d_{1,2}$ bursts of b . Every nonzero row has at least $d_{b,1}$ bursts of b . Let $c_i, i = 1, \dots, d_{b,1}, \dots$ denote the column in which each burst begins. Then for any c_i and $c_{i+1}, c_{i+1} - c_i > b$, where closed-loop measure is allowed. Each column c_i must have at least $d_{1,2}$ nonzero entries. Clearly, a burst of b cannot intersect any given column more than once, hence each column c_i is intersected by at least $d_{1,2}$ bursts. Therefore, every nonzero word in the product code has at least $d_{b,1}d_{1,2}$ bursts of b .

There are two conditions under which equality in theorem 4 holds and one condition under which equality does not necessarily hold. For row transmission, it is always possible for a nonzero word of the product code to have exactly $d_{b,1}d_{1,2}$ bursts of b . To see this, choose exactly $d_{1,2}$ rows each containing exactly $d_{b,1}$ bursts of b such that each nonzero column is a word in the column subcode. Next, assume cyclic transmission. If $d_{b,1} = d_{1,1}$, then it is possible to have exactly $d_{b,1}$ nonzero columns each with $d_{1,2}$ nonzero elements, hence $d_b = d_{b,1}d_{1,2}$. The most interesting case is cyclic transmission and $d_{1,1} > d_{b,1}$. Under these conditions, there is no guarantee, in the general case, that a code word with exactly $d_{b,1}d_{1,2}$ bursts of b exists. It is therefore possible that for certain product codes, $d_b > d_{b,1}d_{1,2}$. To generalize this speculation consider d_{b+1} . There is no general guarantee that $d_{b+1} < d_b$, even if $d_{b+1,1} < d_{b,1}$. It is conceivable then that a cyclic product code may exist such that $d_{b+i} = d_{b,1}d_{1,2}$ for some positive integer i .

Another interesting observation results from considering a column subcode with minimum burst- b' distance $d_{b',2}$ and a row subcode with minimum Hamming distance $d_{1,1}$. For row transmission, this is not interesting, since the burst structure of the rows, not the columns, is essential. For cyclic transmission, however, the same argument can be applied to this case as was applied in the theorem. Thus

$$d_{b'} \geq d_{1,1}d_{b',2}.$$

This does not guarantee that the product code simultaneously corrects $[d'_b - 1/2]$ bursts of b' and $[d_b - 1/2]$ bursts of b . However, that it may be possible for such patterns to be simultaneously correctible indicates the potential usefulness of such product codes.

Theorem 5 relates the minimum burst- r_1 distance d_{r_1} of a product code to r_1 , the number of check symbols of its burst- r_1 -detecting row subcode and to the minimum Hamming distance $d_{1,2}$ of its column subcode.

Theorem 5: $d_{r_1} \geq 2d_{1,2}$.

Proof: We must show that every non-zero word in the product code has at least $2d_{1,2}$ bursts of r_1 . Since the row subcode is burst- r_1 detecting, $d_{r_1,1} \geq 2$. Then by theorem 4, letting $b = r_1$, $d_{r_1} \geq d_{r_1,1}d_{1,2} \geq 2d_{1,2}$.

Exactly the same conclusions can be reached regarding theorem 5 as those following theorem 4. That is, for a cyclic product code it is conceivable that $d_{r_1+i} = d_{r_1,1}d_{1,2}$ for some positive integer i . It is also true that for a cyclic product code with minimum Hamming distance $d_{1,1}$ row subcode and burst- r_2 detecting column subcode, $d_{r_2} \geq 2d_{1,1}$. Again, there is no guarantee that $[d_{r_2} - 1/2]$ bursts of r_2 and $[d_{r_1} - 1/2]$ bursts of r_1 are simultaneously correctible. The importance of theorem 5 results from the ability of a cyclic code with r_i check symbols to detect all bursts of r_i .

Elspas used the burst- r_i -detecting properties of subcodes to correct a spot error.⁷ A spot of r_1 symbols wide by r_2 symbols high occurring in a row-transmitted product code is correctible if the row subcode is cyclic with r_1 check symbols and the column subcode is cyclic with r_2 check symbols. Theorem 6 generalizes this result to the correction of multiple spots. Only row transmission is considered.

Theorem 6: A product code with linear subcodes having burst distances $d_{b_1,1}$ and $d_{b_2,2}$ corrects all patterns of $(d_{b_1,1} - 1)(d_{b_2,2} - 1)$ spots each of dimension $b_1 \times b_2$ or less if the spots fall in an array such that no more than

$$\begin{Bmatrix} d_{b_1,1} - 1 \\ d_{b_2,2} - 1 \end{Bmatrix} \text{ sets of } \begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix}$$

or fewer columns have errors. (See Fig. 3, for example.)

Proof: Detect all column errors and erase the detected errors. All errors are detected since by assumption, no more than $d_{b_2,2} - 1$ bursts of b_2 or less occur in any column. The rows can now be corrected by

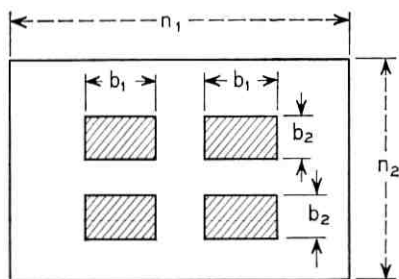


Fig. 3—Example of a multiple spot correctible by a single-burst- b_1 -correcting row code and a single-burst- b_2 -correcting column code. The shaded areas are spot errors.

filling in erasures. All erasure patterns are correctible, since there are no more than $d_{b_1,1} - 1$ bursts of b_1 or less in any row. Theorem 6 reduces to Elspas' result by taking $b_1 = r_1$ and $b_2 = r_2$, since $d_{r_1} = d_{r_2} = 2$.

Theorem 7 states the SBC capability b_1 of a product code with row subcode having block length n_1 and single-burst-erasure-correcting (SBXC) capability r_1 and column subcode which corrects bursts of $b_{1,2}$ and detects bursts of $b_{1,2} + 1$.

