THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLIII JULY 1964 NUMBER 4, PART 1

The SD Submarine Cable System

By R. D. EHRBAR, J. M. FRASER, R. A. KELLEY, L. H. MORRIS, E. T. MOTTRAM and P. W. ROUNDS

(Manuscript received April 3, 1964)

Submarine cable systems of a new design have recently been installed between Florida and Panama, between New Jersey and Cornwall, England, and between Hawaii and Japan. Using a single lightweight armorless cable for most of the route, with electron tube amplifiers encased in rigid containers at 20-mile intervals, this type of system will carry 128 channels in each direction.

More detailed discussion of cable, repeaters, and power equipment will be found in companion articles. This article outlines the system development objectives, gives an over-all system description, and describes the equalization plan and terminal equipment.

I. INTRODUCTION

In 1954, before the installation of the first transatlantic telephone cable, growth studies had shown that submarine cable systems carrying many more channels than the systems then under development would be required to carry future traffic. The development of these wider-band systems was therefore initiated before the development of the SB submarine cable system[1] was complete. Exploratory studies indicated that a new approach to cable and repeater design would be required. The SB system used armored coaxial cables, one for each direction of transmission. The repeaters were long and flexible so they could be handled on shipboard without major modifications to the cable machinery, which

had been developed to lay submarine telegraph systems. While these long, flexible repeaters have proven to be sound in concept and reliable in performance, the form factor is such as to introduce large parasitic inductance and capacitance in the interconnections between the stages of the repeater and in the feedback path from input to output. In consequence, it is not possible to design a wideband repeater using this mechanical approach. For the new broader-band systems, therefore, the repeater circuit was placed in a rigid cylindrical container.

The cable concept was changed, and the strength was placed in the center of the cable rather than in the armor wires on the outside. Such construction makes possible a cable with a very stable transmission characteristic and eliminates the tendency of the cable to untwist under tension. This latter feature makes it easier to lay the large, heavy repeaters in deep water.

The decision to use a new cable and rigid repeaters made it necessary to consider the impact of these new components on the cable laying process. This resulted in the development of new cable laying machinery and techniques, which were combined to produce a new cable laying ship — the C.S. *Long Lines.*

The SD submarine cable system is the result of this development approach. It will transmit 128 channels of 3-kc bandwidth in each direction over a single cable, and experience with initial installations in the Caribbean from Florida to Panama via Jamaica, in the North Atlantic from New Jersey to Cornwall, England, and in the Pacific from California to Japan indicates that system objectives will be met.

II. SYSTEM OBJECTIVES

The basic objective in the development of the SD system was to make available a highly reliable transoceanic facility which would be comparable in quality to the land plant of the continents to be connected, at a cost per channel-mile substantially lower than the earlier relatively narrow-band systems. Consistent with the trend to establish more stringent noise requirements on long continental circuits, noise of 41 dbrn at the zero level transmission point over the life of a 4000-statute mile system was taken as a design objective. This is about 3 db quieter than the transmission noise objective used for many years for continental circuits within the United States.

The system was also designed to carry a greater magnitude of signal per kc of transmitted bandwidth than usual in AM telephone systems. This was necessary for several reasons. In the first place, transatlantic callers have tended to speak at higher volumes than customers do on

local calls within the United States or Europe. Secondly, because trans-oceanic spectrum is more expensive than land plant, it is worthwhile to make more efficient use of it by stacking channels more closely together. This loads the system more heavily. In addition, the probable application of TASI[2] makes it necessary to engineer for a talker activity about three times as great as that normally assumed in engineering land systems. The SD system has therefore been engineered to meet the 41-dbrn noise objective while carrying a signal load about six db higher than the CCITT recommendation for an equivalent eight-group land system, as indicated by Table I.

Many other objectives governed the design and layout of the system. Paramount, of course, was the reliability and stability of all the undersea equipment employed. Convenience of maintenance of the shore terminal stations and a high degree of reliability for the equipment employed at these terminals were also considered important objectives.

III. SYSTEM LAYOUT

As shown by the block diagram of the system in Fig. 1, the length of cable between repeaters is approximately 20 nautical miles (nm). An ocean-block equalizer to correct misalignment is located in the section following every tenth repeater in the direction of the cable lay. The length of cable between ocean-block equalizers and adjacent repeaters is 6 nm. The adjustable equalizer loss is thus approximately equivalent to the loss of 8 miles of cable. The repeaters, equalizers and cable are described in companion articles.[3,4] Transmission in one direction is carried in the frequency band of 108 to 504 kc and in the other direction by the band 660 to 1052 kc. These frequency allocations for 128 channels are shown in Fig. 2. After a system is installed it is possible that an additional 10 channels will meet requirements. When these channels are used the frequency bands are expanded to 90 to 516 kc and 630 to 1052 kc. The two directions of transmission are amplified in a common am-

TABLE I—LOAD ASSUMPTIONS

	SD System	CCITT
Average talker volume (VU)	−10.8	−12.0
Standard deviation (db)	5.8	5.0
Activity	0.75	0.25
RMS power/channel (dbm)	−9.6	−15*
RMS power/group (dbm)	+2.4	−4.2
RMS power/band (dbm)	+11.5	+4.8

* Includes an allowance for power of signaling tones.

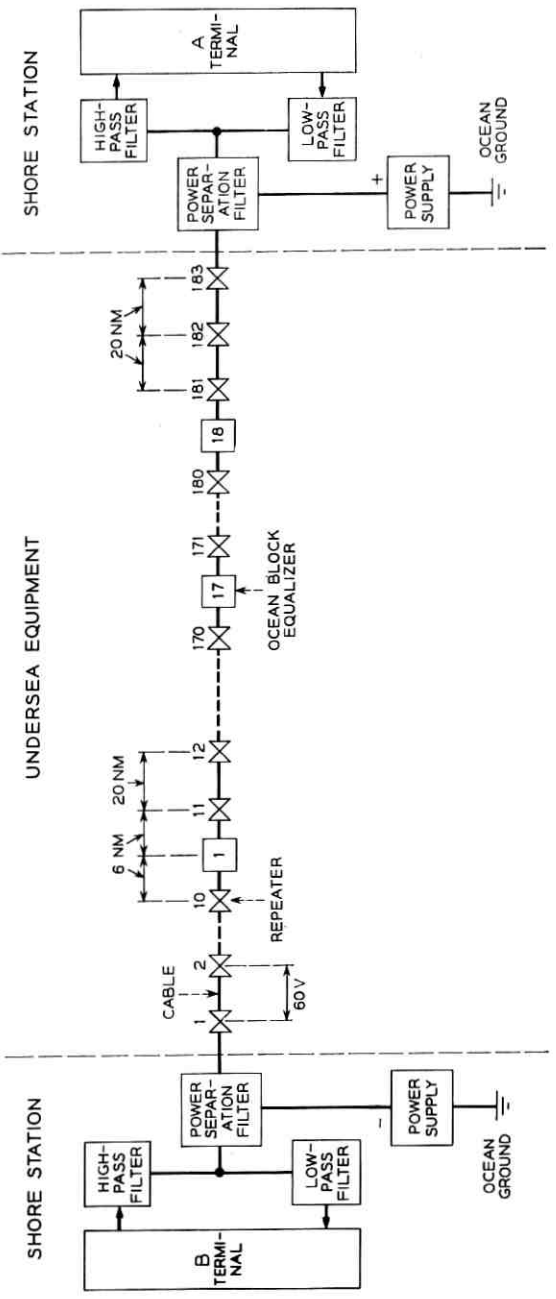


Fig. 1 — Block diagram of system.

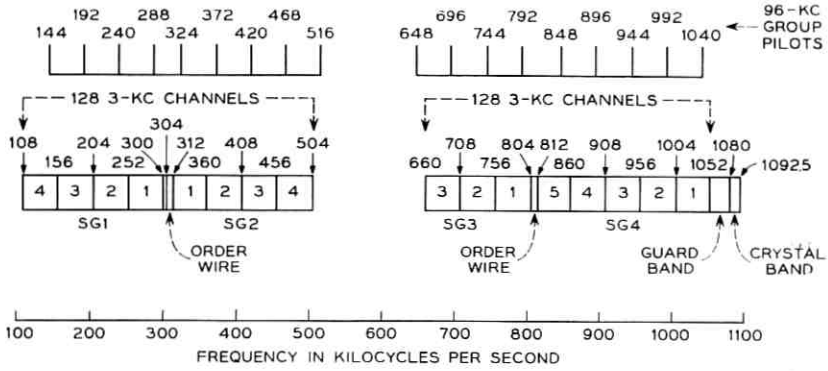


Fig. 2 — Frequency allocations for 128 3-kc message channels.

plifier by the use of directional filters. Parallel amplifiers, each containing three electron tubes, are used to provide protection against tube failure; these amplifiers share a common feedback or beta circuit, as indicated in the block diagram of Fig. 3.

The cable used for the major part of a system has an over-all diameter of $1\frac{1}{4}$ inches and a breaking strength of 18,000 pounds. Armored cable designs are available for use in shallow water to give protection against

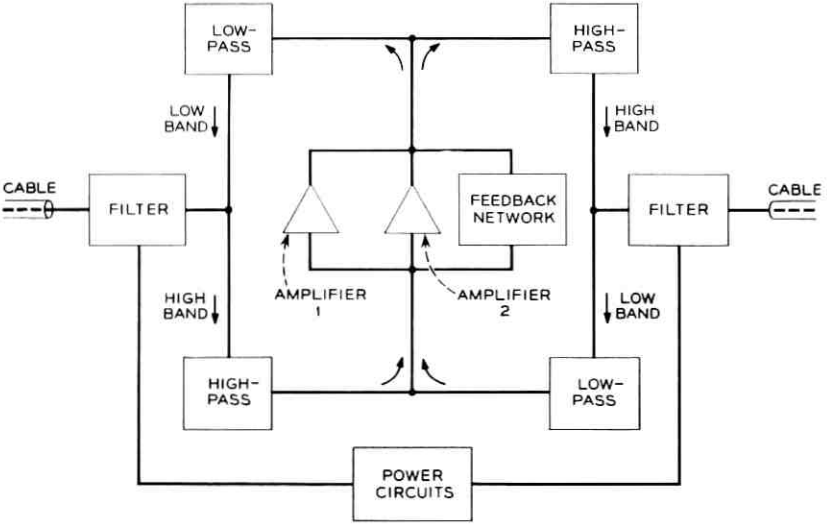


Fig. 3 — Repeater block schematic.

anchor or fishing damage. A shield may also be added to the cable structure at locations where radio or similar electromagnetic interference might be expected. The length of the shore-end section is restricted to the range 5 to 15 nm to minimize noise disturbances.

Power for operating the repeaters is supplied from the shore over the central conductor of the cable. The power system is described in more detail in a companion article.[5] A positive dc voltage is supplied to the cable at the A terminal between the central conductor and ground and a negative potential is supplied at the B terminal. The current path is thus over the central conductor, returning via the ocean. The power supplies provide precise regulation of cable current to a value of 389 ma. The voltage drop in the undersea system is approximately 60 volts per repeater section in armorless cable and 50 volts per repeater section in armored cable. A 3500-nm system will require a nominal supply voltage of 5500 volts at each terminal.

At the shore terminal the signals to be transmitted are frequency multiplexed and pre-emphasized in preparation for transmission through the undersea system. Signals received over the cable are equalized, amplified, demodulated and separated for transmission beyond the shore terminal. Monitoring of performance and trouble location are the other important functions at the shore station.

A description of the shore terminal is given in Section IV, followed by a description of the equalization plan for the complete system in Section V. Facilities for administration and maintenance are covered in Sections VI and VII. The performance characteristics of the first long system between New Jersey and England are given in Section VIII.

IV. SHORE TERMINAL EQUIPMENT

4.1 *Multiplex Equipment*

Multiplex signals to be transmitted over the submarine cable system may be obtained from the land plant on a group connector basis or from channel equipment at the submarine cable terminal. In the latter case, high-efficiency channel banks are normally used. These allocate a 3-kc section of the carrier frequency spectrum to each telephone conversation, instead of the 4 kc common in the land plant. By using two steps of modulation and by imposing more severe requirements on the channel filters, it is possible to obtain a speech bandwidth of approximately 2800 cycles. It is thus possible to transmit 16 channels in the 48-kc carrier frequency spectrum normally allocated to a 12-channel group. The small

degradation resulting from this narrowing of the spectrum of each channel is considered tolerable for submarine cable systems in view of the considerable economy obtained. Channel banks of this type, which are described elsewhere in the literature, are manufactured by several foreign manufacturers.[6]

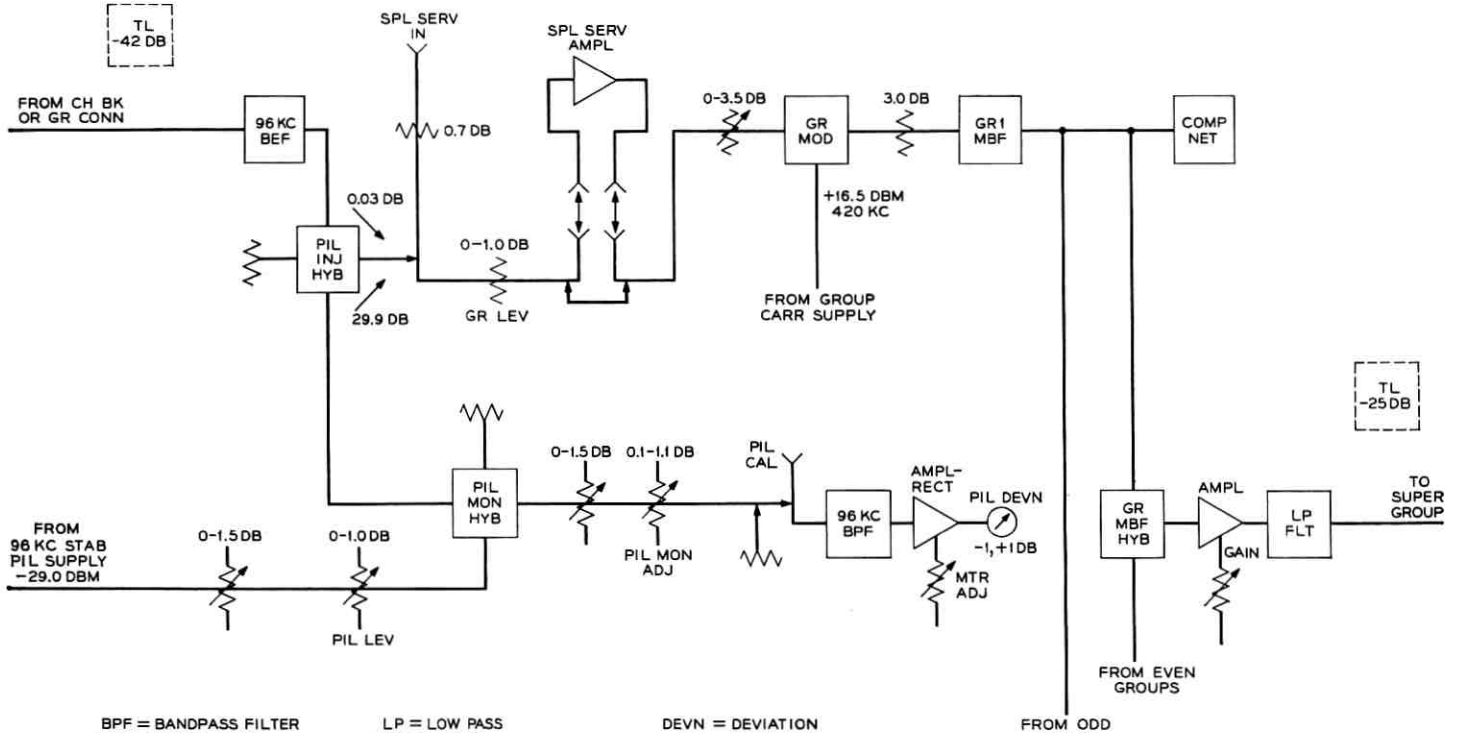
The group and supergroup equipment used is essentially standard L-type carrier multiplex or equivalent equipment supplied by the foreign partner, with minor modifications for pilot insertion and monitoring purposes.

4.2 *Pilots*

As in the land plant, a pilot frequency is transmitted in each group to permit monitoring and adjustment of group transmission at the various terminal stations and switching points. A 96-kc pilot is used to be compatible with the frequency allocation of the 3-kc channel banks. The pilot is inserted in each group modulator as shown in Fig. 4. They appear on the high-frequency line at the frequencies indicated in Fig. 2. The nominal power in each pilot is −20 dbm at zero level. The pilot in any group may be removed if a special service signal requiring a cleared 48-kc spectrum is to be transmitted.

In order to make possible maintenance without service interruption, much of the terminal equipment must be provided in duplicate. It is therefore convenient and desirable to arrange the terminal equipment as two independent transmission paths, one regular and one alternate, each fed by common channel bank equipment, as shown in Fig. 5. With this arrangement it is possible to use the pilots to initiate automatic protection switching from the working to the standby equipment to prevent loss of service in the event of a failure of the terminal equipment. Switching is inhibited if the standby equipment is nonfunctional.

Pilots are monitored at a number of points in the transmitting and receiving terminal. Fig. 6 is a block schematic of one of the monitoring circuits. At each point a narrow-band crystal filter and a narrow-band amplifier, either at 96-kc for monitoring at group input and group output or at line frequencies for monitoring at the supergroups or in the high-frequency line, are used. Meters of high accuracy for maintenance and adjustment purposes, and lower-accuracy relay-type meters to control alarms — and, in some cases, protection switching — are provided. A departure from normal amplitudes of approximately 1.5 db of any pilot will cause a protection switch from regular to alternate terminal equipment.



BPF = BANDPASS FILTER
MOD = MODULATOR
MON = MONITORING
CONN = CONNECTOR
GR = GROUP
BK = BANK
CH = CHANNEL
COMP NET = COMPENSATING NETWORK

LP = LOW PASS
MTR = METER
ADJ = ADJUST
SERV = SERVICES
STAB = STABILIZED
PIL = PILOT
BEF = BAND ELIMINATION FILTER

DEVN = DEVIATION
CAL = CALIBRATE
AMPL = AMPLIFIER
MBF = MODULATOR BAND FILTER
INJ = INJECTION
HYB = HYBRID
SPL = SPECIAL
TL = TRANSMISSION LEVEL

Fig. 4 — Group pilot insertion.

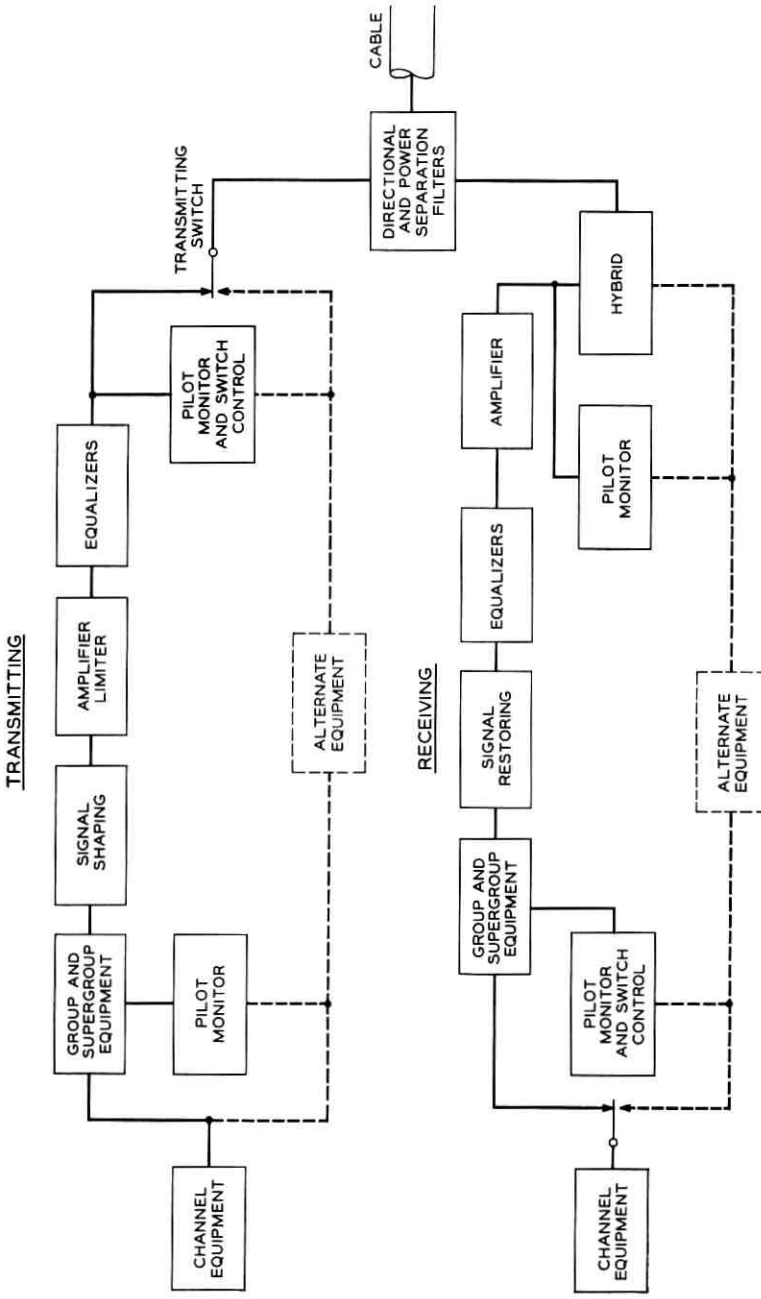


Fig. 5 — Shore terminal block schematic.

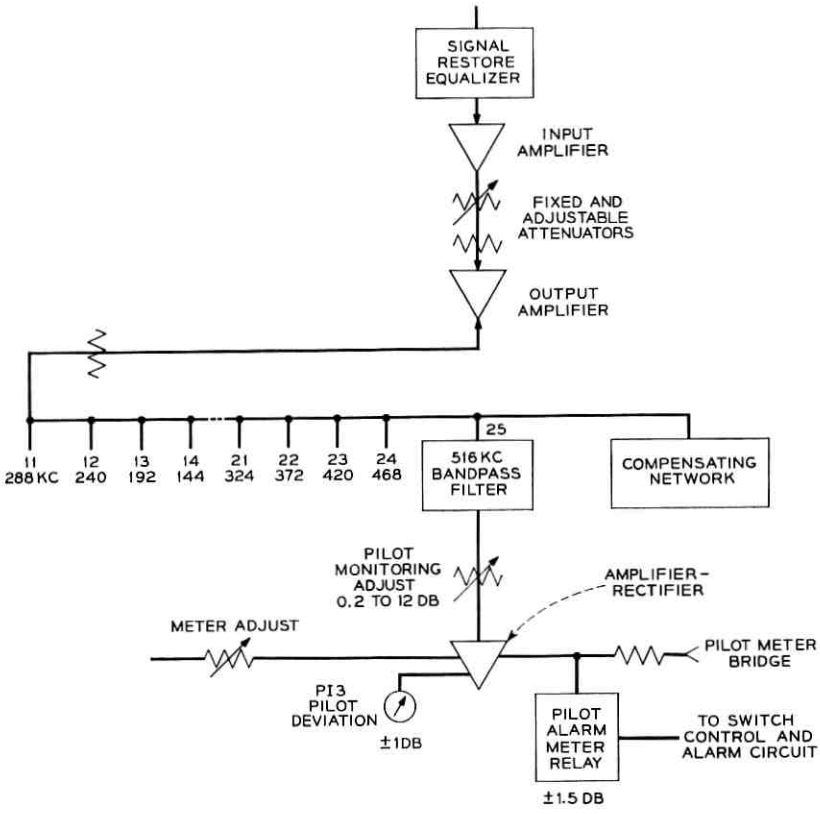


Fig. 6 — Pilot monitoring equipment.

4.3 *High-Frequency Line Equipment*

The final preparation of the signal for transmission over the ocean cable is accomplished in the so-called high-frequency line equipment, which accepts the multichannel load from the supergroup equipment and performs the necessary signal shaping, amplitude limiting and equalization functions.

Fig. 7 shows a block schematic of the transmitting high-frequency line equipment. The output from the supergroups is fed into a hybrid coil which provides for connection of the order wire signal and for pilot monitoring. This is followed by an amplifier and another hybrid that permits maintenance testing on an in-service basis and permits application of test signals for out-of-service adjustment. This is followed by a

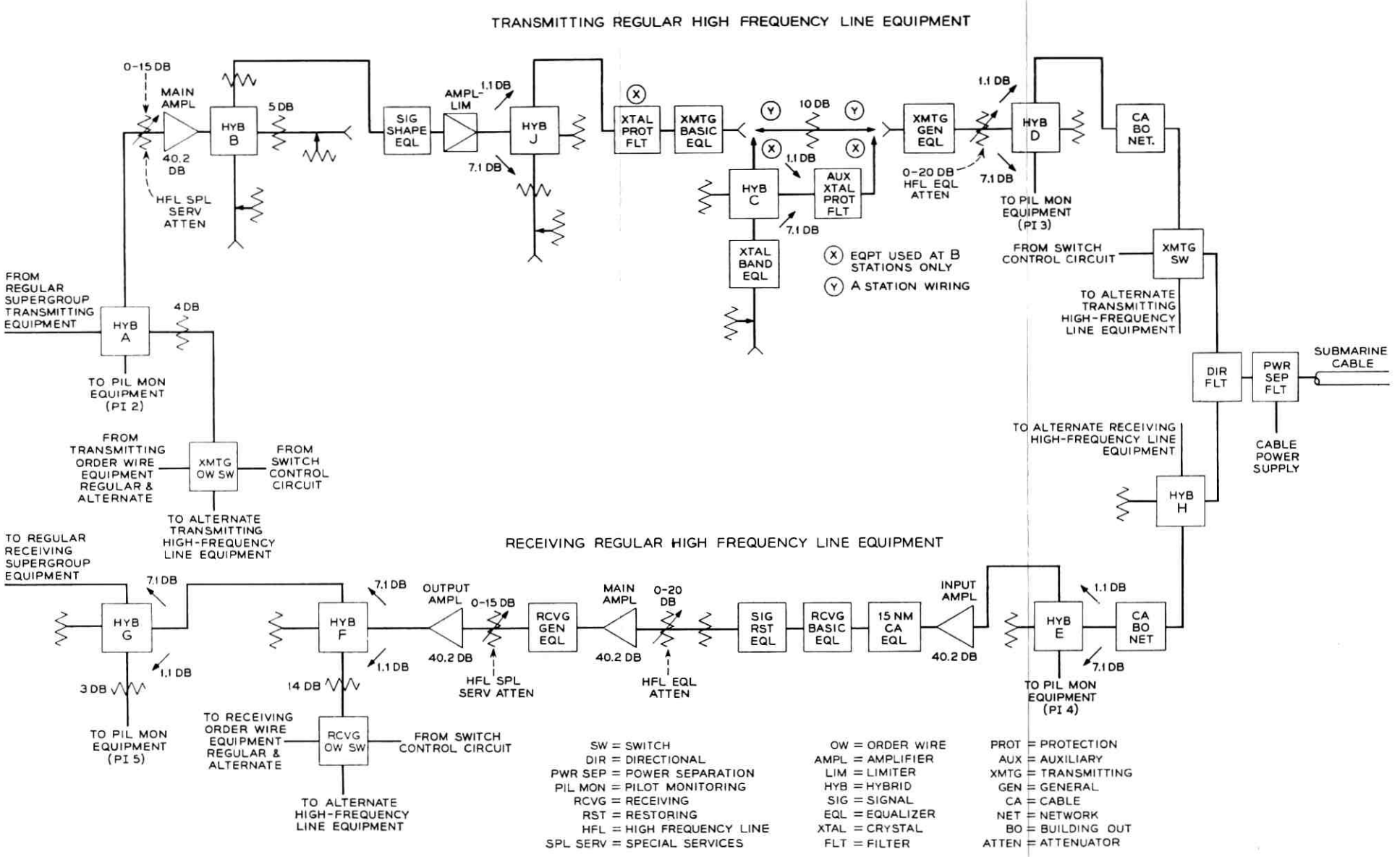


Fig. 7 — High-frequency line equipment.

signal shaping equalizer which predistorts the signal from the multiplex equipment in order to obtain approximately equal noise performance in all channels after transmission over the undersea system. As is commonly the case in cable systems, channels transmitted at high frequencies tend to emerge with a poorer signal-to-noise ratio than those transmitted at lower carrier frequencies unless special provisions — such as predistortion — are made. Such a difference in signal-to-noise performance between various channels is undesirable in any event, since the system performance is limited by the noisier channels. It is particularly undesirable in a system on which TASI may be used, since the customer would notice a sharp difference between noisy and quiet channels as the TASI terminal switched him from one to another. These undesirable effects are eliminated by the signal shaping. The smoothed curves of Fig. 8 indicate the effects of signal shaping on random noise in an SD system. The A curves show the rise in noise with carrier frequency when all channels are transmitted at the same power in the undersea equipment. By lowering the magnitude of all the low-band signals and raising the magnitude of all high-band signals, the noise in the two directions can be made equal (B curves) without increasing the total signal imposed on the system. Finally, the noise for each channel can be made equal by signal shaping (C curves). The characteristics of the signal shaping networks themselves are shown on Fig. 9.

An amplifier limiter which provides 30 db of gain and sharp limiting of the amplitude of the transmitted signal provides protection against overload of the undersea repeaters which might shorten their life. This is followed at B stations by a crystal protection filter to eliminate signals which might fall near the crystal peaks of the undersea repeaters. In this region system gain is so high that additional loss is required to furnish

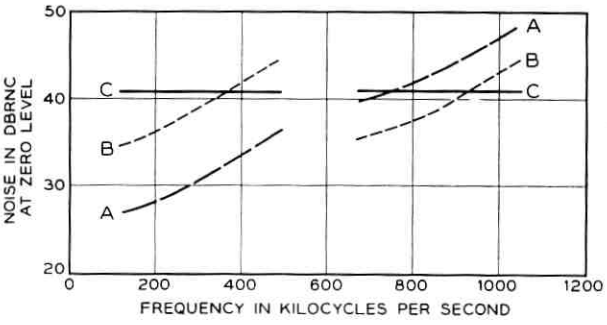


Fig. 8 — Effect of signal shaping on random noise.

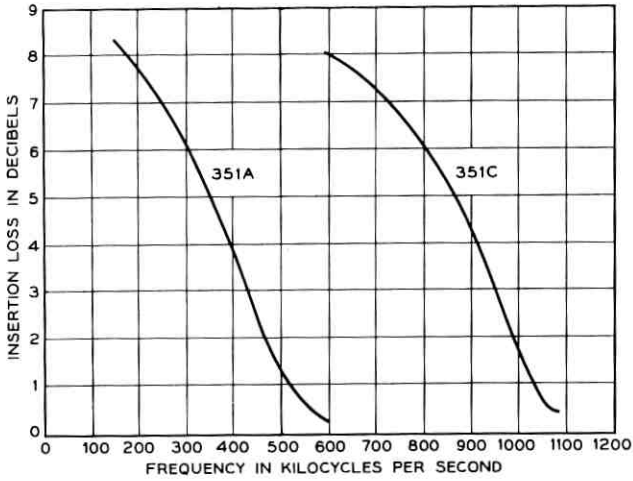


Fig. 9 — Signal shaping networks.

adequate protection. The cable build-out network adjusts for variation in the shore end sections of cable, building out the actual length to an electrical length of fifteen miles plus or minus 0.25 mile. The functions of the directional filter and the power separation filter are similar to those of the corresponding networks in each undersea repeater. Before discussing the various equalizers in the shore terminal equipment, it is appropriate to describe the over-all system equalization plan.

V. EQUALIZATION PLAN

The total cable transmission loss in a 3500-nm system will approach 9000 db at the highest transmitted frequency. This must be matched by the gain of 180 or more repeaters. Obviously, very small deviations in this match must be corrected at intervals along the system to prevent them from accumulating as large misalignments.

Basic equalization of the cable loss is accomplished by shaping the repeater gain to match the loss of the cable at a temperature of 3°C and a pressure equivalent to a depth of 2000 fathoms. The shaping takes place in the input network, the output network, and the feedback circuit of the amplifier.

The lengths of manufactured sections of cable are trimmed at the factory to obtain the desired loss at the highest transmitted frequency. In this way, it is possible to adjust to a first approximation for known departures from the design temperature and pressure conditions.

The accumulated factory cable deviations, plus those arising from uncertainties in ocean-bottom temperature and depth, temperature and pressure coefficients, and transmission effects introduced by the laying process, are equalized at 200-nm intervals during the cable lay with the adjustable ocean-block equalizers. The available shapes, selected by a stepping switch that can be operated on shipboard, are shown in Fig. 10. The choice of shapes is based on the following rationale. The loss of coaxial cable at high frequencies may be approximated as the sum of two terms, one proportional to the square root of frequency (resistance loss) and the other directly proportional to frequency (conductance loss). Experience so far has shown that providing these two shapes in the ocean-block equalizer permits compensation for any cable changes, regardless of source, and is an effective tool for reducing other misalignments as well.

The ocean-block equalizer also forms a convenient point for introducing a fixed equalizer to compensate for accumulated repeater deviations. Without intermediate equalization of this type along the route, the requirements on the repeater characteristic would be so stringent that undue complexity in the amplifier circuitry would be required. This would increase the probability of a failure of a repeater and would increase repeater cost.

During laying, measurements are made between the shore station and the end of the 200-nm ocean block being laid. As the cable reaches the bottom, its loss, which was high while on shipboard because of its relatively high temperature, begins to approach its ultimate value. Shortly before the equalizer goes overboard, it is adjusted for the desired block transmission as nearly as possible. After adjustment, the measuring equipment is connected to the output of the subsequent block and the process repeated until all the cable has been laid.

The choice of the optimum equalizer network is facilitated by using digital computer programs to carry out a "paper lay" in advance of the actual system installation. These programs utilize factory data on cable, repeaters and equalizers (including transmission characteristic and noise figure), depth and temperature information from route surveys, and apply temperature, pressure and laying effect coefficients obtained from a combination of theory and experience. This makes it possible to evaluate the signal-to-noise ratio obtained for the entire system as a result of various choices made in setting the eighteen or more ocean-block equalizers required in a long system. The foresight thus obtained has proven a valuable guide during the actual lay.

Fig. 11 illustrates the degree to which an ocean-block equalizer decreases misalignment. Two curves are shown — one the misalignment

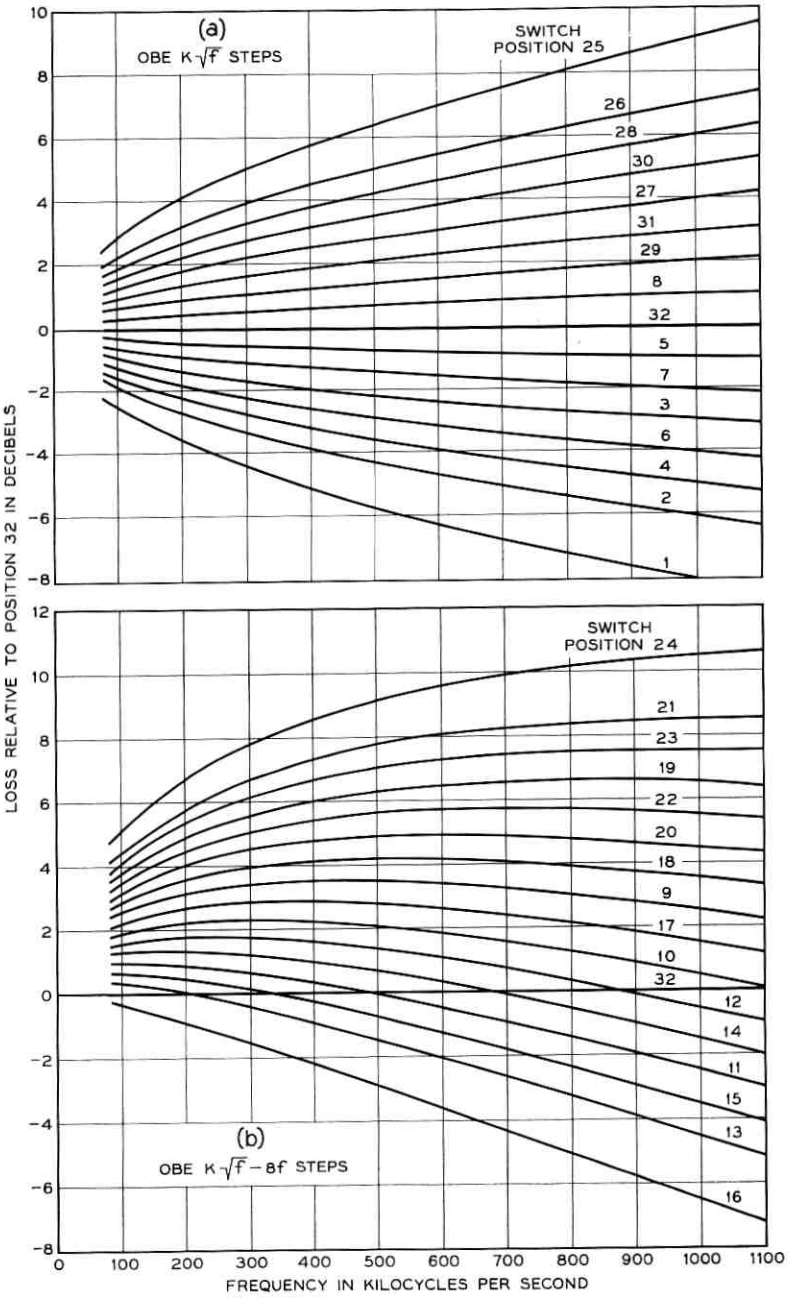


Fig. 10 — Ocean-block equalizer shapes.

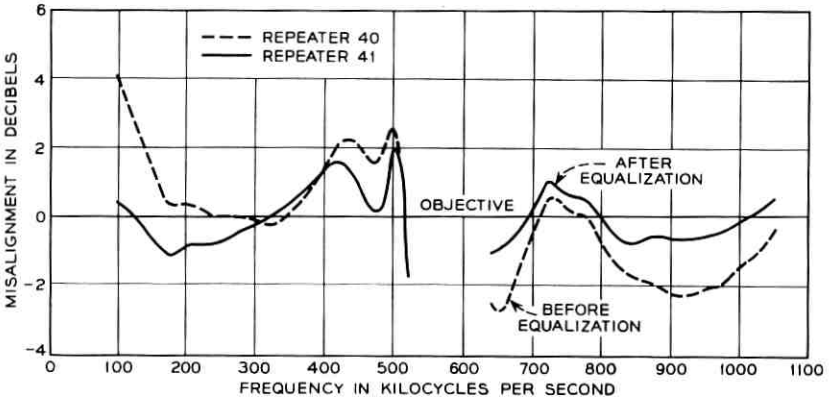


Fig. 11 — Reduction of misalignment by an ocean-block equalizer.

accumulated by the time 40 repeaters had been laid, the other the misalignment at the next (41st) repeater following an ocean-block equalizer. The misalignment has been reduced from a maximum of about 4 db to a maximum of about 2 db. The benefits of repeating such a reduction about eighteen times in a 3500-mile system are obvious.

5.1 *Terminal Equalization*

Once the system has been laid, the misalignments in the undersea system are set. From this point on, careful adjustment of signal levels to obtain maximum signal-to-noise ratios without overloading the repeaters is the only method available to optimize system performance. It is the function of the transmitting terminal equalizers to provide this optimization.

A convenient point of reference in describing the terminal equalization problem is the amplifier-limiter in the transmitting terminal. This is assumed to be an ideal limiter, flat with frequency, which will hold the instantaneous waveform to a fixed predetermined maximum value on both positive and negative peaks. Under gross overload, the output of the limiter will therefore have somewhat the appearance of a square wave. The most important function of the limiter is to prevent overloading the undersea repeaters in case of excessive inputs caused by trouble conditions. To obtain the greatest assurance of protection from overload, only passive circuit elements are installed following the limiter.

The question of what constitutes adequate protection needs to be

answered. It is convenient to consider the problem initially in a perfectly aligned system. The overload point of the repeater is determined by the start of grid current. This occurs at a value of third-grid voltage (output stage grid voltage) which is essentially constant with frequency. Since the limiter operation is independent of frequency, this implies that the transmission between the limiter output and the third grid should be flat with frequency. The waveform out of the limiter will then be reproduced essentially undistorted at the third grid and will appear as a square wave for severe overloads at the input to the system, neglecting waveform distortion caused by nonlinear phase.

In an equivalent four-wire system, the signal voltage at the third grid will be the sum of the voltages supplied from the two ends of the system. It has been assumed that repeaters must not be overloaded even in the unlikely event that these signals add in phase. The repeaters can be protected against this possibility by inserting an additional 6-db loss at each end or by adding unequal values, e.g., 2.4 db for the high band and 12.4 db for the low band, at the two ends, so that the voltage sum adds up to unity. By using the second procedure a 3.6-db signal-to-noise advantage for the high band is obtained.

Operation with only one of the parallel amplifiers working has been assumed in determining the repeater overload, since this is the most limiting condition. An additional 3-db loss has been included in the terminals to cover uncertainties in the repeater levels and variations in the overload point of individual limiters and repeaters.

When misalignments are considered, the point of concern is the highest-level repeater in the system. This is not necessarily the same repeater at all frequencies. In particular, an equivalent four-wire system may have the highest-level repeater at opposite ends of the system for the two frequency bands transmitted in opposite directions. If misalignment is not uniform, the highest-level repeater at a particular frequency may be at some midway point in the system. Theoretically, some advantage might be taken of the fact that simultaneous overloads from the two ends of the system will not have maximum impact on the same repeater. In practice, however, this would lead to unworkable complexity in the rules for adjusting the terminal equalizers, and the signal-to-noise advantages so obtained would be small. Thus, to recapitulate, the loss in the general equalizer and associated attenuator following the limiter at a B terminal is adjusted so that the maximum signal which can be applied to the undersea system is from 2.4 to 5.4 db below the value which would overload the most vulnerable repeater operating with only a single amplifier. The maximum signal which an A terminal can apply is ten db lower than for the B station.

The transmitting *basic* equalizer shown in Fig. 12 has a fixed loss characteristic. It is intended to provide the optimum overload protection as a function of frequency for a nonmisaligned system. The criterion for its design might be stated as follows: when all transmission components between amplifier-limiter and the third grid of the nearest repeater have their nominal values, the basic equalizer characteristic shall shape the signals so that the relationship between power at limiter output and voltage at repeater grid is independent of frequency.

The transmitting *general* equalizer shown in Fig. 13 is intended to provide the additional loss required when the system misalignments are positive. The equalizer will be adjusted at the time the system is initially lined up and will be readjusted subsequently whenever changes are called for because of system gain changes caused by aging, temperature or repairs. Before adjustment during initial line-up of the system — i.e., with all dials at nominal settings — the loss of the equalizer is flat with frequency. The objective of the line-up procedure is to make the amount of loss change from the flat condition numerically equal to

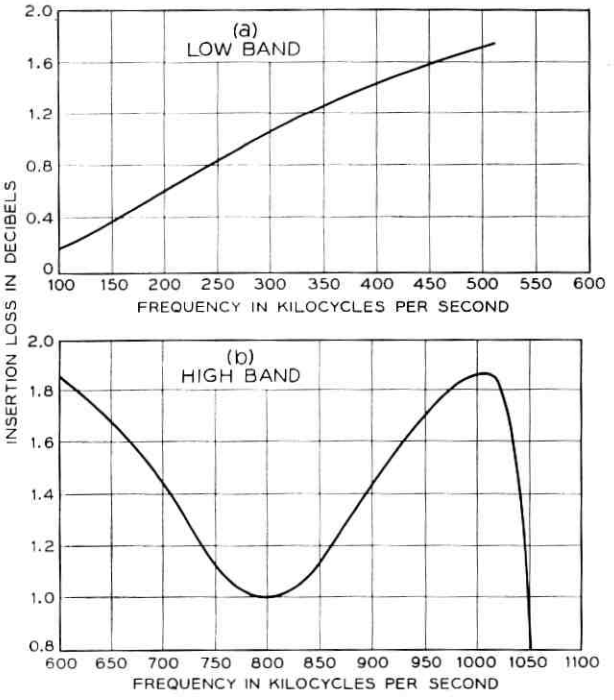


Fig. 12 — Transmitting basic equalizer shape.