Theorem 7: $b_1 = n_1 b_{1,2} + r_1$.

Proof: Correcting the burst of $b_{1,2}$ and detecting bursts of $b_{1,2} + 1$ using the column code leaves no more than r_1 consecutive columns with detected errors. Erasing the r_1 or fewer columns allows correction by SBXC rows.

Theorem 7 applies to row-transmitted product codes as long as the subcodes have the stated properties. For cyclic transmission, however, the additional requirement that $n_1 \equiv 1 \pmod{n_2}$ must be made in order that no more than $b_{1,2}$ or $b_{1,2} + 1$ adjacent rows have errors.

Theorem 8 gives the SBC capability b_1 of a product code with a minimum Hamming distance $d_{1,2}$ column subcode and a $b_{1,1} -$ SBC, $r_1 -$ SBXC row subcode.

Theorem 8:

$$b_1 = n_1 \left(\frac{d_{1,2} - 1}{2} \right) + b_{1,1}, \quad d_{1,2} \text{ odd};$$

$$b_1 = n_1 \left(\frac{d_{1,2} - 2}{2} \right) + r_1, \quad d_{1,2} \text{ even}.$$

Proof: Part 1 (Burton and Weldon¹⁴): $d_{1,2}$ odd. In this case, the column

code is $(d_{1,2} - 1/2)$ —error-correcting. If the burst has length no greater than $n_1(d_{1,2} - 1/2) + b_{1,1}$ then at most $b_{1,1}$ consecutive columns will contain errors after column-correction. Since the row subcode is $b_{1,1} - \text{SBC}$, errors are correctible.

Part 2: $d_{1,2}$ even. The same argument applies as for part 1, except that $d_{1,2}/2$ errors are detectable. Therefore, if the burst is no longer than $n_1(d_{1,2} - 2/2) + r_1$, then no more than r_1 consecutive columns have detected errors. Treating the r_1 or fewer detected errors in each row as an erasure burst provides correction.

Theorem 8 applies for either row or cyclic transmission.

3.4 A Class of MBC Product Codes

Theorems 4 through 8 indicate the potential suitability of product codes for multiple-burst correction. We present a class of codes in this section which illustrates the use of theorems 4, 5 and 8, showing that many error patterns are simultaneously correctible.

We consider a cyclic product code, so cyclic transmission is assumed. The row subcode is $b_{1,1} - \text{SBC}$ and has the constraint: $n_1 \geq 3r_1 - 2$. The column code has $d_{1,2} = 4$, and $n_2 < n_1$. Of course, n_1 and n_2 must be relatively prime. As a final restriction, $r_1 \geq r_2$. We let b_m denote the length of burst such that all patterns of m bursts of b_m are correctible.

Applying theorem 8 gives $b_1 \geq n_1(d_{1,2} - 2)/2 + r_1 = n_1 + r_1$. A second application of theorem 8, reversing the roles of the subcodes gives $b_1 \geq n_2(d_{1,1} - 1)/2 + b_{1,2}$. Since the row code is SBC, its minimum distance $d_{1,1}$ is at least 3 and $b_{1,2}$ is at least 1, giving $b_1 \geq n_2 + 1$. Taking the maximum of these two lower bounds yields $b_1 = n_1 + r_1$. Applying theorem 5, we have

$$d_{r_1} \geq 2 d_{1,2} = 8, \quad \text{hence} \quad b_3 \geq r_1.$$

A second application of theorem 5 gives us $d_{r_2} \geq 2 d_{1,1} = 6$, from which $b_2 \geq 2$. Since $r_1 \geq r_2$, the latter bound is ignored and we take $b_3 = r_1$. Finally, theorem 4 is used to get $d_6 \geq d_{6,1} d_{1,2} = 12$, thus $b_6 \geq b_{1,1}$. To simplify the decoding procedure, we let $b_4 = b_{1,1}$ and $b_5 = 1$, which is using the code in a somewhat suboptimum manner.

It has not yet been shown that all the error patterns defined above are simultaneously correctible. One way to show that two error patterns are simultaneously correctible is to show that the sum of the two error patterns cannot be a code word. Another way, the one which we use here, is to demonstrate a decoding algorithm which corrects any allowable error pattern.

The decoding procedure is to single-error-correct (SEC), double-

error-detect (DED) columns. The pattern of columns with errors is then examined to determine what type of row decoding to employ. In many cases, a more powerful type of row decoding than is necessary will be used, but in no case is an allowable error pattern miscorrected.

The flowchart for decoding is in Figure 4. Let l denote the span of columns with errors, either detected or corrected, measured cyclically. Let p denote the span of columns with detected but not corrected errors, also measured cyclically. Let the number of columns with detected or corrected errors be k . Summarizing the allowable error patterns:

$$\begin{aligned} b_1 &= n_1 + r_1; & b_4 &= b_{1,1}; \\ b_3 &= r_1; & b_5 &= 1. \end{aligned}$$

Since $d_{1,2} = 4$, the column decoding corrects all but r_1 or fewer consecutive digits in any single burst, and detects the rest. All errors in triple bursts are detected, however, a triple error in a column may be miscorrected as a single error or interpreted as a double error. A quadruple burst may cause undetected, miscorrected, or misinterpreted errors, as may a pattern of five single errors.

The flowchart will be explained briefly, then the various error patterns will be tested. Three distinct types of row-decoding will be used:

- single-burst-correcting (SBC),
- single-burst-erasure-correcting (SBXC),
- double-burst-erasure-correcting (DBXC).