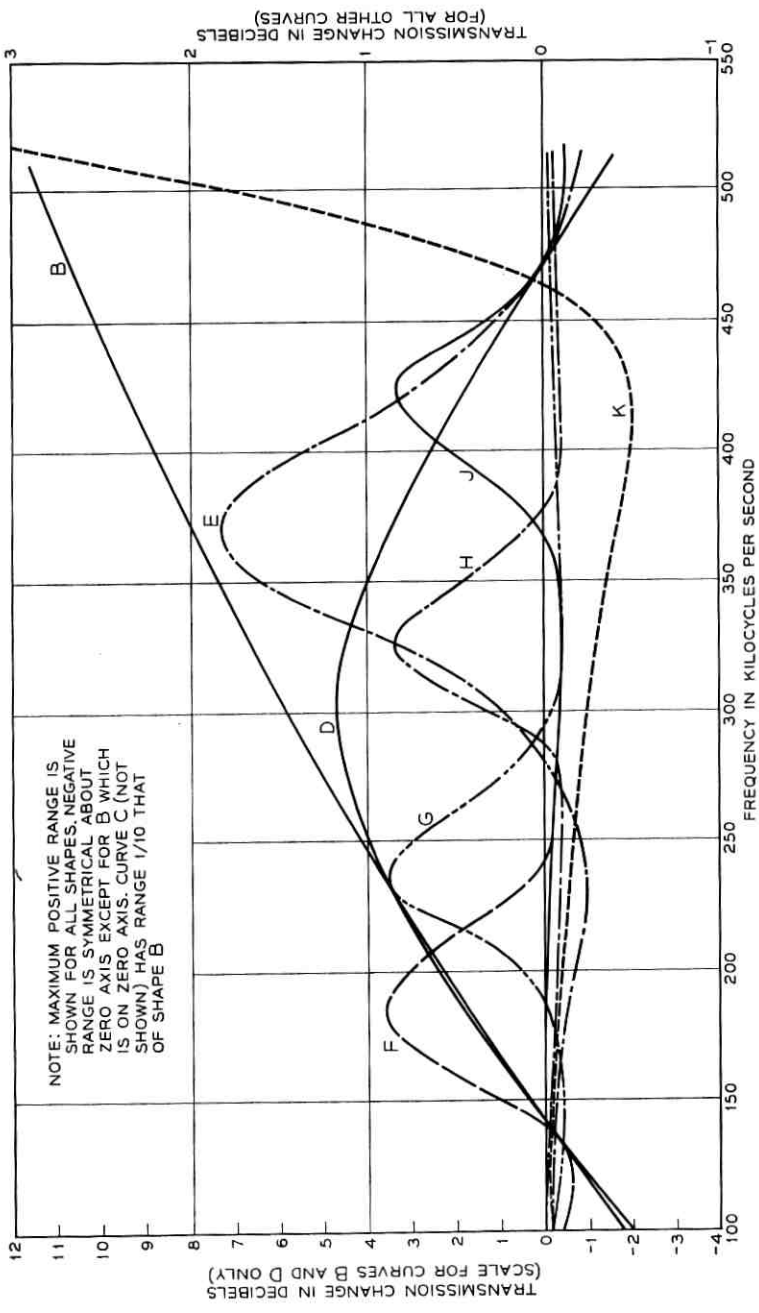


Fig. 13 — Transmitting general equalizer shape.

the maximum positive misalignment at the most vulnerable point in the system.

The function of the receiving terminal equalizers is to provide flat frequency response for the system, end to end. The cable build-out network is intended to bring the electrical length of the shore-end repeater section up to a nominal value of 15 nm. The 15-nm cable equalizer provides a flat transmission characteristic between the output of the last repeater and the input to the receiving basic equalizer. The receiving basic equalizer has a frequency characteristic complementary to the characteristic of the basic equalizer of the transmitting terminal at the other end of the system. The characteristics of the signal-restoring network are complementary to the characteristics of the signal-shaping network in the transmitting terminal.

As mentioned previously, positive misalignments are equalized at the transmitting terminal for overload reasons. Under these conditions, the receiving general equalizer is adjusted to its flat position. For negative misalignments, the total correction is made in the receiving terminal with the transmitting general equalizer set to its flat position. In either case the misalignments in the system will be completely equalized by this process of adjustment.

An exception to these statements occurs when the misalignments are not uniformly positive or negative along the length of the system. In this case the highest-level repeater will, in general, occur at some intermediate point in the system. Where this situation exists, the adjustable transmitting equalizer is used to provide overload protection for this repeater, and the adjustable receiving equalizer is used to compensate for the subsequent misalignment and thus obtain flat transmission for the entire system.

VI. ORDER WIRE

In order to permit communication over the submarine cable system between the terminal shore stations and between these and the gateway stations of the land plants involved, without preempting commercial circuits for the purpose, so-called "order wire circuits" are provided. To economize on the use of the spectrum of the undersea system, special provisions are made to put these maintenance circuits into the gap which normally exists between supergroups by using modified type ON terminal equipment. Fig. 14 shows in block schematic form the arrangements for this purpose. The ON-type channelizing and group frequency equipment is used to take the voice-frequency signals and form a spectrum lying between 100 and 108 kc for one direction of transmission and from

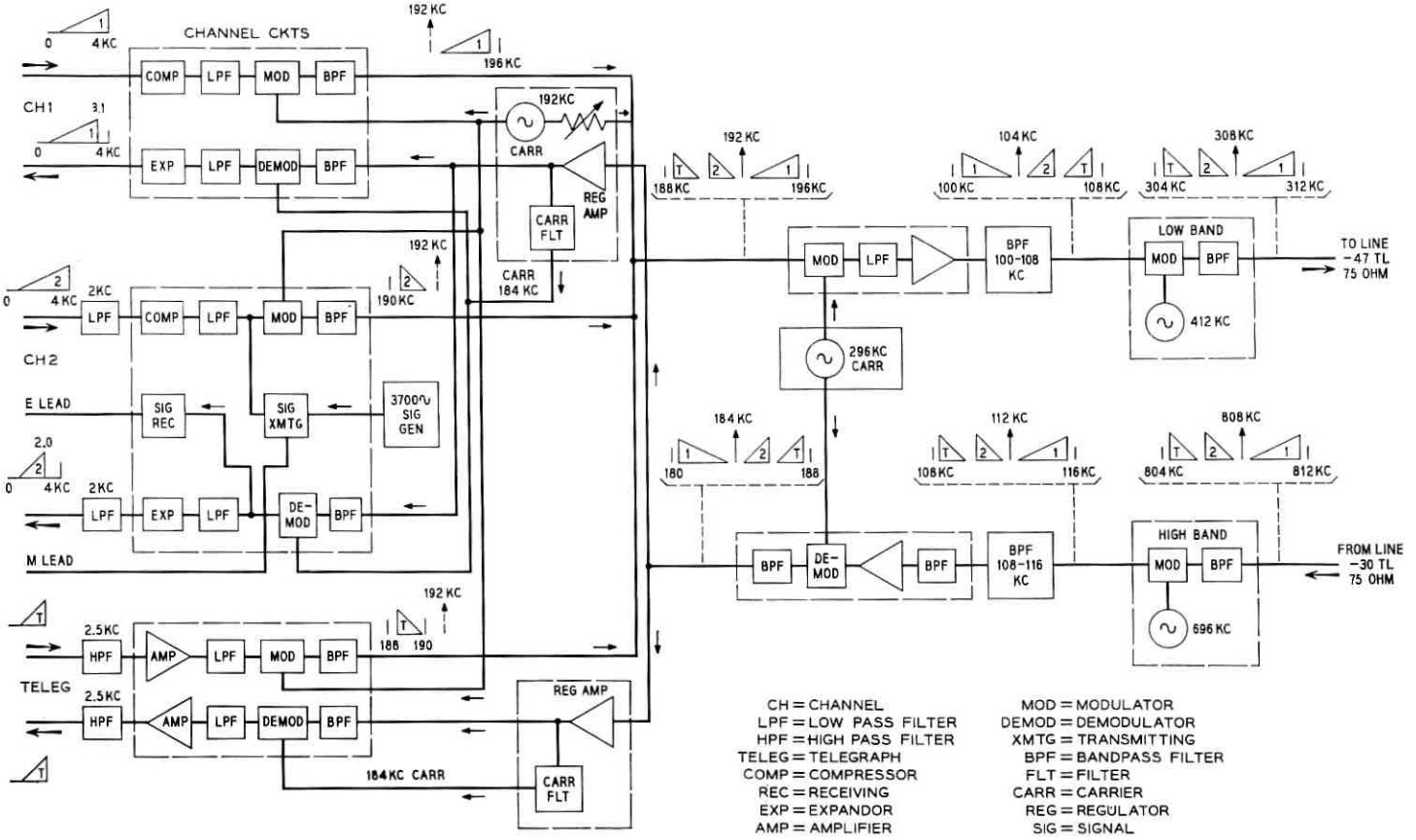


Fig. 14 — Order wire block schematic.

108 to 116 kc for the opposite direction. These signals are further modulated and placed in the 304 to 312 kc deadspace between supergroups 1 and 2, and in the 804 and 812 kc space between supergroups 3 and 4.

The channels thus appear on the line as upper and lower sidebands of 308 and 808 kc. The upper sideband channel is operated at full bandwidth for so-called express order wire service between gateway and land plant cities. The lower sideband channel is operated on a split basis, half of its spectrum being used for voice communication between the submarine cable shore terminal stations and the other half for teletypewriter signals. The use of teletypewriter communications giving a written record is particularly appropriate where language differences exist between the two ends of a system or among the various inland offices administering the circuits.

These signals are introduced into the high-frequency line equipment at a point following the supergroup equipment so that functioning group and supergroup equipment is not required for initial lineup or maintenance of the high-frequency line portion of the system. Switching circuits, interlocked with the automatic switching circuit which protects the multichannel signals, are provided so that the order wire is always connected to the working high-frequency line equipment at the transmitting and receiving ends of the system. Standby order wire equipment is available on a patching basis in the event that the working order wire equipment fails or requires routine maintenance.

Two additional channels are made available as upper and lower sidebands of 300 and 800 kc for use during system installation.

VII. MAINTENANCE TEST EQUIPMENT

In order to continue to obtain satisfactory performance over the life of the system and properly maintain it, various more or less conventional items of test equipment are required. These in conjunction with the pilot monitoring arrangements discussed above permit periodic adjustment of the transmitting and receiving equalizers to compensate for temperature changes, if necessary, and for long-term aging of the system. The items of test equipment required are sufficiently conventional to call for no further discussion here. Of somewhat more interest, in that they are peculiar to submarine cable systems, are the items of test equipment provided for fault locating. Three different types of test sets are employed for three different types of faults presently envisioned. These are:

(*i*) cable fault locating test sets which, in the event that the cable is broken or shorted, permit balancing the physical cable against an arti-

ficial cable to determine with fairly high accuracy the distance from the shore station to the fault;

(*ii*) the crystal band set, which permits measurements of system transmission at the frequencies of the crystals in the undersea repeaters, or measurements of noise peaks originating in these repeaters in the event that transmission from terminal to terminal cannot be obtained;

(*iii*) a repeater low-band fault locating test set which permits identification of a repeater whose transmission of low-band frequencies is faulty although transmission at the crystal frequency is not seriously affected.

7.1 *Cable Fault Locating Test Set*

Cable fault locating test sets of the type employed on the SD system have been described in the literature.[7] The basic philosophy is to provide a network which simulates, section by section, the iterative structure of the undersea system. Low-amplitude direct current, pulsed at a low-frequency rate (1, 2, or 4-second pulses of 50 per cent duty cycle), is sent into a bridge circuit of which one arm is the physical cable and one arm the simulating artificial cable. It is necessary to simulate only those system elements — cable, power separation filters, and repeater dc resistance — which are important at very low frequencies. A short or open (or intermediate resistance fault) can then, in essence, be moved along the artificial cable until the low-frequency impedance of the real and the artificial cable are seen to be equal. From a knowledge of where the fault in the artificial cable must be situated to obtain the best balance, the position of the fault in the real cable can be determined.

7.2 *Crystal Band Test Set*

The crystal band set is also similar to the measuring equipment described for use on earlier submarine cable systems.[8] Each undersea amplifier in the system is equipped with a crystal of unique frequency across the feedback network. At the resonant frequency the feedback path is shunted and the amplifier gain rises to a value determined by its forward gain. The test set essentially consists of a carefully controlled oscillator capable of supplying a signal at the crystal frequency of the repeater or slightly away from the crystal peak in order to determine the gain peak introduced by each repeater crystal circuit.

Alternatively, the receiving components of the crystal band set can be used to measure the noise peaks originating at the crystal frequencies of each repeater even if no signal is applied in the crystal band. In the event of a failure of transmission not caused by a cable fault and there-

fore not detectable by the cable fault test set described above, the noise peaks originating in the submarine cable system can be scanned and the results of this roll-call will indicate the location of the faulty repeater in the system.

7.3 *Repeater Fault Locating Test Set*

The basic concepts of the repeater fault locating test sets are indicated in Fig. 15. From the A terminal (low-frequency transmit, high-frequency receive) short pulses of 350 kc are transmitted. For any particular system, the repetition rate is set so that there is time for a signal to traverse the entire length of the system and return before the next pulse is launched. An answering pulse of 700 kc, produced by second-order modulation, is evoked in each repeater and propagates back toward the A terminal. At the A terminal, these returned 700-kc pulses, which are separated in time by the round-trip delay of 20 miles of cable, are observed. A marker pulse locked to the basic pulse timing circuits permits ready identification of the signal from a particular repeater. The presence or absence — and magnitude — of the pulses from the various ocean-bottom repeaters constitutes a roll-call of the system, and information about low-frequency transmission, high-frequency transmission, and second-order modulation of the repeaters

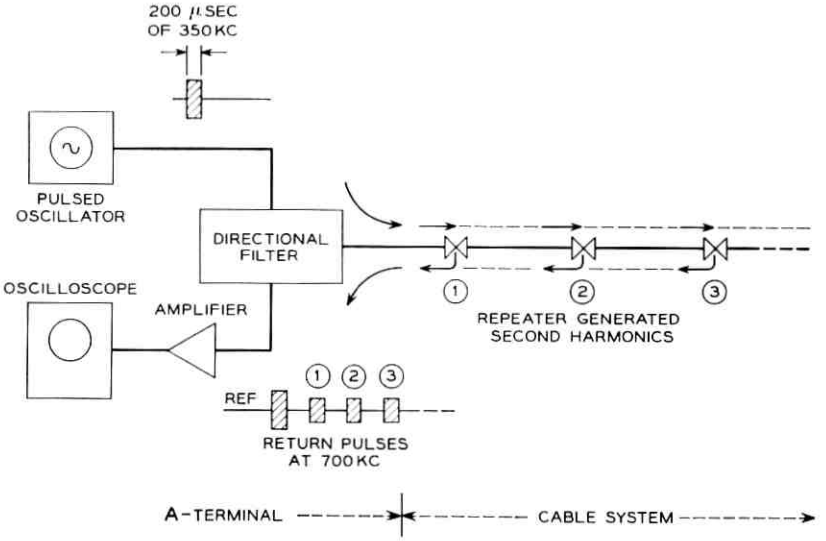


Fig. 15 — Repeater fault locating test set.

which answered the roll-call can be deduced. In a system with relatively little misalignment, the returned pulse trains can be displayed and examined on an oscilloscope. A photograph of a sample display is shown in Fig. 16. Should a system suffer considerable misalignment (6 db or more) the signal-to-noise ratio with respect to the 700-kc pulse would be so poor as to make oscilloscope presentation inadequate. Field experiments have shown that in such a case the presence or absence of a given repeater can be determined, even if the noise (in a 3-kc band) is 30 db greater than the returned pulse, by gating the detector and integrating the return signal over a long period of time — many minutes may be required under the least favorable circumstances.

VIII. SYSTEM INSTALLATION

At present an SD submarine cable system is in operation between Florida and the Canal Zone with an intermediate terminal at Jamaica, W.I. Another is in operation between Tuckerton, New Jersey and Widemouth, England. Installation of a system between Hawaii and Japan has just been completed. Since the system between the United

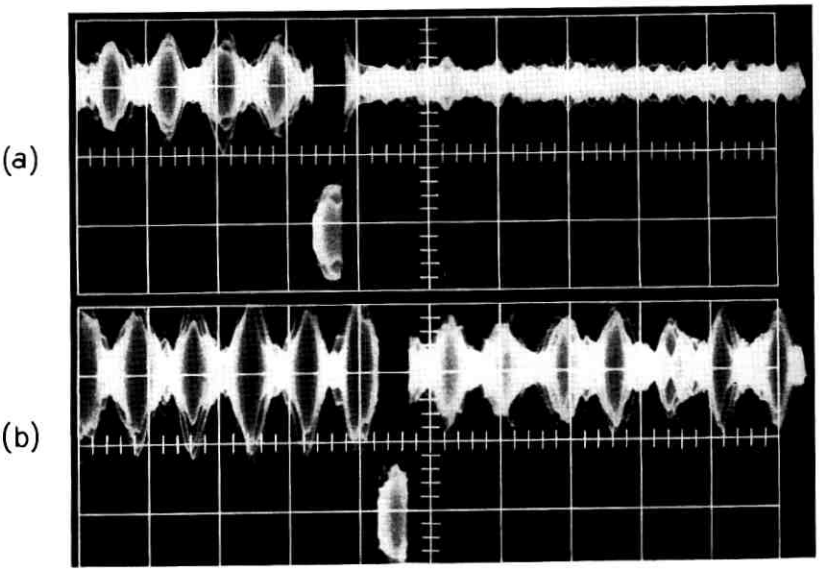


Fig. 16 — Repeater fault locating test set displays: (a) pulse returns from repeaters 178–182 — the pulse from the last repeater in the system (182) has been displaced for identification; (b) returns from repeaters 165–178 — repeater 171 has been displaced for identification.

States and England (commonly called TAT-3) has proved to be typical of long transoceanic systems, its performance will be discussed in some detail.

The length of the TAT-3 cable system is 3511 nm; the system follows a route shown approximately in Fig. 17. Maximum depth is 2800 fathoms. The system uses 182 repeaters and 17 equalizers. Approximately 12 per cent of the total cable is armored.

Laying operations were carried out during the summer of 1963. The British cable ship *Alert* laid the first 612 miles, starting at Tuckerton, New Jersey. Cable Ship *Long Lines* picked up at this point and completed the installation in two operations.

8.1 *Measurements During Laying*

Measurements made during the installation are summarized in Fig. 18 for four frequencies. The measurements show good agreement with the levels predicted from cable and repeater data, using ocean-block equalizer settings chosen during the laying operation.

8.2 *System Transmission*

The net gain of the undersea system (cable, repeaters and equalizers) is shown in Fig. 19. This was measured in October 1963, shortly after completion of laying, and shows similarly good agreement with computations. The transmission characteristic after equalization at the shore stations is shown in Fig. 20. This figure shows the deviation from flat transmission measured at the points where the multiplex signal is connected.

8.3 *Channel Noise*

Random noise measured at channel output with no talker load applied is plotted in Fig. 21. Since cable loss is a function of temperature, the random noise will exhibit a seasonal variation. The measured noise has been corrected to a mean annual temperature for comparison with computed noise under the same temperature conditions. It should be noted that most of the channels meet the 41-dbrn noise objective. Actually, 138 channels are acceptable for service.

8.4 *Modulation Performance*

The computed modulation noise produced by the talker load for which the SD system was engineered is shown in Fig. 22. The method of computation, which — like the equalization and random noise programs

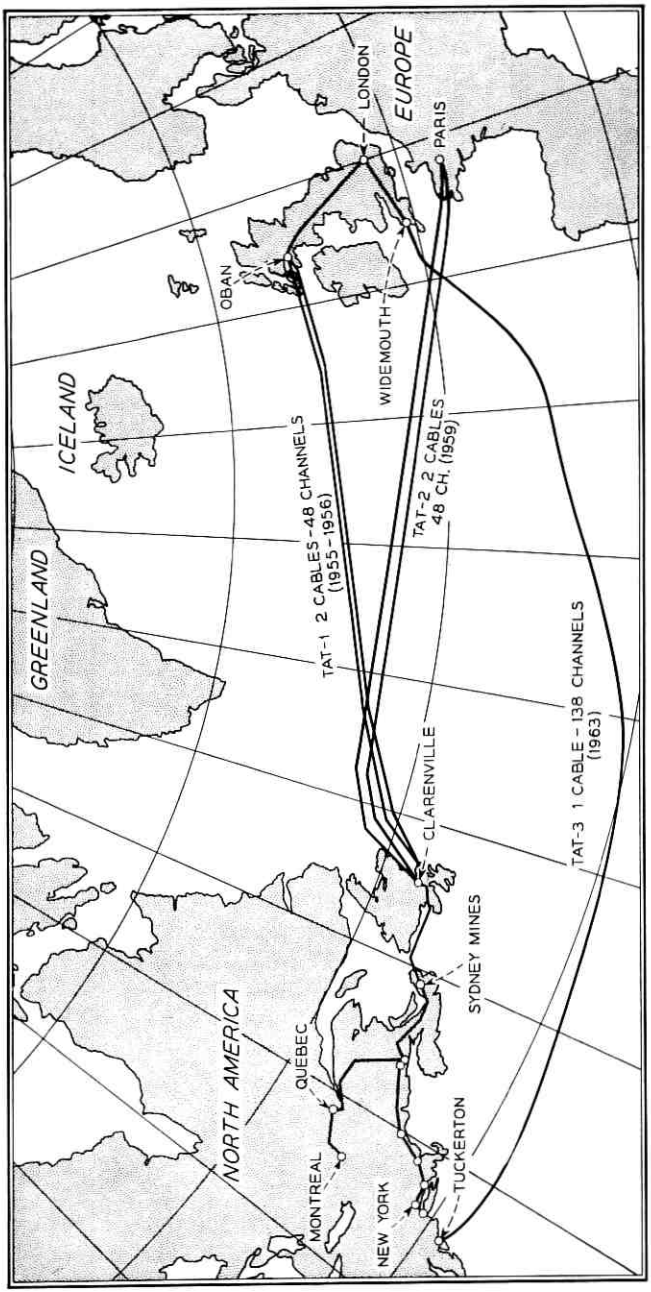


Fig. 17 — Route of TAT-3 system.

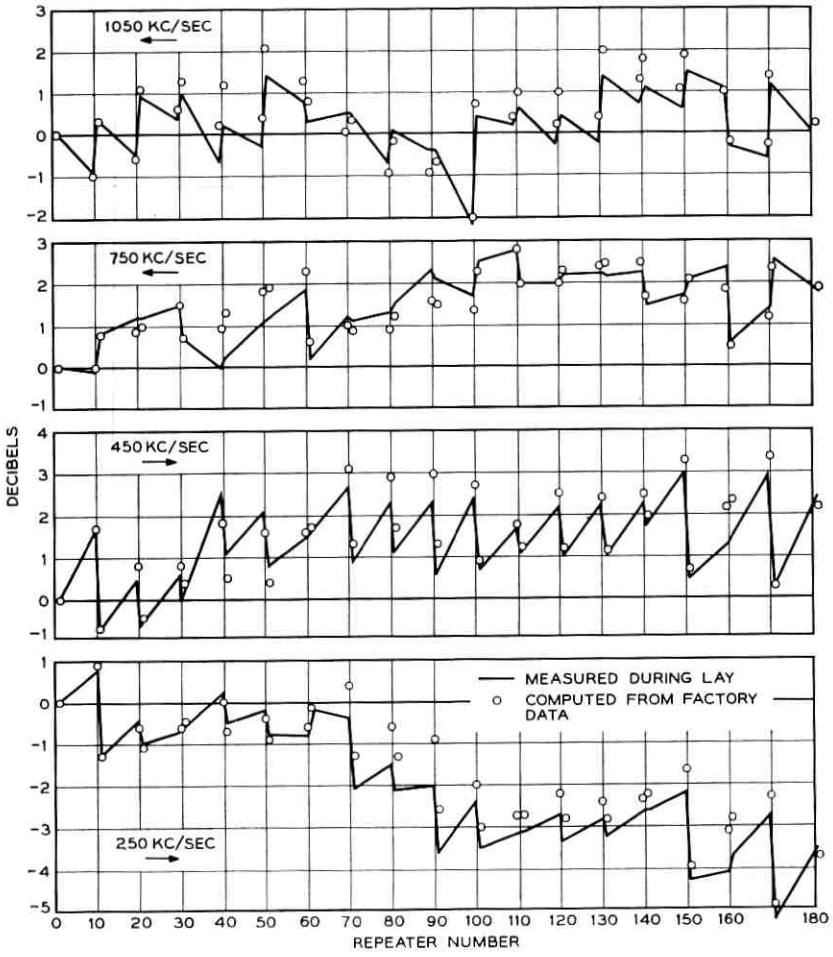


Fig. 18 — TAT-3 level diagrams, showing deviation of repeater output level from nominal.

discussed above — makes use of digital computer techniques, takes into account the effects of directional filter delay distortions on the addition of modulation product contributions from the various ocean-bottom repeaters. Two- and three-frequency intermodulation product measurements made on the undersea portions of the system give results which are consistent with the modulation noise values plotted versus frequency in Fig. 22.

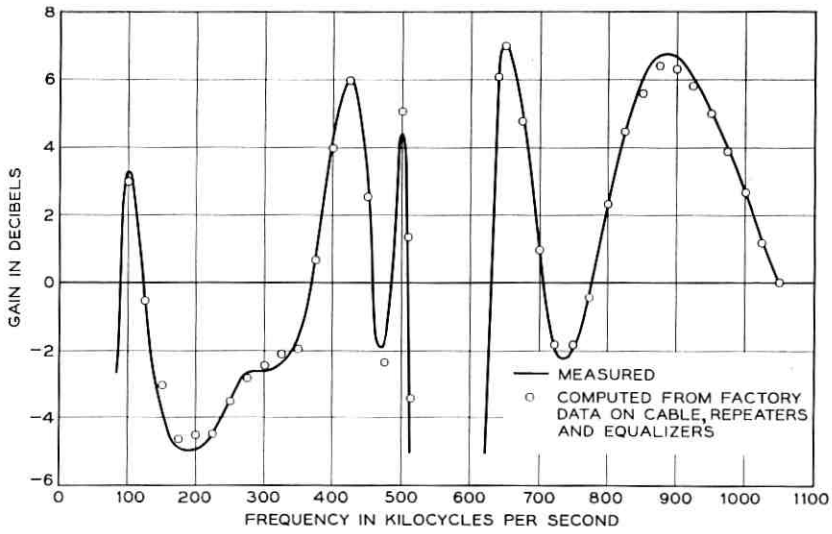


Fig. 19 — Net gain of undersea TAT-3 system.

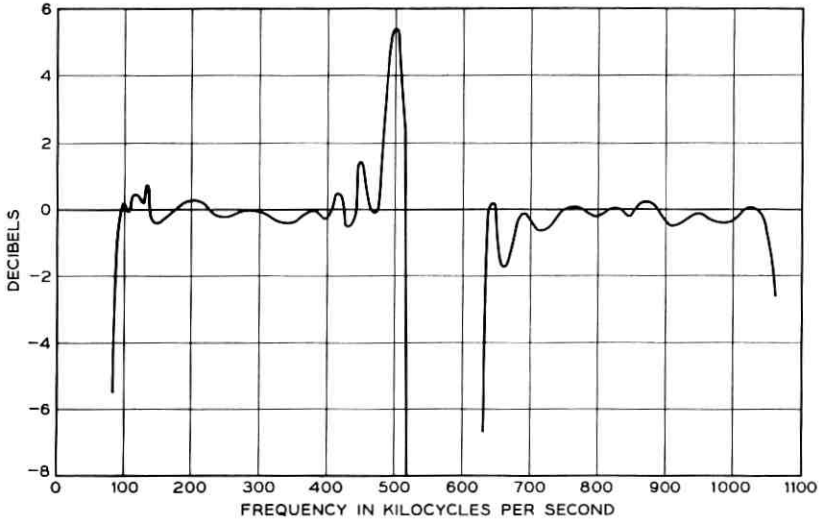


Fig. 20 — Net gain of TAT-3 system after shore terminal equalization.

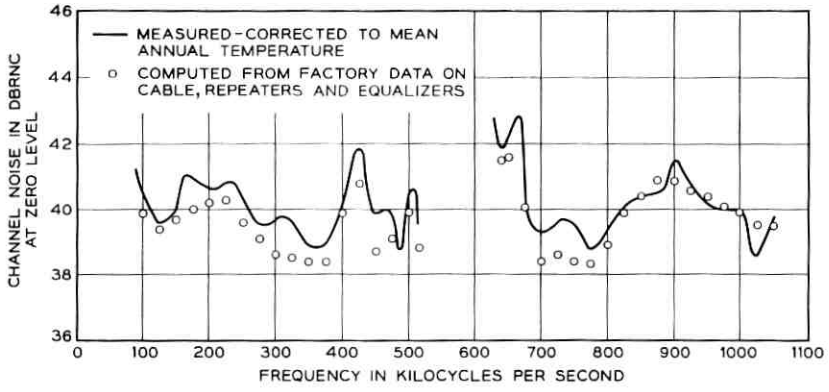


Fig. 21 — Random noise of TAT-3 system at channel bank outputs.

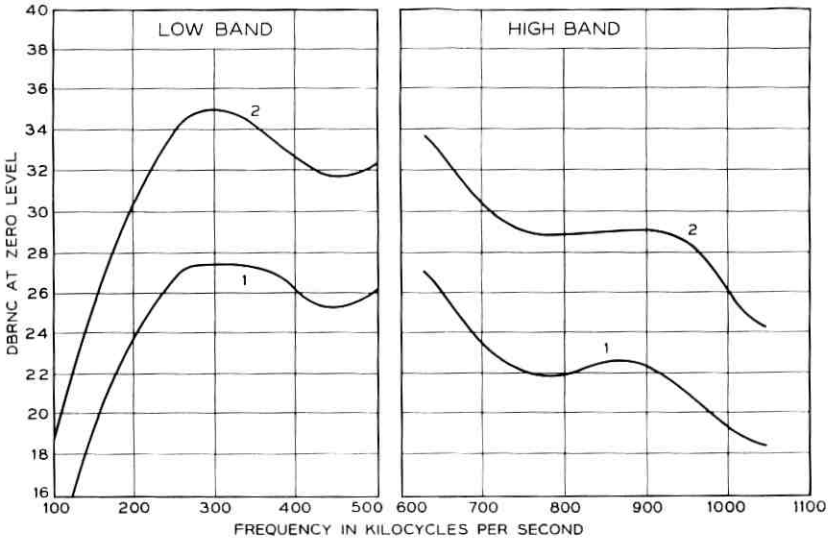


Fig. 22 — Computed modulation noise — TAT-3 system. Load assumptions: talker volume, −10.8 VU; standard deviation, 5.8 db; activity, 0.75. Curve 1 shows average modulation noise; curve 2 shows noise level that will be exceeded only 1 per cent of the time.

8.5 *Change in Transmission with Time (Aging)*

As of March 1964 the Florida-Canal Zone system had been in operation for 13 months and TAT-3 for 6 months. There has been no significant change in transmission that could not be accounted for by temperature changes. It appears that the aging of cable and repeaters will be very small.

REFERENCES

1. Special Transatlantic Cable Issue, B.S.T.J., **36,** January, 1957.
2. Bullington, K., and Fraser, J. M., Engineering Aspects of TASI, B.S.T.J., **38,** Mar., 1959, p. 353.
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., this issue, p. 1243.
4. Bowker, M. W., Nutt, W. G., and Riley, R. M., Design of Armorless Ocean Cable, B.S.T.J., this issue, p. 1185.
5. Bishop, J. D., and Mottel, S., Cable Power Facility, B.S.T.J., this issue, p. 1339.
6. Law, H. B., Reynolds, J., and Simpson, W. G., Channel Equipment Design for Economy of Bandwidth, Jour. P.O.E.E., **53,** 1960, p. 112. See also Tucker, R. S., 16-Channel Banks for Submarine Cable, Bell Laboratories Record, **38,** July, 1960, p. 248.
7. Weber, E. F., Fault Localization Test Set for Submarine Cable, to be published in Bell Laboratories Record.
8. Lewis, H. A., Tucker, R. S., Lovell, G. H., and Fraser, J. M., System Design for the North Atlantic Link, B.S.T.J., **36,** January, 1957, p. 57.

Design of Armorless Ocean Cable

By M. W. BOWKER, W. G. NUTT and R. M. RILEY

(Manuscript received April 15, 1964)

A low-loss coaxial ocean cable has been developed to be used as the transmission medium for the SD system. The strength member of the new cable is located inside the inner conductor. In deep-sea applications only a plastic jacket is required to protect the coaxial; in shallow water, where mechanical hazards are great, armor wires are applied over the coaxial in a more or less traditional manner.

A major concern in the development of the cable was that its transmission characteristics be predictable and be stable with time. This necessitated the consideration of mechanical and electrical requirements as one problem. Over 10,000 nautical miles of the new cable is performing satisfactorily in systems reaching to Europe, Asia, and Central America.

I. BACKGROUND

Development of cable for long, ocean-bottom, repeatered telephone systems is a process of engineering analysis, test, and evaluation. The objective is to provide a transmission medium that is predictable and reliable at reasonable cost. The development of cable for the SD system was carried on simultaneously with development of repeaters and facilities for placing cable in the ocean. Especially critical was the coordination of cable design with repeater design. This required that the cable engineers make early estimates of the cable attenuation characteristic so that repeaters could be designed with a gain characteristic to compensate for the loss of the cable. For instance, a 3500-nautical mile SD system has a cable loss at 1 mc of 8500 db. This means that uncompensated deviations in total cable loss of a few parts in a thousand would result in the received signals varying by tens of decibels.

To achieve an adequate match between cable loss and repeater gain requires first that the designs of both be predictable and stable. It requires that tolerances be placed on manufacturing processes, including precise measurement of final products with possible tailoring of char-

acteristics. It requires that the ocean-bottom environment for each length of cable be determined in advance of manufacture and that the effect of that environment on cable loss be estimated.

Deep-sea cables traditionally have been armored cables which have performed satisfactorily in both telegraph and narrow-band telephone systems. These cables generally have a relatively small central portion of the cross-sectional area devoted to the transmission function. The larger part of the cross section is required for helically applied steel armor wires and jute wrappings. This traditional approach can be characterized as one which first satisfies the electrical need and then adds strength members.

Early in the 1950's, cable designers began to consider larger, lower-loss coaxials which would be needed in broader-band systems. These considerations led engineers from both the Laboratories and the British Post Office to the same general conclusion: future deep-sea cables could be "inside out" cables. The conducting material would need to be only a thin layer because of the elevated frequency band and thus, for a larger coaxial, considerable space would be available in the heart of the structure for a strength member.

Furthermore, an integrated electrical-mechanical approach to the design problem was possible. This approach provided the possibility of minimizing torque-tension coupling, thereby preventing much of the twisting and stretching that had been characteristic of traditional armored cables. The reduced coupling has further advantages during laying operations in that it results in a higher probability of consistent and predictable sea bottom performance and a decreased risk during laying and recovery of kinking at the mechanical discontinuities presented by the rigid repeaters proposed for the system.

The design which evolved is shown in Fig. 1; it consists of an inner strength member, copper inner conductor, low-density polyethylene dielectric, copper outer conductor and an outer jacket of high-density polyethylene.

In deep-sea telephone systems more than ninety per cent of the total distance from shore to shore requires no special protective armor or electrical shielding. This is why most of the design effort concerning transmission predictability, reproducibility and economy has been expended on the basic armorless coaxial structure. The design must be strong enough to survive the rigors of laying and recovery, to resist the crushing pressure of deep water, and to provide a reasonable ratio of strength to weight in water. The armorless cable described herein is not as strong as its armored predecessor, but the above ratio has been main-

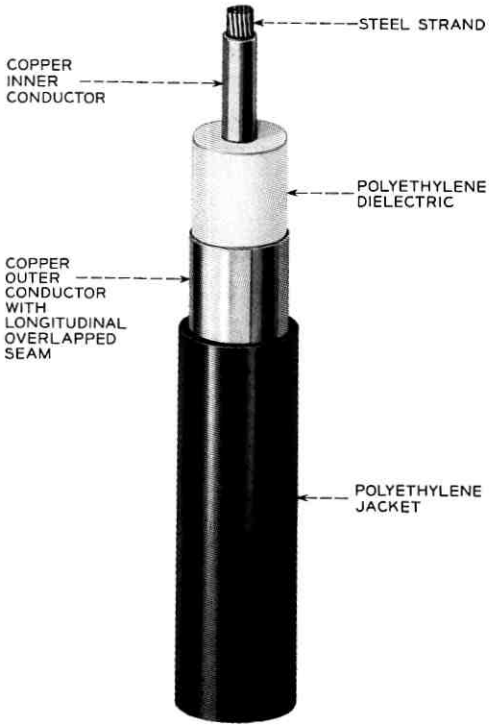


Fig. 1 — Armorless ocean cable.

tained because of the lower weight of the new cable. Once laid, the cable must have adequate resistance to attack by the various organisms that inhabit the ocean bottom and by the adverse chemical environment encountered at shore ends. The jacket of polyethylene seems to be the optimum answer to this protection problem.[1]

Inevitably, the deep-sea cables must be joined to the shore by cable which passes through shallow water. The shallow water environment is such that mechanical protection must be added around the coaxial, and in very shallow water this must be supplemented with electrical shielding. Various amounts of mechanical protection and electrical shielding are provided in special shallow water designs which are shown in Figs. 2–5. Some of the mechanical hazards which the cable encounters are tidal abrasion, tension failure and cutting by trawlers and anchors, crushing by icebergs, and attack by marine life. One or two layers of neoprene-jacketed steel armor wires are used as a protective cage around

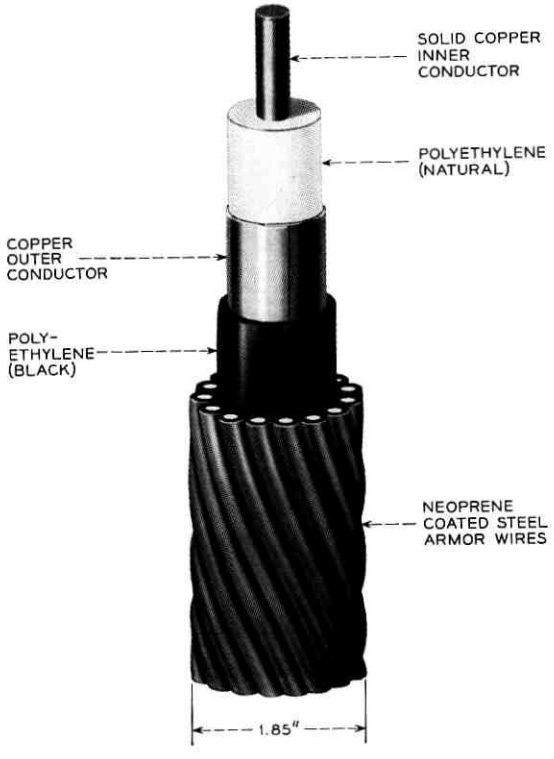


Fig. 2 — SD cable for shallow water.

the coaxial to minimize these hazards. When required, five layers of high-permeability steel tapes are applied over the coaxial to make an effective electrical shield. When external armor wires are used, the inner conductor is made entirely of copper. This is done to provide the ductility required for the conductor to elongate when the armor wires are subjected to tensile loads and concomitant cable elongation.

II. BASIS FOR DESIGN

The basis for the design is the combined electromechanical goal of providing a transmission line that is stable and predictable after laying and for its entire life on the sea bottom. The important electrical properties are the propagation constant and characteristic impedance. The most critical of these is the real part of the propagation constant. The mechanical constraints include those required to achieve predictable and stable electrical performance.

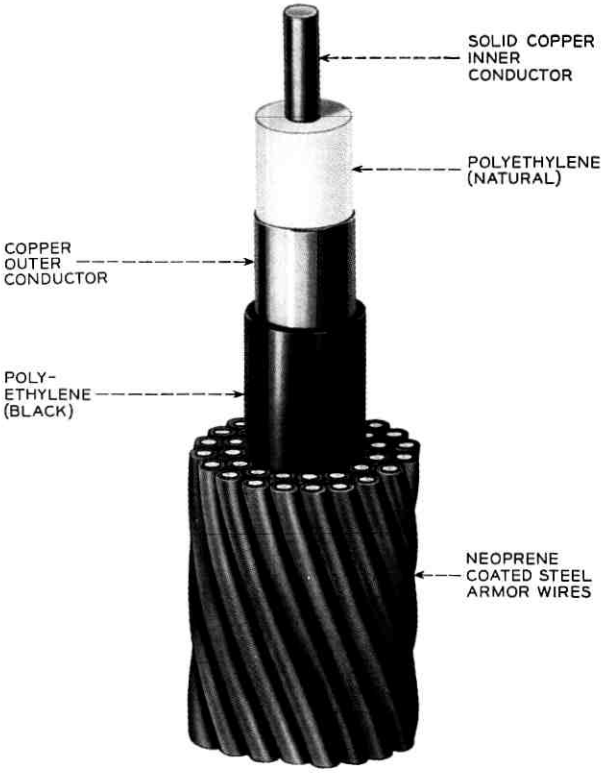


Fig. 3 — SD cable for shallow water, maximum protection.

2.1 *Electrical Diameter of Coaxial*

Economic constraints required selecting a size consistent with minimum over-all annual system costs and compatible with the technology of the time. Roughly, cable loss is inversely proportional to diameter, which means that, over some range, cable costs can be traded for repeater costs. A cable with a diameter of one inch over the dielectric falls into the optimum range for systems with a bandwidth of the order of 1 mc. Because of uncertainties in the ultimate economics of the system, other advantages and disadvantages of making a cable with a diameter of more or less than one inch were considered.

To exceed one inch would aggravate two basic uncertainties of that time. The extrusion of core as large as one inch, free of voids, required a significant step from any previous practice, including the experience of making the 0.620-inch armored cable used in the SB telephone systems.[2]

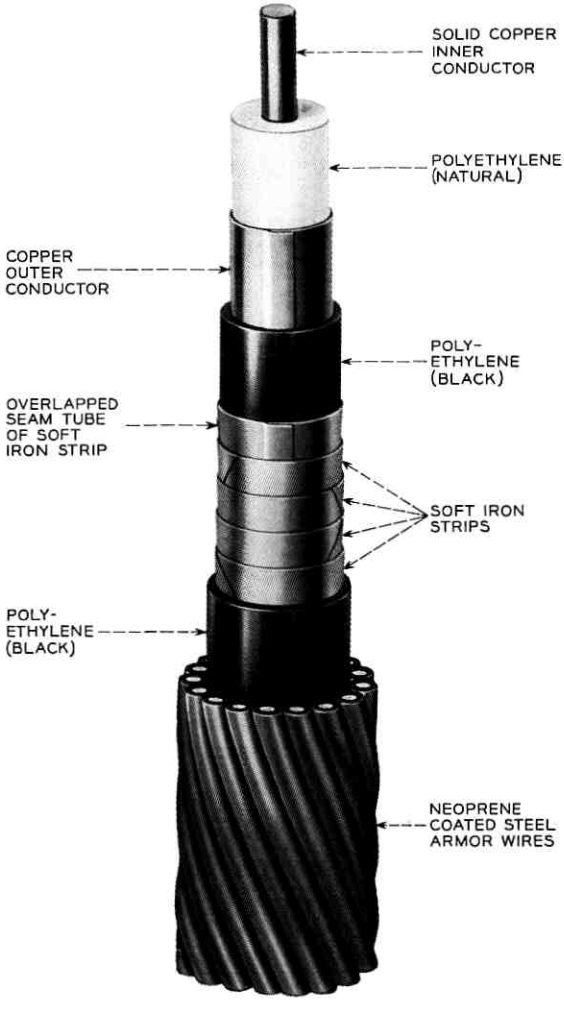


Fig. 4 — SD cable for shore ends.

The capability of a one-inch cable to withstand handling over reels of practical diameters and through cable machinery without serious buckling and perhaps rupturing of the outer conductor was questioned. Conversely, to reduce the diameter and thereby alleviate the above problems would increase the attenuation. This would require more repeaters in a given system and aggravate alignment and reliability prob-

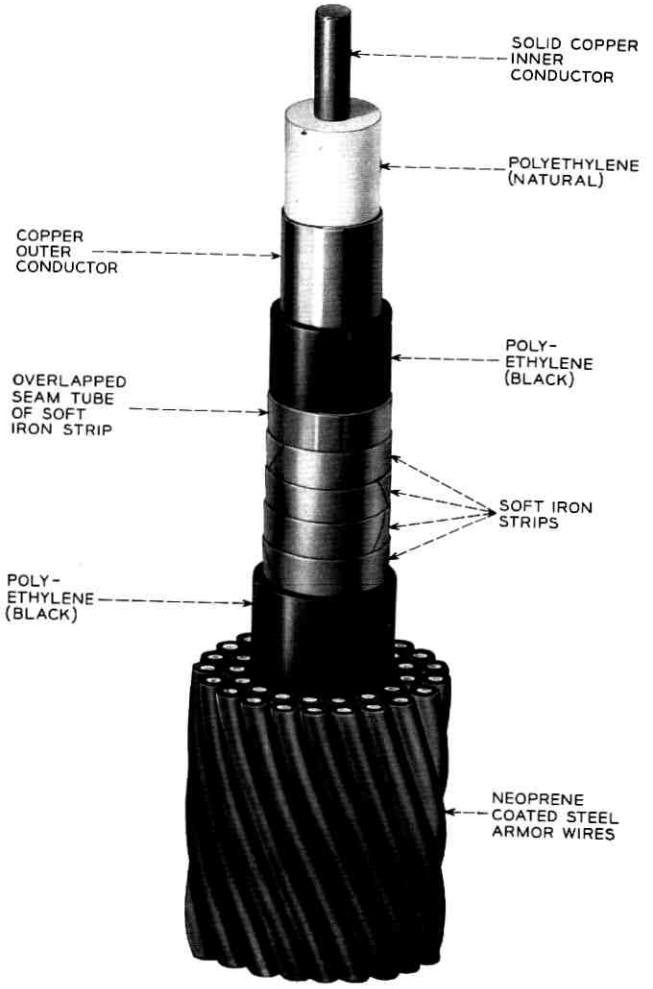


Fig. 5 — SD cable for shore ends, maximum protection.

lems. One inch appeared to be a reasonable diameter from the standpoint of cable mechanical properties and system transmission considerations.