If $l \leq b_{1,1}$, then SBC is used. If $b_{1,1} < l \leq r_1$, then all l columns are erased and SBXC is applied to rows. If $r_1 < l$ and $r_1 < p$, then the doubles (detected errors) are erased, and the rows DBX-corrected. If $p = 0$ and $k = 3$, SBC is used. The next test is to determine whether the columns of errors form a burst of l_1 , $b_{1,1} < l_1 \leq 2b_{1,1} - 1$, followed by $l - (3b_{1,1} - 1)$ or more zeros followed by a burst of $l_2 \leq b_{1,1}$ without detected errors. If so, the first r_1 digits are erased and SBXC is used on rows. Next, a burst of $l_1 \leq b_{1,1}$ followed by $l - (3b_{1,1} - 1)$ zeros followed by a burst of l_2 , $b_{1,1} < l_2 \leq 2b_{1,1} - 1$ is tested for. Erasing the last r_1 columns allows SBXC.

If $l_1 \leq b_{1,1}$ and $l_2 \leq b_{1,1}$, then l_1 and l_2 are erased whence DBXC is used. As a final test, if $l < 2r_1$, the $2r_1 - l$ center columns are erased as are all double errors and SBXC is used. If $l \geq 2r_1$, all doubles are erased and SBXC is used.

The decoding procedure will now be tested for the various error patterns.

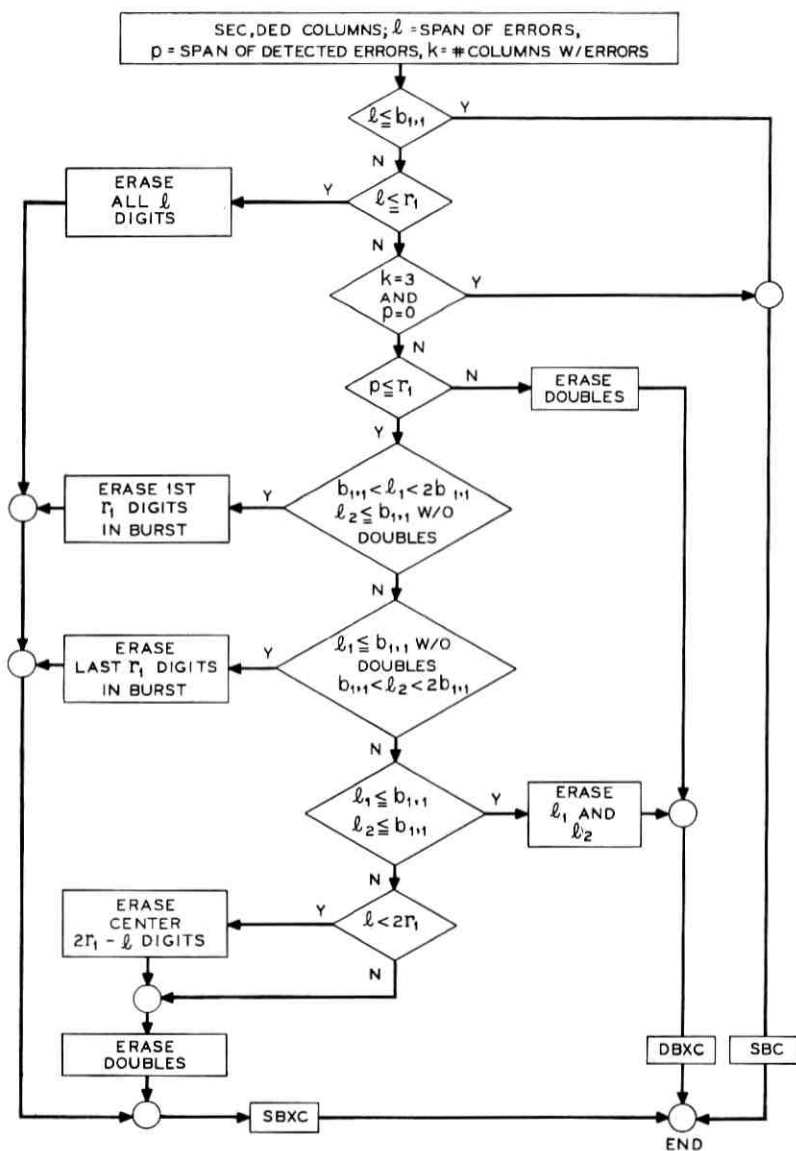


Fig. 4—A decoding procedure.

3.4.1 Five Single Errors

The seven types of error patterns are indicated in Fig. 5.

For case (i), $l \leq 1$, so SBC is successful. For case (ii), if the four errors are undetected, then $l = 1$, so SBC is used. If the four errors are detected or miscorrected, then there result two bursts of $\leq b_{1,1}$. This is also true in case (iii), so that erasing both columns allows DBXC rows. In case (iv), $K = 3$; so if $p = 0$, SBC is employed successfully. If $p = 1$, the triple error is treated as an erasure. In case (v), $K = 3$, but $p = 2$, so SBC is not used. How this pattern is corrected depends on the separation of the columns with the double errors. If $p > r_1$, the 2 columns are erased, then DBX-corrected. If $p \leq r_1$, then SBXC is used. This is possible over any of the five remaining paths. For case (vi), the double will be erased and SBXC used if $l > r_1$. This too can occur over one of the last five paths. Since erasing doubles is a part of each of those paths, and since no more than a burst of r_1 is induced in each row, decoding is successful. Case (vi) is decoded successfully by columns. The row decoding will depend on the relative placement of the five singles, but in no case will decoding fail.

3.4.2 Quadruple bursts of $b_{1,1}$

The five patterns of quadruple burst are shown in Fig. 6.

- (i) For $l \leq b_{1,1}$ SBC is successful.
- (ii) For $b_{1,1} < l \leq r_1$, we use SBXC.
- (iii) For $p > r_1$, simply erasing the doubles allows DBXC. For $p \leq r_1$ there are several possibilities. If either l_1 or l_2 is $\leq b_{1,1}$ and has no doubles, and the other exceeds $b_{1,1}$, then the last or first r_1 digits of the burst

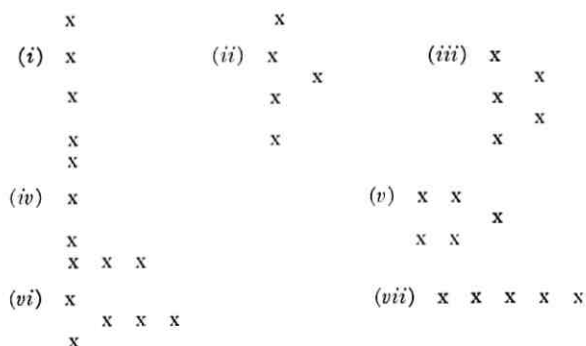


Fig. 5—The seven patterns of five errors.