2.2 *Mechanical Requirements*

From a mechanical viewpoint the basic design requirements are: (1) the ability to withstand the pressure at depths over 4000 fathoms (12,000

psi), (2) sufficient tensile strength to recover cable plus one repeater from depths of 3000 fathoms, and (3) protection from environmental hazards, both mechanical and chemical. The apparent inconsistency between the "pressure" depth and "tensile" depth requirements needs clarification. One is willing to place short portions of a system at depths greater than those from which the cable can be recovered. Laying tension can be limited to safe values, and usually depths greater than 3000 fathoms occur for relatively short distances. Therefore, one is justified in designing for pressures in excess of 4000 fathoms while also designing for a nominal recovery depth of 3000 fathoms.

The first of the above considerations dictates a structure as nearly incompressible as possible. The objective for breaking strength was taken as the sum of the weight in water of 3000 fathoms of cable and one repeater multiplied by a dynamic loading factor of 2.5. This amounts to about 16,000 pounds.

The tensile load of the inner conductor must be transferred to the shipboard machinery through the various cable layers. This requires careful control of the interlayer shear strength.

2.3 *Specific Dimensions and Materials*

Starting with one inch as the diameter of the dielectric, the inner conductor diameter was chosen on the basis of its influence on strength (directly proportional to the cross-sectional area of steel that can be included) and its influence on attenuation (relatively slight over a broad range).

The tensile requirement of 16,000 pounds led to a 0.29-inch stranded strength member consisting of 41 high-strength steel wires (Fig. 6). The lay-up of the strand was chosen for maximum strength in a limited area and has a single direction of lay. The other elements of the cable add to the 16,000-pound strength of the strand so that the cable has a breaking strength of approximately 18,000 pounds. However, only the strength of the strand can be transmitted through a repeater, and hence this becomes the controlling strength of the system. The length of lay of the strand is chosen on the basis of two considerations: (1) it must be short compared to a 90° arc on a 3-foot radius so that when the cable is tensioned over a sheave all wires are equally stressed. (2) it should be long enough so that the untwisting torque produced under tension can easily be restrained by the torque-tube action of the copper inner and outer conductors, and the polyethylene dielectric and jacket. Since the untwisting torque is portional to tension, a single curve may represent the torque produced by different lengths of lay. Fig. 7 shows that increas-

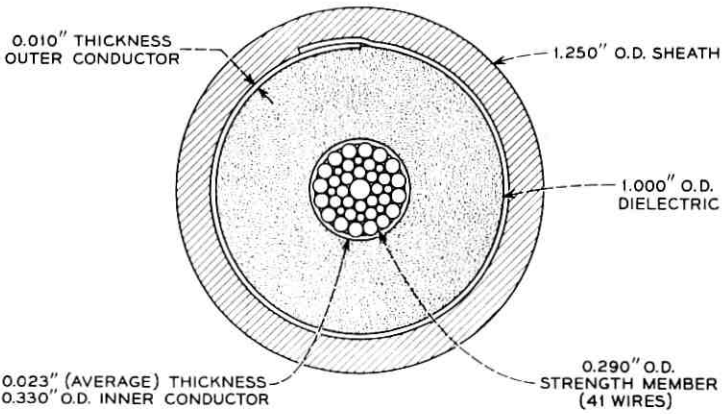


Fig. 6 — Cross section of armorless ocean cable.

ing the lay beyond six inches does not significantly reduce the torque. At normal laying loads, the restraining action of the cable components in the assembly allows a twist of less than one turn in 100 feet of cable length with a 6-inch lay strand. Thus the 6-inch lay meets both objectives. This lay is too long for a strand to hold its pattern by itself

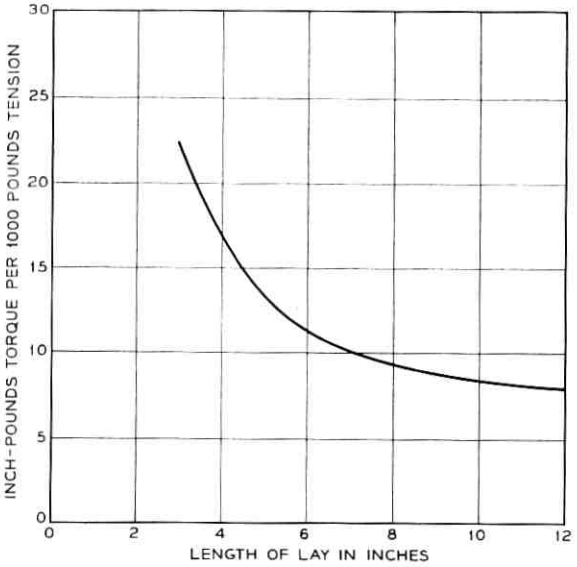


Fig. 7 — Torque vs length of lay for steel strand.

and was one of the factors that dictated the combining of stranding and inner conductor formation into a tandem operation.[3]

The copper conductor which surrounds the steel strand is hermetically sealed by inert gas welding and driven into intimate contact with the strand. The hermetic seal prevents volatiles in the strand from blowing holes in the dielectric during extrusion and also prevents substances on the steel that might be cracking agents for polyethylene from coming into contact with the dielectric. The intimate contact of the copper with the steel assures (1) the necessary transfer of longitudinal shear forces from the steel strand to the copper, (2) the transfer of resisting torque of the tubular members to the strand, and (3) that the tubular conducting member will not be crushed by the high ocean-bottom pressures.

The thickness of the copper portion of the inner conductor is chosen to provide sufficient electrical shielding so that at the lowest frequency very little current penetrates into the steel. Concurrently, there must be sufficient copper to hold the voltage drop for the repeater power current to a reasonable level. These considerations dictated an average copper thickness of about 23 mils with a nominal outside diameter of 0.330 inch.

For a dielectric, low-density polyethylene has a good balance of electrical properties, mechanical properties and cost. It has a low dielectric constant (approximately 2.28) and dissipation factor (approximately 0.0001). It is readily processed, has reasonable handling characteristics and is satisfactory over a wide range of environmental conditions.

Early in the development, it was recognized that for a stable transmission characteristic an ideal outer conductor would be a longitudinal copper one with a thickness of approximately 10 mils. Traditionally, outer conductors greater than about $\frac{1}{2}$ inch in diameter had been made up of multiple helical tapes so that they could be bent at reasonable radii without buckling. Such a conductor has uncertain current paths through changing intertape contact resistances. It also needs additional shielding to prevent crosstalk to any other nearby cable. Crosstalk is especially critical for the period when transmission measurements are being made aboard ship as the cable is being placed in the ocean. Here, interlayer crosstalk is aggravated by the gain of repeaters and may preclude accurate transmission measurements.

The seam in the outer conductor is simply an overlap of about one-quarter inch. Pressure and temperature effects require the outer conductor to expand and contract, changing the circumference of the outer conductor by as much as 30 mils.

To realize the electrical advantages of the longitudinal outer con-

ductor required some means of controlling the buckling when it was handled over reels and sheaves with a radius as short as 3 feet. At this radius of curvature, the outer portion of the 1-inch conductor is stressed beyond its elastic limit. Straightening out the conductor will result in buckling, which can lead to circumferential fatigue cracks unless the copper is constrained in some fashion. Several jacket materials and combinations of tapes and jackets were investigated. Experiments indicated that low-density polyethylene does not develop enough circumferential force to prevent buckling. Polypropylene does, but presents low-temperature and extrusion difficulties. High-density polyethylene $\frac{1}{8}$-inch thick, with properties midway between polypropylene and polyethylene, was found to develop just enough force to prevent any significant buckling for a reasonable number of cycles of reverse bending at a 3-foot radius. Hence, it became the choice for the outer jacket. Carbon black was included to minimize deterioration from photooxidation during the storage period.

The above discussions omitted any consideration of alternatives to copper. Aluminum, with its relatively good ratios of conductivity to cost and weight, was also of interest. However, as an inner conductor material, it imposed too large an attenuation penalty (14 per cent). As an outer conductor material it imposed a modest loss penalty (5 per cent) and additional development problems in the areas of corrosion and buckle suppression. For these reasons copper was selected as the conducting material for both inner and outer conductors.

The nominal attenuation characteristic of the resultant design is shown in Fig. 8.

III. TOLERANCES

Although the cable quite clearly has both electrical and mechanical functions, most of the tolerances are set primarily on the basis of electrical considerations. The exceptions are the stranded strength member and the outer jacket.

3.1 *Tolerances Determined by Mechanical Considerations*

The primary requirement for the strand is to develop the required strength within the allotted space and to be compact and crush resistant. Individual wires are required to have an average ultimate tensile stress of approximately 300,000 psi. There is an ultimate stress tolerance of ± 7 per cent and a diameter tolerance of ± 0.5 mil. Additional requirements are imposed on the selection of combinations of wires to insure

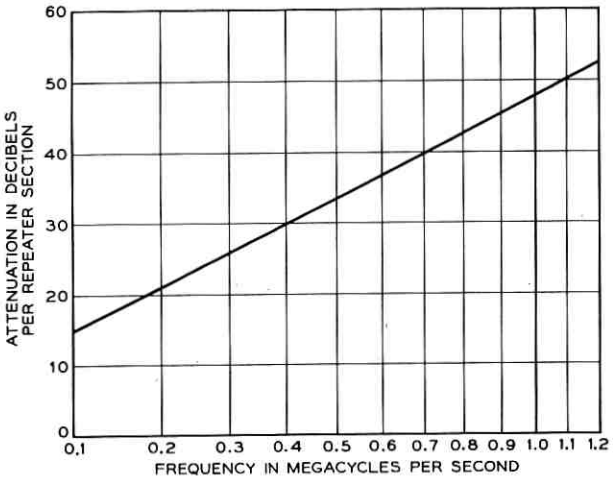


Fig. 8 — Nominal attenuation characteristic.

that the diameter of the completed strand is controlled. The steel wires are stranded in tandem with the copper tube forming, and the two elements are passed through a very tight die which forces the copper into the peripheral vee-shaped spaces between the wires of the outer layer. Thus a compact, concentric and uniform strength member is achieved.

The functions of the outer plastic jacket are protective and structural. Specifications for the properties of the raw material and for the extrusion thereof are chosen to give adequate control, but do not involve close tolerances.

3.2 *Tolerances Determined by Electrical Considerations*

As described elsewhere in this issue,[4] approximately 20 nautical miles of cable constitute a repeater section; ten repeater sections constitute an ocean block. The simultaneous development of the repeater and the cable made it necessary to predict the attenuation of the cable in each ocean block to within ± 2 db, which is ± 1.2 per cent at 0.1 mc and ± 0.4 per cent at 1.0 mc. The ± 2-db allowance includes uncertainties due to tolerances in manufacture, measurement errors, effects of handling and placing, inaccurate knowledge of pressures and temperatures along the actual cable route, and errors in estimating the effects of pressure and temperature. Thus the tolerances on the manufacture of the cable must limit attenuation deviations to considerably less than ± 2 db per ocean block.

Equations for computing the propagation constant and characteristic

impedance of a coaxial transmission line are well known.[5] However, most of the parameters of the cable requiring tolerances are evident from considering an approximate expression for the attenuation per unit length at high frequencies.

$$\alpha = k_1\sqrt{f}\left(\frac{1}{d\sigma_i} + \frac{1}{D\sigma_o}\right)\frac{\sqrt{\epsilon}}{\log D/d} + k_2 f F_p\sqrt{\epsilon}$$

where

f = frequency
d = diameter of inner conductor
D = diameter of the dielectric
ϵ = dielectric constant
σ_i, σ_o = conductivities of inner and outer conductors
F_p = dissipation factor of the dielectric
k_1, k_2 = constants of proportionality.

Omitted from the above expression are second-order or smaller terms which take into account the thicknesses of the inner and outer conductors.

In considering tolerances for the cable parameters, it is convenient to distinguish between tolerances that cause the attenuation to deviate by a constant percentage at all frequencies and tolerances that cause the attenuation to deviate in more complicated ways. A constant percentage deviation is said to have "cable shape" and may be compensated by adjusting the length of the cable. Tolerances that cause deviations having cable shape are those on the dielectric constant and the length measurement itself. Also closely approaching cable shape are the deviations caused by variations in the diameters of the inner and outer conductors, the diameter of the dielectric, and the conductivities of the inner and outer conductors. The percentage deviations, $\Delta\alpha/\alpha$, caused by the parameters having cable shape are plotted versus frequency in Fig. 9. Eccentricity of the inner conductor and deviation from circularity of the cross section are held so that their effects on attenuation are much less than ± 0.1 per cent. To the extent that they exist, however, they also have cable shape.

The remaining tolerances — dissipation factor and thickness of the conductors — do not have cable shape (see Fig. 10). Deviations caused by variations in the dissipation factor increase percentage-wise with frequency and are therefore not well compensated by adjusting length. The tolerances placed were as small as practicable. To reduce the requirements in this area equalizers were designed with a frequency characteristic that could compensate for variations in dissipation factor.[6]

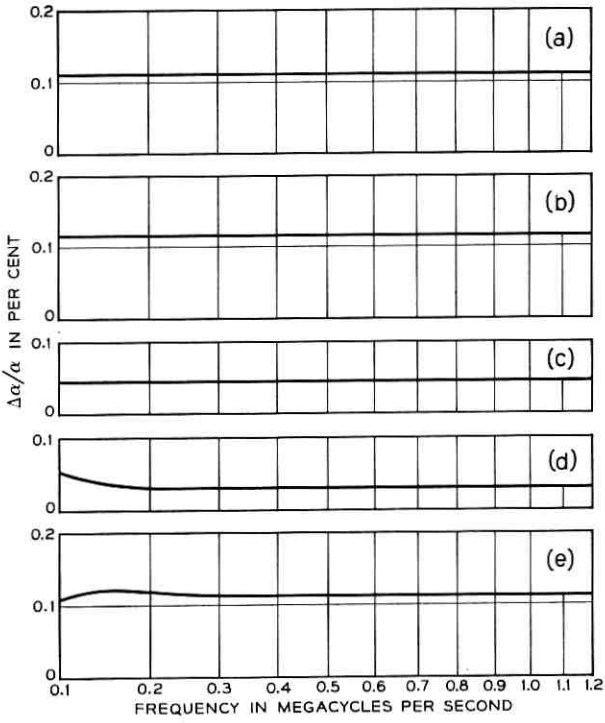


Fig. 9 — Tolerances having approximately cable shape: (a) per cent change in attenuation due to an increase in the dielectric constant of 0.005; (b) per cent change in attenuation due to a decrease in the diameter of the dielectric of one mil; (c) per cent change in attenuation due to an increase in the diameter of the inner conductor of one mil; (d) per cent change in attenuation due to a decrease in the conductivity of the outer conductor of 0.3 per cent; (e) per cent change in attenuation due to a decrease in the conductivity of the inner conductor of 0.3 per cent.

The thickness of both the inner and outer conductors was limited by strength-to-weight and cost considerations. Although the inner conductor thickness is such that very little current flows in the steel, at the lowest frequency there is some sensitivity to thickness variations. For the outer conductor, variations in thickness cause significant variations in attenuation throughout the bottom half of the transmitted band. Concerning the lapped seam of the outer conductor, some variations in contact resistance are possible and cannot be controlled, but fortunately the effect is essentially negligible.

One approach would be to select tolerances so that a length of cable exactly equal to the nominal repeater section length would meet all

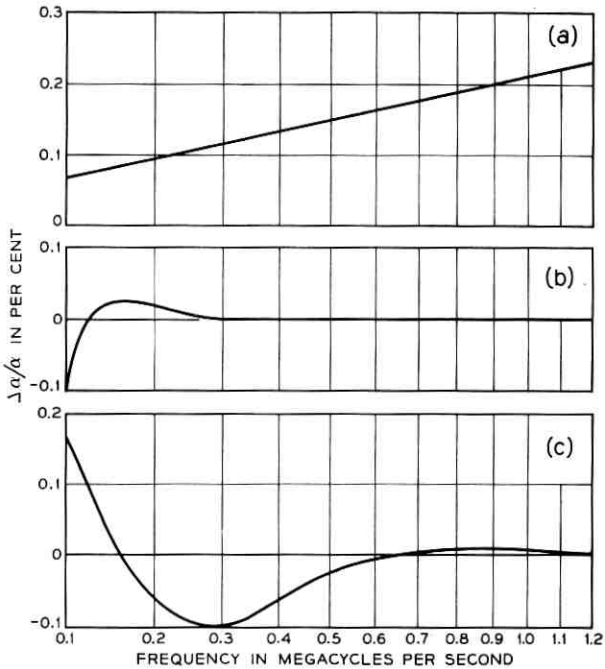


Fig. 10 — Tolerances not having cable shape: (a) per cent change in attenuation due to an increase in dissipation factor of 0.00002; (b) per cent change in attenuation due to a decrease in thickness of the inner conductor of 0.5 mil; (c) per cent change in attenuation due to a decrease in thickness of outer conductor of 0.2 mil.

transmission objectives. However, it was more economical to set the tolerances with cable shape a little wider than this, measure the attenuation of each length of cable after manufacture, and then adjust the length of some of the sections to compensate the cable shape deviations accumulated in previous lengths.

Table I lists the tolerances selected and the attenuation deviations that would be caused by their extreme values. In the extremes, these tolerances would result in deviations in attenuation of an ocean block of ± 1.2 db at 0.1 mc and ± 3.0 db at 1.0 mc. Random combination of the tolerances would result in attenuation deviations of ± 0.5 db at 0.1 mc and ± 1.5 db at 1.0 mc. Adjusting length after manufacture was expected to reduce the range of the deviations at all frequencies to within ± 1.0 db. As a result of favorable experience with the design, the present objective at the manufacturing locations is ± 0.5 db.

TABLE I—TOLERANCES FOR CABLE PARAMETERS

Parameter	Dimension and Tolerance		Effect on Attenuation (Per Cent)	
			0.1 mc	1 mc
Inner conductor diameter	0.330″	±0.001	±0.04	±0.04
Diameter over dielectric	1.000″	∓0.001	∓0.12	±0.12
Thickness of inner conductor	0.023″	±0.0005	±0.10	±0.00
Thickness of outer conductor	0.010″	∓0.0002	∓0.18	±0.01
Conductivity of inner conductor	99.1%	∓0.3	∓0.11	±0.11
Conductivity of outer conductor	100.6%	∓0.3	∓0.05	±0.03
Dielectric constant	2.282	±0.005	±0.11	±0.11
Dissipation factor	0.00012	±0.00002	±0.07	±0.21
Totals				
Algebraic			±0.78%	±0.63%
Root sum square			±0.30%	±0.29%

IV. MODEL MAKING

Armorless cable differs enough from the traditional that attempts to make models on available machinery resulted in samples unsuitable for electrical and mechanical evaluation. For this reason a special laboratory was equipped early in the development period to determine the feasibility of several different armorless cable designs. The cable fabrication laboratory, located at Cambridge, Massachusetts, was operated by the Simplex Wire and Cable Company under contract to Bell Telephone Laboratories. In it were produced the cable samples needed for making the mechanical and electrical tests to verify the design principles and to establish the necessary manufacturing tolerances. Approximately fifty miles of cable were fabricated for this purpose. These experimental lengths permitted Bell Laboratories engineers to determine the feasibility of manufacturing the cable and to establish an optimum balance among the several tolerances. Finally, the experience provided a basis for the study of factory layouts adapted to the manufacture of armorless cable.[3,7]

V. LABORATORY ELECTRICAL MEASUREMENTS

Equipment and methods for making appropriate electrical measurements were developed concurrently with the cable. For the purpose of discussion the measurements can be divided into those made on short, medium, and sea-trial lengths.

5.1 *Short Lengths*

Primary constants measured on short samples are important for the following reasons:

(1) Since the test specimens should be less than one-eighth wavelength at the top measurement frequency, a process not yet refined may be adequate to produce samples.

(2) The quantities measured — resistance, inductance, conductance, and capacitance — are directly related to dimensions and physical properties of materials, and therefore cause and effect may be readily related.

(3) Since the sample lengths are short, the problems of control and determination of environment are easier than for long samples.

The main disadvantage of the measurement of primary constants is associated with the short length, namely that the quantities measured are electrically so small that connecting leads and terminations must be designed and evaluated with considerable care. Resistance and conductance measurements require particular attention to connecting leads. Special facilities were designed and built to meet the accuracy objectives. These included:

(1) a 32-foot long environmental tank capable of simulating pressure and temperature conditions for depths up to 4000 fathoms,

(2) special measurement bridges, and

(3) coaxial comparison standards with electrically thick conductors and a minimum of disc insulators.

To evaluate the effect of pressure seals at the ends of the environmental tank, a two-foot tank was provided that was identical in every respect except length. The bridge and associated leads were located so that lead length and configuration were not changed as the bridge was connected to the standards and to the samples. Special low-resistance coaxial plug and jack connectors were developed to reproduce contact resistance of connections at 1 mc to 0.1 milliohm.

The bridges were maintained at essentially constant temperature. The exteriors of the tanks, including end seals and associated connectors, were well insulated to minimize temperature gradients in the 30-foot test length. The water temperature of each tank was measured at several points along its length to determine the temperature profile of the cable sample. In addition, the average temperature along the length was determined by measuring the dc resistance of an insulated copper wire also contained in the pressure tank.

TABLE II—THEORETICAL VS MEASURED VALUES OF TEMPERATURE AND PRESSURE COEFFICIENTS

Primary Constant	Temperature Coefficients % per °C		Pressure Coefficients % per 1000 Fathoms	
	Measured	Theory	Measured	Theory
$\frac{\Delta R}{R}$	0.20	0.202	0.05	0.05
$\frac{\Delta L}{L}$	0.026	0.021	−0.20	−0.22
$\frac{\Delta G}{G}$	1.5		−4	
$\frac{\Delta C}{C}$	−0.064	−0.060	0.56	0.68

Theoretical and measured values of pressure and temperature coefficients at 1 mc are compared in Table II.

The corresponding temperature and pressure coefficients of attenuation may be derived from the measured primary constant coefficients. With regard to temperature coefficients, $\Delta R/R$ accounts for over three-quarters of the effect on attenuation; with regard to pressure coefficients, $\Delta C/C$ accounts for over half of the effect on attenuation, with much of the balance due to $\Delta L/L$. The derived attenuation coefficients are given in Table III, as well as the coefficients observed in placing transoceanic systems. The experience with long systems indicated a temperature coefficient 6 per cent smaller than predicted and a pressure coefficient 8 per cent smaller than predicted.

The small conductance loss of the cable made the determination of its pressure and temperature coefficients less exact than desired. The tests did indicate, however, that the effective conductance at ocean bottom conditions would be less than at factory conditions. Experience with

TABLE III — TEMPERATURE AND PRESSURE COEFFICIENTS OF ATTENUATION

Attenuation	Temperature Cofficients % per °C		Pressure Coefficients % per 1000 Fathoms	
	Derived from Primary Constants	System Experience	Derived from Primary Constants	System Experience
$\frac{\Delta\alpha}{\alpha}$	0.17	0.16	0.38	0.35

actual ocean systems has confirmed this, observed decreases in conductance being as much as 50 per cent larger than these coefficients would indicate.

It was also observed in the measurements on short lengths that the effective dissipation factor of the cable increased with increasing frequency, from about 0.00012 at 0.1 mc to 0.00014 at 1.0 mc. When 6-foot samples were carefully dried in a vacuum, the dissipation factor decreased and was essentially constant with frequency. One theory advanced to explain the behavior was that a small gap existed between the dielectric and the outer conductor, and that in this space a conducting film formed. Such a film, being in series with the coaxial capacitance, would cause an apparent variation in dissipation factor with frequency. Either removing the film by thorough drying or bringing the outer conductor into intimate contact with the dielectric would reduce the conductance and therefore the attenuation.

5.2 *Measurements on Intermediate Lengths*

To reduce the uncertainties of making connections, measurements were made on samples several hundred feet long where the attenuation at one mc might be at least a good fraction of a db. Two types of measurement were in this category.

Chronologically, the first of these was a resonant-type measurement made on samples from 1000 to 3000 feet in length. The measurement was made using a symmetrical bridge and consisted of measuring the input impedance of a short- or open-circuited sample. The reactive component was eliminated by adjusting frequency and the real component was balanced using deposited carbon resistors mounted on special plugs. The measurement showed an excellent agreement between theoretical and measured phase delay (see Fig. 11).

The second type of measurement of intermediate-length samples was an aging test to determine the stability of the attenuation of the cable.* In this test 600-foot lengths of cable were subjected to a simulated laying cycle, after which they were maintained at a typical sea-bottom pressure and temperature. The attenuation was monitored for any change. The facility was arranged so that the cable could be tensioned and water pressure applied simultaneously to simulate the pressure-tension cycle of a laying operation. The attenuation was

* This test was the primary responsibility of T. Slonczewski, then of Bell Telephone Laboratories.

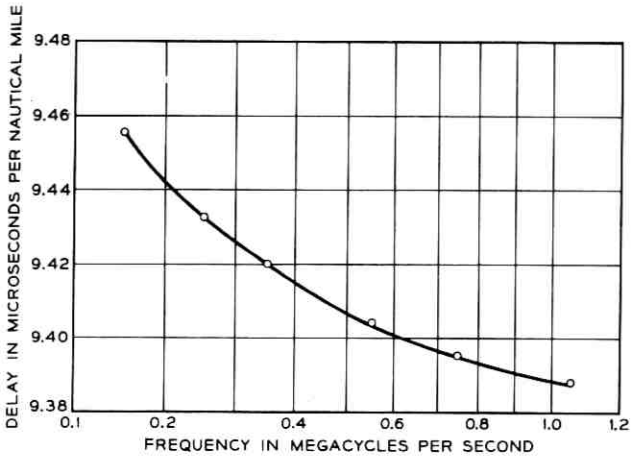


Fig. 11 — Measured phase delay (open circles) compared to theoretical (curve).

measured on an absolute basis to an accuracy of ± 0.00002 db, temperature maintained at 3°C and measured to ± 0.01°C and pressure was maintained at 5000 psi and measured to ± 10 psi. The over-all accuracy of the measurement facility permitted detecting a change of 0.02 per cent at one megacycle. The test has been in progress for more than three years and no change has been observed.

The question of possible aging was so critical that shore-controlled equalizers were then being developed. With the favorable results from the aging tests, it was concluded that shore-controlled equalizers were not necessary.

5.3 *Sea Trials*

The final experiment of the design portion of the development program was to fabricate 35 miles of cable in five-mile lengths. Three of these were placed and measured on sea bottom — one each at approximately 500, 1500, and 3000 fathoms.

The cable was placed by the cable ship *John W. Mackay*. The procedure in placing the sea-trial cable was to first establish stable laying conditions by paying out wire rope and scrap cable. This was followed by the test sample, a repeater housing, and a test lead coming on board ship (see Fig. 12).

The test method utilized echo techniques in that a long single-frequency pulse was reflected from open-circuit terminations. The sea end of the five-mile test sample was open circuited. In the repeater housing

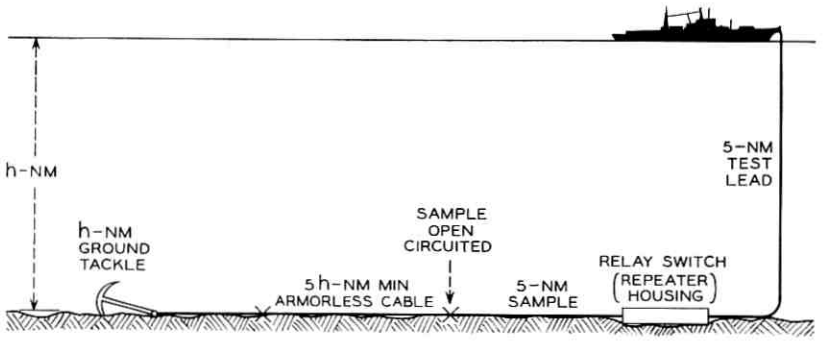


Fig. 12 — Cable arrangement for sea trial.

between the test sample and the test lead, switching was provided to either connect the sample to the test lead or to open circuit the test lead. By alternately measuring the test lead and the test lead plus sample, the attenuation of the sample was determined.

The three lengths as measured at the pilot plant had losses at 1 mc that averaged $\frac{1}{2}$ per cent less than had been predicted. The attenuation at 1 mc after placing in the ocean averaged $\frac{1}{4}$ per cent greater than had been predicted (see Fig. 13). This pattern of behavior is reasonably characteristic of all the cable that has been manufactured and placed.

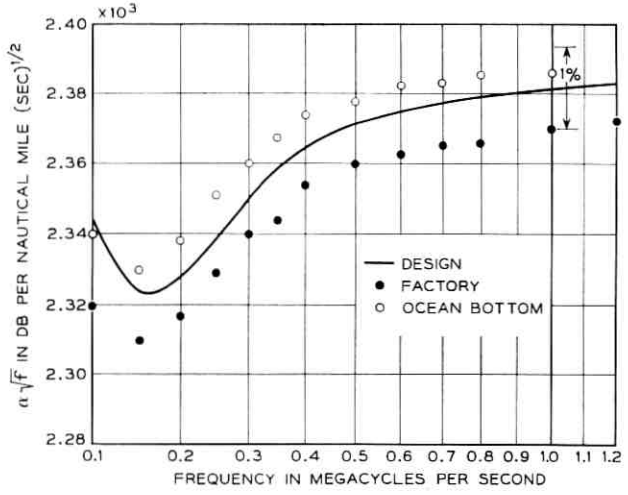


Fig. 13 — Attenuation of sea-trial cable design, as manufactured and on ocean bottom.

The predictions of the cable attenuation at atmospheric pressure were based upon intimate contact of the outer conductor with the dielectric. Such a situation certainly obtains at sea-bottom pressure. However, under factory conditions, in spite of the hoop stresses of the outer jacket, it is conceivable that a minute air space exists between the copper and the dielectric, particularly in the vicinity of the overlap seam. Therefore, it is believed that the $\frac{1}{4}$ per cent difference at sea-bottom conditions is the true prediction error and that the $\frac{1}{2}$ per cent difference in the factory is due to uncertainties in the precise dimensions of the outer conductor and the concomitant effects on conductance mentioned above.

5.4 *Laboratory Mechanical Tests*

The calculated mechanical properties of the cable were confirmed experimentally. In the case of tension, torque, bending, and reasonable combinations of these, the cable was tested until one of its conducting or protective components failed.

One of the mechanical properties studied was the torque-tension characteristic, which can be approximated by

$$\text{twist/unit length} \equiv \theta = \alpha T + \beta M$$

where T is the tension in pounds and M is the applied external torque in inch-pounds. The constant α is in radians per foot per pound of tension; the constant β is in radians per foot per inch-pound of torque. Both α and β are functions of cable geometry and material properties, some of which are nonlinear. In addition, they are combined in such a way that the resulting structure is nonisotropic as well as nonlinear. By making certain simplifying assumptions, it was estimated that α and β would have values of 2×10^{-6} and 3.3×10^{-4}, respectively.

Values of α and β were measured on 100-foot lengths of cable hanging from a tower. To the bottom end of the cable was fastened a large container that could be filled with water. Tension was applied free of torsional restraints or, alternatively, tension was applied and the torque measured that was required to prevent twist. The resulting values of α and β were 1.2×10^{-6} and 2.5×10^{-4}, respectively.

VI. CABLE JOINING

In order to transfer the tension between two lengths of cable or join a length of cable to a repeater, it is necessary to make a high-strength

splice to the steel strand. This requires that all of the polyethylene and cable layers be removed and the steel strand carefully cleaned. A copper-plated steel ferrule is then swaged onto the strand to bridge the strength (and conductivity) to a repeater housing or to another length of cable. Polyethylene layers are restored by molding, and the outer conductor is restored by brazing in new material.

Cable repairs are made so that no external bulge exists and therefore there is no problem in coiling or handling repaired cable. The use of the steel ferrule gives an electrical impedance discontinuity, but this is tolerable so long as splices are infrequent and at random locations.

VII. SUMMARY

The objective of the development program was to design an ocean cable having lower loss than previous cables. The cable program was concurrent with a program to develop a new repeater. Thus it was necessary to predict the attenuation-frequency characteristics of the cable as it would be on the ocean bottom after all changes due to factory and shipboard handling, the placing operation, and ocean-bottom pressure and temperature had taken effect. It was also required that this attenuation characteristic should not change with time.

By now, experience with armorless cable systems placed between Florida and Jamaica, Jamaica and Panama, plus systems from the United States to England and from Hawaii to Guam, indicates that the attenuation has been predicted to better than 1 per cent over the frequency spectrum of interest. This achievement has permitted the use of relatively simple and inexpensive procedures to maintain system alignment. Furthermore, in the first year of experience there has been no detectable aging in the transmission characteristic. Thus the most important goals of predictability and stability appear to have been attained.

VIII. ACKNOWLEDGMENTS

The authors wish to acknowledge the many contributions made by the members of the staff of the Simplex Wire and Cable Company and the Bell Laboratories groups involved in the cable design and development.

REFERENCES

1. Snoke, L. R., Resistance of Organic Materials and Cable Structures to Marine Biological Attack, B.S.T.J., **36**, September, 1957, p. 1095.

2. Lebert, A. W., Fischer, H. B., and Biskeborn, M. C., Cable Design and Manufacture for the Transatlantic Submarine Cable System, B.S.T.J., **36,** January, 1957, p. 189.
3. Lerch, B. W., and Phelps, J. W., Armorless Cable Manufacture, B.S.T.J., this issue, p. 1209.
4. Ehrbar, R. D., Kelley, R. A., Morris, L. H., Mottram, E. T., and Rounds, P. W., The SD Submarine Cable System, B.S.T.J., this issue, p. 1155.
5. Schelkunoff, S. A., Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields, B.S.T.J., **13,** October, 1934, p. 532.
6. Brewer, S. T., Dickinson, F. R., and von Roesgen, C. A., Repeaters and Equalizers for the SD Submarine Cable System, B.S.T.J., this issue, p. 1243.
7. Phelps, J. W., and Riley, R. M., Laboratory for Ocean Cable, Bell Laboratories Record, **40,** March, 1962, p. 89.

Armorless Cable Manufacture

By B. W. LERCH and J. W. PHELPS

(Manuscript received April 17, 1964)

The major portion of the cable used in an SD system is of an armorless coaxial design, with a strength member within the inner conductor structure. It is fabricated in continuous 20-nautical mile repeater sections and is stored, when finished, in individual pans. Every section is therefore available in any sequence to facilitate the most nearly ideal ship loading schedule.

This paper summarizes the process experience gained during the cable design and development program, details the processes involved in the purchase of cable raw material, discusses the manufacturing operation, and includes a measure of cable reproducibility for the Baltimore plant of the Western Electric Company.

I. MANUFACTURING BACKGROUND

It had been the intent, during the design and development program for cable for an SD system, that the specification defining and covering the cable would be written in terms of end product requirements, insofar as the necessary guarantees of long and stable cable life would permit. For this reason, laboratory work on the cable design, including the production on a semi-pilot plant basis of perhaps 50 or 60 nautical miles (nm) of cable samples, was conducted with primary attention placed on the product and not on the process or machinery. It was deemed sufficient to demonstrate that cable of adequate reproducibility and stability could be produced on at least one combination of cable making machines.

It was apparent, however, that an appreciable amount of factory planning time might be saved and some duplication of effort eliminated if a record of the experiences obtained in the cable laboratory was made available to prospective cable manufacturers. The intent was not to make process a specification requirement but merely to outline diagrammatically the processes used in the experiments; to list certain types of machines known to have adequate capacity and reliability; to delineate the need for machinery for which there was no prototype; to show ap-

proximately the lengths of the several manufacturing lines needed to achieve realistic manufacturing speeds; and to estimate adequate material and in-process storage areas.

There had been, additionally, several pieces of test equipment devised specifically for the measurement, recording and control of certain process parameters. It was not thought necessary to specify the use of these pieces of test equipment, since other devices could be found that would work adequately well. However, it was felt that information on the experience with these devices should also be made available.

As a result of the above, a study was made of an idealized ocean cable factory layout by an outside engineering organization in consultation with Bell Laboratories employees who were familiar with the operation of the cable laboratory. The report emanating from this organization gave a summary of laboratory experience and a point of departure for factory design and equipment purchase and installation programs.

A second, and equally important, by-product of the operation of the cable laboratory was the opportunity for appraisal of preliminary specifications for the several materials used in cable fabrication and a study of the changes wrought in material properties due to processing. Initial operation of the laboratory was accomplished through use of standard materials, with only minor emphasis on the level and dispersion of the critical material parameters. As cable samples were produced with more finesse and more meaningful experience, more attention was paid to the procurement of material and to the properties of that material before and after processing. Toward the end of the operation, the laboratory materials were purchased under tentative issues of material specifications. Through comparison of the data obtained from the basic raw material, from cable produced from this material, and from the basic computational analysis of cable parameters, it was possible to assess changes that occur to the material in process and to improve upon the predicted cable characteristics for matching to the amplifier and equalizer designs.

The overriding requirement for successful operation of an ocean cable system is continuous, uninterrupted life. Experience in the cable laboratory demonstrated that cable making machinery must be of such quality and stability that continuous operation can be achieved in the production of a 20-nm length of cable (a repeater section) to assure that the structure will be strong and stable and without discontinuities. Obviously this implies the use of quite massive machinery throughout and very precise coordinated drive systems for the several pieces of powered equipment in each of the production lines. In addition, the usual methods of product

inspection, such as quality control involving sampling for attribute averages, cannot be used. Instead, measurement of each of the critical parameters in cable processing must be done on a continuous basis, and the results obtained must be put into a continuous permanent record. In many cases a strip chart recorder is used.

The Baltimore plant of the Western Electric Company* was selected for production of armorless cable in March, 1961, and factory planning began immediately. Ground was broken for plant construction in August, 1961, and the first cable footage was made in April, 1962. Manufacturing capacity is 5000 nm per year.

The factory building is a windowless, steel-sheathed structure with building columns and heavy machine loads set over approximately 1000 piles. The 170,000 square feet of floor space comprises four basic areas. Inner conductor, dielectric, and jacket operations are all done in an atmosphere of filtered air maintained at a slight positive pressure, and the concrete floor is non-dusting. The temperature-controlled room for dielectric sizing and repair and outer conductor application is maintained at 72°F and at a dust count level of less than 6000 parts per million. The floor is tiled. The terminating and test rooms are temperature and humidity controlled, the floor is tiled and the walls are vinyl coated. The finished cable storage area is arranged for the stacking of pans four high with a floor load under each stack of 256 tons, and has minimal light and heat requirements.

An access channel has been dredged in the waterway adjacent to the plant to a depth of 34 feet, sufficient to accommodate any existing fully loaded cable ship.

The cable factory and loading facilities are shown in Fig. 1, along with C. S. *Long Lines*.

II. RAW MATERIAL PROCUREMENT AND INSPECTION CONTROL

There are five materials used in the structure of the armorless cable. They are:

(1) steel wire
(2) inner conductor copper strip
(3) polyethylene dielectric
(4) outer conductor copper strip
(5) ethylene plastic jacket.

---

* Cable to the same specification, though not necessarily by an identical process, has also been made by Standard Telephones and Cables, Ltd., Southampton, Hampshire, England, and by Ocean Cable Co., Ltd., Yokahama, Japan.

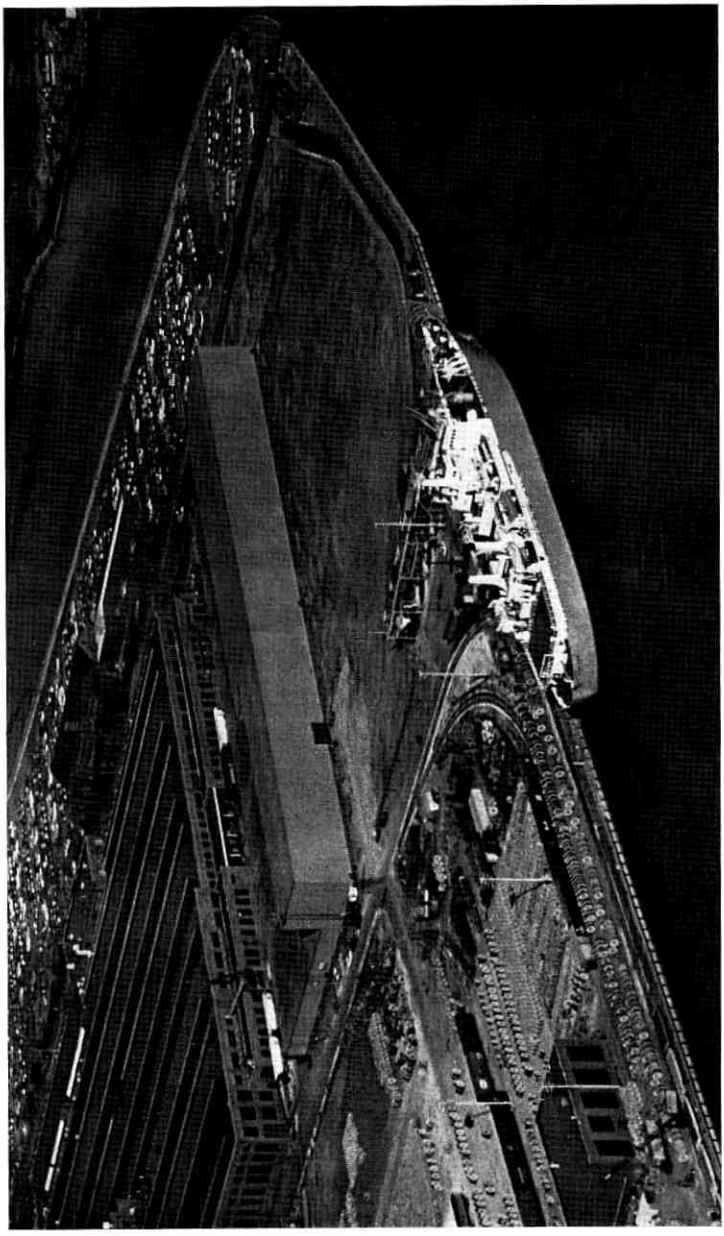


Fig. 1 — Cable factory and loading facilities, with C.S. *Long Lines*.

2.1 *Steel Wire*

The strength member of the cable consists of forty-one high tensile strength, medium to high carbon steel wires of five different diameters, 0.069 inch, 0.047 inch, 0.041 inch, 0.039 inch, and 0.030 inch, each having a tolerance of ±0.0005 inch. The tensile strength tolerance for the 0.069 inch wire is 280,000 psi to 320,000 psi; for the other sizes it is 290,000 psi to 330,000 psi. The supplier conducts 100 per cent inspection for wire diameter and tensile strength. After the wire is received at the cable factory, Western Electric inspectors randomly select and check twenty samples for diameter and ten samples for tensile strength from normal lots of fifty reels.

To achieve a firm, well fitting pattern of strand wires, it is necessary not only to hold tight tolerances on individual wires, but also to specify that the diameters of a lot be uniformly distributed about the nominal diameter. Fig. 2, showing the effect of die wear on a 0.069-inch wire, illustrates that a 4000-pound lot would have diameters individually within tolerance limits, but with an average diameter appreciably greater than nominal. This is not tolerable and accordingly it was specified that, in any given lot inspected, the diameters of twenty randomly selected samples had to meet the following additional tolerance:

$$x = \frac{\sum_{1}^{20} (d_x)}{20} = d \pm 0.0001$$

where d = nominal wire diameter and
d_x = measured diameter.

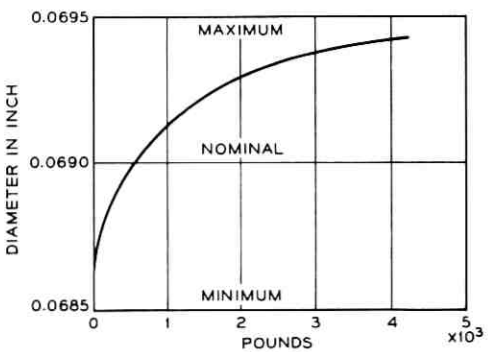


Fig. 2 — Wire diameter changes due to die wear.

It is not necessary to make a similar type of analysis for tensile strength. If all measured samples fall within the tensile strength tolerances, the lot is acceptable.

2.2 *Inner Conductor Copper Strip*

This material is rolled from oxygen-free copper cake having a conductivity of 101.4 per cent ± 0.3 per cent I.A.C.S., measured on a 0.0808-inch wire drawn from the cake at the copper refinery. The width and thickness dimensions on finished strip are 1.600 inches ± 0.005 inch and 0.023 inch ± 0.0003 inch, respectively. In order to assure that the finished lot of material is uniformly distributed about the nominal dimension, a sample size of twenty is selected, and charts of grand averages and deviations from the nominal are generated. The grand average is computed from the following formula:

$$\bar{x} = \frac{\sum_{1}^{20} (x)}{20}$$

where $\bar{x}$ = grand average and x = measured dimensions.
The deviation from the nominal is then computed as follows:

$$\sigma = \left[\frac{1}{19}\sum_{1}^{20} (x - \bar{x})^2\right]^{\frac{1}{2}}.$$

Typical charts are shown in Fig. 3.

2.3 *Polyethylene Dielectric*

Polyethylene is delivered to the factory in 100,000-pound capacity hopper cars to minimize the number of times the material is exposed to contamination. A representative sample of material is obtained as the hopper car is being loaded at the supplier's plant. This sample is molded into sheets and extruded into tapes for inspection of electrical, mechanical, chemical and cleanliness properties. The most stringent inspections are for dielectric constant, dissipation factor and contamination. Failure to meet any one of the requirements is cause for rejection of the entire hopper car section.

2.4 *Outer Conductor Copper Strip*

This material is electrolytic tough-pitch copper having a conductivity of 101.2 per cent ± 0.3 per cent I.A.C.S., determined by conductivity tests at the copper refinery. The width and thickness requirements on

processed strip are 3.430 inches ± 0.005 inch and 0.010 inch ± 0.0002 inch, respectively. Deviation charts similar to those in Fig. 3 are designed around these dimensions.

2.5 *Ethylene Plastic Jacket*

The jacket material is also received in the factory in 100,000-pound capacity hopper cars. It does not have to meet the same electrical and cleanliness requirements as the polyethylene insulation. However, quite stringent requirements are placed on the material physical properties to assure that it will do an adequate job of protecting the cable outer conductor under conditions of environmental stress.

III. CABLE FABRICATION AND STORAGE

Fabrication of ocean cable for SD system use is a four-step process. First, and most difficult, is fabrication of the inner conductor. This is

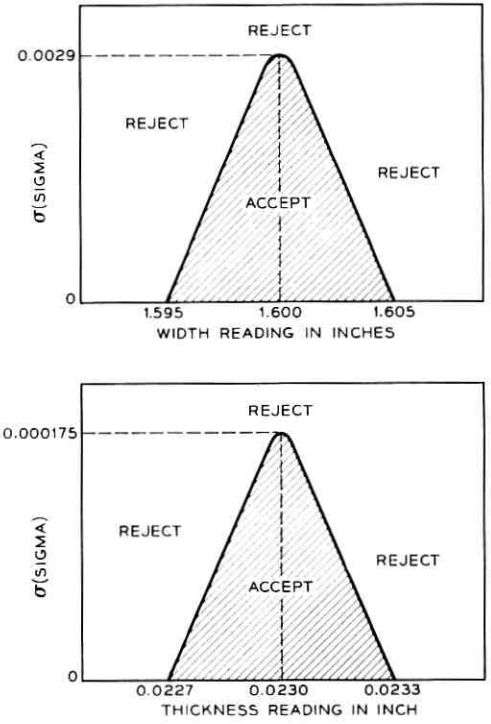


Fig. 3 — Copper strip quality control — acceptance ranges for width and thickness variations.

followed by the extrusion of polyethylene dielectric. As it is not possible in the present state of the art to control dielectric diameter variations to the very small tolerance dictated by transmission requirements, it is necessary to follow extrusion with a core sizing operation. At this point it is possible also to improve upon the as-extruded core concentricity, if necessary. The fourth and final operation in cable fabrication is the tandem application of outer conductor and jacket. In some sections a fifth operation, dielectric repair, is necessary.

Completed sections of cable are used in groups approximately 200 nm in length, called "ocean blocks." As each section made varies slightly from the nominal design characteristic of attenuation, it was concluded that more uniform blocks could be assembled if sections were selected from the total lot available rather than taken in order of manufacture. For this reason, each completed section is taken up and stored in an individual pan capable of holding 20 nm of cable with both ends accessible. Any cable ship loading sequence can therefore be specified and followed.

3.1 *Formation of Inner Conductor*

The inner conductor (see Fig. 4) is a composite copper-jacketed steel wire rope. The process by which it is made is shown diagrammatically in Fig. 5. Steel wire is purchased in nominal lengths of 124,000 feet wound on the bobbins used in the tubular strander. Wire is drawn from the 41 bobbins in the strander and formed into the strand pattern in a

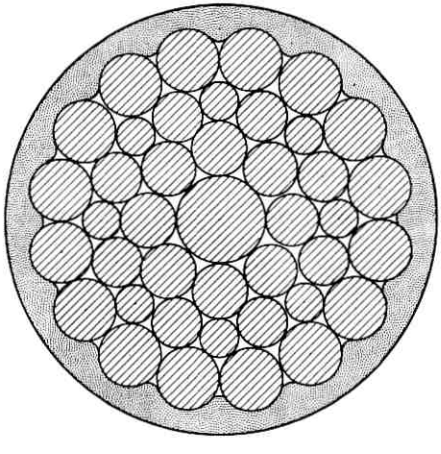


Fig. 4 — Cross section of finished conductor.

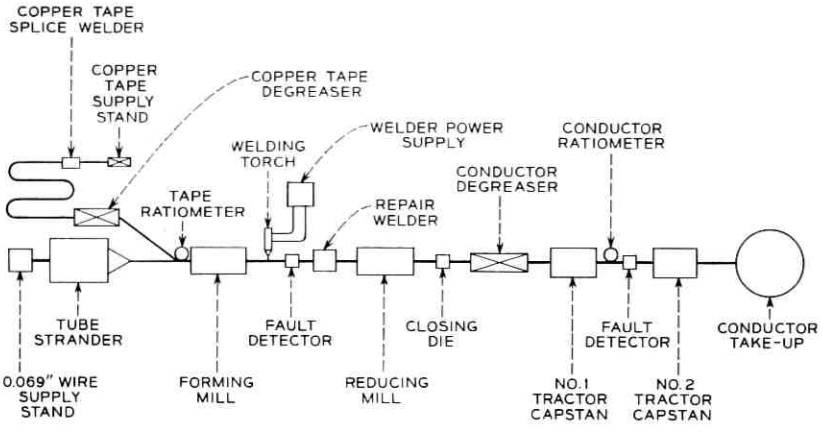


Fig. 5 — Inner conductor line.

special strander closing die. In a parallel operation, a strip of oxygen-free copper 1.6 inches wide by 0.023 inch thick is fed from the pay-off stand into an accumulator, then to a vapor degreaser, and then into a conventional tube forming mill in which certain rolls are grooved to straddle the strand. The edges of the copper strip are sheared to 1.533 inches to form a tube of precise diameter with clean abutting edges for welding. The strip is then formed around the steel strand and positioned so that the strand rests in the bottom of the tube, as shown in Fig. 6, to keep the steel strand remote from the heat of the welding arc and so prevent degradation of its tensile properties.

The butted edges of the formed copper tube are welded in a continuous seam, using tungsten inert gas arc welding techniques. The welded tube is then reduced in successive steps with reducing rolls until the inner diameter is a close fit over the steel wire strand. The copper tube and steel strand are then drawn through a die. This results in a tightly compacted structure with close control on conductor diameter and with copper forced into the interstitial strand spaces. This assures that the inner conductor will be structurally firm under the forces of ocean-bottom pressures. The operation of the copper die is as much an extrusion operation as a drawing operation because of the pressures involved in forcing copper into the interstices.

The copper exerts a tight grip on the steel strand, and it is therefore necessary that the strand be formed with essentially no differences in the lengths or tightness of the 41 wires. Bobbins of steel wire are selected prior to loading the machine with wire sizes that minimize the

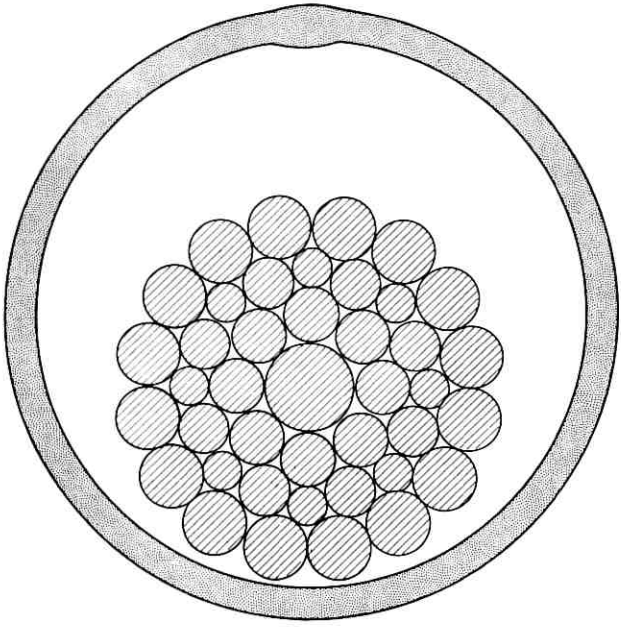


Fig. 6 — Steel strand and welded tube.

strand diameter deviation from nominal. These bobbins are then positioned in the strander to make a most nearly round and compact strand. Steel wire size and position control, along with a special strand closing die, assures that the strand, as it is formed, will have no incremental differences in length in the several wires that would subsequently be squeezed back at the copper closing die position. The special die is shown in Fig. 7.

Alternative methods of tube welding had been tried in the laboratory, but tungsten inert gas welding was shown to be superior at that time. Four items need to be controlled to have a long-life welding operation:

(1) The inert shielding gas supply must be very pure and dry and be uninterrupted. Gas impurity affects the tungsten electrode, promoting excessive erosion and the formation of "whiskers" on the electrode tip.

(2) The surfaces of the copper strip must be free from deposits of oil, water or other contaminants, as the volatilization of these materials results in contamination of the emitting electrode tip and increases the level of porosity in the welded seam.

(3) The butting edges of the tube must be freshly slit, level and parallel, and under a slight amount of positive pressure when passing

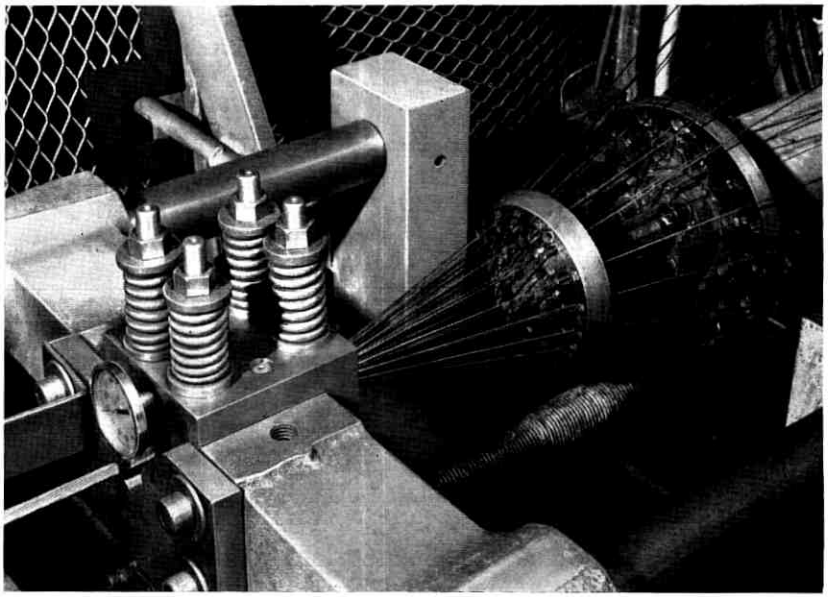


Fig. 7 — Strander point and closing die.

under the welding arc. To maintain seam alignment, the tube must be in slight tension in the welding rolls and the tube forming and reducing rolls must be in precise alignment to eliminate tendencies to roll the tube.

(4) The welder power supply must have ample capacity for 100 per cent duty and control must be precise to track tube speed with weld current. It must be stable over long periods of operation and be capable of repeating current values at particular settings corresponding to particular line speeds.

Successful operation for long periods of time is partly a matter of tube forming and seam alignment and partly a function of precise control of weld current, electrode position, and electrode tip configuration. By selecting a mixture of helium and argon gases, a maximum weld penetration-to-width ratio with minimum electrode tip erosion is obtained, making possible welding runs of from 60,000 to 124,000 feet, depending upon the purity of the shielding gases. With careful attention to the tube forming process and with automatic arc length control and precise speed relationship of all the machines in the system, the welding operation need be interrupted only when defects in the copper strip occur or a steel wire breaks.

Experimentation has shown that in a sufficiently clean shielding gas atmosphere and with a sufficiently clean metal, variation in the composition of the electrode has little effect on electrode life. Electrodes are normally made of sintered tungsten with 1 or 2 per cent of thorium added. The 2 per cent thorium content electrodes do not last as long as those of 1 per cent, but their characteristics enable an arc to be more easily established.

Oxygen-free copper was selected early in the development program as the material to be used in the inner conductor. The principal reason for its selection is that in the oxygen-free condition there are no oxide or gas pockets formed under the heat of the welding torch to weaken the tube structure. Additionally, oxygen-free coppers can be obtained with high levels of ductility and conductivity. These are important, as there is an appreciable amount of work hardening of the copper during the reduction and die drawing operations.

Inner conductor copper is purchased in coils of strip as large as possible, weighing approximately 350 pounds per coil. The transition from coil to coil is made without stopping the line by storing approximately 400 feet of strip in the accumulator, thereby providing sufficient time to substitute a new coil for the exhausted one. The copper strip is spliced by joining ends which are sheared at an angle of 20° from the transverse, overlapped $\frac{1}{32}$ inch, and welded with a tungsten inert gas arc. The weld penetrates through both sections, fusing them together and leaving a rounded section somewhat thicker than the original strip. As this section will not form properly, nor pass through the reducing die, the weld is peened to uniform thickness. The peened section is then annealed to recrystallize the grain structure. This operation produces a joint very much like the original strip in dimensions and metallurgy.

Both of the welding operations in the inner conductor line — i.e., tube welding and coil-to-coil welding — are in some respects self-checking processes. The reducing and die-sinking operations will, in most cases, rupture the copper tube either longitudinally or circumferentially if the weld is not of the proper size or density. In order to assure complete success in these operations, certain qualifications of operators and equipment are necessary. Periodic checks of coil-to-coil welds are made with selected samples that are subjected to tensile and elongation tests. Checks are made of the tube welding process at the beginning and end of each manufactured length. These involve samples of unreduced tube that are subjected to a flare test using a 60° included angle steel cone. Properly welded tube will take elongations of 35 per cent.

It is necessary that the copper jacket of the inner conductor be continuous and fault-free throughout its length to assure isolation of the polyethylene dielectric from the steel strand. Consequently no holes, slits, or unwelded seam portions are allowed in the copper jacket. By arrangement of the equipment in the inner conductor line, space has been provided, between the tube welding area and the first pass of the reducing mill, in which partially welded or unwelded tube can be rewelded while the line is moving. If a hole or a slit not on the welded seam occurs, the line must be stopped after the fault has passed through the die and a strip of 0.005-inch copper brazed over the fault under controlled conditions. The repair, smoothed and polished, will withstand polyethylene extrusion temperatures and subsequent bending stresses.

A testing device, shown schematically in Fig. 8, was developed at Bell Laboratories to detect faults in the welded tube. A pair of coils surrounding the tube are mounted approximately 2 inches apart. These coils form two sides of a very sensitive bridge circuit. In effect, the welded copper tube is a shorted secondary winding to each of these coils, and changes in the tube unbalance the bridge circuit. By phase comparison

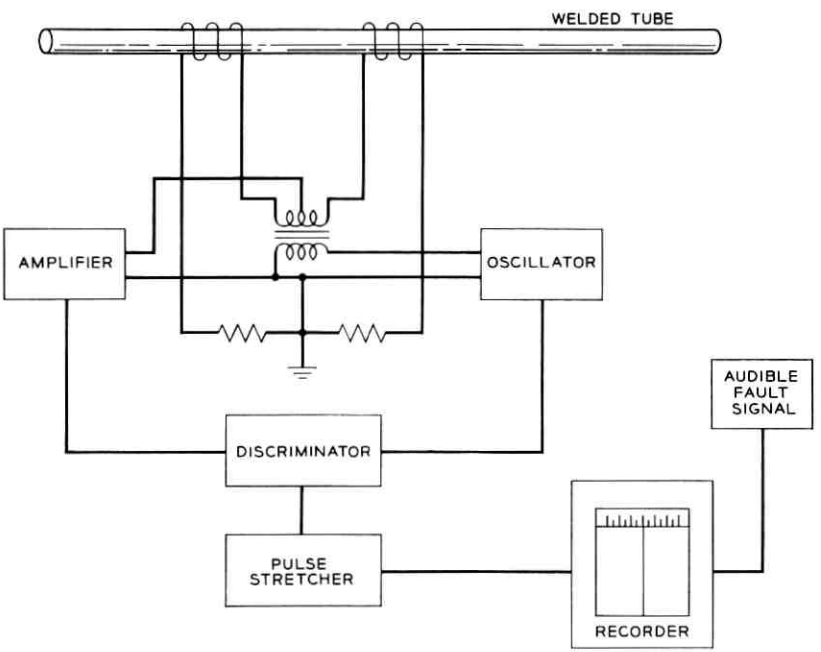


Fig. 8 — Weld integrity test set.

of the input and output bridge signals, it is possible to ignore changes in copper conductivity, wall thickness and tube surface smoothness and to detect faults in the tube. This allows the operator of the inner conductor line to survey over-all operations without paying constant attention to the weld seam, and thereby reduces operator fatigue while certifying the soundness of the welded tube.

The copper strip, formed in an oversized tube, can be formed most readily and handled safely in the reduction process if the tube wall is approximately 0.023 inch thick prior to reduction. It is apparent then that the cross-sectional area of the tube prior to reduction is appreciably greater than that required in the finished conductor, and that the relative speeds of the steel strand and the copper are not equal until they join at the copper sinking die. This difference in speed is used to determine the effective wall thickness of the copper in the finished product. The ratio of the speeds of the two elements is determined by monitoring the copper strip and the finished conductor with two photo tachometers. The differential count on the strip tachometer is compared to a thousand counts supplied by the conductor tachometer and is "read-out" on a decade counter display. The system resets itself and repeats periodically. The ratio is varied as the conductor diameter and tape thickness vary to maintain the required effective wall thickness.

3.2 *Extrusion of Dielectric*

Extrusion of the polyethylene dielectric material on SD system cable is much more difficult than conventional insulation or wire jacketing processes in that the ratio of extrudate wall thickness to conductor diameter is appreciably greater. The cable dielectric must remain for a long period of time with a constant potential gradient applied, and voids in the material might initiate corona discharges that would introduce noise into the system. The prevention of voids in the polyethylene is achieved through careful control of the heat extracted from the material as a function of time. The outer fibers of the material must not be allowed to solidify in too great a depth while they are at too large a diameter and while the inner fibers are still in a plastic state. If this occurs, the dimensional contraction of the material upon cooling must extend radially outward and, hence, pull the dielectric away from the conductor in the center. The rate then which can be approached, but not surpassed, is one in which the dielectric heat is extracted radially outward such that the entire mass approaches the crystalline phase state with small temperature differential between the inner and outer layers.

In order to extract heat from the dielectric at a maximum safe rate and thereby achieve high production speed, a trough system with graduated water temperatures is used. In essence, the hot extrudate is run first into water close to the boiling point. It is kept in this medium until the average temperature of the material and the temperature differential between inner and outer surfaces has fallen to a point where the rate is somewhat lower than optimum. The material is then passed into a trough with a water temperature cooler than the first trough. The temperature difference between the first and second troughs is obviously a function of the state of the extrudate at the time of transition, and can be greater if the extrudate has approached the water temperature in the first trough. This scheme of successive changes in cooling rate can be repeated any number of times until the entire mass of polyethylene material passes through the crystalline phase change. From this point onward, the material is solid; it changes dimension as a unit mass and may therefore be led into cold water so that it may be handled around sheaves without damage. A schematic layout of the dielectric extrusion line is shown in Fig. 9.

The polyethylene used as the dielectric is an electrical-grade material with a very low (0.3 maximum) melt flow index and is characterized by a very narrow spread of dielectric properties and by extreme purity. As the end product desired is a transmission line whose electrical properties are essentially invariant throughout its entire length, and as the raw material dielectric properties are ever so slightly altered by the extrusion process, it is necessary that variations in extrusion conditions be minimized through control of the entire extrusion and cooling process. This is accomplished partly by the design of the extruder and the extruder screw and partly through automatic control of temperatures in each of the extruder zones, in the extruder die and in each of the trough zones. Typical temperature changes in the troughs, for instance, are $\pm 3°F$.

Control of the extrudate dielectric properties is also maintained by rigorous attention to possible sources of material contamination. The resin, as produced by the polyethylene supplier, is kept within a closed system from the time of packaging until it is extruded on the inner conductor. It is shipped from the supplier's plant in nominal 100,000-pound quantities in railroad hopper cars. These cars discharge into outlets at the bottom of each of the three car sections. Compressed filtered air is used to blow the material from the hopper car to an in-building storage bin and from the bin to the extruder hopper. The duct system is aluminum throughout. Classifiers and separators are located in the pipelines just above the extruder hoppers to remove

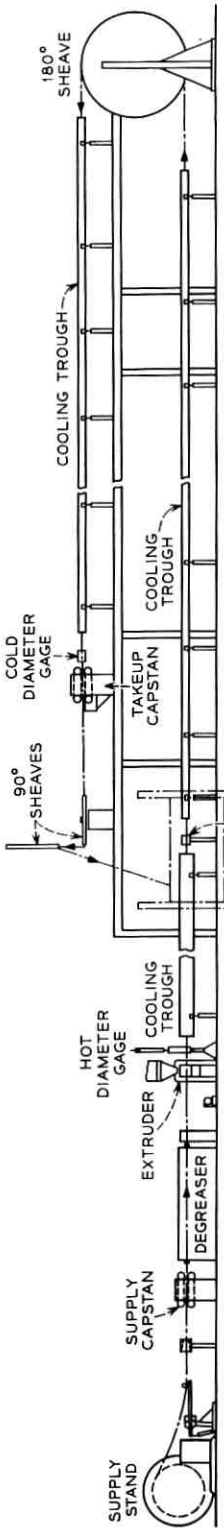


Fig. 9 — Elevation of ocean cable dielectric line.

smaller than normal polyethylene granules and dust (called fines) generated during conveying.

The sizing operation that follows extrusion is somewhat sensitive to changes in the depth of material to be removed and is therefore more accurate if tight control is held on the as-extruded diameter. The hot material is measured between the extruder die and the input to the first trough section by a gauge having a swept light beam that travels across the extrudate. Indications from this gauge permit the operator to make corrections in core size by manually adjusting the extruder screw speed.

It is essential also that the dielectric be extruded as concentrically as possible. Concentricity is normally maintained within 0.007 inch and cannot exceed 0.020 inch without jeopardizing the sizing operation. Eccentricity is minimized in the extrusion process by manually adjusting the relative positions of the extruder tooling according to information received from an eccentricity gauge. This gauge, operating on the principle of capacitance unbalance as measured between the grounded inner conductor and pairs of electrodes contacting the dielectric, is placed between the first and second trough zones, where the material has solidified sufficiently to allow a slight pressure on the surface. Adjustments to improve concentricity are not required very often, as the drive systems coordinating the several elements of the line will remain constant for long periods.

3.3 *Dielectric Repair*

Among the several ways in which the extrudate may be damaged while in a heated, softened condition are line stoppage due to power failure and machinery malfunction. The most common type of damage is caused by imperfections in the inner conductor resulting from an imperfect weld or poorly rolled copper strip. As the inner conductor passes through the heated zone of the extruder head, air and other gases within the conductor expand. Gas escaping from any hole in the copper will swell the extrudate and form a permanent void pocket next to the conductor.

The principal mechanism for the detection of voids is the eccentricity gauge on the extrusion line, although the occurrence of most voids can be predicted by continuous surveillance of the inner conductor during production, both visually and with the conductor integrity tester. Voids that occur are repaired in an operation subsequent to core extrusion, at which time the inner conductor fault is also repaired.

Soldered sleeves, to provide a gas-tight seal over small copper cracks, are made by first tinning both the inner conductor and a presized 0.012-

inch thick copper sheet. The sheet is tightly formed around the inner conductor, with nearly butting edges, and heat is applied to fuse the bond. The applied temperature is held under 450°F to prevent annealing of the steel strand wires.

Copper-plated steel sleeves are inserted at points where the strand may have been damaged or where more than 2.0 inches of missing copper is encountered. They are approximately 6 inches long, with a 0.750-inch outer diameter before crimping. As such repairs are potential weak points in the cable strength member and are also undesirable from a signal echo standpoint, no more than two splice sleeves are permitted within any 20 nm length of cable. No splice sleeve is permitted within 3 nm of either end, and splices may be no closer together than 1.5 nm. Fig. 10 shows the sleeve application and steps in dielectric repair.

In order to transfer the full strength of the steel through the splice with a sleeve of reasonable length, it is necessary to prepare the strand wire ends so that each of the strand wires will contribute to tensile strength within the sleeve area. The bonding agent used to hold all of the strand wires together as a unit is an epoxy that is cured at a temperature of approximately 385°F. The strand wires are parted,

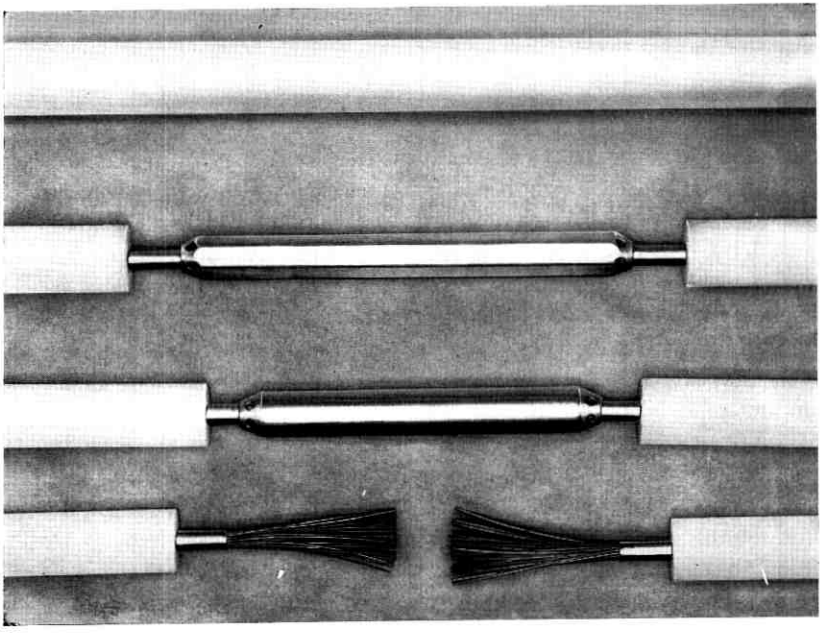


Fig. 10 — Steps in repair of inner conductor, and dielectric remolding.

cleaned in an ultrasonic bath of trichloroethylene, coated with epoxy and then re-laid in pattern. The two prepared strand ends are inserted into the sleeve and the sleeve is pressed with dies, hexagonal in shape, that produce a finished sleeve approximately 0.640 inch across the flat faces. The elongation of the sleeve after pressing is sufficient to capture the inner conductor copper at the ends. A heat fixture is used to cure the enclosed epoxy.

Restoration of the dielectric is done with an extrusion molding technique. A $1\frac{1}{4}$-inch vertical extruder, crosshead and die are used to convey a homogeneous melt of compound into the die cavity. Extruder barrel, head, and die inner and outer zone heats are individually temperature controlled. The crosshead is fitted with a bleeder plug so that compound can be continuously purged, thereby preventing oxidation of compound within the extruder barrel. Water-cooled dies up to 30 inches long for 15-inch repairs are used. After molding, the flash and sprues are trimmed, and the joint is X-rayed to check for unacceptable inclusions, voids, and eccentricity.

3.4 *Dielectric Sizing*

It is possible, through close attention to the extrusion process, to maintain a diameter variation about the nominal of approximately ± 0.005 inch. Such variation would be intolerable in this design, as the effects on attenuation of positive and negative diameter excursions are not self-cancelling, and it would therefore be impossible to achieve the computed cable characteristics. For this reason, it is necessary to put the extruded dielectric through a supplementary sizing operation.

The machine used, shown in Fig. 11, works like an oversized pencil sharpener. The polyethylene is cut with three tool steel cutter blades arranged symmetrically on a rotating head. Finished diameter is measured with a diameter gauge, and the measured diameter is displayed on a strip chart recorder. The sensitivity of the gauge is such that ± 0.0001 inch can easily be read; adjustments in diameter, when necessary, are made manually. The concentricity of the dielectric surrounding the inner conductor is also monitored, as shown in Fig. 12. Measurements are made of capacitance unbalance between two pairs of electrodes and the inner conductor. The electrode pairs, acting independently, are arranged in horizontal and vertical planes. The capacitance information is fed into amplifying and recording circuitry, where the unbalance is displayed on two pens on a strip chart recorder and is also fed into a two-channel servomechanism system. The servo signals actuate motors that cause

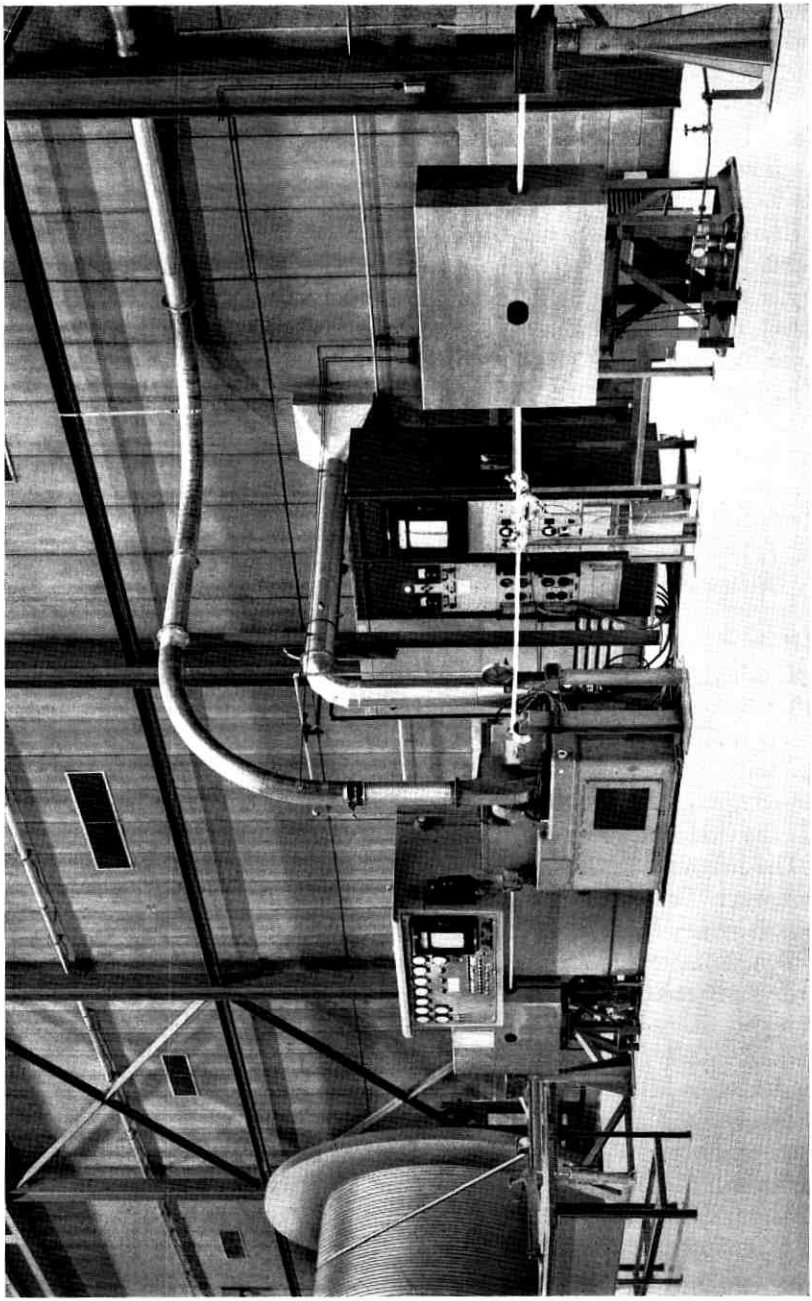


Fig. 11 — Dielectric sizing machine.

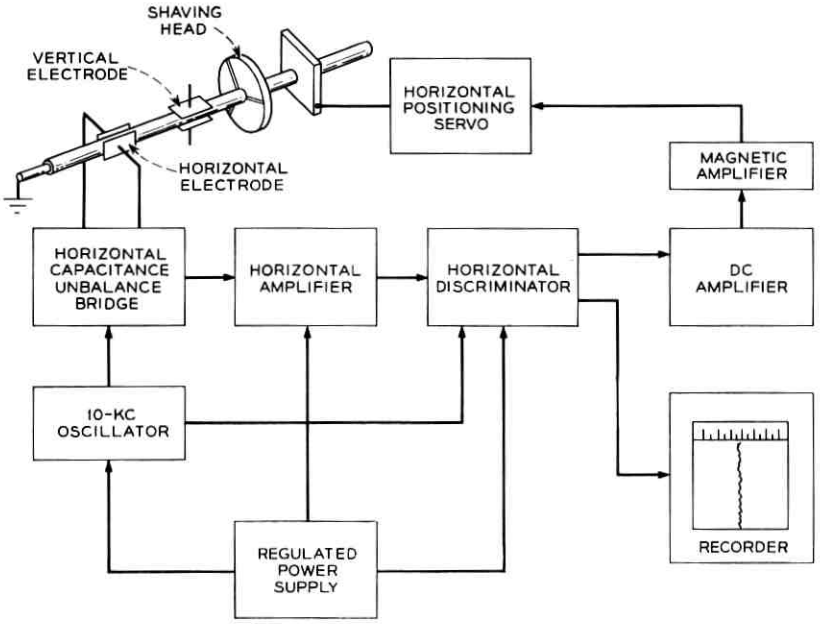


Fig. 12 — Shaver eccentricity control system — horizontal half.

the movement of pairs of positioning fingers bearing on the unsized material. The servo systems are zero-seeking devices and as such are insensitive to diameter variations.

The nominal diameter of the as-extruded dielectric is 1.050 inch. Experience has shown this to be the optimum depth of cut for the machine used. This also allows corrections in eccentricity up to 0.025 inch to be made without departing from a circular cross section.

3.5 *Application of Outer Conductor*

The outer conductor structure is ideal electrically. It is of high-conductivity material, is cylindrical in shape, and has an overlap in the longitudinal seam to reduce signal radiation to a tolerable order of magnitude. Mechanically, the outer conductor requires special treatment both in fabrication and in subsequent cable handling. It is not possible to apply the outer conductor and store the completed coaxial before jacketing, as the outer conductor would be wrinkled. Consequently, outer conductor forming and cable jacketing are done in a straight line, uninterrupted, tandem operation.

Outer conductor copper is purchased in coils weighing approximately 430 pounds. Coils of copper are added to the system as required, with an overlap braze connection between coil ends. The braze is made at an angle of 65 to 70 degrees with respect to the strip edge to distribute the double thickness of copper axially along the cable after forming. Forming is done in a special pull-through forming machine that applies the conductor strip with minimal stretch. The path of the strip and the sized core through the mill is not horizontal but follows a parabolic curve having the general formula $y = kx^a$, as shown in Fig. 13. This is done in an attempt to cause the center elements and edge elements of the strip to travel equal distances between the point at which the strip is first bent and the point where forming is completed.

A peculiarity of the forming process is inserted in the last set of tooling in the machine. The top sector of the roll, that portion which embraces the overlapped joint, is cut to provide a nonsymmetrical amount of relief to the overlapped edge as shown in Fig. 14. The heat and pressure of the jacket extrusion process are transmitted through the copper conductor, causing the underlapped edge to be pressed into the heat-softened dielectric. The amount and position of the indentation of the underlapped edge are controlled by the roll to minimize any tendency to produce an air gap in the structure.

Outer conductor forming and jacket extrusion are shown in Fig. 15.

3.6 *Extrusion of Jacket*

The jacket material is a high-density ethylene plastic requiring higher extrusion temperatures than those used with the dielectric material. The presence of carbon black in the material makes necessary the use of a hopper dryer system. This is because the material is highly hygroscopic

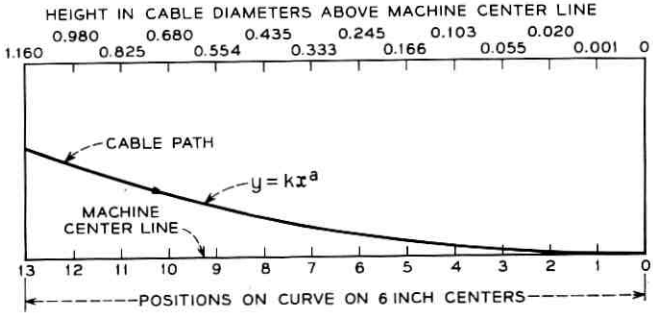


Fig. 13 — Path of cable through forming mill.

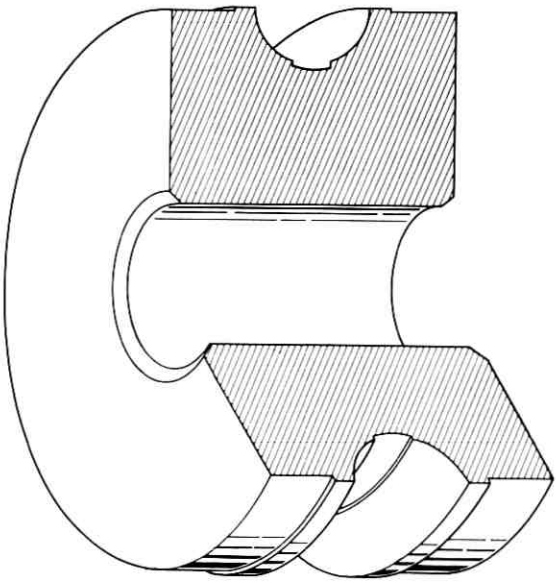


Fig. 14 — Last-pass forming roll.

and water in concentrations greater than 0.04 per cent weakens the extruded material to a degree sufficient to prevent the necessary outer conductor protection.

Relaxation times of the jacket extrudate are appreciably longer than those for the dielectric material. They typically run several minutes at the crystalline melt temperature, as compared to less than a second for the dielectric. It is not possible, therefore, to apply the jacket in the relaxed state typified by low measures of material retraction. It is possible, however, to achieve shrink-back levels sufficiently low to assure a service life expectancy in excess of 40 years.

The completed cable structure is shown in Fig. 1 of Ref. 2.

IV. ELECTRICAL TESTING

As the cable is jacketed in 20-nm lengths, it is coiled into a storage pan and subsequently immersed in a test tank of circulating water held at 10°C ± 0.2°C. About twenty-four to thirty hours in the water tank is required for the cable to reach temperature equilibrium. Fig. 16 shows the test tank with pans of cable in place.

The two ends of the cable to be tested are brought out of the tank

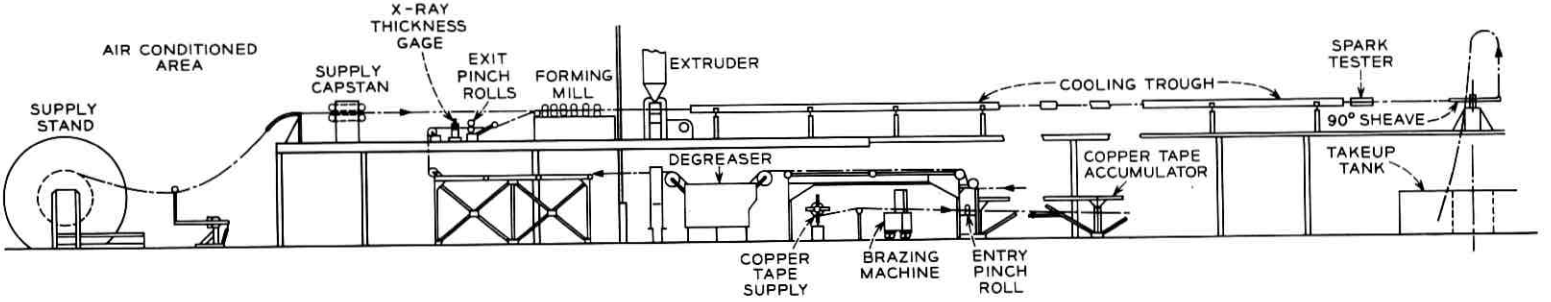


Fig. 15 — Elevation of outer conductor forming and jacketing line.

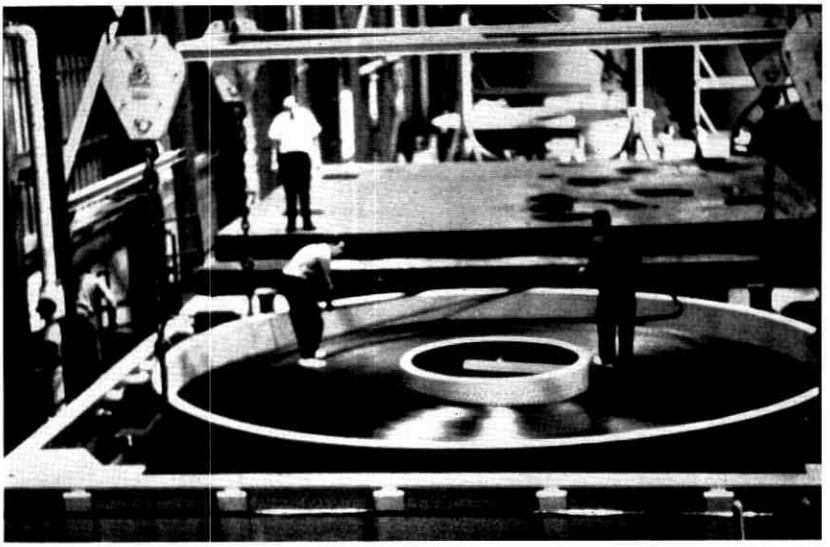


Fig. 16 — Test tank.

and into the test room (See Fig. 17) through an access port. In the test room, temperature is maintained at 73°F and humidity at 55 per cent. The measurements made here include inner and outer conductor dc resistance, insulation resistance, dielectric high-voltage breakdown, pulse echo, delay, low-frequency capacitance and attenuation.

Conductor resistance readings are made using a Wheatstone bridge of 0.01 per cent accuracy. Measurements are made initially of the inner conductor to determine when the cable temperature has stabilized. After three consecutive readings of the same values are obtained at one-half hour intervals, temperatures are considered stabilized and the outer conductor resistance is measured and recorded.

Insulation resistances of both the dielectric and jacket are obtained at 500 volts using a megohm bridge. Minimum insulation resistance requirements are 100,000 megohm-miles for the dielectric and 15,000 megohm-miles for the jacket.

Breakdown tests are made at a potential of 35,000 volts dc between the inner and outer conductors. Both positive and negative potentials are applied at both ends of the cable for periods of one minute each.

The pulse echo test set, shown schematically in Fig. 18, is used to determine the magnitude of reflected signals due to impedance mismatches within the cable. The magnitude of the reflected signals is

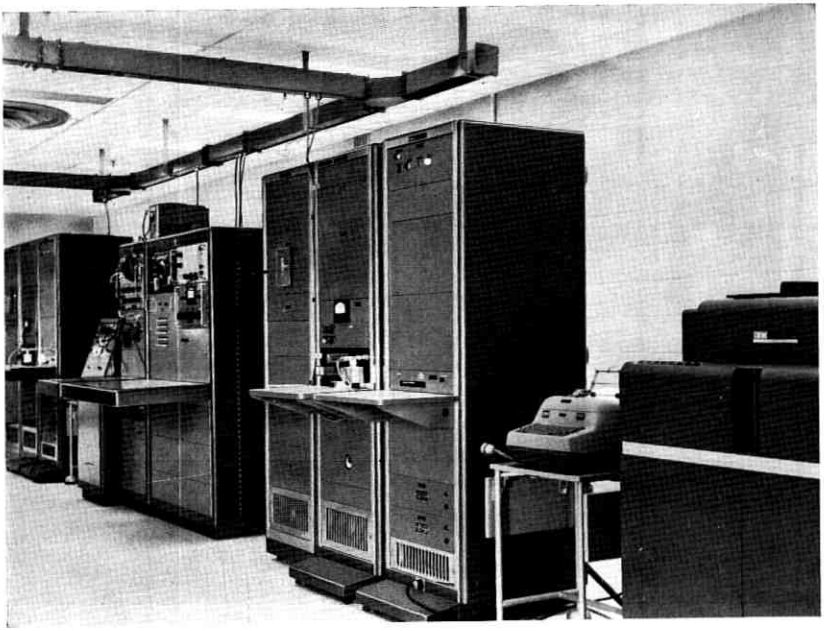


Fig. 17 — Test room.