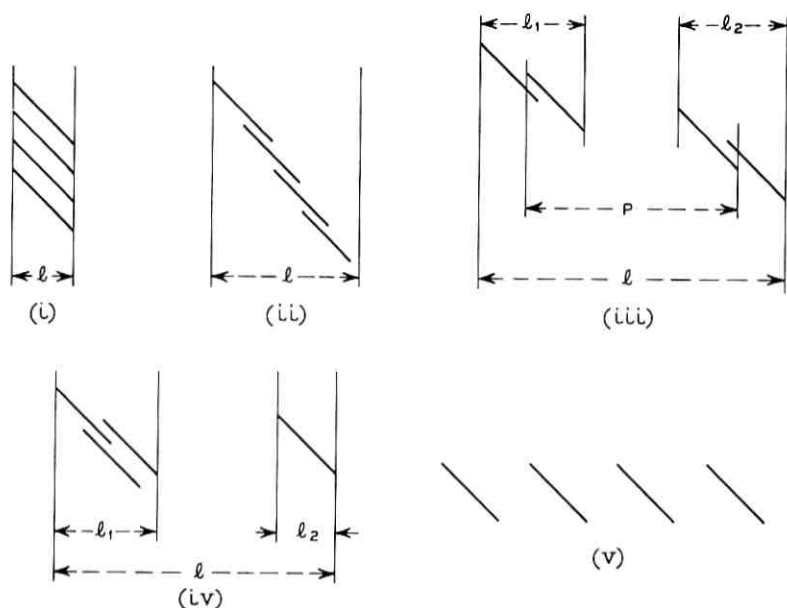


Fig. 6—The five patterns of quadruple bursts.

are erased and SBXC is used. If both l_1 and l_2 are $b_{1,1}$ or less, then both are erased and DBXC is used.

(iv) In this case, care must be taken since the overlap of the three bursts may cause miscorrected errors. Since $l_2 \leq b_{1,1}$ and has no doubles, if l_1 is such that $b_{1,1} < l_1 < 2b_{1,1}$ then the first r_1 digits of l are erased. If the roles of l_1 and l_2 are reversed, the last r_1 digits of l are erased. In either case, any miscorrected columns are erased, then SBXC is used. If both l_1 and l_2 are $b_{1,1}$ or less, then both are erased and DBXC is used. If $l_1 > 2b_{1,1} - 1$, then no miscorrection occurs, and one of the last two paths is followed, hence SBXC rows.

(v) All are corrected by column decoding, and again, no more columns are erased than the row subcode can decode.

3.4.3 Triple Burst of r_1

See Fig. 7 for the five types of triple bursts.

- (i) This is handled by SBC or SBXC, depending on l .
- (ii) Care must be taken to erase any possible miscorrected columns. If the pattern falls into the category of a quadruple burst, case (iv),

then the first r_1 or last r_1 digits are erased and SBXC is used. If not, and if $r_1 < l < 2r_1$, the center $2r_1 - l$ digits and all double errors are erased. It is easy to see that erasing $2r_1 - l$ center digits erases any miscorrected columns, and that the total span of erased columns does not exceed r_1 . Then SBXC can be used.

- (iii) It is for this case that the block length restriction $n_1 \geq 3r_1 - 2$ is necessary, for without that constraint, there could result three bursts of double errors. The doubles here are erased, then SBXC is used.
- (iv) This is decoded as a previous case as is case (v).

3.4.4 Single Burst of $n_1 + r_1$

A single burst will leave no more than r_1 consecutive columns with double errors after column decoding. The decoding used varies in accord with the pattern of column errors, but it is easy to see that all single bursts are corrected by the row decoding.

3.5 Examples and Comparison with Other Codes

Table I lists some sample codes from this class. Their performance is discussed below.

It is difficult to compare the MBC product codes of Section 3.4 as an

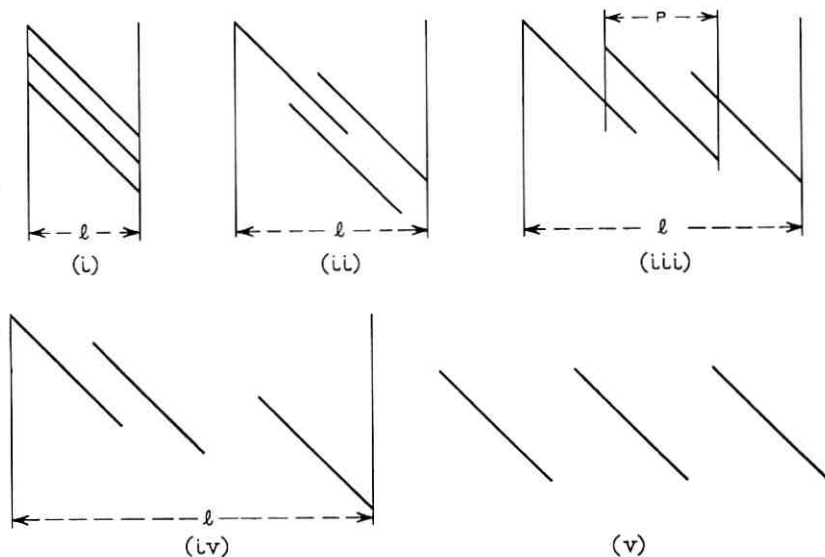


Fig. 7—The five patterns of triple bursts.

TABLE I—MBC PRODUCT CODES

n_1	k_1	$b_{1,1}$	n_2	k_2	$d_{1,2}$	n	k	b_1	b_3	b_4	b_5	ROW CODE	COL. CODE
35	25	4	31	25	4	1085	625	45	10	4	1	$\frac{(x^5 + 1)(x^7 + 1)}{x + 1}$	BCH*
39	27	5	7	3	4	273	81	51	12	5	1	[7] p. 12 13617†	BCH
51	35	7	7	3	4	357	105	67	16	7	1	[7] p. 13 304251	BCH
30	20	4	7	3	4	210	60	40	10	4	1	[17]	BCH
45	35	4	7	3	4	315	105	55	10	4	1	[17] p. 25	BCH
			31	25	4	1395	875	55	10	4	1	[17] p. 25	BCH
31	21	4	7	3	4	217	63	41	10	4	1	[17] p. 28	BCH
31	21	4	15	10	4	465	210	41	10	4	1	[17] p. 28	BCH
31	20	5	15	10	4	465	200	42	11	4	1	[17] p. 32	BCH
			21	15	4	651	300	42	11	4	1	[17] p. 32	[5] p. 11 123
31	25	2	21	15	4	651	375	37	6	2	1	BCH	[5] p. 11 123
28	19	3	15	10	4	420	190	37	9	3	1	[17] p. 23	BCH

* Bose-Chaudhuri-Hocquenghem.

† Generator polynomial in standard octal notation.

entire class with other MBC codes. Therefore, we select two fairly representative codes from Table I and compare them with roughly similar codes from three other classes. Table II provides the comparison. Under the heading "code", the 2.0 means code from Table I, "B and C" = Bahl and Chien, "I" means interleaved, and "R-S" stands for Reed-Solomon.

TABLE II—A COMPARISON OF FOUR TYPES OF MBC CODES

CODE	(n, k)	k/n	b_1	b_2	b_3	b_4	b_5	b_6
2.0	(651, 375)	.58	37		6	2	1	
B&C	(660, 420)	.64	44	8				
I	(630, 390)	.62	40	20		10		
R-S	(511, 403)	.79	46	19	10			1
R-S	(511, 421)	.82	37	10			1	
2.0	(465, 210)	.45	41		10	4	1	
B&C	(455, 288)	.61	35	9				
I	(441, 273)	.62	28	14		7		
R-S	(511, 421)	.82	37	10			1	

The two Bahl and Chien codes were selected from Table 1 of Ref. 9. The (630, 390) interleaved code is a (63, 39) BCH*² 4-error-correcting code interleaved ten times. The (441, 273) code is the same (63, 39) code interleaved seven times. The Reed-Solomon codes are over GF (2⁹). The (511, 403) code is 6-error-correcting, while the (511, 421) code is 5-error-correcting.

In the first grouping of codes, the product code is about the same as the R-S (511, 421) code in error-correcting ability, however, has a lower rate. It has slightly lower rate than the B and C code, but performs somewhat better. The interleaved code and the R-S (511, 403) code perform somewhat better than the other codes in the first grouping. The product code and the B and C code (also a product code) are the easiest to decode in the table. While the interleaved and R-S codes are somewhat superior in performance to the product codes, they have decoding disadvantages. The decoding operations in the R-S code must be made over GF (2⁹), and furthermore, a 6-error-correcting decoder is not trivial. The interleaved code must decode ten words of length 63 each containing up to four errors, also not trivial.

In the second grouping, the product code is at least as good in performance as any of the others except for its lower rate.

* Bose-Chaudhuri-Hocquenghem.

IV. CONCLUSIONS

The class of codes presented in Section 3.5 indicates the potential usefulness of product codes for multiple-burst correction. The parameters of the codes were chosen to make the decoding easy. For example, r_1 was assumed to be less than r_2 , thus avoiding the problem of correcting a double burst. The double burst case necessitates an iterative use of the decoding algorithm. It can easily be shown (although none of the theorems indicates this) that $b_2 \geq n_2/2$ if the minimum distances $d_{1,1}$ and $d_{1,2}$ are at least four (see Fig. 8). The double burst shown can be corrected by SEC, DED columns, then rows, then columns again, and so forth, until no errors exist. Such an iterative procedure is too slow to be very useful except perhaps as a proof of double-burst-correcting ability.

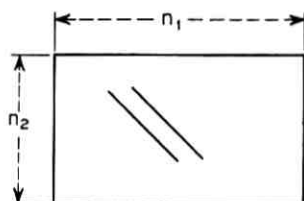


Fig. 8—Double burst of $n_2/2$.

Other parameter selections were made to make the decoding scheme work. One such parameter is $b_s = 1$. Another is that the SBC ability $b_{1,2}$ was not specified; the column code was assumed to be SEC, DED. Still one more point is that the burst parameters b_1 , b_3 and b_s were just lower bounds. It is likely that certain cyclic product codes from this class have greater burst-correcting abilities than these lower bounds. Moreover, it is possible that a simpler decoding procedure exists. Considering all these factors, the usefulness of product codes for multiple-burst correction seems clear.

As a final point, the problem of correction of up to four bursts or five single errors has been reduced to single-burst correction or to single or double-burst-erasure correction. This does not say that double-burst-erasure correction is always easy. Bahl, Chien and Tang derived a DBXC procedure which is simple for some codes and not so simple for others.¹⁰ Gilbert codes are codes for which DBXC procedure is simple.