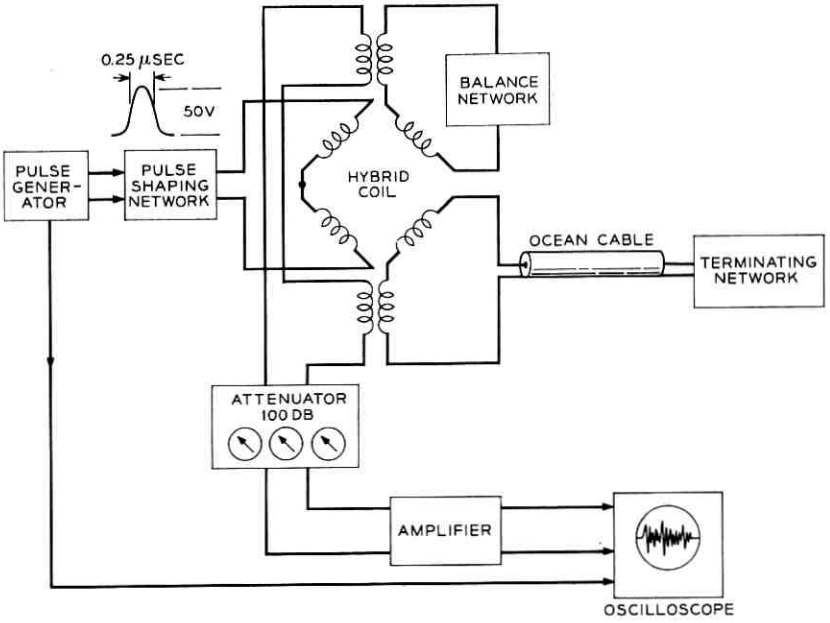


Fig. 18 — Pulse echo set.

measured in db below the test pulse, and is referred to the point of incidence. A 0.25-microsecond raised cosine pulse is used, and the reflected pulses or echoes are required to be 55 db or more below the incident pulse. Tests are made from each end of the cable to a point 13 nm in from the end. Echo magnitudes are corrected for the distance from the end, and photographic records are made of the patterns obtained from each cable. The accuracy of the echo test set is ±1 db.

The delay test set measures the delay encountered by a 1-mc signal traveling through a 20-nm cable section. The set uses an extremely accurate oscillator with a short-term stability of ±1 cps at 1 mc, a narrow-band receiver, and an electronic frequency counter, as shown in Fig. 19.

Capacitance measurements are made at a frequency of 21 ± 1 cps to avoid errors which result from standing wave and resonance effects that occur if the wavelength of the signal is not long with respect to the cable length. This frequency also avoids subharmonics of the 60-cps power line frequency, but is high enough to be insensitive to dielectric absorption effects. Cable capacitance is measured to an accuracy of ±0.1 per cent.

The most important electrical test is the cable attenuation measurement. This test is made using an autobalance transmission measuring set, designed by Bell Telephone Laboratories and built by Western Electric at Kearny, New Jersey. A block diagram of the set is shown in Fig. 20. It measures the attenuation of a 20-nm cable length as a

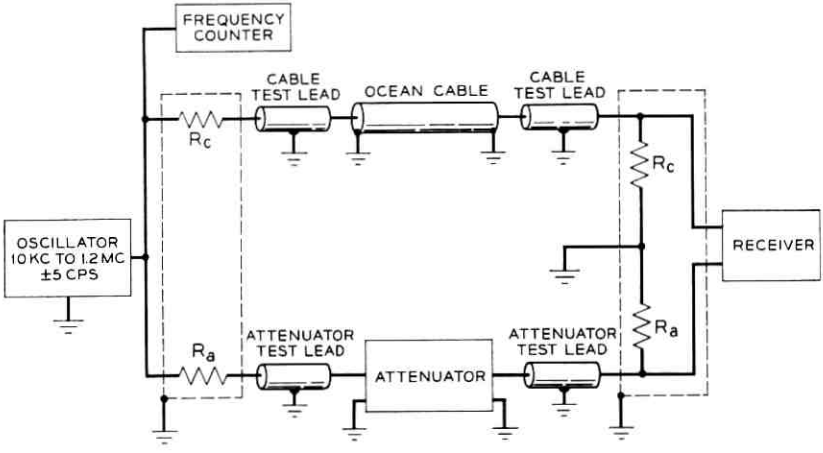


Fig. 19 — Delay set.

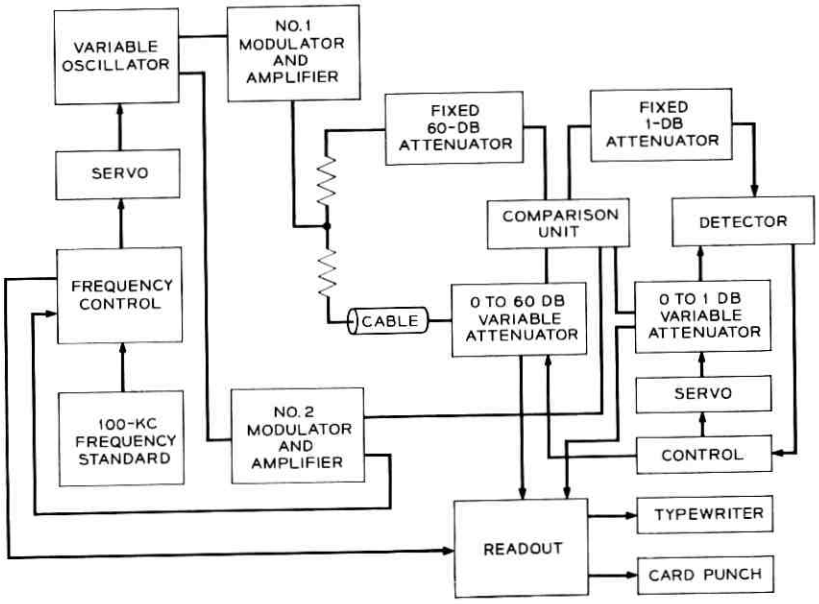


Fig. 20 — Auto-balance attenuation test set.

function of frequency with an accuracy of ±0.4 per cent over the frequency range from 50 kc to 1200 kc, in increments of 50 kc ± 25 cps. The set is automatic, with servo-driven frequency control of the signal source and self-balance of the loss measuring circuit. Readout is automatic and, upon completion of each loss measurement, the data are printed by an electric typewriter and punched on cards by an IBM summary punch. Subsequent calculations are made by processing the cards in an IBM calculating punch. Finally, the cards are fed through an IBM interpreter which prints the punched data across the top of the card, making the results readily available.

Orders for cable are placed with the factory, specifying the nominal length and desired 1.0 mc attenuation in db. Finished cable is accepted if the attenuation values are within 0.4 per cent of the ordered values.

V. CABLE TERMINATING

Part of the cable manufacturing process is the assembly of the cable termination.[1] The space in which this operation is conducted is a dust-free air-conditioned room with good lighting, manned by operators garbed in nylon dress. Slots in the wall enable the cable ends to be brought into the room. The following sequence of operations is followed:

(1) The cable end is cut back and a threaded steel strength member is crimped onto the steel strand.

(2) A T-shaped anchor assembly is screwed, then pinned, to the threaded steel sleeve. The anchor has an insulated conductor attached for subsequent joining to a similar conductor coming from the repeater.

(3) The gap in the dielectric between the cable and the anchor is replaced by injection molding.

(4) A housing, which becomes the outer conductor structure, is joined to the cable and assembled over the anchor.

(5) The ethylene plastic jacket is replaced by injection molding.

(6) A gimbal joint which enables the cable to be flexed at angles up to 45° with respect to the repeater axis is assembled.

After each of the two molding operations, sets of radiographs are taken to inspect for voids and/or contamination.

When the terminations are completed on both ends, the cable is tested a second time for high-voltage breakdown, resistance of inner and outer conductors, and dielectric and jacket insulation resistance.

VI. ARMORED CABLE

Cable to be used in shallow water, where it may be subjected to abrasion or fouled by trawling equipment, is protected by either one or two layers of neoprene jacketed steel armor wire over the basic cable structure.[2] The armor wires are applied helically to provide a degree of flexibility to the composite structure. However, as a result of the stresses in the unidirectional helices, the cable will elongate and twist when subjected to tension. For this reason, an annealed solid copper conductor is substituted for the composite inner conductor used in the armorless structure.

For those cables which will be placed close to shore and which may be exposed to radiation from radio, Lorac or Loran stations, supplemental shielding of the cable is achieved by application of five layers of soft iron strip, one applied longitudinally, the remaining four helically, over the cable jacket. An outer jacket is applied over the strips to a diameter of 1.5 inches. The neoprene-jacketed armor wires are then applied over the outer jacket in either one or two layers, depending upon the degree of abrasion or fouling the cable may be subjected to.

Except for the modified inner conductor, cable operations and tests are identical to those used for the armorless design.

Cable, before armoring, is moved in the twenty-foot diameter pans by trailer from the ocean cable factory to the armoring building. (See Fig. 21 for a view of the plant layout.) After armoring, the cable is

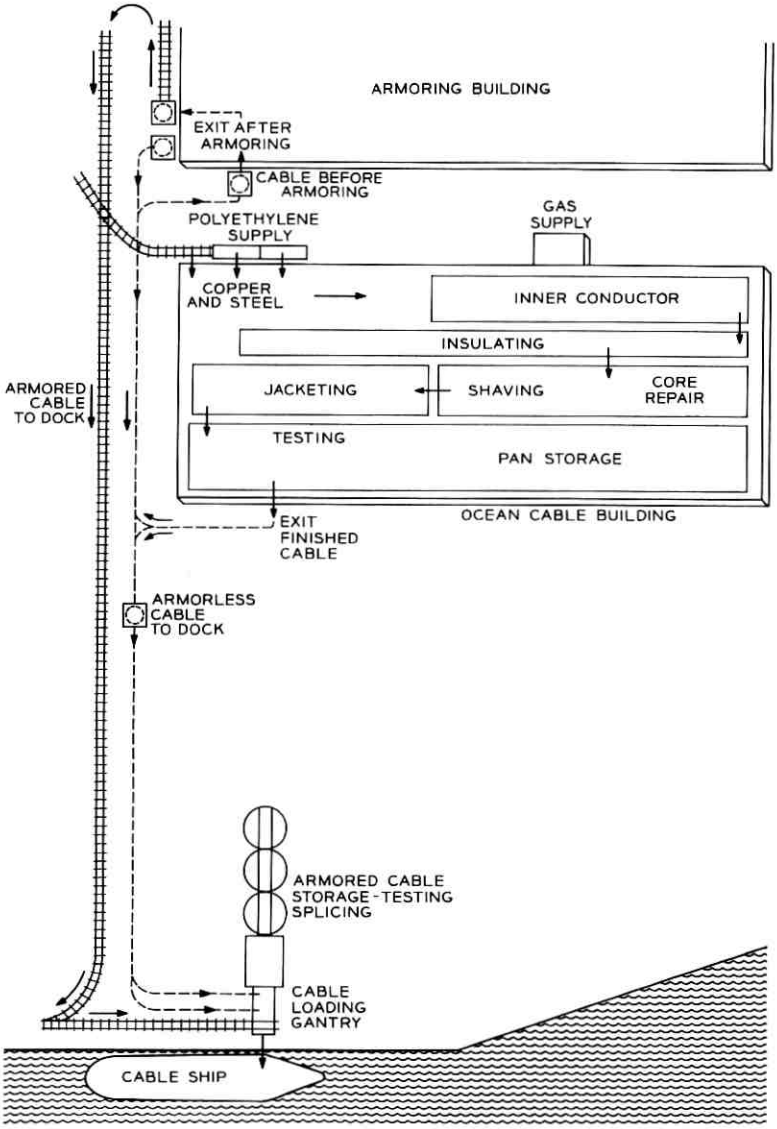


Fig. 21 — Ocean cable manufacturing facility, Baltimore, Md.

coiled into pans located at the exit of the last operation and moved to test and dockside on trailers or railroad flat cars. At dockside the several sections of armored cable may be spliced together or terminated with special couplings, and then stored in concrete tanks.

VII. TRANSPORT AND HANDLING OF CABLE

All cable is manufactured, transported and stored in nominal 20-nm lengths with gross weights ranging from 20 tons for a full reel of inner conductor to 64 tons for a pan of finished cable. Fig. 21 also shows the cable processing paths. The controlled-temperature, ultra-clean room shown in Fig. 22 is the site where extruded dielectric take-up, dielectric repair, dielectric sizing and outer conductor application are performed as the cable is wound from one 14-foot diameter reel to another. The cable is bent to the minimum drum radius of three feet, but no twist is imparted to the cable. All reels are moved into and out of position by crane. Rotation of the massive reels, coupled with the heavy driving and braking forces, generates high stress concentrations in the reel structures, necessitating the use of bolted construction where possible.

After the last manufacturing operation, the application of the outer conductor and jacket, the cable is distributed into steel pans approxi-

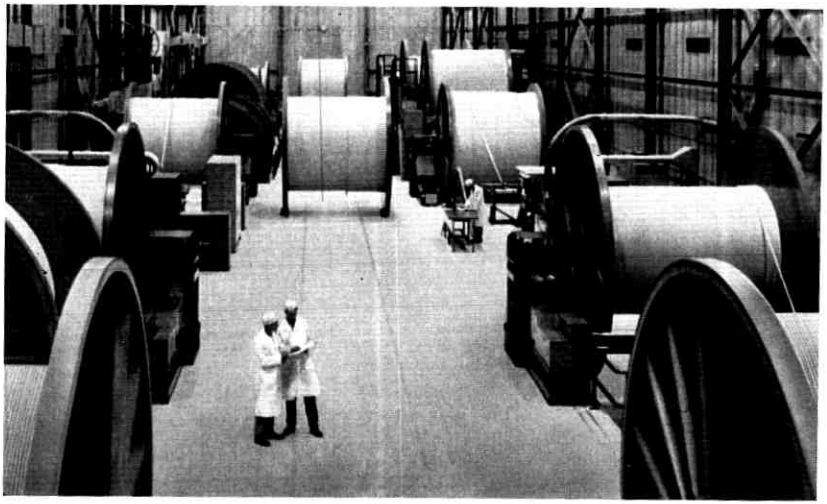


Fig. 22 — Controlled-temperature clean room.

mately twenty feet in diameter. From the jacketing line straightaway to the cable ship tanks the cable is normally bent and straightened eight times around radii of no less than four feet.

Handling of the 20-nm lengths of cable in pans provides maximum flexibility in terminating, testing and storage. Three thousand miles of cable can be stored in the factory (see Fig. 21) and can be transported, as required, to the ship's side in any sequence for loading. From the pan at shipside the cable is uncoiled and transported to the ship tanks through short, smooth troughs. Physical protection of the cable and terminations from abrasion and excessive flexure is thus provided for.

Storage of panned cable at 60°F minimum temperature is essential when ship loadings are made in freezing weather. In several loadings the relatively warm cable was towed to shipside, placed in heated tents (See Fig. 23) and pulled through the troughs without encountering a helical set in the cable despite the low temperature of the atmosphere. If the cable had been stored outside the building, the cable would have been more difficult to load because of increased stiffness in the structure.

VIII. PROCESS INSPECTION

In addition to final cable electrical requirements, tests are made on all cables for other specification requirements such as lack of contact

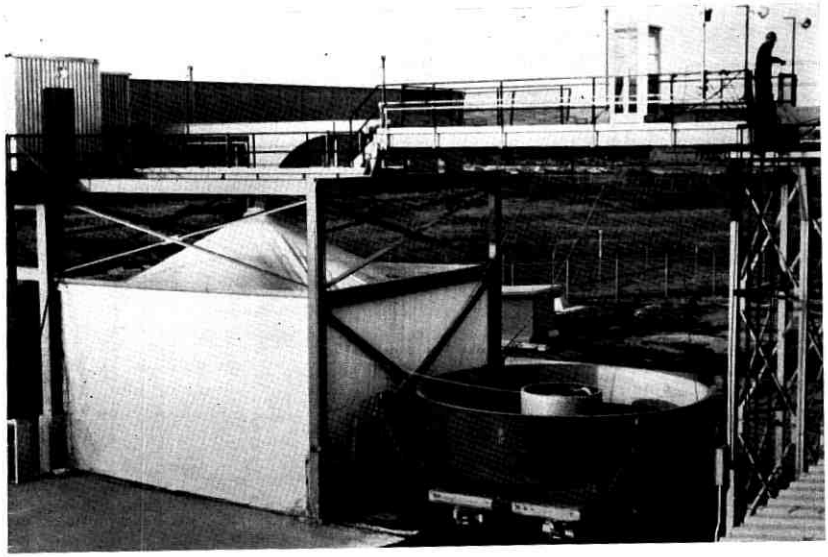


Fig. 23 — Heated loading tent.

between the dielectric and outer conductor, number of reverse bends the cable will withstand before outer conductor failure occurs, and ethylene plastic jacket retraction.

All operators are periodically checked for qualification. No unqualified operator is permitted to work on the product.

There are also many in-process recorder charts which are examined for specification compliance. Typical of these charts are:

(1) inner conductor fault chart,
(2) eccentricity and diameter charts on the dielectric extrusion,
(3) water trough temperatures on the dielectric extrusion lines,
(4) sized dielectric eccentricity and diameter charts,
(5) outer conductor thickness chart,
(6) jacket eccentricity and diameter charts, and
(7) jacket cooling trough temperatures.

All applicable charts become part of the cable history. If a defect is revealed, it is authenticated carefully, and if necessary the cable is rewound to that position and repairs are made, or the cable is discarded.

IX. PROCESS RESULTS

At the beginning of this article, it was stated that the specification covering SD system cable was written in terms of end product requirements in the hope that cable would be produced to a greater degree of uniformity than had previously been possible. Several things can be enumerated that made this a worthwhile choice. Chief among these is the structure itself. Only three materials, steel, copper, and polyethylene, are used in the basic design, and with the exception of the outer conductor overlap, all of the elements are simple, cylindrical shapes. Perhaps of equal importance are the specifications covering those materials. They assure the procurement of material of a homogeneity rarely seen

TABLE I — CABLE ELECTRICAL PARAMETERS: TEST RESULTS FOR 100 CONSECUTIVE CABLE SECTIONS

Measurement	Measured Values in Units per nm				
	Average	Maximum	Minimum	Standard Deviation	Pct. Std. Dev.
1.00-mc loss in db	2.3900	2.4013	2.3801	0.004803	0.2010
I.C. R_{dc} in ohms	1.7600	1.7807	1.7388	0.009285	0.5276
O.C. R_{dc} in ohms	1.3754	1.3831	1.3672	0.003042	0.2212
21-cps capacitance in μfd	0.2135	0.2148	0.2125	0.000513	0.2405
1.0-mc delay in μsec	9.4063	9.4329	9.3849	0.009568	0.1017

in a large-scale manufacturing operation. Other factors worthy of mention include operator selection, training and qualification, use of large but precise machinery, and maintenance of plant cleanliness.

How well this goal has been achieved can best be determined by reference to Table I. In this table are listed several of the critical cable electrical parameters along with results obtained on 100 consecutive cable sections (approximately 2000 nm) made in the latter half of 1963 at the Baltimore plant of the Western Electric Company.

REFERENCES

1. Brewer, S. T., Dickinson, F. R., and von Roesgen, C. A., Repeaters and Equalizers for the SD Submarine Cable System, B.S.T.J., this issue, p. 1243.
2. Bowker, M. W., Nutt, W. G., and Riley, R. M., Design of Armorless Ocean Cable, B.S.T.J., this issue, p. 1185.

Repeaters and Equalizers for the SD Submarine Cable System

By S.T. BREWER, F.R. DICKINSON and C.A. VON ROESGEN

(Manuscript received April 17, 1964)

An equivalent four-wire repeater and an adjustable equalizer are described. The repeater incorporates a low-aging design to reduce the effect of electron tube gain changes. The equalizer contains a number of bridged-T networks, some of which can be switched in or out by a selector.

The high-pressure container which houses either unit is made up of a beryllium copper cylinder with welded dome covers. Electrical connections are made through polyethylene-metal seals.

I. INTRODUCTION

The decision to replace the flexible repeater of the SB submarine cable system[1] with a rigid structure of more conventional length-to-diameter ratio permitted the design of a repeater of increased bandwidth with equivalent four-wire operation. The objectives relating to stability and reliability remained the same as for the SB repeater.

In the design of repeaters for submarine cable systems, attention must be paid to the smallest detail, and long testing programs are required to realize the stability and reliability objectives. Generally, basic circuit and mechanical concepts are well proven, but must be refined to an extremely high order of perfection to meet the system transmission and life requirements.

The SD system[2] for which this new repeater was designed operates over a frequency range of 108-504 kc (low band) and 660-1052 kc (high band). The repeaters are spaced at 20-nautical mile (nm) intervals. The repeater gain matches the cable section loss within +0.30 or −0.10 db in the low band and ±0.10 in the high band. Excess gain or loss of a repeater section is called misalignment. To avoid a substantial buildup of misalignment, equalizers are inserted every 10 repeater sections. This keeps the misalignment down to ±0.20 db per 192 nm. The equalizers also compensate cable loss changes that occur during laying. For this

purpose they are adjustable from outside their housings until a short time prior to overboarding.

II. REPEATER CIRCUIT

The repeater block schematic, Fig. 1, shows the feedback amplifier, the two sets of directional filters, and the two power separation filters. The cable connects to the power separation filter, which directs the dc cable current to the electron tube heaters and the high-frequency signals to the directional filters. These filters combine both low-band and high-band signals entering from opposite power separation filters and connect them to the amplifier input. After amplification the two bands are split again by the directional filters and sent out through the proper filters to the cable.

2.1 *Power Separation Filters*

The power separation filters (PSF) shown in Fig. 2 are of a series high-pass, low-pass design. A coaxial choke is inserted in the signal path between the PSF and the directional filters to avoid ground loop couplings. The gas tube is shunted across the signal path output to protect components from high-voltage surges in cases of an accidental cable short. The dc output lead of the PSF connects to the repeater unit con-

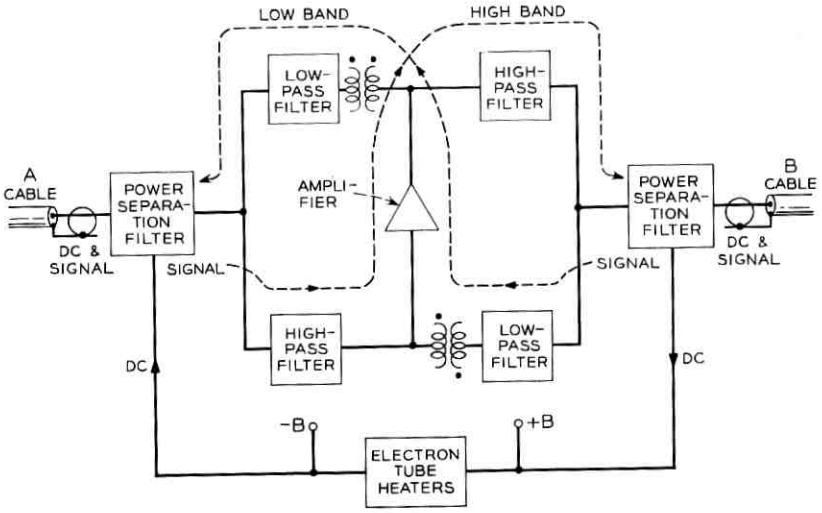


Fig. 1 — Repeater block schematic.

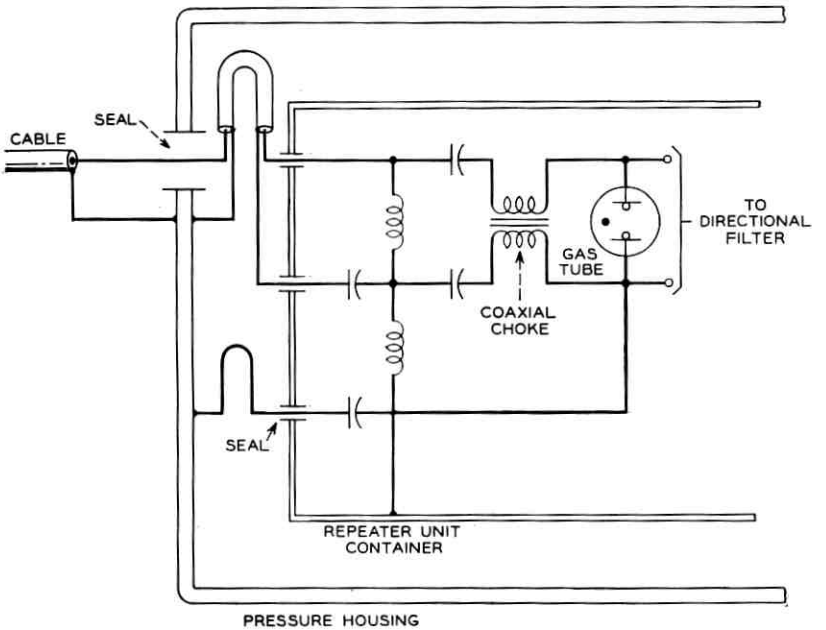


Fig. 2 — Power separation filter.

tainer. As a result, this container is at the potential of the cable inner conductor.

2.2 *Directional Filters*

The directional filters are designed as a constant-resistance four-port network.[3] Special high-Q, adjustable air core inductors are used. Design and adjustment of these filters is critical, since there will be many filters in tandem in a long system. It is important that in-band transmission variations be held to a very small value.

The power separation filters and the directional filters create several spurious feedback paths around the amplifier that would affect the repeater gain characteristic were it not for the high loop losses. The directional filter loop (see Fig. 1) is made up of two symmetrical paths; a 1:1 transformer appears in one path, and a $1:-1$ transformer in the other. Ideally then, the two unwanted signals cancel at the amplifier input port. This balance, however, can be maintained only if all ports are terminated by good impedances. Therefore all filter and amplifier impedances exceed a minimum return loss of 27 db.

2.3 *Amplifier*

The amplifier (Fig. 3) is a three-stage, double-μ circuit, feedback device. Major features of interest are the double-μ circuit and the low-aging design.

The input and output coupling networks are identical. The gain is shaped by means of a shunt *RL* network which attenuates low frequencies and by a transformer resonance which peaks at high frequencies. The gain slope of each network is 10 db. The high-side impedance is 3200 ohms at 1 mc, which is the optimum impedance for maximum power output with single-μ circuit operation. An impedance balancing network is used with the hybrid type transformer to provide good amplifier input and output impedances.

The β network is essentially a two-terminal network consisting of one *RL* and five *RLC* branches in parallel. Its impedance varies from about

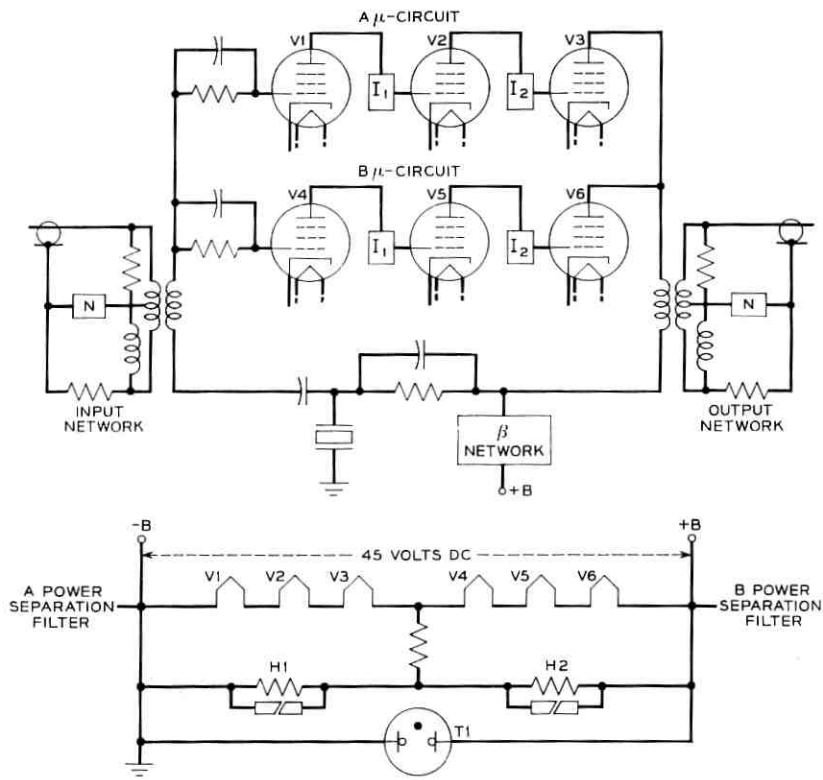


Fig. 3 — Amplifier circuit.

70 ohms at 100 kc to 7 ohms at 1100 kc. The RL branch carries the plate current to the two output tubes. The high side of this network connects through a resistor in parallel with a capacitor to a resonant quartz crystal. As in the SB repeater, this crystal permits monitoring of μ circuit gain changes and identification of a defective amplifier.

The amplifier in-band feedback design takes advantage of phase control to minimize the effect of electron tube aging on amplifier gain. In the gain expression μ appears as $(\mu\beta)/(1 - \mu\beta)$ or $1/(1/\mu\beta - 1)$. The absolute value of the denominator $[(1/\mu\beta) - 1]$ is a function of both magnitude and phase of $\mu\beta$ or

$$\left|\frac{1}{\mu\beta} - 1\right| = \sqrt{1 - \frac{2}{|\mu\beta|}\cos\theta + \frac{1}{|\mu\beta|^2}}.$$

For example, the $\mu\beta$ effect[4] of an amplifier with 40 db of feedback has the same value as that of an amplifier with only 20 db of feedback provided $\mu\beta$ phase is 86.9° in both cases. Therefore, by controlling the phase of $\mu\beta$, a substantial reduction in the sensitivity of the $\mu\beta$ effect to a decrease of tube transconductance can be achieved. Based on currently available electron tube aging data, system misalignment from this source over a

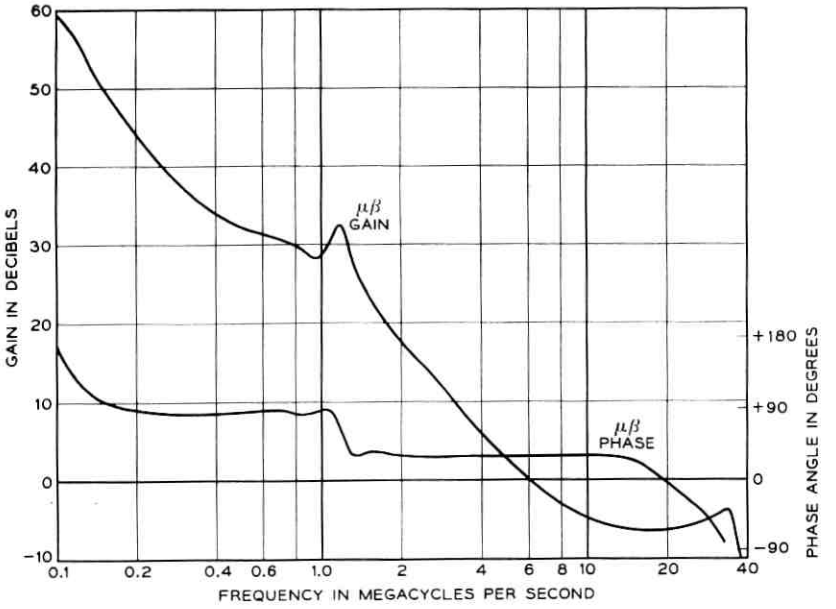


Fig. 4 — Amplifier feedback characteristics.

20-year period is expected to be not more than 4 db for a 200-repeater system.

A special feature of the high-frequency feedback cutoff (Fig. 4) is the loop gain peak at 34 mc. At this frequency, the β network resonates with the parasitic capacitance of the β loop. The resulting favorable phase at lower frequencies makes it possible to achieve maximum available feedback at band edge. The minimum stability margins are: 5 db gain margin at phase crossover and 25° phase margin at gain crossover. Transit time is 2.7° per mc for the whole loop including electron tubes.

The double-μ circuit is used to improve amplifier reliability as far as electron tube failures are concerned. The most likely tube failures are open heaters and electrode shorts. Therefore, to realize the full advantage of this redundancy, it is necessary to be able to operate a functional μ circuit after its twin has failed. Grid-to-cathode shorts, if they occur, will not disable the functional μ circuit, because they are either isolated by grid isolation networks in the first stage or completely independent in the second and third stages.

An open heater, on the other hand, will disable the amplifier, since the gas tube will fire and remove power from the undamaged string. After the repeater in question has been identified by measurements of the crystal noise peaks, power can be turned down to extinguish the gas tube. The current can then be raised to 140 ma. At this point, the amplifier voltage is not sufficient to fire the gas tube. However, under these conditions sufficient heat is generated in the heat coil to melt solder that bridges its resistance (see Fig. 3). Thus a resistor previously not carrying current is substituted for the heaters of the failed amplifier circuit. The equivalent space current of one amplifier circuit is carried by the intact heat coil.

III. REPEATER PERFORMANCE

The SD repeater is primarily characterized by its insertion gain (Fig. 5), its noise figure, and its power output (Fig. 6). As already pointed out, the insertion gain closely matches the cable loss.

Since high frequencies are attenuated by the cable considerably more than low frequencies, they are more vulnerable to noise originated at the repeater input. For this reason both repeater noise figure and power output capabilities are optimized at high frequencies.

During the manufacture of components, networks, filters, and amplifiers, great control is exercised to make sure not only that requirements are met, but also that the averages do not wander beyond tolerable limits. Fig. 5 shows the manufacturing deviations of the repeater insertion gain based on 80 units.

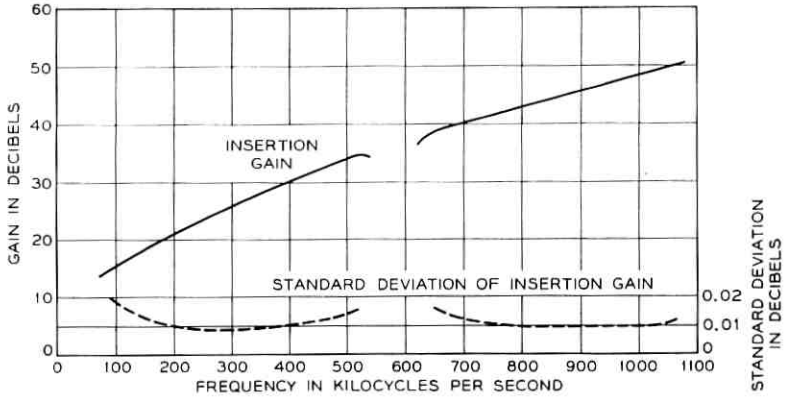


Fig. 5 — Repeater insertion gain.

IV. COMPONENTS

The majority of the electrical component types used in the SD repeater are similar to those used in the SB repeater. However, to fill new needs and meet new requirements several new types of components were introduced after extensive testing. In all cases, similar designs had been used satisfactorily over long periods of time in regular telephone plant applications. The new component types are: polystyrene capacitors, composition resistors, heat coils, and ferrite cores.

Polystyrene capacitors are used in most low-voltage applications for

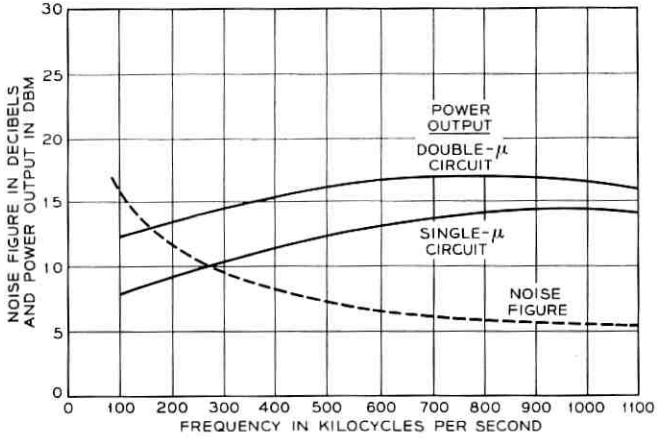


Fig. 6 — Repeater power output and noise figure.

values of capacitance above 5000 pf, where Q and tolerance requirements would not permit the use of paper capacitors.

The composition resistors are used for resistance values above 2000 ohms, where tolerances are liberal. Although these resistors are manufactured by conventional processes, special precautions are taken to produce a uniform product of the highest quality. They are subjected to further screening and stabilization before use in a repeater.

The heat coil is a new device used in the power path to bypass an open heater circuit. It consists of two insulated masses of lead-antimony alloy arranged inside a vitreous enameled resistor so that when sufficient power is applied, the masses of alloy melt, thereby shorting the resistor.

Ferrite cores are used on the control grid and screen grid leads of all electron tubes to provide adequate margin against high-frequency sing of an individual tube.

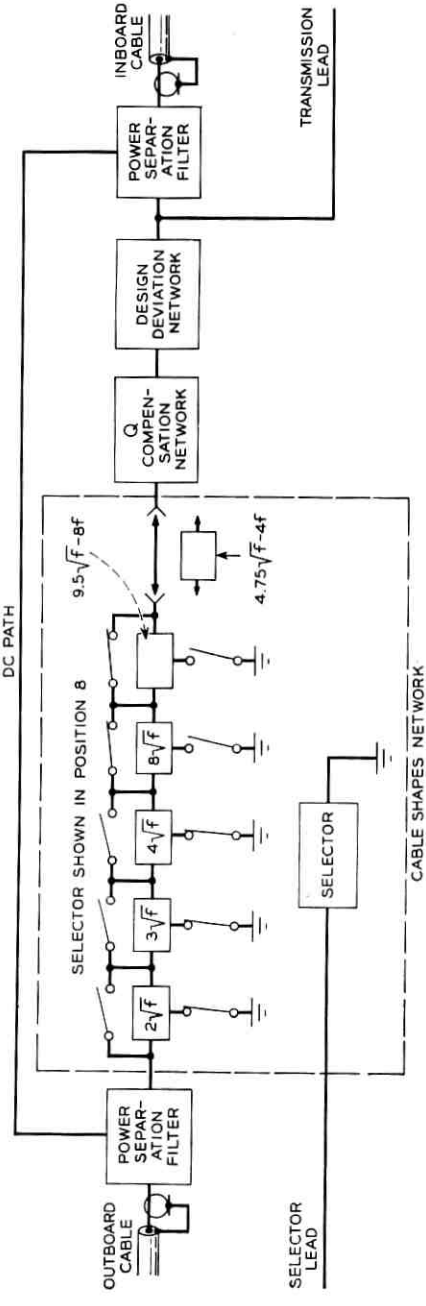
V. EQUALIZER CIRCUIT

Block-end equalizers are located at the end of every group of 10 repeaters. They are inserted between two cable sections, each 6 nm in length. The difference in loss between the nominal 20-nm section and this 12 nm of cable provides the equalizer loss range (constant loss and adjustable loss). The equalizers are designed to compensate the misalignment which has accumulated over 10 sections because of differences in average characteristics of cable and repeater; they also provide an adjustment to compensate the cable laying effect of the block, that is, the difference between predicted and measured sea-bottom cable loss.

The equalizer (see Fig. 7) uses PSF's to bypass the cable current; the inner housing in this case is at sea ground. Since all networks are designed to cover the frequency range of both transmission bands, directional filters are not required. All equalizing networks consist of one or more bridged-T constant-R sections.

5.1 *Design Deviation Network*

This network compensates the mismatch between the average repeater and the average cable section at average depth and temperature. The objective is to keep this misalignment within ± 0.2 db per block. It is more practical as well as more economical to do this in the equalizer than to tighten repeater requirements to ± 0.02 db.



NETWORK SECTIONS	SELECTOR POSITIONS											32
	1	2	3	4	5	6	7	8	9	10		
$2\sqrt{f}$	0	1	1	0	0	0	1	1	1	0		0
$3\sqrt{f}$	0	0	1	1	1	0	0	1	1	1		0
$4\sqrt{f}$	0	0	0	0	1	1	1	1	1	1		0
$8\sqrt{f}$	0	0	0	0	0	0	0	0	1	1		0
$9.5\sqrt{f}-8f$	0	0	0	0	0	0	0	0	0	0		1

0 = SECTION OUT
1 = SECTION IN

Fig. 7 — Ocean-block equalizer.

5.2 *Directional Filter Q-Compensation Network*

The directional filter inductors are the only repeater components whose manufacturing variations regarding dissipation exceed the limits that system requirements impose. It is for this reason that three Q-section networks are available, one for low-, one for average-, and one for high-Q inductors.

Directional filters are assigned to particular repeaters before the repeaters are assembled. Their insertion losses are added and compared to an objective for the 10-repeater block. Depending on the magnitude and the sign of the deviation, a high-, average- or low-Q section compensating network is picked for the equalizer at the end of this block. Any residual deviation is taken into account in the next equalizer.

5.3 *Cable Shapes Network*

This network consists of five switchable sections and one section that is wired in on an optional basis. The switchable sections can be switched in or out in any combination by a 32-position selector,[5] activating five contact pairs, with a make and a break contact in each pair. Each contact pair controls one network. The make contact is in series with the shunt branch of a bridged-T section and the break contact is in parallel with the series branch. In the nonoperated state, the network is therefore removed from the circuit.

The dominant contributors to the loss of the cable are conductor resistance and dielectric loss. Loss due to conductor resistance varies as the square root of frequency; dielectric loss varies directly with frequency. The bulk of cable loss is $\sqrt{f}$. The linear component of loss is small by comparison and is dependent on the power factor of the polyethylene, a parameter which is difficult to measure and control. The procedure for equalization during laying calls for equalizing a linear f characteristic with $\sqrt{f}$ loss until the value of the former has built up to a considerable magnitude. This will tend to produce a misalignment shape of the form $\pm K_1 f \mp K_2\sqrt{f}$. The equalizer is designed to handle $\pm K_1 f \mp K_2\sqrt{f}$ or $K_3\sqrt{f}$ shapes. $K\sqrt{f}$ means a loss of K db at 1 mc with a $\sqrt{f}$ shape, and Kf has an analogous meaning.

The nominal characteristics of the six sections are $2\sqrt{f}$, $3\sqrt{f}$, $4\sqrt{f}$, $8\sqrt{f}$, $9.5\sqrt{f} - 8f$, and $4.75\sqrt{f} - 4f$. The last is supplied on an optional basis. The range of $\sqrt{f}$ provided corresponds to ± 3.3 nm or $1\frac{3}{4}$ per cent of the loss in a block. The range of linear shape provided is sufficient to cover a variation of ± 80 per cent in the nominal conductance loss. With these shapes and step sizes, both $\sqrt{f}$ and linear cable

loss deviation can be equalized to ± 1 db at 250 kc and ± 0.5 db at 1 mc. This is consistent with the rule for making the tolerances over the band roughly inversely proportional to $\sqrt{f}$.

The nominal setting of the cable shapes network ideally is $13.25\sqrt{f} - 4f$. Since this shape is not available, a nominal block would be equalized by inserting $9\sqrt{f}$. It should be noted that such a nominal block would be left with a residual misalignment of $-4.25\sqrt{f} + 4f$. At the end of a second nominal block this would be wiped out by using $17.5\sqrt{f} - 8f$. The misalignment introduced by the equalizer in the case of nominal cable is necessary to make it possible to handle linear misalignment of either sign using only one switchable section having linear frequency loss. In order to be in a position to handle linear factory deviations if they occur, the optional $4.75\sqrt{f} - 4f$ section is provided. It can be included or omitted at the time the equalizer is assembled in the factory.

Two leads are brought out of the equalizer in addition to the cable leads. One of these connects to the solenoid which operates the selector. The other bridges onto the transmission path between the networks and the inboard power separation filter. It is used to measure transmission during the laying. Both leads are sealed after equalizer adjustment and prior to overboarding.

VI. REPEATER MECHANICAL DESIGN

There are basically three major facets to the physical realization of the SD repeater circuit. These are the repeater unit, the pressure housing, and the cable-to-repeater coupling. (See Fig. 8.) The nature of the individual requirements was such that each could be developed independently as long as there was sufficient coordination to complete an over-all integrated design.

6.1 *Pressure Housing*

The development started at the outside, the rigid repeater housing. At the time, circuit proposals were not complete enough to permit active work on component layouts. However, previous experience with broadband amplifiers permitted a fair judgment of the volume requirements. Since maximum compactness was desirable from many standpoints, a cylindrical design was chosen with the largest diameter which seemed practical for handling in factories and on shipboard. The final length was adjusted to the requirements of the circuit configuration.

Proper and economical design of the housing is important. It must not leak over the expected life span. It must not be overstressed at the

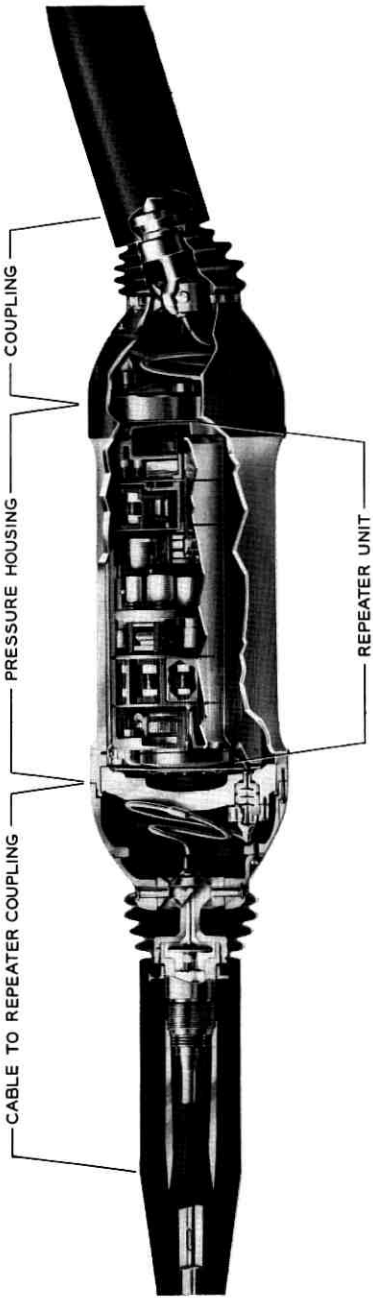


Fig. 8 — Repeater — cutaway drawing.

greatest expected ocean depths, but it must have a minimum factor of safety so that it will not become unmanageable. Exposure of organic materials to sea water pressure must be kept at a minimum. The housing finally utilized to meet these needs consists of a cylinder and two domed end covers, as in Fig. 9.

A variety of materials were considered for use in the pressure housing. In 1955 the list was narrowed to two: steel and a heat-treated copper beryllium alloy. Steel has a slight strength advantage over copper beryllium but has the great disadvantage of a high corrosion rate. If steel is used, protective coatings of some kind must be provided. If the galvanic protection of zinc were utilized, the coating would necessarily be applied after all machining operations and any metal joining operations which might be performed in final closure. The integrity of such a zinc coating could not be reliably ascertained. Furthermore, any joints between the copper return tapes in the cable, or other copper or copper

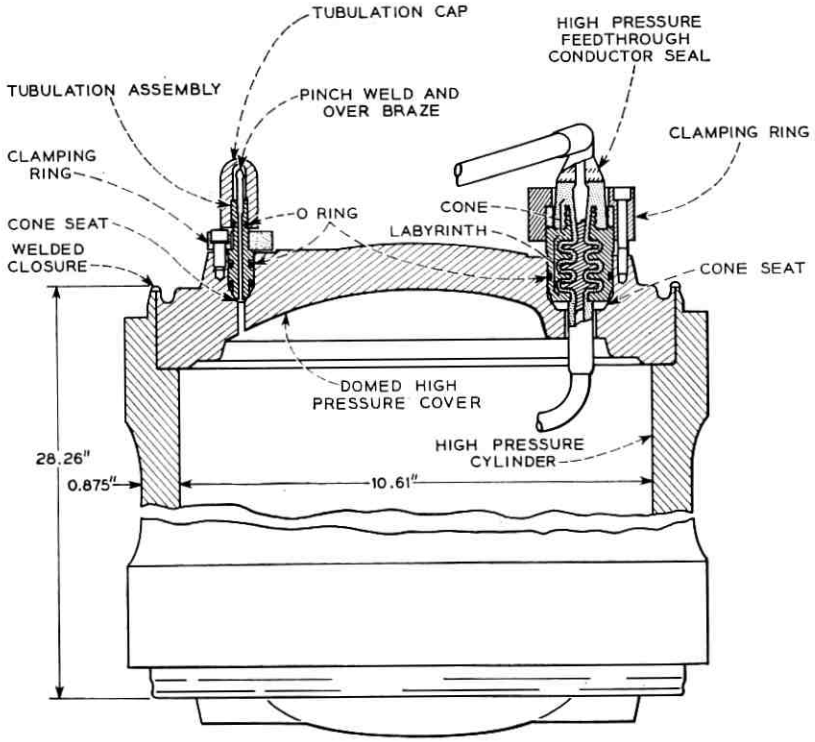


Fig. 9 — Pressure housing.

alloy parts would need to be carefully protected from electrolysis due to copper-steel couples. The most satisfactory protection for the steel would therefore result from a complete skin of copper, or copper alloy, sealed against the entry of sea water. A protective skin of any type would be subjected to abrasion on shipboard and sea bottom and would probably require mechanical protection from such abuse.

Machining costs of steel would not be less than those for copper beryllium, so that the only economy would be in the cost of the material itself. Such economy would be more than offset by complications in assembly to afford corrosion protection.

Further consideration of steel was abandoned in favor of a new copper beryllium alloy specifically developed to meet the special requirements of this application. This alloy has been demonstrated to have the low initial corrosive rate of 0.001 inch per year in salt water, decreasing as patina develops. It is abrasion resistant, and when heat treated to a Rockwell hardness of C-36 develops strength equivalent to many of the steels which could have been chosen. This material can be cast by several methods, it can be wrought, and it can be readily welded. However, unless special techniques are employed, casting processes result in a dendritic grain structure which is not necessarily impervious to helium or water vapor.

6.2 *Cylinder*

The cylinder is fabricated from a semicontinuous cored cast billet, which is forward extruded into a cylinder approximating final dimensions. The extrusion process is performed through special dies at proper temperature in a hydraulic press. Press thrust is in the order of 7000 tons, resulting in a wrought cylinder of high strength and reliability.

The machining operation provides a lip used for final closure welding, a means for attaching cable terminations, and an internal shoulder which accommodates the thrust load of the end covers. The allowable stress for this material in the wrought condition is 120,000 psi at the 0.01 per cent yield point. The design was based on a calculated combined stress value of 100,000 psi under a hydraulic load pressure of 12,000 psi (4400 fathoms). Strain gauge examination of several models indicated the actual combined stress to be 93,000 psi, satisfactorily confirming the analytical work.

6.3 *Cover*

The domed covers used to close the cylinder ends are made of the same copper beryllium alloy as the cylinder. In this case, however, the

method of fabrication is different in that another special casting technique is used. This process, known as pressure casting or liquid forging, consists of closing shaped dies under high pressure on molten metal which has just begun to freeze, resulting in a dense casting of approximate size and shape. The force used to form these covers is in the order of 500 tons.

The design of these covers is specially tailored to meet the needs of the final welded assembly. The inner surface is made up of a large spherical radius in the center, blended into a smaller radius at the outside. The outer surface is made up of a large spherical radius in the center blended into a smaller but upturned radius at the outside. The two large radii are not parallel, but are arranged to produce a tapered wall thickness with the thinnest point at the center.

This design was not amenable to straightforward analysis because of the discontinuities associated with feedthrough seals. Therefore a "best approximation" was used to build early models. These were made of aluminum, and the results of strain gage examination and destruction testing were extrapolated to predict copper beryllium performance. The final copper beryllium design had a maximum bending area about half the distance from center to rim. This minimizes any expansion due to bending from hydraulic pressure. The allowable stress for this material in a pressure cast condition at 0.01 per cent yield is 105,000 psi and design objectives were set at 85,000 psi. The measured maximum combined stress in the final copper beryllium models was 49,000 psi under a hydraulic load of 12,000 psi. A greater amount of conservatism was acceptable in the cover design, as any change would be of a minor consequence in the size or weight of the repeater. Furthermore, in this complex configuration it was not certain that strain gauge examination accurately disclosed all points of high stress.

Machining produces a close fit of the cover in the cylinder bore (maximum clearance on the diameter 0.007 inch) and a flat rim to seat on the shoulder in the cylinder. A welding lip to match that in the cylinder is also provided.

Both cylinder and cover must, as stated above, be impervious to water vapor. The practical measurement of the leak-proof quality is by means of helium gas and a mass spectrometer type of leak detector. Each part must meet a maximum leak rate requirement of 5×10^{-8} std. cc of helium per second at any pressure to 12,000 psi.

6.4 *Welding*

The cover design minimizes bending and distortion at the edges. Thus the cover and cylinder lips which are matched and adjacent in assembly

may be welded together for final closure in an edge weld virtually free of either tensile or shear stresses (see Fig. 10). The fully automated weld process employs an inert-gas-shielded, nonconsumable electrode, arc technique. No filler material is added; thus only parent metal appears in the weld. Welding is done at a speed of 12 inches per minute with a welding current of 290 amperes. The welder is fed from a balanced-wave ac power supply to minimize the effect of oxidation. The gas shield is argon supplied at the rate of 15 cf/h. An automatic head with slope control up and down minimizes the effect of puddle build-up and eliminates arc blow.

Although the process yields a reliable and quite consistent weld, there are times when, at spots, the required weld depth of 0.06 inch is not attained. Thin spots are not detectable by leak detection or radiographic methods, and it is therefore required that each weld be examined by ultrasonic scanning to accurately determine actual weld depth penetration. A transducer alternately transmits and receives ultrasonic pulses (5 mc) at a rate of 1000 pps to measure reflections.

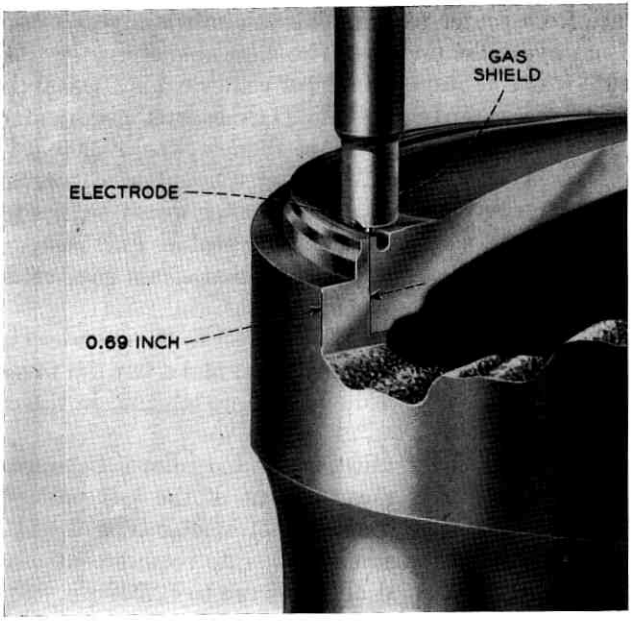


Fig. 10 — Weld process.

6.5 *Feedthrough Seals*

Unfortunately for the mechanical designer, the very function of the repeater housing requires the provision of feedthrough conductor seals. Seals must be free of corona noise at voltages as high as 6000 volts dc and must meet the same leak rate requirements as the rest of the parts making up the pressure housing.

Preferably, the seal should be of a vapor barrier type made of glass or ceramic. Anticipated difficulties with glass at these voltages and pressures, plus complications necessary to effect a transition between the basic seal and the polyethylene insulation of the cable, made the development of a polyethylene seal an attractive prospect. While it is impossible to prevent the diffusion of water vapor through an organic material such as polyethylene, it was felt that a low rate of diffusion would be acceptable if a second barrier could be provided to protect the circuit components. The use of a soft plastic material such as polyethylene presents a further problem: a design with low leak and diffusion rates will still extrude unless a method is devised which distributes the high sea bottom pressure load properly.

The high-pressure seal design (see Fig. 9) consists of a conical load section on the high-pressure side, followed by a labyrinth in which three antiextrusion disks are set in series. These disks are of phosphor bronze and form an integral part of the center conductor. The labyrinth is made up of a series of split machined rings set into an over-all copper beryllium casing. With the center conductor held in proper relation to the casing, the polyethylene insulation is molded directly into the cavity under pressures of about 9000 psi.

Molding polyethylene into a fixed cavity of this type inevitably results in shrinkage of the material away from the cavity walls during the subsequent cooling cycle. Thus it was impossible during the early development stages to produce a seal that did not leak at low pressures. To overcome this difficulty a new bonding process was developed, taking advantage of the ability of polyethylene to bond to a copper oxide under proper pressure, time and temperature conditions.

Bonding is limited to certain critical areas of the cavity in order to establish satisfactory electrical performance. The configuration of [illegible] design is such that bonding is required only at low pressures. At [illegible] pressures the self-energizing qualities take over the sealing fu[illegible]

High-pressure seals of the current design do not extrude [illegible] pressures, they meet a leak requirement of 5×10^{-8} std. [illegible] second, and they perform satisfactorily electrically

designed to assure that no direct leakage paths, such as cracks or incomplete polyethylene-to-metal bonds, exist. Even in the absence of cracks, there is some water permeating the polyethylene. This phenomenon, called "diffusion," is largest in shallow tropical waters. Under these conditions, the water diffusion rate has been found to be one order of magnitude lower than the over-all water leakage corresponding to the final helium leak test requirement on the assembled repeater. Under deep-sea conditions, the effect of the lower temperature on the water permeability of polyethylene, and on the water vapor pressure, as well as the compression of the polyethylene by the high hydrostatic pressure, reduces the water diffusion rate by at least another order of magnitude.

The seal casing is set into a counterbore in the end cover, coned at the bottom. The casing is machined with a conical end differing in angle by one degree from that in the cover. This arrangement provides a metal-to-metal cone seat, in a manner regularly used for high-pressure piping. The conical seat is protected from corrosion by an O ring on the sea water side. Preloading of the cone seal is accomplished by a clamping ring and bolt circle, while the O ring assures further loading from sea bottom pressures.

Each repeater housing must be leak tested after final closure over the entire pressure range up to 12,000 psi before leaving the factory. The leak rate requirement is 5×10^{-8} std. cc of helium per second. This rate has been chosen for the over-all assembly as well as individual parts as the limit of meaningful sensitivity obtainable in mass spectrometers. It corresponds to a water leak rate of at most 3×10^{-12} grams per second under deep-sea conditions, which would raise the relative humidity within the repeater housing by about 15 per cent over a 20-year period.

To provide access of leak detector sensing to the inside of the housing while the outsid[illegible] mall tubulation is mounted in one
[illegible] separate assembly mounted in the
[illegible]als. After completion of the leak
[illegible]he tubulation itself is tested by the
[illegible]sotope and subsequent searching
[illegible]milar to that used for the flexible
[illegible]so that less than a thimbleful of
[illegible]be safely made a more sensitive
[illegible]ssible.

With this material, test time is
[illegible]s as well as manufacturing costs.
[illegible]e design of the tubulation, which

[illegible]he
[illegible]igher
[illegible]ection.
[illegible]at deep-sea
[illegible]ce of helium per
The leak test is

After isotope testing, the pinch welded tubulation is overbrazed and finally protected against corrosion and mechanical damage by an O-ring sealed cap.

6.6 *Repeater Unit*

The purpose of the repeater housing is to provide a means whereby the circuit components can operate at atmospheric pressure and low humidity indefinitely. Certain requirements, such as shock absorption, 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 equipment.

Copper would have been best electrically for the inner housing but was abandoned primarily because of the cost of forming it in the complex patterns required by the design layout. The heavy weight of a housing of copper was also a factor.

It was found that plaster mold castings of aluminum made by the Antioch process* could be reliably expected to be helium-tight without additional treatment. The aluminum alloy used is known commercially as 355 alloy and is heat treated to a T5 condition. Castings to this specification are dimensionally stable and subject to little or no creep with age.

All castings used are tested at one atmosphere of helium and must meet a maximum leak rate of 5×10^{-8} std. cc He/sec. Stress was not an important factor since minimum wall thickness permitted by the casting process resulted in a unit easily capable of withstanding pressures of 100 psi. The normal pressure load in service is 3 psi positive pressure within the housing.

The complete inner unit, called a repeater unit, is made up of five convenient subassemblies, each with a major circuit function: two similar power separation filters, two similar directional filters and an amplifier, as shown in Fig. 11.

6.7 *Amplifier*

The amplifier is the most critical of these major designs and requires very careful layout. The redundancy of the circuit design, with two paralleled μ circuits and a common β circuit, pointed to a preferred basic layout of two similar chassis separated by a space required for the β

* A proprietary process involving special methods for handling plaster and sand for the molds. In addition to gas tightness, this process results in castings of die cast quality with respect to surfaces and dimensional accuracy, requiring very little machining.

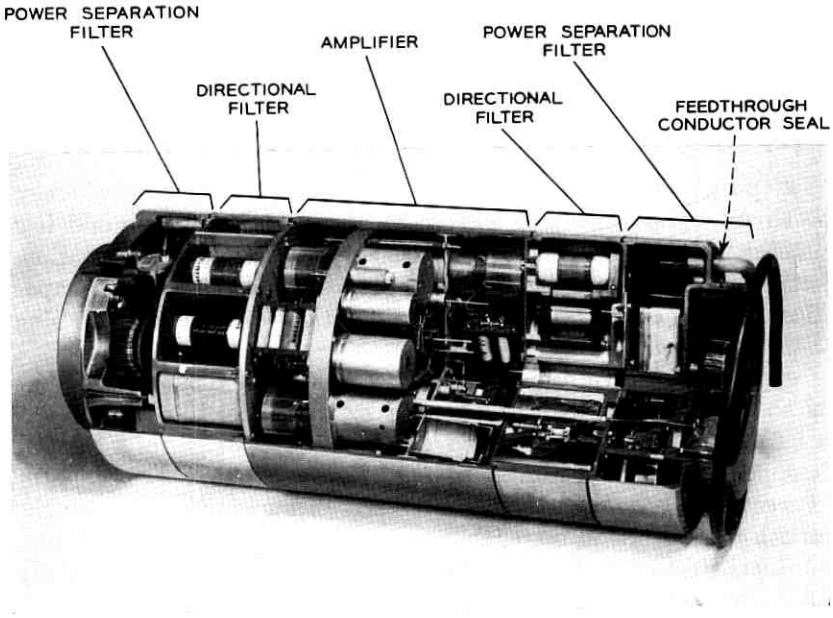


Fig. 11 — Repeater unit — cutaway view.

network. The space thus created is utilized to accommodate other components, such as large bypass capacitors. In manufacture, the basic layout permits separate assembly of each amplifier chassis with a maximum of accessibility and inspectability. Fig. 12 shows the two amplifier chassis before assembly into a complete amplifier, shown in Fig. 13.

Maximum reliability and uniformity have been designed into the amplifier based on the following rules: components must not be lead supported; component leads must not be bent, as the bending process could result in nicks or strains and therefore incipient ultimate breakage; components as well as wiring must be held in position; and their locations must be fixed and uniform from assembly to assembly. To meet these requirements, all amplifier networks, such as interstage and coupling networks, are of the "cordwood" design, with each component supported between two cavities in molded plastic plates. Wiring strips of known quality and thickness are used to interconnect components. All wiring strips are spaced 0.080 inch away from the plastic plates to afford visibility for the inspection of both sides of the soldered joints and to minimize burning hazards to the plastic during soldering. The strips are plated with 40 microinches of gold to facilitate soldering with a minimum of flux.

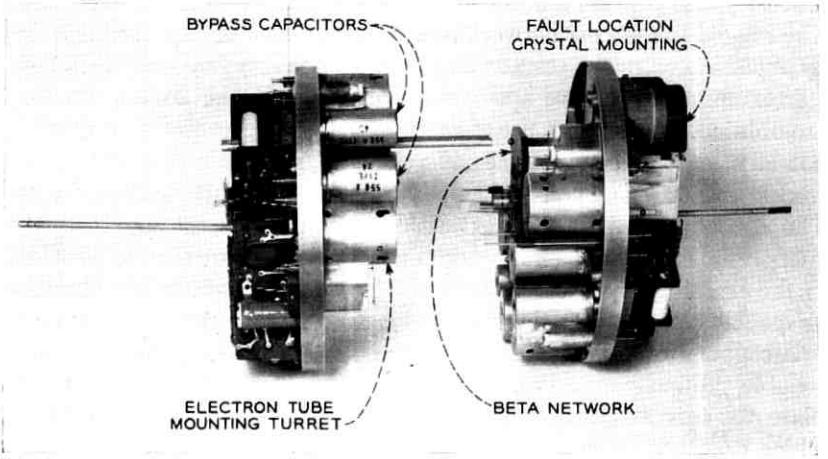


Fig. 12 — Amplifier before final assembly.

The plastic material, diallyl phthalate, used for the network plates was chosen for best combined mechanical and electrical properties and a minimum amount of corrosive outgassing, a necessary attribute for the protection of components in a closed housing for extended periods of time. Spacers are aluminum, and in many cases are dual purpose in

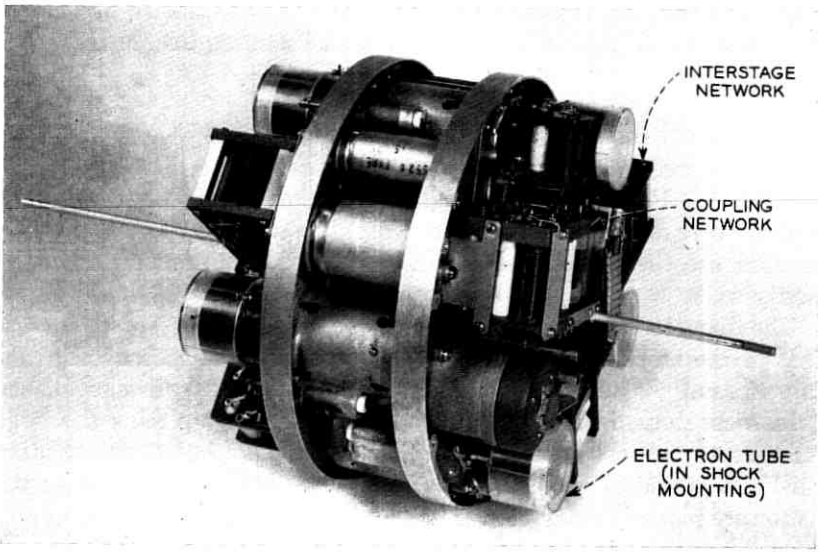


Fig. 13 — Assembled amplifier.

that they also provide a means for mounting the network on the chassis. The ruggedness of this network assembly has been amply demonstrated by punishing vibration tests at 50 g over a frequency range of 20–100 cps.

Maximum ruggedness and design flexibility in the amplifier chassis are obtained through the use of castings. The two chassis are similar but not identical. There are some differences in the layout of coupling networks, and only one chassis is equipped with the fault location crystal. The electron tubes are supplied with individual housings equipped with rubber shock mountings. To accommodate these, turrets are provided in the chassis casting. These serve a dual purpose, as they are also used as spacing bars between the two chassis, affording a large area of support.

A copper grounding plate superimposed on the chassis is used to realize reliable ground connections to the aluminum chassis. The grounding plate also serves as a convenient means for tying down the power wiring, which does not require rigid control.

The final twin chassis assembly is very rugged, as attested to by no failures in wiring or assembly when subjected to the same rigorous vibration tests as individual networks.

The complete amplifier assembly is mounted into a cast "barrel." Special means are used to mount the unit with inside fastenings. It is necessary to maintain a smooth exterior on the over-all repeater unit assembly, and the fully occupied amplifier chassis left no room for internal bosses in the barrel. Consequently, "butterflies" are set in internal grooves in the barrel and equipped with screws to assure a tight fit regardless of the varying length due to manufacturing tolerances.

6.8 *Directional Filter*

Another major subassembly of the repeater is the directional filter (see Fig. 14). Each unit consists of a high- and a low-pass filter. The very large air core inductors must be surrounded by as much space as possible to minimize the reduction of Q due to the eddy current losses in the cavity walls.

The diameter of the directional filter containers must exactly match that of the amplifier. There are ten filter sections in each of the containers, and the ultimate configuration chosen was a division of the cylindrical shape into the ten cylindrical sectors. By placing the large inductor toward the base of the sector, maximum space to surrounding shields is provided, while the capacitors which make up the rest of the sector are placed toward the apex.

All ten sections are mounted on a base plate equipped with grooves

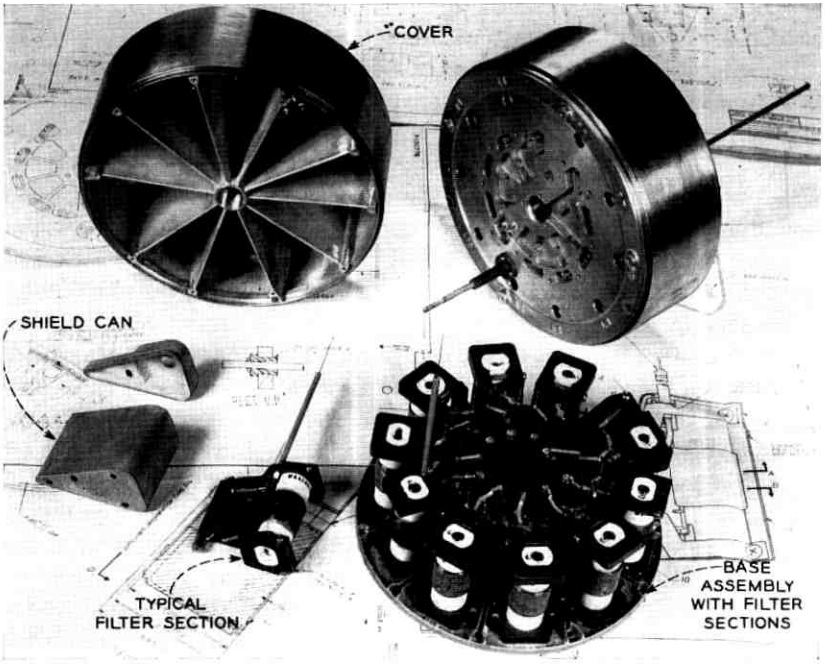


Fig. 14 — Directional filter.

between adjacent sections. The separating ribs in the deep cover match the grooves of the base. This tongue and groove arrangement, supplemented by a woven wire mesh insert, provides shielding between sections.

In the circular design the input low-pass filter section is adjacent to the input high-pass filter section. Crosstalk requirements are severe and are not met with the container shields alone. Additional copper shielding is therefore provided for both of these sections.

A unique feature is the arrangement of the interconnecting wiring. In normal filter designs, sections which must be tuned individually are tuned in equivalent cavities and then transferred to the final assembly. In this design terminals are carried through the housing and are not connected until after each section is individually tuned. Thus the effects of shifts in position with respect to cavity walls are avoided.

6.9 *Power Separation Filter*

The third major subassembly required for the repeater unit is the power separation filter. Because of large size and irregular shape, these

components were not amenable to the "cordwood" type of construction. Instead, components are fastened directly to the casting.

The circuit layout for the repeater unit is arranged so that high potentials appear only in the power separation filter. The components are physically arranged to provide a minimum of exposed high voltage wiring.

The power separation filter is the unit into which transmission and grounding leads are fed. Since the repeater unit forms a second barrier to water vapor, it is necessary to provide helium-tight seals for these leads. This seal consists merely of a center conductor, bulged slightly in the middle, and a copper beryllium shell with an outward bulge in the middle to match that of the center conductor for uniform thickness of insulation. Polyethylene is molded into the cavity between center conductor and shell.

The five major subassemblies are assembled into a repeater unit. Since it is necessary to keep the exterior of the unit smooth, all fastenings must be internal, except for those fastenings which can be put in from the end. The major problem in assembly is the attachment of the directional filters to the amplifier assembly. This is accomplished by providing an expansible stainless steel ring mounted in an internal groove in the barrel. This ring, which is inserted after the amplifier is in place in the barrel, is equipped with welded-on bosses, substituting for the bosses which could not be cast in place. The ring is kept in the expanded position by a double wedge, driven and locked in place. Three rods are threaded into the bosses and carried through the directional filter, which is held fast by nuts on the far end of the rods.

The two directional filters are cross connected by shielded coaxials which are carried through the amplifier. Careful coordination of the two designs was necessary to effect a straight line path for these rigid pipes. The shields (copper tubes) are mounted in the amplifier section and serve as assembly guides for the coaxials themselves, which are necessarily appendages to the directional filter. Since the expanded mounting band may be oriented as required, it can be positioned to the requirements of any individual assembly, avoiding the difficulties resulting from manufacturing tolerances. Other coaxials from the amplifier (input and output) are merely carried through a center hole in the directional filter, as are the power leads. These input and output coaxials must be kept straight during assembly, but must be bent to make connections on the cover side of the directional filter. To preserve the integrity of the solid return on these leads, an electroformed bellows is interposed between two rigid tubes, permitting bending to assembly

requirements. The amplifier, assembled to both directional filters, is a major subassembly of the repeater unit.

6.10 *Repeater Unit Assembly*

Inasmuch as the over-all repeater unit is an assembly of five major units and two end caps, there are six joints between castings which must be sealed. O rings are used to effect these seals. Here the O ring is not used in a normal application but is used as a gasket confined in an open dovetail groove (see Fig. 15). The design permits sufficient rubber flow to assure adequate electrical contact between units and sufficient gasket pressure to assure sealing under either internal or external pressure loads.

The sealed unit is dried and tested for leaks, but since the plastic materials will continue to outgas even after the drying operation, precautions are taken to insure low humidity for the entire life of the unit. The possibility of condensation as a consequence of a temperature drop when the power is removed makes such a precaution important. A glass-sealed desiccator mounted directly below the tubulation in the power separation filter provides the required control.

The final seal-off of the tubulation is done while the unit is pressurized with nitrogen (see Fig. 16). A cup-like aluminum plug inserted into the tubulation hole is used for this purpose. A tapered steel pin is forced into the cup to expand the walls. The hollow seal-off plug design includes a pin at the forward end. This pin is arranged to pierce the glass seal, thus activating the desiccant simultaneously with seal-off.

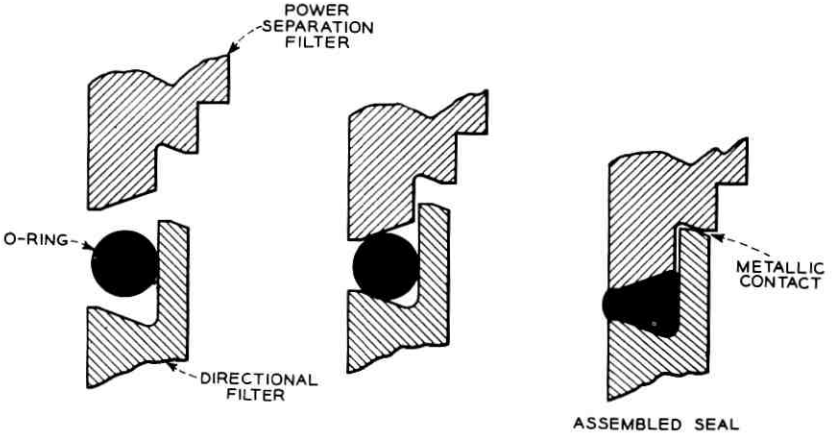


Fig. 15 — O-ring assembly.

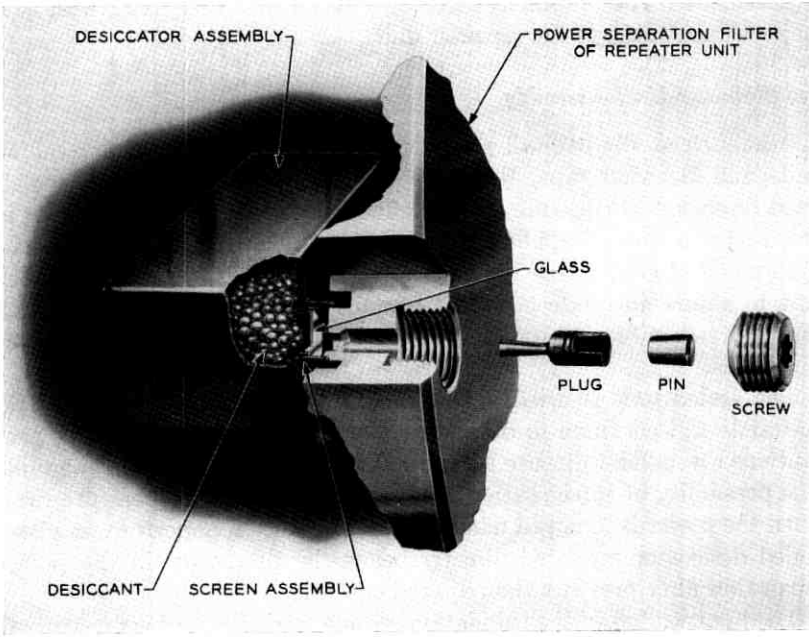


Fig. 16 — Final sealing of tubulations.

To provide electrical insulation, the entire repeater unit is covered by a wall of epoxy 0.140 inch thick, cast in place. The epoxy is mica-filled and the mixture is adjusted to have a coefficient of expansion approximately equal to that of aluminum. A flexibilizer is added to prevent damage to the coating as a result of temperature cycling during manufacture and subsequent handling. The exothermic heat generated during the epoxy curing is not sufficient to raise the temperature of components within the repeater unit beyond safe limits.

The over-all dimensions of the completed repeater unit are:

length	23.30 inches
diameter	9.54 inches
weight	65 lbs.

6.11 *Shock Mounting*

It is not desirable that a heavy repeater such as this require handling with more than ordinary caution. Therefore, the repeater unit is spring mounted within the pressure housing for shock absorption.

There are twelve single-leaf springs running the entire length of the repeater housing. Each spring is curved in section and placed between pressure housing and repeater unit with the edges toward the pressure housing. The springs are made of 5-ply epoxy-bonded fiber glass, cured to the required curved shape. The diameter of the circle tangent to the inserted springs is 0.05 inch less than the diameter of the repeater unit, so there is a preloaded condition when the repeater unit is inserted.

Endwise shocks are absorbed by flat multi-ply epoxy-bonded fiber glass springs mounted in the high pressure cover. These are positioned so that when the covers are in position against the shoulders in the high pressure cylinder, they also are preloaded.

With the repeater unit inserted into the spring assembly, the leads from the high- and low-pressure seals are joined by welding the conductor 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 against the springs during shock conditions.

6.12 *Complete Repeater*

After the high-pressure covers are set and welded the repeater is complete. Two small polyethylene-insulated flexible ground leads are then spot welded to the high-pressure cover. Insulated leads are used in lieu of a copper shield braid which might be attacked by corrosion. These leads are ultimately run directly alongside the center conductor lead and approximate the performance of a coaxial structure.

Completed repeaters are shipped from the factory in a special package designed for protection against severe shock. The repeater body is carried in two end sockets made of rigid foam and mounted on a palette. The density of the foam is controlled so that under shock conditions greater than 25 g the foam will deform, limiting the shock to the repeater. Any shocks to the repeater in transit from factory to shipboard are recorded on an impactograph mounted at one end of the repeater.

The permissible temperature range of the repeater is limited by the oil-filled high-voltage capacitors. It is therefore necessary to specify the range of temperature exposure during transit or storage. To assure that the allowable range has not been exceeded, a test tube indicating thermometer has been developed and mounted in the other end of the repeater. This test tube thermometer will burst from freezing below 0°F or from expansion above 150°F. A delay time is built in so that short exposures at either extreme will not cause breakage.

The foam-blocked repeater on its palette is covered by a conventional package of sufficient strength to permit stacking.

6.13 *Cable Connection*

In order for the repeater to become part of a system it is necessary that it be mechanically and electrically joined to the cable. The bulk and weight make handling in normal cable loading lines difficult, and the design therefore permits the loading of repeaters as cargo separate from the cable. Thus a termination applied at the cable factory may be joined to the repeater on board ship to provide an orderly and systematic arrangement for subsequent overboarding.

Since most of the cable is not armored, the design of the termination or coupling is primarily directed at this type of cable. The coupling design for armored cable is similar with few changes in parts (see Fig. 17).

The center strength member of the cable is a strand comprising 41 high-strength steel wires. In order to effect a satisfactory mechanical cable termination, it is first required that this strand be gripped to its ultimate strength without slip. A sleeve of AISI 1141 steel, $4\frac{13}{16}$ inches long by $\frac{3}{4}$ inch OD, 0.333 inch ID, has been developed for this purpose. This sleeve is pressed over the strand into a hexagonal exterior shape 0.658 inch across the flats. Press force is in the order of 500 tons. An epoxy coating over all of the wires prevents slip before ultimate strength is developed. The assembly develops a strength of 17,000 pounds ultimate.

The copper overlay of the steel strand is stripped back for most of the sleeve length. The internal diameter of the sleeve is contoured at the cable end to provide a good grip to this copper without crushing or tearing. The opposite end of the sleeve is threaded for further assembly into the coupling anchor.

The terminated steel strand is the center conductor of the cable and as such must be insulated. The polyethylene used as cable insulation is desirable to avoid discontinuities. Polyethylene, however, is a soft material without much shear or tensile strength. It will also readily cold flow. Therefore the strand termination has been arranged to compressively load the insulation and encapsulate it so as to restrict cold flow. A large disk or anchor, is threaded onto the pressed sleeve. The anchor is premolded in polyethylene and ultimately closely confined in a copper beryllium housing which forms the return conductor of the coaxial structure. A flexible polyethylene-insulated lead is attached to the repeater end of the premolded anchor assembly for later attachment to the repeater.

The greatest cable tensions, which could cause cold flow of the polyethylene, are applied only for short periods during laying and recovery. There are, however, ocean-bottom conditions which could result in con-

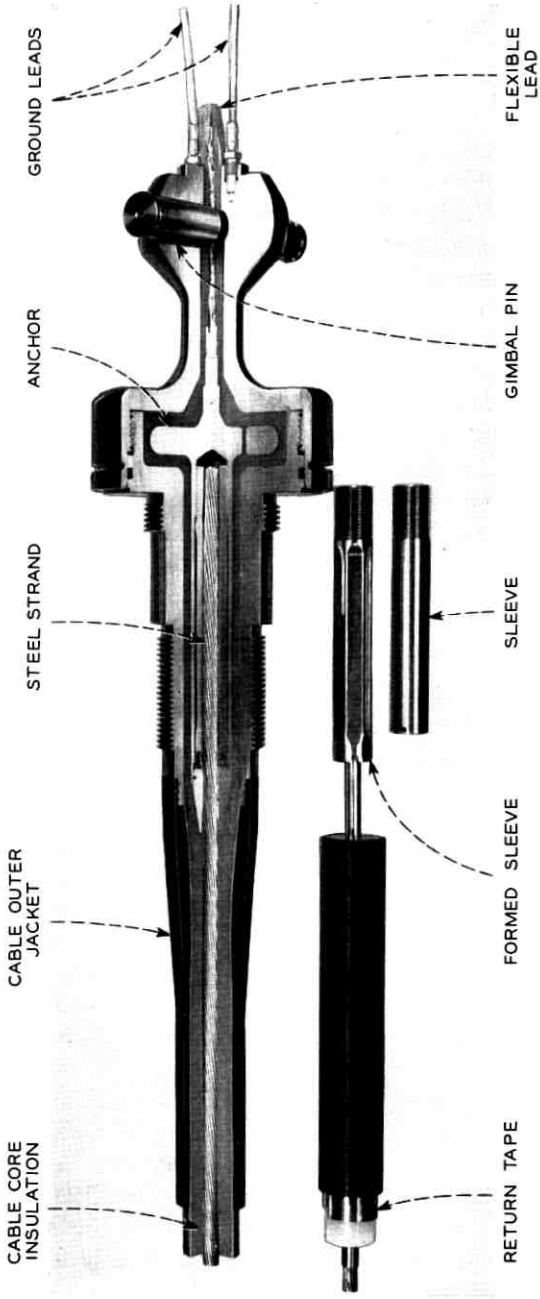


Fig. 17 — Repeater-cable coupling.

stant cable tensions regardless of the care with which the cable is laid. Should a repeater with a specific gravity of about 4.0 be laid in silt, the lighter cable (sp. gr. 1.1) would float while the repeater would sink, resulting in tensions up to an estimated 1,500 pounds. The method used to confine the polyethylene has been found to be effective against cold flow under this value of tension.

Armorless cable with a polyethylene outer jacket and rather delicate copper tape return conductor is subject to damage from any severe bending. To best accommodate normal shipboard operations, flexibility has been added by the provision of a gimbal ring as a coupling between repeater and cable. The gimbal ring has been designed to provide 45 degrees of free motion in any direction from the axis of the repeater. Thus the cable is maintained in a straight line when the repeater-cable assembly is bent to any diameter of 8 feet or larger.

Further protection for the cable against damage is afforded by a bell-mouthed gland nut and a rubber boot which provides a tapered stiffness type of support. Tests have shown that this assembly combination may be satisfactorily run over sheaves with pitch diameters as little as 7 feet under laying or recovery tensions.

The gimbal ring is mounted in a copper beryllium casting provided with a bolt circle matching that furnished in the repeater end attachments. The gimbal ring assembly is protected from the incursion of large stones or shells by the addition of a surrounding rubber bellows. This arrangement assures freedom of operation in any later recovery operation.

Couplings are assembled to precut cable lengths at the cable factories. Two polyethylene patch molding operations resembling cable splicing molds are required. In the first, the cable core insulation is patched to the insulation around the anchor, providing continuous polyethylene center conductor insulation from cable to flexible pigtail.

Subsequent to the core patch mold, the return tape of the cable is joined to the anchor housing by a brazing operation. Finally, the outer jacket is restored in the second molding operation. The anchor housing is directly exposed to sea water, and termination of the restored cable jacket must be designed to protect the cable return conductor from corrosion. Annular grooves in the anchor housing into which the molded material flows and shrinks provide this protection.

The armored cable required for shallow water applications moves the strength member from the center to the outside of the cable. The basic coupling design used for armorless cable is readily adapted to this difference in cable structure by a few part changes.

Sufficient resistance against bending is provided by the stiffness of the armor wires, and both gland nut and rubber boot are eliminated. A slotted ring threaded onto the anchor housing is substituted for both parts. A single armor wire, terminated in a crimped sleeve, is laid in each slot, providing a termination equal in strength to that of the armored cable.

Inasmuch as the armor wires are wound helically on the cable, and the center conductor is solid copper, tension in the cable will result in a longitudinal displacement of the core and center conductor with respect to the armoring. There will also be relative twist. This condition requires another departure from the basic concept used in the armorless coupling. The anchor is eliminated and the cable is terminated in a molded assembly including a center conductor and two ground leads. The assembly can slide freely in the neck of the housing. The two ground leads are run in a spiral in the anchor cavity. The spiral configuration accommodates any relative twist, while the ability to slide accommodates longitudinal displacement.

Terminated cable lengths of either armored or armorless type are ultimately connected to repeaters in a like manner on shipboard, to complete the system.

VII. CONCLUSION

The first SD submarine cable system was put in service between Florida City, Florida and Jamaica in February 1963.

Several other systems have since been laid. By early April, 1964, 502 repeaters and equalizers were operating at sea bottom; at that time 640 units had been manufactured.

All repeaters and equalizers are performing as predicted and, as indicated by measurements on several installed systems, system requirements were met.[2]

REFERENCES

1. Special transatlantic cable issue, B.S.T.J., **36,** Jan., 1957.
2. Ehrbar, R. D., Fraser, J. M., Kelley, R. A., Morris, L. H., Mottram, E. T., and Rounds, P. W., The SD Submarine Cable System, B.S.T.J., this issue, p. 1155.
3. DeSoer, C. A., Four-Port Constant R Network Required by a Two-Way Single Amplifier Repeater, Trans. I.R.E., **CT-5,** Dec., 1958, pp. 276–283.
4. Bode, H. W., *Network Analysis and Feedback Amplifier Design*, Princeton, D. Van Nostrand, 1945, p. 33.
5. Bohman, V. F., and McKenzie, R. W., Selector for Underwater Cable Equalizers, Bell Laboratories Record, **42,** April, 1964, p. 124.

Manufacture of Rigid Repeaters and Ocean-Block Equalizers

By S.G. JOHANSSON

(Manuscript received January 22, 1964)

Manufacture of two-way rigid repeaters and ocean-block equalizers at a rate equivalent to approximately 7000 miles of cable a year required the establishment of a new plant. A suitable building was constructed and equipped to maintain the closely controlled environmental conditions essential to attaining the quality required to assure a minimum product life of 20 years. An organization was established and trained to operate the plant. Special facilities were developed for manufacture and testing, including automatic readout and transmission of the test information to a data center where the results are recorded in a punched card system and analyzed by machine.

The facilities, their use, production methods from procurement of materials to packing of completed repeaters and equalizers, and precautions taken to attain and assure the required high quality are reviewed on the following pages.

I. INTRODUCTION

Early in 1959 a decision was made to start production of a new rigid repeater for SD submarine cable telephone systems.[1]

Since repeater manufacturing know-how was available at the Kearny, New Jersey, plant of the Western Electric Co. from earlier production of flexible repeaters for SB cable systems,[2] the manufacture of the SD repeater was also allocated to Kearny. The same general philosophy of building integrity into the product to the limit of practicability was again to apply. The same manufacturing requirements on environment, personnel, wage payment, training program, and inspection on a 100 per cent basis were also to be used.