REFERENCES

1. Mandebrot, B., "Electromagnetic Turbulence in Communication Systems," Int'l. Conf. on Microwaves, Circuit Theory and Information Theory, Tokyo, Japan, September 1964.
2. Peterson, W. W., *Error Correcting Codes*, Cambridge, Mass.: M.I.T. Press, 1961, pp. 168-182.
3. MacWilliams, J., unpublished work.
4. Chien, R. T., and Tang, D. T., "On Definitions of a Burst," *IBM J. Research Dev.*, 9, No. 4 (July 1965), pp. 292-293.
5. Berlekamp, E. R., *Algebraic Coding Theory*, New York: McGraw-Hill, 1968, pp. 338-346.
6. Wolf, J. K., "On Codes Derivable from the Tensor Product of Check Matrices," *IEEE Trans. on Information Theory*, *IT-11*, No. 2 (April 1965), pp. 281-284.
7. Elspas, B., "Design and Instrumentation of Error-Correcting Codes," Stanford Research Institute, Menlo Park, Calif., 1961.
8. Kasahara, M., and Kasahara, Y., "Pseudo-Cyclic Codes for Double Burst Error Correction," *Electronics and Communications in Japan*, 50, No. 1 (July 1967), pp. 9-18.
9. Bahl, L. R., "Correction of Single and Multiple Bursts of Errors," Ph.D. Thesis, University of Illinois, 1969.
10. Elias, P., "Error-Free Coding," *IRE Trans.*, 1954 Symposium on Information Theory, *IT-4* (1954), pp. 29-37.
11. Slepian, D., "Some Further Theory of Group Codes," *B.S.T.J.*, 39, No. 5 (September 1960), pp. 1219-1252.
12. Reddy, S. M., unpublished work.
13. Massey, J. L., *Threshold Decoding*, Cambridge, Mass.: M.I.T. Press, 1963, p. 6.
14. Burton, H. O., and Weldon, E. J., "Cyclic Product Codes," *IEEE Trans. on Information Theory*, *IT-11*, No. 3 (July 1965), pp. 433-439.
15. Abramson, N. M., "Encoding and Decoding Cyclic Code Groups," unpublished work.
16. Bahl, L. R., Chien, R. T. and Tang, D. T., "Correction of Two Erasure Bursts," *IEEE Trans. on Information Theory*, 15, No. 3 (May 1969), p. 431.
17. Schmandt, F. D., "Single Burst-Error-Correction-capabilities of Binary Cyclic Codes," RADC-TDR-63-301, Rome Air Development Center, Griffis Air Force Base, N. Y., August 1963.

Contributors to This Issue

R. C. ALLEN, A.B., Geology, 1957; M.A., 1958, Boston University; Western Electric Co., 1959-1960; Bell Telephone Laboratories, 1960—. Mr. Allen worked briefly in quartz crystal engineering at the Western Electric Company. Since joining Bell Telephone Laboratories he has been engaged in applied marine geology and oceanography. He is presently concerned with route engineering problems related to millimeter waveguide installations.

CLEO D. ANDERSON, B.S.E.E., 1960, University of Idaho; M.E.E., 1962, New York University; Bell Telephone Laboratories, 1960—. Mr. Anderson has been mainly concerned with systems analysis of submarine cable systems. He is now supervisor of the High Frequency Radio Group. Member, IEEE, Sigma Tau, Phi Kappa Phi, Eta Kappa Nu.

J. E. BERRANG, B.S.E.E., 1965, Valparaiso Technical Institute; Bell Telephone Laboratories, 1952—. Mr. Berrang is a member of the Opto-Electronics Research Department where he is currently participating in experimental studies of camera systems for color *Picturephone*[®] visual telephone service. Member, IEEE.

S. THEODORE BREWER, B.S.E.E., 1937, M.S., 1938, Purdue University; Bell Telephone Laboratories, 1937—. In his early assignments, Mr. Brewer contributed to the development of broadband coaxial systems and video feedback amplifiers, including the design of measuring equipment associated with these developments. Later, he designed circuits of electronically controlled automatic switching systems. More recently, he led a group responsible for the electrical design of undersea repeaters of the SD and SF Systems. He currently heads a department devoted to system and terminal design used in undersea communications and development of improved overseas communications by high frequency radio. He holds patents on control and feedback systems, switching networks, and repeater circuits. Member of I.E.E.E., Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

JOHN D. BRIDWELL, B.S.E.E., 1965, Kansas State University; M.S.E.E., Massachusetts Institute of Technology, 1966; Ph.D., Poly-

technic Institute of Brooklyn, 1969; Bell Telephone Laboratories, 1965—. Mr. Bridwell has worked in digital circuit design for the El Telemetry System. He is currently involved in the design of a data network for telemetry systems. Member, Eta Kappa Nu, Phi Kappa Phi, IEEE.

ROBERT G. BUUS, B.S.E.E., 1959, University of North Dakota; M.E.E., 1961, New York University; Bell Telephone Laboratories, 1959—. Mr. Buus was initially engaged in submarine cable measurements and repeater design. He has taught a course at Bell Laboratories on Transmission System Design and contributed to the textbook, *Transmission Systems for Communications*. Mr. Buus currently supervises a group responsible for the system analysis and application of TASI on overseas circuits.

EDWIN T. CALKIN, B.S. (Eng.), 1961, The Cooper Union; M.S.E.E., 1963, New York University; Bell Telephone Laboratories, 1961—. Mr. Calkin worked on data set and data processing power supply circuit designs until 1964. Since 1964, he has been involved in the circuit development of submarine cable power supplies.

ALAN T. CHAPMAN, B.S. (Chemistry), 1929, Washington State College; M.S. and Ph.D., 1932, Ohio State University; National Research Council Fellow, California Institute of Technology, 1932-1934. E. I. DuPont Company, 1934-35; Western Electric Company, Engineering, 1935—. Mr. Chapman initially worked in the Electron Tube Shop in New York City and subsequently worked on mica capacitors and crystal units in Kearny, New Jersey. He became department chief in Materials, SQC, Automation in 1947, and assistant manager, Switchboards, Key Equipment and Cable in 1955. Since 1957, he has worked on submarine cable repeaters, except for the period 1963-1964 when he worked on exchange cable engineering. Member, American Association for the Advancement of Science.

ROBERT L. EASTON, B.S. (M.E.), 1953, M.S. (M.E.), 1954, California Institute of Technology; Bell Telephone Laboratories, 1954—. Mr. Easton conducted analyses which led to the SD Submarine Cable System. In more recent studies he has contributed to the design and equalization of the SF system. He currently supervises a group responsible for economic, performance and circuit studies of systems going beyond SF in capacity. Member, Tau Beta Pi.

IGOR GOLIOTO, M. E., 1961, Stevens Institute of Technology; M.S.M.E., 1963, New York University; Bell Telephone Laboratories, 1961—. Mr. Golioto has been a member of the Power Systems Physical Design Department since joining the Laboratories. He has been responsible for equipment design of the carrier, microwave, coaxial, submarine, ESS and general use power plants. Presently, he is involved in the design of a military submarine cable power plant.

HERMANN K. GUMMEL, Diplom-Physiker degree (1952), University of Marburg, Germany; M.S. (physics), 1952, Ph.D. (physics), 1957, Syracuse University; Bell Telephone Laboratories, 1957—. He has worked in semiconductor electronics and presently heads a department responsible for design analysis. Member, American Physical Society, Sigma Xi.

W. B. HIRT, Associate Degree in electrical technology, 1962, Westchester Community College; Bell Telephone Laboratories, 1962—. Mr. Hirt has aided in the development of the SF amplifier and fault location oscillator. His most recent work is concerned with the computation coincident with the installation and equalization of the SF Carrier System.

J. J. KASSIG, B.S.E.E., Massachusetts Institute of Technology, 1952; M.S.E.E., Rutgers University, 1955; Bell Telephone Laboratories, 1955—. Mr. Kassig is working on a repeater design for higher capacity submarine cables.

ANDREW W. LEBERT, B.S.E.E., 1932, New York University; Cornell Dublier Corporation, 1932-36; Bell Telephone Laboratories, 1936—. For his first five years at Bell Laboratories, Mr. Lebert worked on transmission engineering on open wire and cable carrier systems. He then was concerned with fault location problems. During World War II he turned to military communications on cable and open wire after which he spent eight years on coaxial cable systems development. Since 1952 he has been connected with the design and development of land coaxial cables, the SB, SD and SF Ocean Cable Systems, and the development of military ocean cable and facilities for antisubmarine warfare systems. He was made a supervisor in 1954 and a department head in 1961. Member, IEEE, Tau Beta Pi, Psi Upsilon.

NATHAN G. LESH, B.S., 1943, Lehigh University; M.S., 1960, Stevens Institute of Technology; Bell Telephone Laboratories, 1956—. Mr.

Lesh has worked on passive components for the Bell System and military applications. In 1960, he was appointed supervisor of the Resistor Development Group. He presently supervises a group in the Film Circuit and Component Development Department. Member, Eta Kappa Nu, Tau Beta Pi.

ROBERT L. LYNCH, A.S. 1950, Kansas City (Missouri) Junior College; B.S.E.E., 1957, Kansas University; M.E.E., 1959, New York University; Bell Telephone Laboratories, 1957—. Mr. Lynch has worked on the development of cable machinery on the Bell System Cable Ship Long Lines and the design of transmission equipment for the shore terminals of SD Submarine Cable Systems. He assisted in the cable laying and installation of several SD systems. Mr. Lynch was involved in the design of the transmission equipment for the shore terminals of the SF Submarine Cable System and later supervised a group responsible for the physical design of SF Submarine Cable Terminals and the TASI B System. Since 1968, he has supervised a group responsible for the physical and electrical design of the SF System shore terminals. He also assisted in the installation, including the cable laying, of the first SF Submarine Cable System between Florida and St. Thomas, Virgin Islands.

JAMES MCKENNA, B.Sc. (Math), 1951, Massachusetts Institute of Technology; Ph.D. (Math), 1961, Princeton University; Bell Telephone Laboratories, 1960—. Mr. McKenna has done research in quantum mechanics, electromagnetic theory and statistical mechanics. He has recently been engaged in the study of nonlinear partial differential equations which arise in solid state device work, and in the theory of stochastic differential equations.

WILLIAM McMAHON, B.S., 1942, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1926—. During his early career, Mr. McMahon engaged in research on the preservation of wood and other organic materials. He later conducted studies of rubber compounding and did research on electric insulating materials. He is presently in charge of a group concerned with the development of insulation materials for electrical capacitors. He holds eight patents, including those on wood preservatives, rubber compounding techniques, special types of electrical transmission lines, insulating materials and capacitor structures. Member, American Chemical Society.

D. O. OLDFATHER, B.S.E.E., 1960, Oregon State University; M.S.E.E., 1964, University of Maryland; Bettis Reactor Engineering School, 1961; Naval Reactors Division of the Atomic Energy Commission, 1960 to 1964; Bell Telephone Laboratories, 1964—. Originally, he designed networks used in the SF System undersea equalizers and then prepared detailed test procedures for the installation, equalization, and line-up of the SF System. He is presently a member of the Terminal and Systems Design Department, engaged in circuit development for the SF Submarine Cable System terminal equipment. Member, Phi Kappa Phi, Tau Beta Pi, Sigma Tau, Eta Kappa Nu, Pi Mu Epsilon.

DONALD H. ORT, B.S.M.E., 1957, M.S.M.E., 1960, Rutgers University; Bell Telephone Laboratories, 1960—. Mr. Ort completed the communications development training program in 1962. He was first engaged in conducting route studies and economic analyses as part of the circular waveguide development project. Later he worked on route studies, economic analyses, and outside plant cost estimates for a variety of PCM and analog transmission systems. He joined the Ocean Cable Protection Group in 1965, conducting the initial feasibility studies for the burying project and serving as project coordinator for the burying of the Jacksonville-St. Thomas SF System. Mr. Ort currently is an Assistant Engineering Manager in the Outside Plant Section of the American Telephone and Telegraph Company in New York. Member, Pi Tau Sigma, Sigma Xi.

H. C. POON, B.S.E.E. and M.S.E.E., 1962, Massachusetts Institute of Technology; Ph.D., 1967, Harvard University; Bell Telephone Laboratories, 1966—. Mr. Poon works in the field of semiconductor physics and device analysis. Member, Eta Kappa Nu, Tau Beta Pi, Sigma Xi, American Physical Society.