In the following information reference will occasionally be made to flexible repeater manufacture at Hillside, New Jersey, with the assumption that interested readers have previously read or now have access to the information published in 1957.[2]

II. PLANNED CAPACITY AND DELIVERY SCHEDULES

The manufacturing planning was based on a capacity of six repeaters per week on a one-shift five-day basis. It was estimated that this capacity would permit production of up to eight repeaters per week on an all-out basis when all processes had been shaken down and operators fully trained.

The initial estimated manufacturing interval was 90 weeks, which should be reduced to between 60 and 65 weeks when production got underway with fully-trained operators. In reality, operations were started by the middle of September, 1960, and the first repeater was shipped by the end of June, 1962. This initial interval amounted to 92 weeks. The normal manufacturing interval is about 63 weeks, the greater portion of which is used for temperature cycling and aging of components.

Schedules called for delivery of 66 repeaters by the end of 1962 and, during 1963, 108 in the first quarter, 110 in the second quarter, 83 in the third quarter, and 123 in the fourth quarter. To meet these schedules required spurts in the output rate of up to ten enclosures per week. This was accomplished by earlier start of apparatus production and more rapid build-up to a rate corresponding to eight repeaters per week. By providing some additional testing capacity and extending enclosure operations into a second shift, it was possible to assemble repeater units and enclose repeaters at the higher rate required to meet over-all schedules. Nearly a year was required from the time of shipment of the first repeater to build up to maximum production capacity.

III. HIGHLIGHTS OF CLARK BUILDING

From the design information available at the time authorization to go ahead was received, a rough shop layout was made. Based on Hillside experience and the nearly sevenfold increase in quantity of apparatus to be manufactured, a minimum need for 85,000 square feet of floor space was indicated.

After inspection of several buildings contractual agreement was entered with a contractor to build and lease to Western Electric a 97,000-square foot building at Terminal Avenue, Clark, N. J. This location was deemed satisfactory from the standpoint of availability of the required type of labor within the nearby areas. The building was designed and built to meet the specifications for a repeater shop and other Western Electric standards.

Construction of the building was started in June, 1959, and completed in July, 1960. Only the land and building shell were leased from the

contractor. All installations of lights, drop ceiling, special rooms, floor tile, heating boilers, air-conditioning equipment, and cafeteria facilities were provided by Western Electric. Although the building was originally acquired on a ten-year lease with option to buy, it was purchased by Western Electric in July, 1962.

The building is of one-story steel and masonry construction, 360 × 270 feet, with 40 × 30 foot bays. In office areas and "white" (i.e., super-clean, controlled-environment) areas, ducts for power, telephone, and other communication circuits were installed on ten-foot centers, with knock-outs every two feet, in the concrete floor slab. A sheet rock drop ceiling was installed 11 feet above the floor. All air-conditioning ducts, piping, and other services were installed in the six-foot space between the roof and drop ceiling. In office areas and cafeteria, acoustic ceiling tile was used. To avoid pipes in the open, water, compressed air, deionized water, dry nitrogen outlets, and outlets from the central vacuum cleaning system were, wherever possible, installed in partition walls and inside the covering over the building columns. The back wall of the building was constructed with future expansion in mind. The columns are provided with plates for connecting joists to them, and the wall is cinder block construction for ease of removal in part or whole. A general block layout of the building is shown in Fig. 1.

To meet the New Jersey manufacturing building code without having to erect internal firewalls which would have interfered with the general layout, ten-foot-wide main aisles are provided. These aisles would allow movement of standard mechanized fire-fighting equipment and thus provide the required protection.

The state code also specified heat vents in the roof, which would have required undesired openings in the ceilings of the white areas. However, mechanical fire baffles in the air-conditioning ducts and means for exhausting the return air completely to the outside in case of smoke satisfied this requirement and eliminated the need for openings in the ceilings.

Since eating in places other than the cafeteria is undesirable in a plant of this type, the size of the cafeteria had to be increased somewhat over the normal requirements for the plant. The cafeteria seating capacity at Clark is 354, with facilities for cooking, dishwashing, etc.

The conditions of temperature, humidity, and ventilation are provided by five separate air-conditioning systems with a total capacity of 600 tons and with one common chilled water supply for cooling. Two of the systems, each with a large plenum chamber which includes the filtering equipment and a circulating blower, cover the main portion of the building. The filtering equipment consists of a mechanical filter

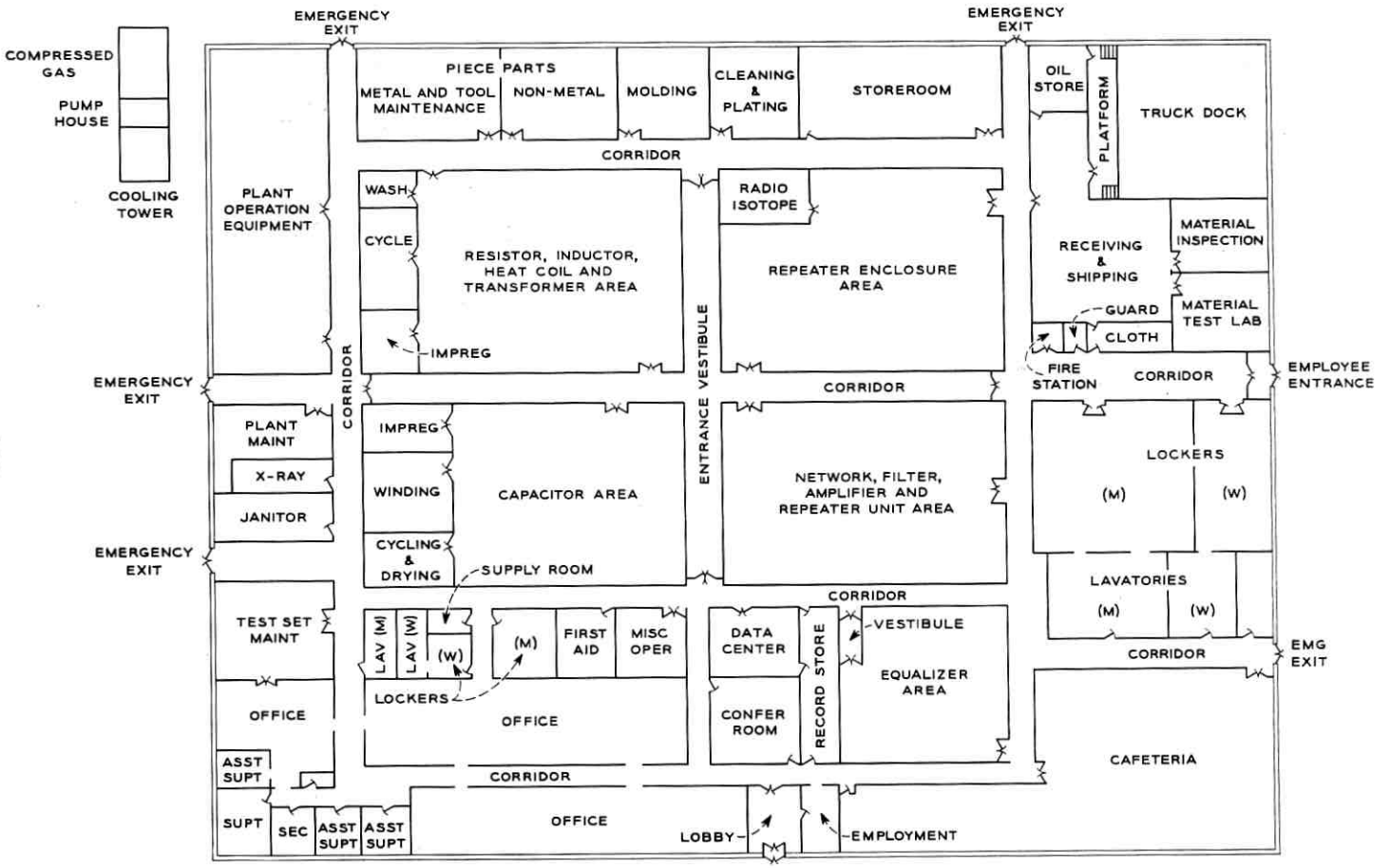


Fig. 1 — Plant layout.

followed by an electrostatic filter and a second mechanical filter. Originally the design called for a cold water spray for cooling the air, which would have been simple and efficient and would have stabilized the humidity at the desired value. However, this proved unsatisfactory due to minerals and algae in the water, which were precipitated upon evaporation and circulated out through the ducts in the form of fine dust. To attain the desired low dust count, a change was made to the use of cooling coils in place of the water spray; the change eliminated direct contact between the chilled water and the air and the associated evaporation and precipitation. The air is circulated by two blowers in series: one 40-hp unit in each plenum chamber, which provides positive pressure in the supply duct, and a 15-hp unit at the end of the return duct in a penthouse on the roof above the entrance to each plenum chamber, which provides a negative pressure in the return duct. The latter blowers also provide the means of exhausting directly to the atmosphere any smoke or fumes from fire or other causes through a system of louvers operated by a photoelectric smoke-detecting device. One of these main systems serves the four large white areas in the center of the building, shown in Fig. 1. The second system provides air conditioning for all of the peripheral areas, including offices but excluding the cafeteria. Means are also provided to interchange these systems.

Since smoking is permitted in the cafeteria, a separate system is used, which recirculates and provides the necessary make-up air for the cafeteria only. The air supplied to the locker and toilet rooms is not recirculated. To maintain uniformity of temperature during severe outside climatic conditions, means are provided for heating or cooling the air above the drop ceiling. In case of extreme cold, the space above the drop ceiling is heated by steam coils installed in several locations. To attain the desired low humidity of 20 per cent maximum in the capacitor winding area, a secondary cooling system is used which takes the approximately 55-degree air from either of the large plenum chambers and cools it down to approximately 22°F to precipitate the moisture. Since this is below the freezing point of water, dual cooling coils are provided: one is defrosted while the other one is in use, and vice versa. This switching is done by an automatic timing device.

The cooling facilities consist of three separate 200-ton centrifugal refrigerator units with chilled water outputs connected to a common header. The chilled water is retained in a closed-loop circulating system. A common cooling tower is used for cooling the water from the individual condensers. This arrangement permits one, two, or three of the machines to be used as required, resulting in good efficiency for the varied load

encountered during the year. Except during the very few days of extreme high temperature and humidity encountered during the year, two of the units are sufficient to carry the load. This provides the required standby capacity for maintenance and repairs.

Two 200-hp boilers provide the necessary steam for heating and humidifying during the cold season and for dehumidification during hot weather. (To remove excess moisture, the air is cooled to approximately 45°F and then reheated to meet temperature requirements in the different areas.) One of these boilers is sufficient to carry the load, except possibly during a few days in the year when extreme temperatures may be encountered. During the winter, steam heat is also provided under the outside windows throughout the plant to offset cold air circulation and draft.

Approximately $2\frac{1}{2}$ miles of ducts are installed above the ceiling for circulation of the air. The supply air is admitted to the rooms through the outer opening of coaxial diffusers, and the return air is pulled back through the center opening. The circular return ducts from the diffusers run right through the supply ducts to the return ducts, which are located directly above the supply ducts. Sufficient blower capacity is provided to change the air in the building completely every five minutes. Up to 25 per cent of make-up air can also be provided. This is adjustable to the comfort needs of the building population.

The control circuit for temperature and humidity is air-operated. To ascertain that the air is sufficiently dust-free, a dust count is taken every morning in a number of different locations. A special Bausch and Lomb microscope is being used for this purpose. In making a test, air is pumped past a slide where any dust particles stick to a predetermined surface area. From the amount of air that passes this surface and the number of dust particles found on it, the average number of particles per cubic foot is obtained. In general, the dust count at Clark is comparable to the conditions obtained at Hillside and is considered quite satisfactory.

The over-all illumination of 85 foot-candles is provided by means of fluorescent lights in the ceiling. Certain of these lights are on a special emergency lighting circuit supplied automatically by a motor generator set in case of power failure. Provision has not been made for operating the plant in case of general power failure, since the power requirements are nearly 1500 kva. However, to assure maintenance of power, the plant is connected to a power loop which can be fed from several points.

The floor covering consists of one-foot squares of vinyl asbestos tile, which is a compromise with the original plan of using long lengths of wide material to avoid the possibility of dirt catching in the cracks

between tiles. The tile has, however, proven quite satisfactory, since it has expanded and nearly eliminated all openings between tiles.

The walls or partitions for the different rooms in the building are commercial standard steel partitions having a baked enamel finish. Since these partitions were not of sufficient height to reach the ceiling, they were extended by an upper fabricated wall section of mineral boards. The partitions are insulated with rock wool, and the seams between panels are caulked to eliminate seepage of dust. Any openings at the floor are sealed by a flexible plastic base. To provide a more desirable environment, better possibilities for supervision and inspection, and, to some extent, an easy way to view operations, large windows were included in the partitions. For safety, this is wire-type glass.

To attain the best possible clean conditions in the white areas, the ceilings were covered with a washable vinyl fabric cloth. The same type of covering is also used on the portions of the walls above the metal partitions in these areas. For cleaning purposes, a central vacuum system has been included. Outlets are installed throughout the building for connection to the cleaning and scrubbing equipment. The acoustic ceiling tile in the office areas and cafeteria is a washable type which has a smooth surface without perforations. In aisles and locker rooms, the sheet rock ceiling is finished by painting. This also applies to the upper portions of the walls above the metal partitions in these and the non-"white" areas.

IV. CLARK ORGANIZATION

The same basic pattern as used at Hillside on the flexible repeater project, of having the organization report to the engineer of manufacture, is followed. However, since the operation is considerably larger, it is a full-time job for one superintendent, who is assigned to this project only.

There are four levels of supervision: superintendent, assistant superintendent, department chief, and section chief. At Clark there are three assistant superintendents — one responsible for engineering, one for inspection, testing and plant maintenance, and one for operating and production control. The responsibilities in each of these areas are, in turn, broken down into three departments with up to seven section chiefs in each, excluding engineering, which has no sections. On the average there are 16 operators reporting to each section chief.

The total population at Clark is 550, of which approximately 75 per cent are performing actual shop operations. Of the latter, approximately 40 per cent are female and 60 per cent male.

V. REPEATER ASSEMBLY

An outline of the major assembly steps from apparatus to completed repeater is shown in Fig. 2.

VI. TYPES AND QUANTITIES OF APPARATUS

In each repeater there is a total of 201 items of apparatus such as capacitors, inductors, transformers, resistors, heat coils, crystal units, electron tubes, and gas tubes. Crystal units are manufactured at Western Electric's Merrimack Valley Works and electron and gas tubes at the Allentown Works. Composition-type resistors and vitreous enameled resistors are purchased from outside suppliers. The remaining items, 86 codes, are manufactured at Clark; samples are shown in Fig. 3. A tabulation of apparatus used in the repeater is shown in Table I.

In addition to the items used in the repeater, there are 36 capacitors of 33 codes, 35 inductors of 31 codes, 53 resistors of 42 codes, and a stepping switch for the equalizer.

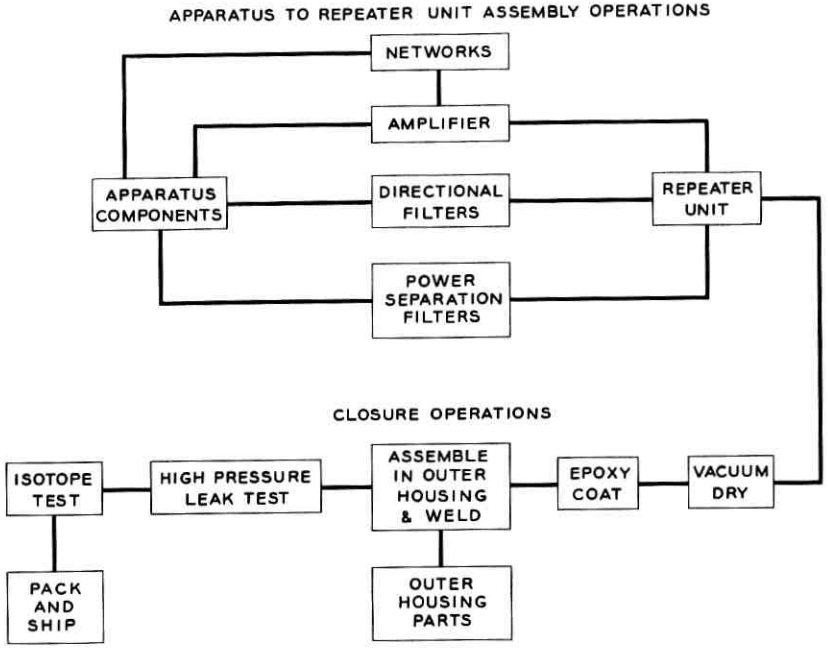


Fig. 2 — Major repeater assembly steps.

In general, the type of apparatus is similar to that used in the flexible repeater. However, the method of manufacture and the detail design in a number of cases are quite different.

VII. TYPE OF MANUFACTURING FACILITIES

A number of new manufacturing facilities were developed.

7.1 *Machinery*

Of the machinery for producing apparatus, capacitor winding machines, mica lamination silver coating machines and washing facilities, and some of the coil winding machines were new. After requesting bids from several companies, a manufacturer specializing in capacitor winding machines was selected to design and build machines to our specifications. Two types of machines, one for high-voltage capacitors and one for low-voltage capacitors, were built to these specifications, which also embodied some of the manufacturer's own outstanding features. The latter included a magnetic braking device to apply the proper tension to the paper spools, dual winding spindles, and electronic means for determining the foil length. The high-voltage capacitor winding machines, shown in Fig. 4, have as many as 18 spindles for paper and foil, which require, among other things, very close mutual line-up and tension adjustment to produce the required quality of winding.

An adaptation of a commercial machine was used for silk screening the silver coating on mica laminations. Better mechanical alignments were required, the right squeegee material had to be found, and the silk screens had to be made to closer tolerances and standards than ordinary. The originally planned number of openings in the screen and the number of parts coated at one time also had to be decreased, because it was not possible to maintain close alignment between the different openings in the screens.

New washing facilities for mica laminations were also developed. At Hillside, acetone was used for washing laminations, but with the larger quantities involved at Clark, the greater amount of acetone needed would have produced too great a fire hazard. A method using hydrogen peroxide and centrifugal drying was therefore selected. The new washing facilities are safe and require less space.

The design of coils for the one-megacycle repeater permitted all of them to be machine wound, whereas for the flexible repeater the majority had to be hand wound. Commercial machines are used for straight

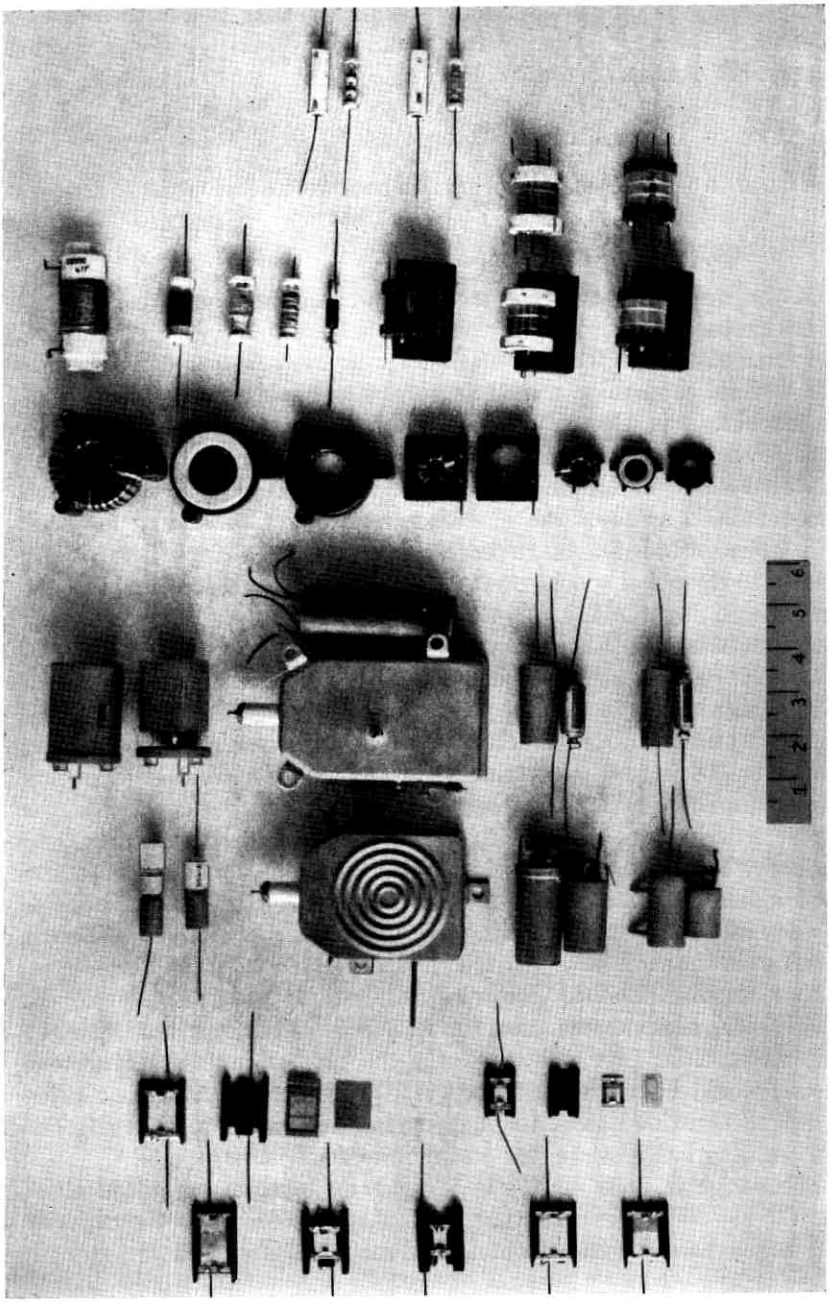


Fig. 3 — Sample parts.

TABLE I — APPARATUS USED IN REPEATERS

Part	No. of Different Codes	No. of Units	Value Range	Accuracy Range
Capacitors				
High-voltage paper	2	3	0.074–0.120 mf	3–5%
Low-voltage paper	9	22	0.001–2.0 mf	2.5–5%
Polystyrene film	6	12	0.0075–0.0192 mf	1–2%
Mica	24	46	0.000050–0.010350 mf	0.2–2%
Inductors				
Air core solenoid type—adjustable	15	24	3.45–415 μh	0.7–1.5%
Air core solenoid type	11	16	0.34–1000 μh	0.5–3%
Dust core toroidal type	4	8	115–9200 μh	1–2%
Transformers				
Ferrite core	4	6	—	—
Resistors				
Bifilar wire wound	14	25	15–600 ohms	0.1–1.1%
Mandrellated wire wound	4	10	1100–3800 ohms	0.4–0.8%
			73.2 ohms	1.5%
Vitreous enameled type	2	3	890.0 ohms	3.5%
Composition type	4	16	11,000–500,000 ohms	2%

solenoid, duolateral, progressive duolateral, and toroidal windings. However, a number of refinements were required to provide the proper wire tension, wire guide, and uniformity of winding.

Operations connected with enclosure of the repeater also required a number of new type facilities. Among these are vacuum drying stations, epoxy coating facilities, repeater assembly machine, welding machine for end covers, ultrasonic testing facilities, high-pressure helium leak testing facilities, isotope testing facilities, repeater handling facilities including trucks and hoists, and polyethylene molding facilities.

Considerable simplification of the vacuum drying stations and lower cost were accomplished with the use of a heating blanket to provide the specified bake-out temperature for the assembled repeater unit. The blanket, which completely surrounds the repeater unit, is equipped with thermostatic control to maintain the temperature.

Considerable development work was done by Bell Laboratories on a process for welding the beryllium copper end covers into the beryllium copper housing cylinder. Based on the process criteria established by Bell Laboratories, specifications for a production welding machine were prepared.

The machine, shown in Fig. 5, has two turntables to hold the repeater in a vertical position and means for centering it on the turntable so that the welding area runs true. The pedestal, which holds the welding head and a milling head, can be turned to either of the two turntable

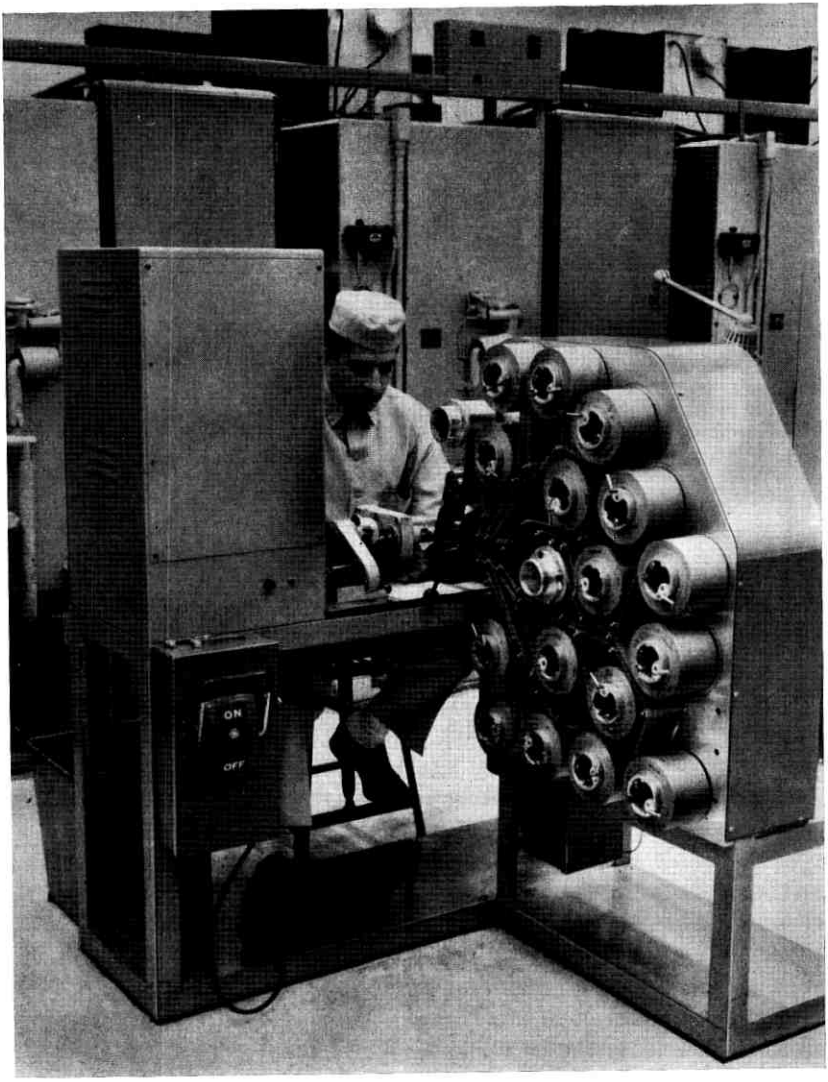


Fig. 4 — Capacitor winding machine.

positions. The milling head can be swung in position over the repeater welding area for removal of the surface layer to expose a clean metal surface for welding. The milling head is also used for removing the weld when repair is necessary.

The welding head uses nonconsumable tungsten electrodes with

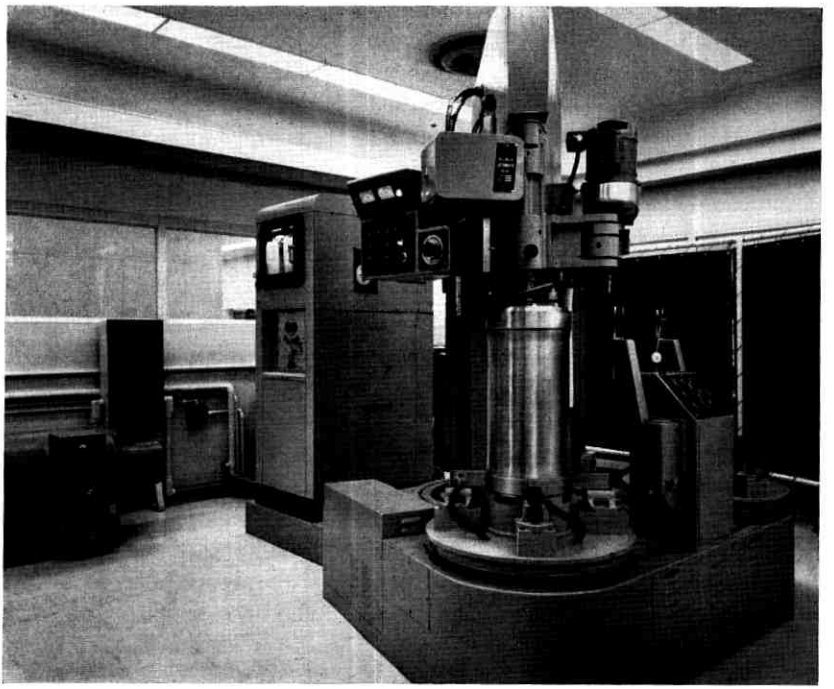


Fig. 5 — Repeater welding machine.

means for automatic control of the electrode gap. The current to the arc is furnished by a balanced-wave ac power supply which eliminates the dc component produced by the rectifying action of the beryllium oxide formed from beryllium in the copper alloy. A large dc component will reduce the stirring action produced by the ac current. A 30 kva regulated motor generator set is provided to eliminate the variations caused by the usual fluctuations in line voltage. The welding arc is shielded by an argon gas curtain.

The whole welding cycle, which takes approximately three minutes, is automatic. The cycle is started by pressing a button, and at the completion of the weld the current is tapered off to eliminate lap in the weld.

The high-pressure helium leak testing facilities will permit testing at pressures up to 12,000 lbs per square inch. Three-stage piston pumps are used with velocity and glass wool filters to keep the helium gas free of oil vapors from the pump lubricant. The pressure vessels are made from alloy steel and all gas-tight seals are made with O rings. The closure plugs for the vessels are held in place by heavy threaded sleeves which screw into the mouths of the vessels. To prevent extrusion of

the O rings as the clearance between the vessel and the end plug increases due to expansion of the vessel at the higher pressures, a double O ring arrangement with an auxiliary steel ring is used. The steel ring fits inside the vessel and rests against the closure plug. Between the ring and the vessel is an O ring, and between the end of the steel ring and the closure plug is another O ring. As the pressure inside the vessel is increased, the steel ring expands and is also pushed up against the closure plug by the pressure, thus maintaining a tight gas seal. The lower part of the high-pressure vessels for testing complete repeaters is below floor level in a pit, whereas smaller vessels for testing covers, etc. are above floor level. The high-pressure pumps and test vessels are surrounded by heavy boiler plate enclosures as a safeguard for personnel in case a fitting or pipe should give way. The enclosures are open at the top to permit rapid escape for gas. Access to the vessels is through heavy doors, also of boiler plate, which are interlocked so that pressure cannot be applied while the door is open and so that when pressure is on the door cannot be opened. The application of pressure, the operation of pumps, and the flow of helium are controlled from a common control console. All high-pressure piping interconnecting units in the test set-up is run in troughs in the floor which are covered with heavy steel plates. Hoists on overhead tracks are used for handling repeaters and the heavy fixtures used in the high-pressure testing operations. The hoists are arranged to cover the entire test area.

A lead-shielded room has been provided for isotope testing of the pinch weld on the tubulation which gives access to the inside of the repeater housing. The operation has been automated to a fair extent. The application of isotope solution under pressure and removal of it after a specified pressure holding period and following washing cycle are programmed and controlled by push buttons, eliminating the need for the operator to be close to the source of radiation. The set-up is shown in Fig. 6.

Special molding presses were provided for molding polyethylene in seals and anchor details for repeater couplings. These presses are of three sizes: 10, 15, and 50-ton clamping pressure. Means for programming the molding cycle and recording temperatures and pressures and timing for injection of the polyethylene have been included. Arrangement for stabilizing the temperature of the cooling water for the molds has also been incorporated.

7.2 *Electrical Test Equipment*

To avoid operator error in writing down or transcribing test results, test sets are equipped with automatic readout, and, in some cases,

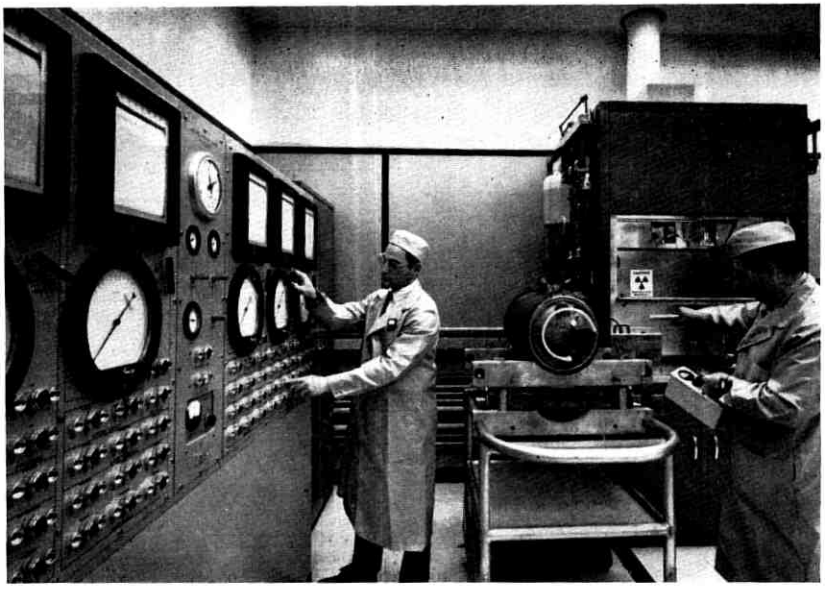


Fig. 6 — Isotope test facilities.

automatic balancing features. For example, in testing capacitors, the operator inserts a section of a punched card, punched with the type of test to be performed, apparatus code, and test limits, in the data transmitter. A smaller card section with the operator's identification and the test set identification number is also inserted in the data transmitter. The capacitor to be tested has a tag attached in the form of a small section of a punched card, punched with the serial number and the apparatus code. This card is inserted in a second data transmitting section after the capacitor has been connected to the test terminals. Next, the operator pushes a button which causes the bridge to balance itself rapidly and show the result on the dials. At this point the operator pushes a button on the transmitting panel which transmits the information on the three card sections, plus the reading of the test set, to a data center, where a receiver connected to a punched card punch receives the information and creates a card punched with the complete information transmitted, plus the time the data were received. In the data center the data card is compared to a standard card punched with the test limits and an indication whether the capacitor met requirements or not on that particular test. Upon completion of all the tests on an item of apparatus, the data from the individual cards are collected on a master card which is used for verification that the apparatus item

has gone through all of the process steps and has met all requirements. A computer is available in the data center for making calculations where needed and performing the more complicated verification operations. A diagram of the flow of information from the test sets to the data center and processing of the data cards is shown in Fig. 7.

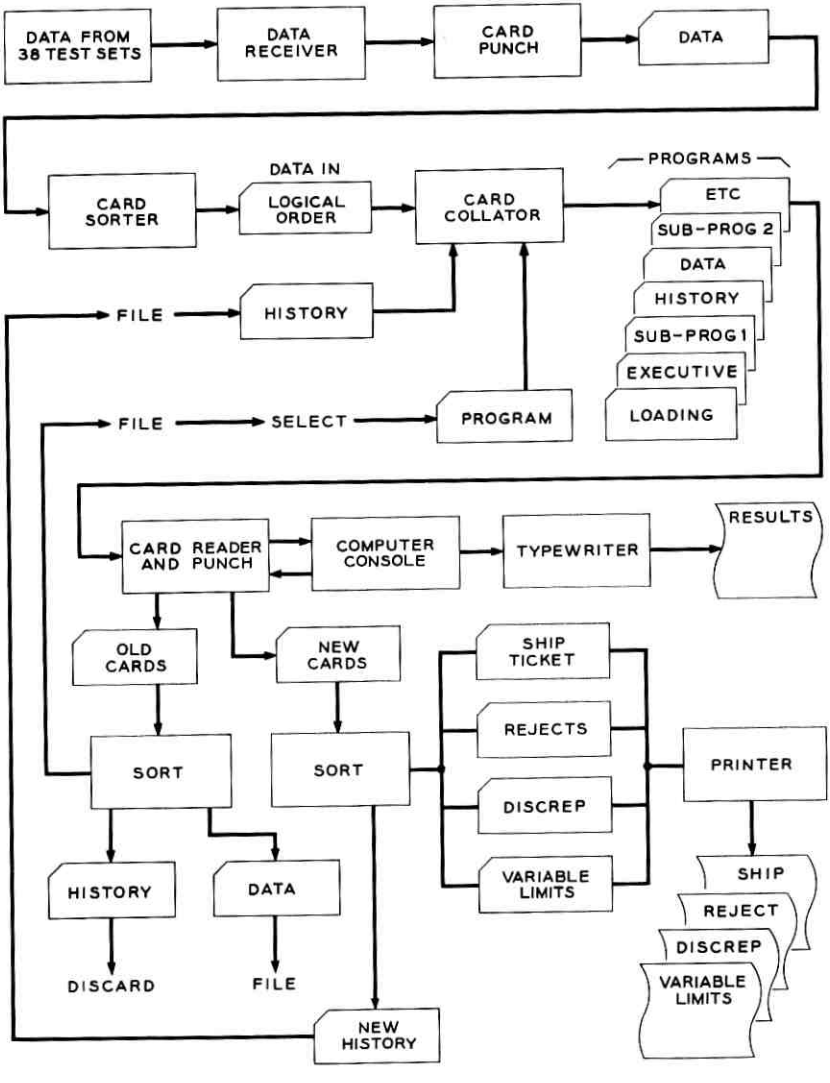


Fig. 7 — Flow of information from test set to data center, and processing of data cards.

A digital voltmeter method is used for measuring the insulation resistance of capacitors, since this method lends itself well to the required conditions of the test and is adaptable to read-out and transmittal of information to the data center. The insulation resistance measurements must, to be consistent, be measured at a predetermined time after application of the voltage, which must also be held precisely at the desired value.

Maxwell-type bridges capable of accuracies of ± 0.02 per cent are used for inductance and effective resistance measurements. This accuracy of measurements is obtained by limiting the range of the bridge and by careful selection of components and their over-all assembly in the bridge. Since only a limited number of frequencies in the 20-kc to 2-mc range were required, it was possible to use the "difference frequency" from fixed crystal heterodyne oscillators, which are simple and quite stable. The detectors used to indicate bridge balances are selective to the test frequency to eliminate errors from harmonics in the oscillator output.

Measurements of resistance on resistors are made by passing a known current from an exceptionally well-regulated current supply through the resistor and then measuring the voltage drop across it with a digital voltmeter. This accomplishes resistance measurements to within ± 0.01 per cent with a simple arrangement for transmittal of the resistance readings to the data center. A standard cell is used as reference in the voltmeter, and Zener diodes serve this purpose in the current supply.

For measuring the transmission characteristics of networks, filters, amplifiers, and completed repeaters, automatic transmission sets covering the 50-kc to 2-mc range are used. These sets are programmed to test at a certain number of frequencies over the desired transmission band. The results are sent in to the data center and recorded, and are also typed out on an automatic typewriter connected to the set. The programming of the set can be changed as required to perform tests at any frequency desired within the range of the test set. The test set, shown in Fig. 8, is housed in three floor cabinets, one containing the heterodyne oscillator and associated power supplies, the second the measuring circuit, and the third the indicating and readout equipment.

Phase and loss test sets capable of outstanding accuracy for measurements on networks, amplifiers, and repeater units outside the transmission band, at frequencies up to 30 mc, are constructed using a number of standard purchased units. These units, the majority of which are of German make, have performed very satisfactorily

For adjusting filter and other network circuits, visual return-loss

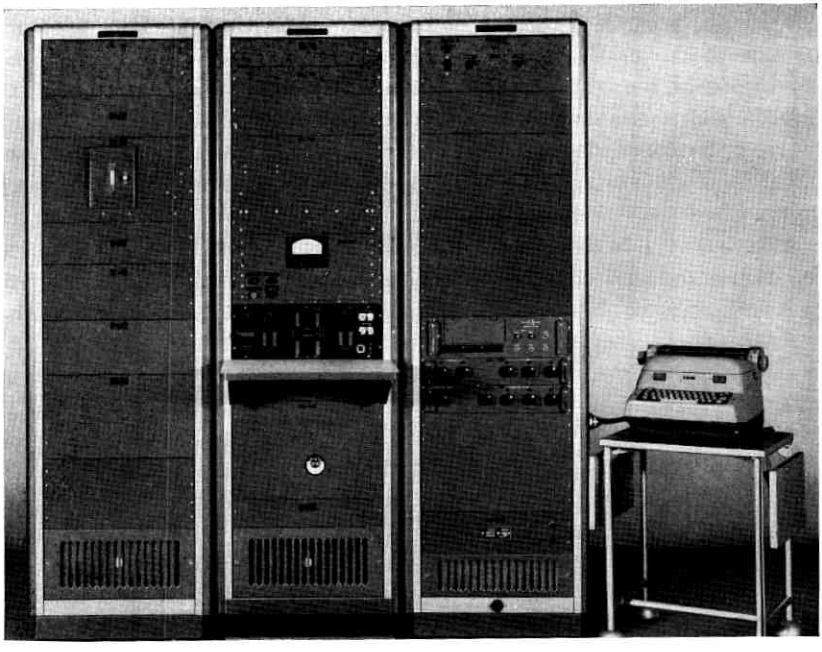


Fig. 8 — Auto-balance transmission measuring test set SID-100799, group 1 (front view).

sets have been provided. On these sets the transmission characteristics with frequency for the apparatus under test are displayed on a 17-inch picture tube.

Practically all test sets are provided with automatic readout, which requires a 150-conductor cable connection to the data center. The cables, which are run in the floor ducts mentioned earlier, are terminated in plug-in sockets under the test sets on the benches or the floor, depending on the type of test set.

VIII. JOB LEVELS AND PERSONNEL

Assignments among the people performing work directly on the repeater have, as at Hillside, been divided into nine different job levels, depending upon the requirement on skill and tour of duties. The lowest level includes assignments such as helper and materials handler, whereas the highest level entails duties which are much more complicated, such as those of the layout operator, whose assignments include operations requiring a long training period and a high degree of skill, and also include

some instruction and work assignment duties. At Clark the greater portion of the population is in the four lower grades, with a peak in the third level. This is quite different from Hillside and can be accounted for by a greater breakdown of operations at Clark. The considerably higher production rate of individual components at Clark permits full-time occupancy for one person on just a few operations.

The key personnel — such as engineers, supervisors, and a number of the skilled trades and layout operators — were obtained by transfer from Hillside or Kearny. Other personnel were hired locally, usually within the first three grades. As they acquired skill, they were upgraded into higher grades and trained for the new assignment. The training program used for operators at Clark is similar to the one employed at Hillside.

In establishing the job grades, the requirements and the description of the assignments to be covered were recorded. Based on this and the general knowledge of elementary operations and their "value," the jobs were classified within the nine assignments mentioned. General judgment was used, and the levels and grades were discussed among several supervisors in order to establish a grade which was as realistic as could be obtained.

In hiring personnel and evaluating them, they were interviewed by supervision, in most cases at several levels, and also given standard tests devised for screening applicants for Western Electric employment. In general, the selection has been good and there have been very few cases where employees unsuitable to the assignment have been hired. There did not seem to be any serious objections to the specific work rules applying at Clark with respect to clothing, environment, and other items. As at Hillside, there seemed to be a genuine interest in the over-all project and its success, or what we might call "team spirit."

Communication between operators and supervision in different areas is extremely important. The larger the work force and operation, the more difficult is the task of maintaining these lines of communication. It is believed that the Clark organization of between 500 and 600 people is approaching the maximum which could be safely, from a quality standpoint, and efficiently utilized at one location on a complicated project of this type.

IX. PROCUREMENT AND TESTING OF RAW MATERIAL AND PARTS

Raw material and parts are purchased according to KS specifications and drawings. Ninety-one suppliers are involved, covering approximately 200 different items of raw material and 1,125 different parts. The raw

material includes items from metal bars and wire to plastic material and solvents. The parts are practically all of the metal, molded, and ceramic types.

Most of the raw material specifications are written around the A.S.T.M. standards, but nevertheless a good percentage of the material has failed to meet these general specifications and cannot be used in the product. This applies specifically to metal rods and sheet metal, which were purchased in smaller lots from jobbers. The philosophy of 100 per cent testing, as used at Hillside, is followed. This has paid off in many cases, because individual items in lots of material were frequently found to be entirely different from the rest. A special laboratory is set up at Clark for this testing purpose. It is equipped with facilities for chemical analysis and physical tests on material, including tensile, Rockwell, and microscopic examination.

In a number of cases it was necessary, in conjunction with the supplier, to develop materials to meet repeater specifications. For example, Teflon rod was purchased from two suppliers who did their utmost to produce rod to meet specifications. The main reasons for rejections were minute inclusions of foreign material, probably carbon, and cracking of rod materials during a bend test. To obtain satisfactory material, it was necessary to work with the suppliers and make changes in their processing technique to produce this material in reasonable lengths, free from imperfections. This development work enabled the supplier, in many cases, to place a higher-quality product on the market. This also was the situation on the ceramic items which, although they are parts, were mainly considered as raw materials.