G. J. SCHAIBLE, B.S.E.E., 1941, New York University; Bell Telephone Laboratories, 1928—. Since joining Bell Laboratories, he has engaged in development of exchange area, toll, and land coaxial cables. During World War II he was involved in measuring transmission characteristics of flexible coaxial leads for radar equipment. He was Resident Cable Engineer at a domestic and foreign manufacturer's plant during production of the TAT-1 and TAT-2 ocean cables for the SB System, and later supervised groups responsible for the design of SD and SF Ocean Telephone Cables, and for military underwater facilities. Presently supervisor of the Molding and Splicing Group in the Ocean Cable

Department at the Baltimore Laboratory, he is responsible for the development of splicing techniques for ocean telephone cables and military underwater facilities.

DONALD L. SCHARFETTER, B.S., 1960, M.S., 1961, and Ph.D., 1962, Carnegie Institute of Technology; Bell Telephone Laboratories, 1962—. Mr. Scharfetter's fields of interest have included metal-semiconductor contact theory, p-n junction diode and transistor theory, avalanche diode oscillator analysis, and computer-aided design. Member, IEEE, Tau Beta Pi, Eta Kappa Nu, Pi Mu Epsilon, Sigma Xi.

L. M. SCHINDEL, B.S.E.E. 1928,, Iowa State University; American Telephone and Telegraph Company, Long Lines Department, 1928-1955; Western Electric Co., 1955-1961; American Telephone and Telegraph Company, Long Lines Department, 1961—. As cable engineer for ocean cables, Mr. Schindel is responsible for making ocean-bottom surveys for transoceanic projects and the preparation of plans and specifications for the on-land, shore end and deep sea portions of such projects. Member, National Society of Professional Engineers; senior member, Institute of Electrical and Electronics Engineers.

JOHN L. THOMAS, B.S.E.E., 1957, University of Maine; M.E.E., 1960, New York University; Bell Telephone Laboratories, 1957—. Mr. Thomas engaged initially in circuit design work associated with special applications of submarine cable systems. He worked on systems analysis and supervised a group responsible for the circuit design of shore terminal transmission facilities associated with the SF System. He later supervised a group responsible for repeater, equalizer and special test set circuit design for submarine cable systems. He is presently responsible for the design of transmission surveillance and fault location circuitry for the L5 Coaxial System. Member, Phi Kappa Phi, Tau Beta Pi.

WILLIAM J. THOMPSON, B.S., 1929, University of California, Berkeley; Bell Telephone Laboratories, June 1929—. Mr. Thompson's work was concerned with the development of power transformers and voltage regulators and, later, transmission transformers and inductors for various carrier and radio systems. During World War II, he worked on sonar systems. Since 1956, he has been concerned primarily with the development of high reliability components for Bell System submarine cable systems and, more recently, for military applications.

Since 1945, he has supervised a group responsible for transformer and inductor development. Senior Member, IEEE.

CHARLES A. VON ROESGEN, Dipl. Ing., 1952, Swiss Federal Institute of Technology; Bell Telephone Laboratories, 1953—. Mr. Von Roesgen worked on development of automatic test sets, repeaters and multiplex equipment for submarine cable systems. He is a supervisor in the Digital Multiplex Department. Member, IEEE.

A. J. WAHL, B.S., 1942, University of Kansas; Ph.D., 1950, Princeton University; United States Air Force, 1951–1953; Bell Telephone Laboratories, 1950–1951, 1953—. Since 1953, Mr. Wahl has been engaged in various phases of transistor development. His early work was in the area of surface effects relating to device behavior, particularly in regard to device reliability. Later he was concerned with providing the semiconductor devices for the Telstar satellites. He has supervised the Bell Laboratories efforts in providing the transistors and diodes for the SF Submarine Cable System. He is currently supervisor of a group concerned with the development of microwave transistors and submarine cable devices. Member, IEEE, Sigma Xi.

E. WASSERSTROM, B.Sc., 1956, M.Sc., 1960, Technion-Israel Institute of Technology; Ph.D., 1964, Brown University; Division of Sponsored Research at M.I.T., 1962–1964; Department of Aeronautical Engineering at the Technion, 1964–1968, 1970—; Bell Telephone Laboratories, 1968–1970 (on leave of absence from the Technion). He is currently engaged in numerical analysis.

JACK K. WOLF, B.S.E.E., 1956, University of Pennsylvania; M.S.E., 1957, M.A., 1958, and Ph.D., 1960, Princeton University; Bell Telephone Laboratories, 1968–1969. Mr. Wolf is a Professor of Electrical Engineering at the Polytechnic Institute of Brooklyn; for the academic year 1968–1969 he was on a leave of absence to the Communications Theory Department at Bell Laboratories, Murray Hill, New Jersey. His main interests are in information theory, algebraic coding theory, and detection theory. Member, Tau Beta Pi, Sigma Xi, Sigma Tau, Eta Kappa Nu, Pi Mu Epsilon, IEEE, American Association for the Advancement of Science, American Association of University Professors.

PAUL A. YEISLEY, JR., B.S. (Physics), 1952, Lafayette College; Bell Telephone 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 and SF Submarine Cable Systems. He has more recently been involved in the physical design and development of the SD-C and SG Submarine Cable Systems. He is a consultant in numerical control.

Erratum

"Radiation Losses of Tapered Dielectric Slab Waveguides," B.S.T.J., Vol. 49, No. 2, February 1970, pp. 273-290, by Dietrich Marcuse
Equation (44) of my paper should read

$$\exp \left[-i \int_0^z (\beta_0(s) - \beta) ds \right].$$

The factor $\exp [-i(\beta_0 - \beta)z]$ appearing in equations (45), (53) and (54) should be replaced by this correct phase factor. All other equations remain valid.

The error has the effect of letting the radiation loss of a linear taper shown in Fig. 8 appear too high. The point at $L = 0$ remains the same but the loss at $L/d_1 = 40$ is reduced from 0.036 to 0.014. All other curves remain correct.

I am grateful to Mr. A. L. Jones of the IBM Corporation for pointing this error out to me. (D.M.)

5358-17-6¹⁴⁵