Another unusual material is the singly-oriented polystyrene film for capacitors. As far as we have been able to determine, there is only one supplier in the country for this type of material. This supplier has a small shop with only one other employee. A considerable number of trial lots of this material were made by him and he, too, had to devise and learn a number of new tricks and precautions based on Western Electric's tests before he could produce a satisfactory product.

Of the approximately 200 different types of raw material, over 6,000 samples were tested during the first year, which amounted to over 25,000 different determinations. A large number of these were chemical tests which required considerable time to perform and evaluate.

One difficulty which was encountered was caused by the general lack of knowledge and understanding of our needs for unusual quality. In discussing specifications and drawings with suppliers, there was often misinterpretation of requirements and lack of understanding of the

difficulties involved. This resulted, in many cases, in too low a bid and inability to deliver satisfactory parts on time. When additional lots of parts were to be ordered, new suppliers were involved, who had to be indoctrinated and developed to produce the required quality of parts. Although this resulted in considerable engineering effort, it did generally produce parts at lower cost of equal and, in some cases, better quality. On critical parts, where the efforts to produce them were extensive, and where the production of the parts would take up a considerable portion of the capacity of the supplier's facilities, dual sourcing was arranged wherever possible. Also taken into consideration was the possibility of human or natural interference with the flow of material. For example, on the beryllium copper housing and end covers for the repeater enclosure, a second source was developed which is about to produce after nearly a year of effort.

Where the suppliers are located some distance from the Clark shop, the Western Electric supplies inspection organization is called upon to do the inspection at the source, to ascertain that the quality is maintained at the desired level. This arrangement involved training of supplies inspection personnel in different fields to realize new types and levels of quality.

In some cases the suppliers were not able to produce to our original specifications even after considerable assistance. In these instances, detailed analysis of their abilities and present-day manufacturing techniques was necessary before changes in the requirements were undertaken to make it possible to produce parts. This, in general, resulted in changes and tighter requirements on associated parts.

X. MANUFACTURING EXPERIENCE AT CLARK

10.1 *Paper Capacitors*

High-voltage capacitors which must be capable of operating at potentials up to 6600 volts dc are wound with eight layers of 1-mil thick paper between aluminum foil. Proper alignment and tension of the large number of papers and foil did initially present problems, which have been greatly reduced as improvements in winding machine design and winding methods were made and operator skill increased. Some unusual difficulties were discovered after cycling and the six-month life test on the early lots of these capacitors. It was found that the weld between the foil and the flag terminals had ruptured, which caused variation in the impedance of the capacitor. The weld rupture apparently

was caused by mechanical forces, including expansion and contraction with temperature and vibration, during the temperature cycling and life testing processes. The difficulty seemed to be connected mainly with insufficient flexibility of the internal leads from the capacitor units and insecure wedging of the capacitor units within the can. The problem was solved by increasing the flexibility of leads, improving mounting of the capacitor units, and changing the welding cycle to improve the weld.

Another difficulty was indication of denser metallic particles within the capacitor assembly on X-ray examination. In a number of cases this was found to be inclusions in the steel of the can caused by brazing alloy splashed during the assembly of mounting lugs. To avoid unnecessary rejection, the cans are X-rayed before assembly. Where these inclusions are found, a notation is made in the record so that in the final X-ray of the completed unit inclusions which are not harmful can be disregarded. The selection of paper, winding technique, and impregnation followed the earlier Hillside pattern.

The low-voltage paper capacitors use 0.3-mil thick paper in three and four layers between the foil. The low-voltage capacitors are also manufactured in a manner similar to that used at Hillside. However, one difficulty which was experienced might be worth mentioning. Some of these capacitors have a combination enclosure consisting of a ceramic cup and a metallic cup. The metallic cups were made from steel rods by turning. It was found that the steel, although it met all test requirements, had minute inclusions in it which were dissolved during the chemical cleaning and plating processes, producing a porous condition. This did not in many cases show up before the leak test of the completely assembled capacitor. The difficulty was overcome by a change to a cup drawn from sheet steel.

During impregnation, special precautions had to be taken to assure that the oil level in the capacitor was of the proper height. If the capacitor cooled down too rapidly, the excess oil would not have an opportunity to escape before the final sealing, causing the pressure to build up during cycling to a point where it would rupture the enclosure. By providing an infrared heating lamp over the capacitor in the sealing operation, the temperature is maintained and the desired oil level is obtained.

Based on the experience at Hillside and further study and work on causes for rejects, the over-all yield has been improved to the point where it exceeds that at Hillside. A general picture of the yield is shown in Fig. 9.

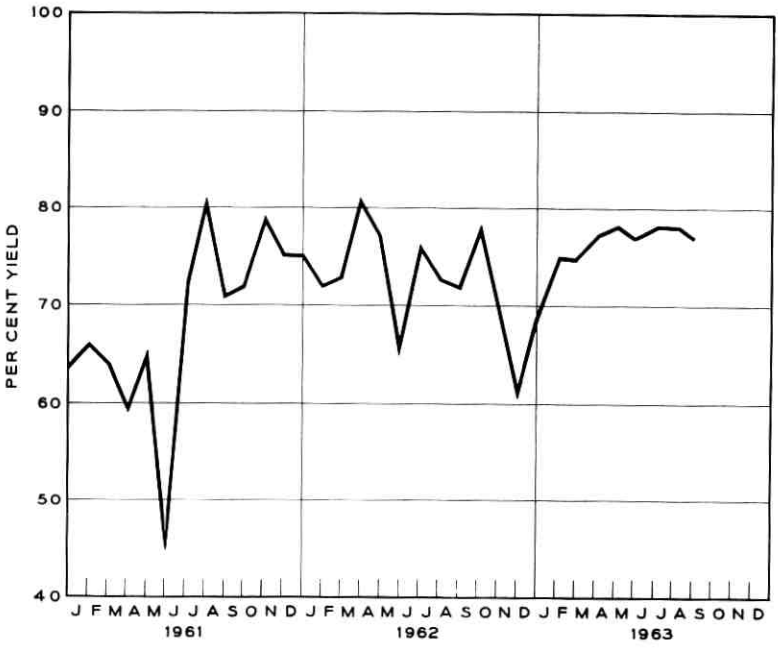


Fig. 9 — Paper capacitor yield.

10.2 *Polystyrene Capacitors*

This is a new type of capacitor used for the first time in Bell System repeaters. A singly-oriented polystyrene film is used. Since this film has a very high insulation resistance and easily becomes statically charged, it has a tendency to pick up dust and foreign particles during winding. It was necessary to enclose the spools on the winding machines with Plexiglas covers and to perform the winding in an area of 35 to 40 per cent humidity. The wound capacitor unit is heat treated to shrink or coalesce the polystyrene film around the foil to exclude air pockets and produce a homogeneous unit. The capacitor units are then mounted in metal cans having ceramic button insulators at one end. No air drying or special treatment is given the capacitor units during mounting.

After manufacturing these capacitors for nearly a year, using the original lot of polystyrene film, difficulties with cracks in the film were encountered when the leads were soldered on the units. Several changes were made in the soldering technique and some in the winding processes to avoid undue strains in specific areas. However, this did not entirely solve the difficulty, since it apparently became inherent in the poly-

styrene film after storage for a period of time. A new supply of film solved the difficulty, and presently the supply is being held to a maximum storage interval of six months. In reviewing the manufacture of this type of capacitor, a most important aspect seems to be the supply of good polystyrene film. A graph of the inspection yield of polystyrene capacitors is shown in Fig. 10.

10.3 *Mica Capacitors*

In designing the mica capacitors for the SD repeater, greater margin was provided between the silvered areas and the edge of the mica laminations than for the flexible repeater. This has made possible a sizable increase in yield, as has a change in design to omit cementing operations.

The silk screen method is still being used for application of the silver coating to the laminations. However, in place of the hand method, the machine described earlier is used. With this tooling, greater accuracy of size and location and fewer blemishes and other similar defects of the coated areas have been obtained. This also applies to the firing method, in which the silver is fused to the mica. It was found that minute particles in the firing furnace from carbonized silver paste vehicle and lint from operators' clothing and other sources would settle down on the coated laminations before the coating had dried. By predrying

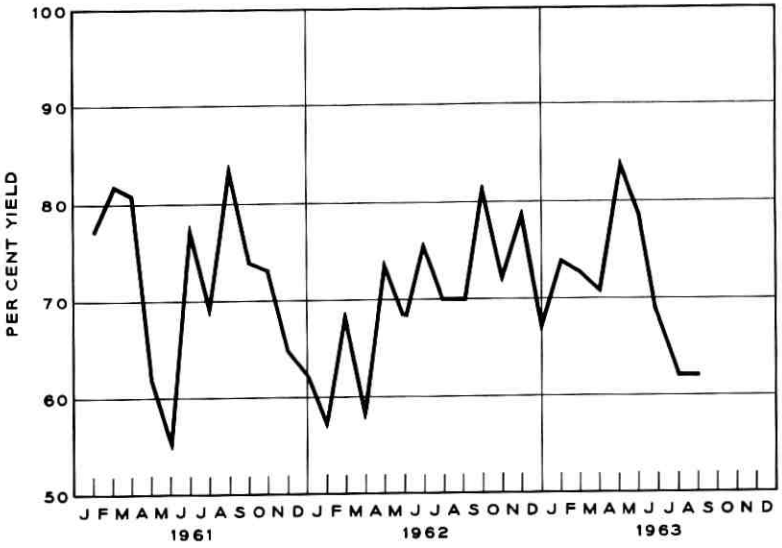


Fig. 10 — Polystyrene capacitor yield.

the laminations under infrared lamps before they enter the tunnel furnace, and providing better exhaust of the smoke from the vehicle, rejects for this cause have been reduced considerably.

Cracking of mica during the crimping of terminals has also been reduced by improvements in method and, to some extent, terminal design. Approximately half of the mica laminations are actually used in repeaters. Fig. 11 is a graph of the yield that has been obtained.

10.4 *Inductors*

All inductors are machine wound, a variation from the flexible repeater procedure, where the majority were hand wound. The winding time per unit has been decreased considerably for inductors with such strict requirements. Most types are wound in an hour or less. The change to polyurethane-insulated wire has made it easier to strip and tin without necking down the wire. A change in the other direction is the extensive use of Litz wire having up to 90 strands of No. 40 wire. However, it has not presented any unusual problems from the standpoint of cleaning and soldering. The Litz wire is used mainly in the coils for directional filters where a high and uniform Q value is important. To obtain the desired uniformity, all wire needed for foreseeable projects was manu-

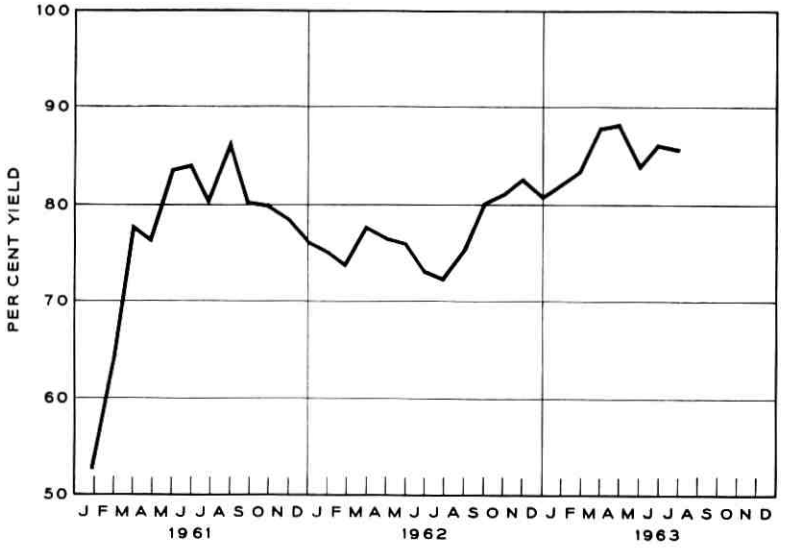


Fig. 11 — Mica capacitor yield.

factured at one time. This is important, since the type of equalizer used in the cable system is determined by the directional filter characteristics which, in turn, depend to a great extent on the Q value of the inductors.

As far as winding is concerned, there were, in general, no unusual problems that required extensive development work to solve. One exception to this was the progressive duolateral winding of some of the filter coils. Here the adjustment of tension, speed of the machine, play in the machine, and several other factors required considerable cut-and-try effort and development before a satisfactory winding could be produced.

When we speak about inductors, we may separate them according to the winding type: (1) single layer, (2) double layers in parallel, (3) progressive duolateral, (4) straight duolateral, and (5) toroidal with Permalloy dust cores. A number of the solenoid inductors are adjustable over an average range of ± 1 per cent by means of carbonyl iron cores. Difficulty in cementing the cores into the threaded carrier was resolved by applying the cement to the bottom of the carrier and then inserting the core. The most tedious coil to wind is a toroidal coil, where Mylar tape insulation has to be applied under the winding and between layers of the winding. This particular coil requires close to a full day for winding, although it is small and requires relatively few turns.

Practically all of the solenoid coils are wound on Mycalex cores which are held to close dimensional tolerances. The majority of the coils therefore require adjusting only by the adjustable core, since the number of turns per inch is also held to close limits. Fig. 12 shows the yield obtained on the inductors in general.

10.5 *Transformers*

The six transformers used in the repeater have ferrite cores which are manufactured at Hawthorne. Two of the transformers have Mycalex winding forms and the remaining four have diallyl phthalate forms. The secondary winding is wound on an inner spool and the primary on an outer spool in single layers. Some difficulty was experienced in obtaining the desired uniformity of winding and placing the shields between the windings properly to obtain the required direct and distributed capacitance values, which are important to meet the desired transmission characteristic. Cracking of the ferrite cores was also experienced. This was connected with the process of cementing the core sections and the winding form to the core. By changing the cementing technique, undue stress in the cores from the shrinking of the cement while drying was avoided and the difficulty cured.

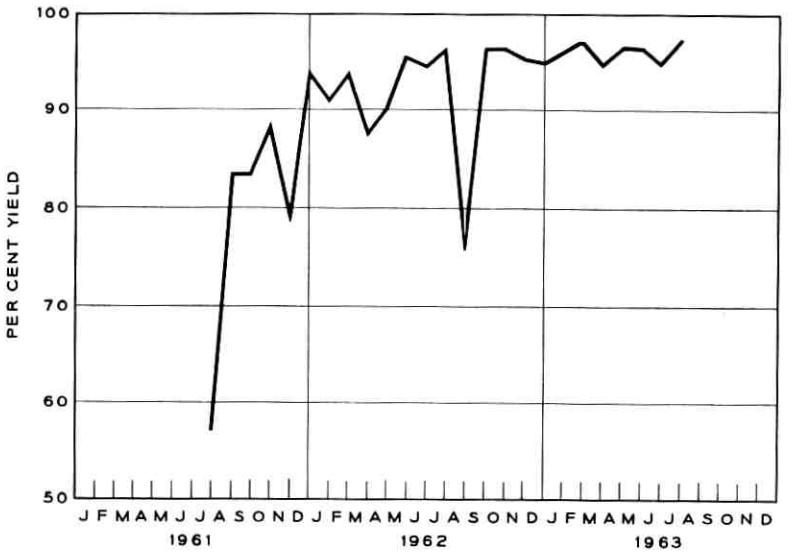


Fig. 12 — Inductor and transformer yield.

10.6 *Resistors*

Two types of wire-wound resistors are made at Clark, one using straight resistance wire, the other mandrellated wire made at Buffalo. Both types are wound on Mycalex spools which have gold-plated nickel pigtail leads. The ends of these leads at the spool winding area had to be flattened and notched to produce a fork-like terminal to hold the resistance wire. Originally this was incorporated in the lead wires before molding into the Mycalex spools. However, difficulty with closed slots and dirt in the slots from the molding operations caused considerable trouble. To overcome this, a special tool was developed for flattening the leads and producing the notch after molding of the spools. No unusual difficulties in brazing the resistance wire, such as experienced at Hillside, were encountered. This could be accounted for by the knowledge gained on facilities, materials, and operators' skill from Hillside. The yield on resistors is now running above 90 per cent, as shown in Fig. 13.

10.7 *Composition Resistors*

The composition-type resistors are purchased to specifications which stipulate special precautions on an otherwise standard product. Purchased resistors are then subjected to a selecting and screening process,

followed by a six-month life test. Special precautions are taken with some of these test procedures to avoid secondary effects caused by a change in moisture content, which affects the resistance values considerably. In the low-temperature cycling test, where the resistors are brought down to −100°F, the humidity of the air in the test chamber was reduced to less than $\frac{1}{10}$ of 1 per cent. Four different resistance values are used, from 11,000 to 500,000 ohms. The yield, surprisingly enough, varied considerably among codes and lots of resistors. Over-all, approximately 28 per cent of the resistors purchased end up in repeaters.

10.8 *Enamel Resistors*

The two codes of vitreous enameled resistors used in the heater circuit and heater protection circuit are purchased. The manufacture of these resistors required special facilities, including air-conditioned space, at the supplier's plant. Considerable change in his methods and procedures was also necessary.

10.9 *Networks, Filters, Amplifiers, and Repeater Units*

From the standpoint of manufacture, the mechanical assembly of networks (shown in Fig. 14), amplifiers (shown in Fig. 15), and repeater

Fig. 13 — Precision resistor yield.

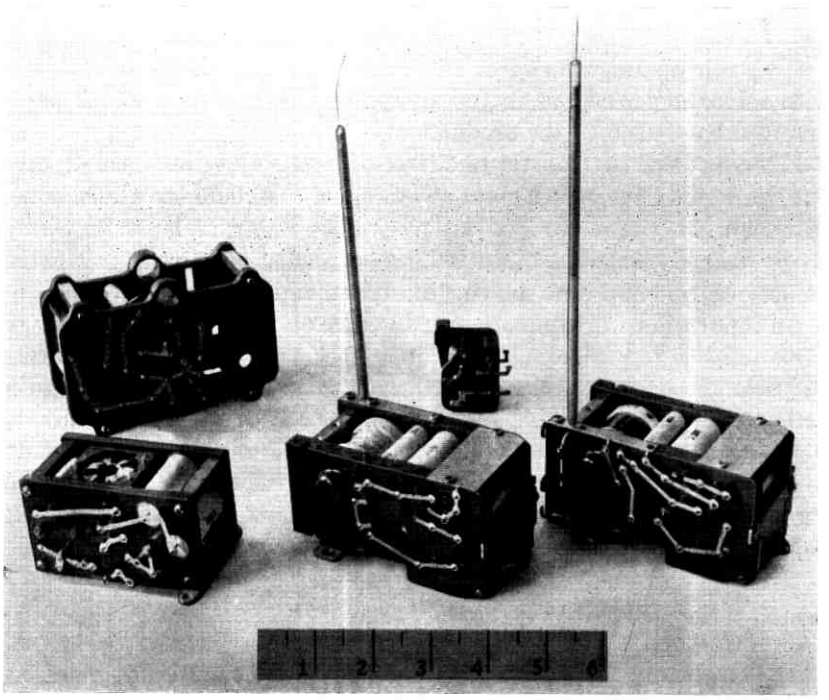


Fig. 14 — Networks.

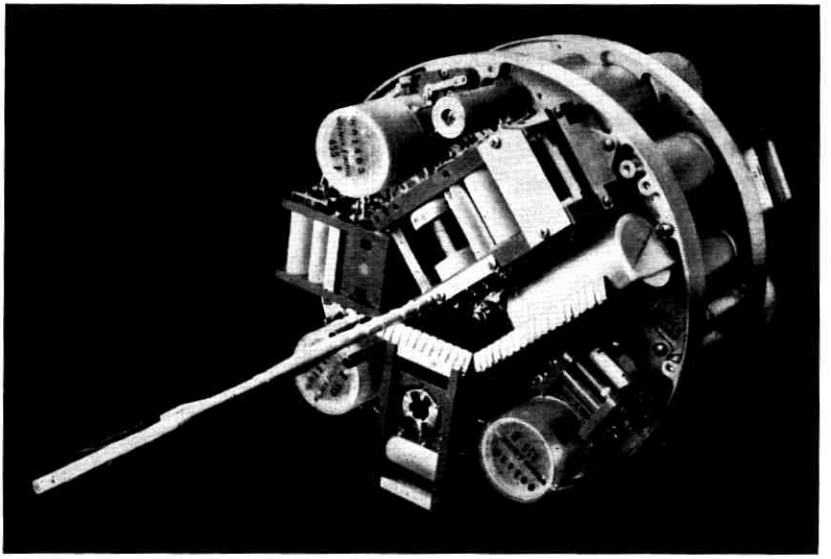


Fig. 15 — Amplifier.

units is fairly straightforward. However, from time to time difficulties with minor imperfections in the aluminum castings and molded parts involved have plagued the production.

Establishment of realistic electrical test limits at the start of production would have been a great advantage if this could have been done. Although the well controlled mechanical assembly and use of wiring strips defined parasitics quite well, small deviations from anticipated values occurred and were difficult to explain and trace.

An innovation worth mentioning in connection with the assembly of these units is the soldering technique developed for connecting wiring strips to the apparatus leads. Preformed rosin-core solder rings and a carbon electrode for heating are used. The solder rings are placed around the leads on top of the straps. A metal electrode is placed on the strap to be soldered, and a current is passed through the carbon electrode and the solder ring to the strap. A timing device limits the heating to a safe value.

10.10 *Repeater Closure*

After assembly of the repeater unit, it is tested for gas tightness, using a helium mass spectrograph leak detector, and then vacuum dried approximately 100 hours while heated to 130°F. This is considered to be the start of closure operations.

After vacuum drying, the repeater unit is covered with an epoxy coating to provide the necessary electric insulation between it and the outer housing which surrounds it. The housing is at ground potential, while the repeater unit is at cable potential. The coating is about $\frac{5}{32}$ inch thick and covers the repeater unit completely. The epoxy, a mica-filled compound, is carefully mixed immediately before use and degassed in a bell jar under vacuum. In applying the coating, small epoxy areas or lands are first cast on each end of the unit. These lands are used to locate the final mold, which completely surrounds the repeater unit. Each end requires approximately 24 hours for the epoxy to cure. The mold consists of two cast-aluminum half cylinders which are locked together around the repeater unit and have attached to them end covers matching the repeater unit shape. The end covers rest on the epoxy lands and line up the rest of the mold with respect to these lands. In filling the mold, the epoxy compound is poured into an injection cylinder, the lower part of which is connected to the bottom of the mold by means of a flexible hose. Air pressure is then applied to the cylinder above the epoxy compound, which is forced into the mold from the bottom. Since a considerable amount of heat is developed in the compound from the

curing, which starts as soon as the compound is mixed, the injection cylinder is water cooled to limit the temperature and slow down the curing. Two transparent plastic tubes are provided on top of the mold as risers for the compound, which shrinks to some extent as it cures. To eliminate voids in the coating, a slight air pressure is applied through the risers on the epoxy in the mold. This will permit the epoxy to flow slowly as it cures from the bottom up. To favor curing from the bottom up, the mold is heated by a warm air heater placed externally at the bottom. Any voids that appear after the casting process are filled in, using a tinker's dam.

After epoxy coating, the repeater unit is inserted in the outer beryllium copper cylinder, where it is held in place by Fiberglas springs resembling venetian blind slats. These springs serve as additional electrical insulation and at the same time as a cushion against shocks to which the repeater may be subjected during laying operations, such as hitting the sea bottom. A hydraulic jack is used to push the repeater unit into the outer cylinder. The cylinder is then placed in the assembly machine, Fig. 16, where it is held in place in the center of the machine while two end covers with the required seals in them are held in the fixture just outside the cylinder and lined up with the bore in the end of the cylinder. While the end covers are held at some distance from the cylinder, the leads are spliced together, and the joint is over-molded with polyethylene to insulate it and X-rayed. The X-ray, which is taken in two planes, determines the concentricity of the polyethylene insulation and ascertains whether it is free from foreign particles or voids.

After making all connections to the repeater unit and shielding and taping the leads, the leads are coiled up inside the end of the cylinder while the cover is being pushed in place. Coiling of the lead is done automatically in the assembling machine while the cover is pushed in place. As the heavy circular-disk Fiberglas spring attached to the end cover of the repeater locates the repeater unit in an axial direction inside the cylinder, a certain amount of pressure is applied to the repeater unit from the end covers. The covers are secured with steel clamps, which are bolted to the cylinder using the threaded holes in the periphery of the cylinder. The repeater is then removed from the assembly machine and is ready for welding the end covers in place.

After the repeater is trued up on one of the turntables on the welding machine, the milling head is swung out over the repeater and the surface of the edges of the cover and cylinder lips is milled off to remove the fine layer of beryllium oxide which forms fairly rapidly on beryllium copper. Next, the milling head is swung out of the way, the welding head is moved in place over the repeater, and the starting button is pushed.

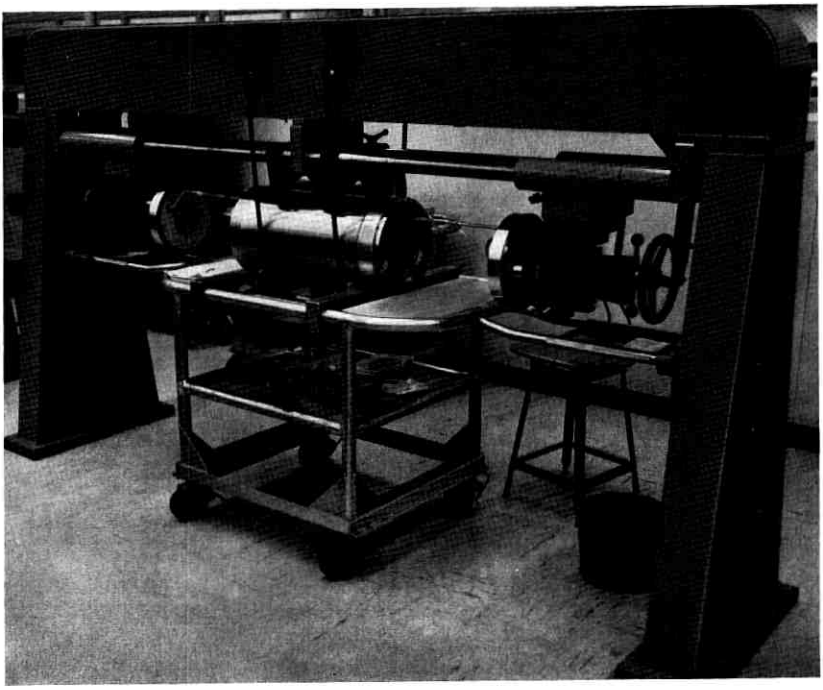


Fig. 16 — Repeater assembly machine.

This starts the turntable, turns on the argon shielding, and starts the arc by means of a high-frequency discharge between the electrode and the repeater. The arc is played on the lips of the cover and cylinder and melts both to form them into a rounded bead over the area while they are being rotated. An ac voltage of about 14 volts and a current of approximately 270 amperes is used, and the repeater is rotated at $\frac{1}{3}$ rpm.

To ascertain that the weld is of sufficient depth, an ultrasonic test is performed in a water bath at a number of points around the circumference (see Fig. 17). The required minimum is $\frac{1}{16}$ inch, but the welds usually run close to twice this.

The repeater is now ready for a high-pressure helium leak test, during which it is subjected to helium at a pressure of 11,600 pounds per square inch on the outside, while the inside is evacuated and then connected to a mass spectrograph leak detector. Before the repeater is placed inside the large steel test vessels, it is surrounded by a "dead man." This consists of an aluminum cylinder in two halves placed on the outside of the repeater cylinder and blocks at each end, which are hollowed out to fit

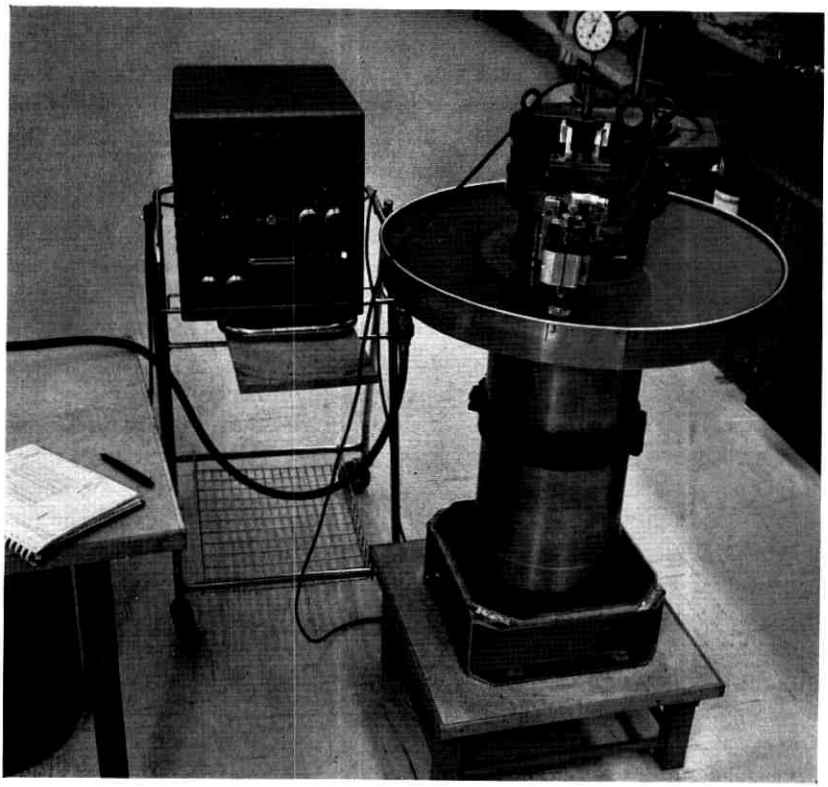


Fig. 17 — Ultrasonic test of weld.

closely over the covers and seals. This reduces the free volume in the test chamber and thus keeps the total amount of energy stored in the compressed helium to a minimum for the sake of safety. The pressure on the repeater in the vessel is increased gradually to the test pressure, held there for close to an hour, and then decreased over a period of about an hour to avoid damage to the polyethylene leads. Before inserting the repeater in the test vessel, the polyethylene-insulated pigtail leads are wrapped with polyethylene tape to avoid penetration of helium into the insulation, as this may produce minute blisters as the pressure is reduced.

The helium gas is reused; only a small amount is exhausted for each test. From the regular gas cylinder, two larger storage vessels are filled at a pressure of 2,000 psi. From here, the gas is pumped into a smaller

high-pressure storage vessel at 13,000 psi, from which the gas is controlled through a console and slowly bled into the test cylinders until the desired pressure is obtained. During this operation, helium is pumped into the storage vessel to maintain pressure. When pressure is released, the gas is first bled back to the 2,000-psi storage vessels and below this pressure into a low-pressure storage tank, where it is held at approximately 100 pounds pressure.

When a repeater has passed the high-pressure helium test, the tubulation which connects the leak detector to the inside of the repeater is still open and must be closed. This is done by pinch-welding while the inside of the repeater is filled with dry nitrogen to a pressure of approximately 2 pounds. To ascertain that the pinch-weld is satisfactory, a test using a radioactive isotope is performed. A small cylinder is placed over the tubulation and held in place by a heavy fixture. The cylinder is filled with an isotope solution on which pressure is applied to bring it up to 11,600 pounds per square inch. This pressure is maintained for six hours; then the isotope solution is drained off and the tubulation and the cylinder thoroughly washed. Any penetration of the isotope solution would be detected by a high count on a measuring instrument. This whole process, which has a fairly involved filling and washing cycle, is done semiautomatically. Valves are opened and closed automatically, permitting wash water, air, and the isotope solution to be moved in and out of the small test chamber as required to perform the test.

Electrical tests are performed on the repeater before welding and after the complete leak test to ascertain that it meets requirements. The last operations are forming and annealing of the pigtail leads, mounting of the end cones, and painting. After this the repeater or equalizer is ready for packing and shipping.

Included in the closure operations is manufacture of the seals for the repeater unit and the repeater. The seals use polyethylene for insulation and permit electrical connections to the inside of the repeater or equalizer circuit. In the case of the repeater a seal against sea pressure is provided. The repeater unit seal is a barrier to water vapor. In molding the seals the metal outer housing and center conductor, which have a layer of oxide applied chemically to assist bonding them, are held in place in a mold, and polyethylene is injected to fill the spacing between, and bond to, these metal parts.

The molding of satisfactory seals has been the most difficult operation on the whole project to master and control. There are still unknown factors which are difficult to control and evaluate.

The main difficulty is obtaining a good bond between the polyethylene and metal and at the same time attaining the required electrical charac-

teristics. Nearly a year was required to bring this process under reasonable control. Numerous small changes were made during this period to improve the design. They consisted mainly of controlling the bonding areas by restricting bonding to the outer shell to certain limited areas to avoid internal stress in the polyethylene. Also, a period of rest in the order of four to ten days was required for the seal after molding, to permit relaxation of stresses in the polyethylene.

XI. VERIFICATION OF COMPLETED REPEATER

As in the case of the flexible repeater, the rigid repeater is verified before shipment to ascertain that every operation has been performed and all requirements have been met. This includes every raw material item, piece part, and subassembly.

All electrical test results except simple go or no-go tests are punched on cards. The results of the go or no-go tests and mechanical inspection are entered manually on the punched card section attached to the apparatus by the inspector performing the test. This inspector also verifies that prior operations have been performed satisfactorily as indicated by the entries on the card section. When an apparatus item is completed, inspection visually verifies from the card section that all operations have been performed and accepted. A final-acceptance punched card, indicating that all operations have been completed, is originated and forwarded to the data center for verification of the electrical requirements before releasing the apparatus for use in the product.

In the data center the test data cards are sorted by group code, serial number, and operation number, to put them into logical order. The test data cards and unit history cards are now collated by group code and serial number. Next, the data and history cards are collated with computer program cards by group code. Finally, the combined cards are placed in an IBM 1620 computer where the data cards are read and the limits checked, as called for by the program card, and a punched output of verified test data is produced. Any errors in the data cards or in the processing of them are detected by the computer and a listing is automatically typed. Before proceeding with the verification, the items listed are corrected. After removal from the computer, the punched cards are run through a printer, where the components are listed as satisfactory to ship or to be rejected, depending on which is to apply.

All data pertaining to the assembly of a repeater are compiled in two data books — one covering mechanical items and the other electrical information. The data are entered by hand in these books by the individual operators and inspectors. All component apparatus assembled

into a repeater is listed in the mechanical data book and identified by code and serial number. Pertinent electrical test data, which are compiled on punched cards, are printed or typed on a sheet which becomes part of the data book. These two books are jointly associated with the repeater assembly until completion of the repeater, at which time the books are merged into one repeater data book for a specific repeater. This data book is verified by Western Electric inspection representatives and then pertinent data (primarily electrical) are check-verified by the customer's representative before the repeater is accepted and released for shipment.

For each repeater and its components there are about 3,750 separate electrical tests involving 38 test sets of 20 different types, in addition to all other tests and inspections. Approximately 3,000 punched cards, punched with up to 40 items and with up to 40 written-in items, are required, along with approximately 17,200 entries in the data books.

XII. PACKING

In preparing for shipping, the open end of each of the end cones on the completed repeater is closed with a Lucite cover. An impactograph to register any shock to which the repeater may be subjected during handling and shipping is mounted on one cover. On the other cover is mounted a min.-max. temperature indicator which will show whether the repeater has been subjected to temperatures in excess of 120°F or below 0°F during storage and transport. Next, the repeater, which weighs around 500 pounds, is mounted on a $5 \times 2\frac{1}{2}$-foot flat between two rigid foam plastic housings which fit over the ends of the repeater and are held together with the repeater between them by four $\frac{3}{4}$-inch steel tie rods. In case of excess shock, the foam plastic will crush and reduce the shock on the repeater. The housings, in turn, are bolted to the flat and a wooden cover is placed over the assembly and strapped to the flat. The cover has a sliding door at each end for reading the impactograph and temperature indicator. The packing arrangement, which is reusable, is identical for both repeaters and equalizers and is designed for handling by a fork lift truck. It is also relatively inexpensive, which is of particular advantage in cases where it would be impractical to return it for reuse.

REFERENCES

1. Lamb, H. A., and Heffner, W. W., Repeater Production for the North Atlantic Link, B.S.T.J., **36**, Jan., 1957, p. 103.
2. Brewer, S. T., Dickinson, F. R., and von Roesgen, C. A., Repeaters and Equalizers for the SD Submarine Cable System, B.S.T.J., this issue, p. 1243.

Electron Tubes for the SD Submarine Cable System

By V. L. HOLDAWAY, W. VAN HASTE and E. J. WALSH

(Manuscript received April 16, 1964)

Three new codes of electron tubes of the high reliability that is required for submarine cable service have been developed for use in the type SD systems. The basic philosophy was much the same as that which proved successful in designing tubes for the earlier cable systems. However, in order to meet the one-megacycle bandwidth requirement of the new system, it was necessary to extend such amplifier tube design considerations as transconductance, cathode current density and interelectrode spacing. The emphasis in this paper is on designing for reliability, since the amplifier tube is electronically quite conventional. In the field of protective devices, the new gas-filled cold cathode tubes also offer conservative and reliable applications of established design techniques.

I. INTRODUCTION

Three codes of electron tubes have been designed and developed specifically for use in the SD submarine cable system: the 455A-F, 456A, and 458A. The 455A, B, C, D, E, and F are pentode-type amplifier tubes manufactured as one type and subsequently designated for the six amplifier sockets to obtain optimum repeater performance. The 456A is a gas-filled cold cathode tube used as a power bypass device to (1) protect the heaters of the amplifier tubes, and the low-voltage components in parallel with them, from transient high-voltage surges on the cable and (2) to provide a continuous path for the cable current in case of an open heater in any of the amplifier tubes. The 458A signal path protector is a high-speed gas-filled cold cathode tube used at both the input and output of the repeater to protect the transmission path from voltage surges.

For a period of approximately two years engineers of the Western Electric Company were resident at Bell Telephone Laboratories to participate in the final development stages of the tubes. As a result of

this endeavor the final tube designs and processing schedules represent the combined thinking of both the development and manufacturing organizations.

II. RELIABILITY CONSIDERATIONS

2.1 *Objectives*

Reliability requirements for electron tubes for use in submarine cable systems were formulated for the SB (flexible repeater) system[1] and have not been materially changed for the new broader-band system. Lifting cables for the purpose of changing faulty repeaters will always be an expensive operation. Revenue loss during the out-of-service period increases with bandwidth because of the additional voice channels.

The reliability objective for the tubes may be stated in several ways. Loosely, one hears "no tube failures in twenty years." Actually, the objective adopted is that the probability of system failure due to a tube failure shall not exceed 50 per cent for a twenty-year service period for a 3000-mile system. This objective corresponds to a mean time between system failures of twenty-nine years.

The amplifier tubes (455A-F) are probably limiting with respect to reliability, because they have closer interelement spacings, tight requirements on stability, and are operating continuously, as compared to the gas tubes (456A and 458A), which are on standby duty. Because of this and the large number of repeaters required for a long system, the amplifier includes two parallel strings of three tubes each with the circuit so arranged (see Fig. 5, Section 3.4.1) that the most probable kinds of tube failure will cause a system failure only if both strings are involved. Such defects would be permanent open circuits or short circuits between tube elements. There is no protection against plate shorts in the amplifier output stages, and it is also recognized that a single noisy tube could make the system unsatisfactory for commercial service.

The objective stated above corresponds to a tube failure rate of 0.08 per cent* for twenty years for defects not minimized by circuit redundancy. A failure rate of 2.4 per cent* is the corresponding value where circuit redundancy is provided. In the reliability analysis, it has been assumed that the deterioration due to decreasing thermionic emission would be negligible. This assumption is justified in Section 3.4.2, *Thermionic Life.*

---

* See Appendix for details.

2.2 *Background and Prospects*

The reliability objective and the prospects for adequate stability for the 455A-F tubes for twenty years were based largely on the experience gained from the 175HQ tube used in the SB system where, to date, no tube failures have occurred during more than 80 million tube-hours of sea-bottom operation. In 1955, when development work was started on 455A-F tubes, 18 of the 175HQ tubes had been in sea-bottom service in the Havana-Key West system for about five years with no failures and no significant change in tube performance. Some 50 175HQ tubes of essentially the final design had been life-tested for lengths of time up to about sixteen years at various operating conditions. The results indicated that the transconductance would decrease less than 1 per cent per year at the operating conditions adopted for cable use. Additional life tests on 4800 tubes[1] showed that a 5000-hour aging period was sufficient to cull out substandard tubes and allow selection of those tubes most likely to meet the reliability objective. A 5000-hour aging-in and screening period was therefore adopted for the 455A-F tubes.

Since no gas tubes were used in the Havana-Key West system, no field data were available. However, considerable life testing had been done on the gas tube subsequently used in the first transatlantic system.

This background offered promise that the development of a higher transconductance vacuum tube and suitable gas tubes was feasible, with the stated reliability as a reasonable objective. However, it was evident that unproven materials, processes, and design features should be avoided where possible. Also, the basic philosophy which had been used in the fabrication and selection of 175HQ tubes was adopted for the new tubes. Every significant deviation from the "expected" by an individual or a batch is viewed with suspicion. If the cause can be found and understood, the cloud of suspicion may be removed; otherwise, the affected tubes must not be used. The methods of implementing this philosophy are described in a later section.

III. THE 455A-F AMPLIFIER TUBES

3.1 *Design Considerations*

The 455A-F electron tubes are pentodes of the indirectly heated cathode type using fine-pitch frame-type grids. With the emphasis on reliability and reliance on proved techniques, most of the design features are conventional, some perhaps even old-fashioned. Since optimizing pen-

tode performance characteristics by geometry is an established art, the emphasis of this article is on designing for reliability.

Because of the excellent performance of the 175HQ tube it was used as the datum point for designing the new tube. The most important parameters and operating conditions for the 175HQ and the 455A-F tubes are compared in Table I. This comparison provides a basis for explaining the design considerations involved in the development of the 455A-F tubes.

3.1.1 *Cathode Temperature*

The most important operating factor affecting the thermionic life of the cathode is the cathode temperature. It was decided at the outset that, in spite of the higher current density required, the cathode temperature for the 455A-F tubes would be no higher than that of the 175HQ. Use of a special cathode alloy and improved processing and storage techniques for tube parts made this possible.

3.1.2 *Cathode Current Density and Transconductance*

The higher cathode current density of 10 ma/cm^2 is necessary in the 455A-F tube to achieve the higher transconductance. Life tests on 175HQ tubes which had run about fourteen years at a cathode temperature of 710°C showed no significant difference between tubes operated at 0.2 ma/cm^2 and those operated at 2.8 ma/cm^2. These and other similar results indicated that 10 ma/cm^2 was a reasonable operating level. The higher transconductance also requires an improved cathode alloy which develops less interface resistance than the nickel used in the 175HQ tubes.

3.1.3 *Grid-to-Cathode Spacing*

The 0.0055-inch grid-cathode spacing of the 455A-F is also necessary to achieve the higher transconductance. This is perhaps where the great-

TABLE I — COMPARISON OF PARAMETERS AND OPERATING CONDITIONS

	175HQ	455A-F
Cathode temperature (true)	670°C	670°C
Cathode current density	0.7 ma/cm^2	10 ma/cm^2
Grid-cathode spacing	0.024 in.	0.0055 in.
Maximum element voltage	51 v	45 v
Transconductance	1000 μmhos	6000 μmhos

est extrapolation from the 175HQ (0.024-inch spacing) is involved. Spacings of only 0.0025-inch are successfully employed in several telephone tubes for readily accessible land systems. For the less accessible submarine cable tubes, it was felt that the greater care in fabrication and inspection which is necessary to minimize particles can economically be justified. This would make the 0.0055-inch spacing feasible, even with the high level of reliability required.

3.1.4 *Maximum Element Voltage*

Although there was no indication that the maximum element voltage of 51 volts was high enough to affect the life of the 175HQ tube, it was found possible to operate the 455A-F at 45 volts maximum, which was considered desirable.

3.2 *Structural Features*

3.2.1 *Interelectrode Spacings and Foreign Particles*

The design of the tube was strongly influenced by the "particle problem." The reduction in grid-cathode spacing is a serious factor because of the risk of failure due to a particle of foreign matter becoming lodged in this critical area. A conducting particle can cause a short circuit or it can cause noise. A nonconducting particle can cause noise, particularly after it has been subjected for some time to the deposition of the normal vaporization products from the cathode. Noise is particularly serious, since the parallel-path redundancy feature of the amplifier circuit actually increases the risk of trouble due to noisy tubes, because there are twice as many tubes in positions followed by substantial amplification.

The tubes have been designed to minimize particle generation, are fabricated in extremely clean areas using the highest practicable degree of housekeeping measures, and employ selection criteria designed to minimize the hazards due to particle contamination.

3.2.2 *Tube Mount Assembly*

A phantom drawing of the based tube is shown in Fig. 1. The glass envelope, the ceramic base, and the braided gold-plated beryllium copper flexible leads are essentially the same as those used in the earlier cable tubes. The stem of the glass envelope is made from a molded glass dish into which are sealed eight pre-beaded two piece nickel-dumet leads

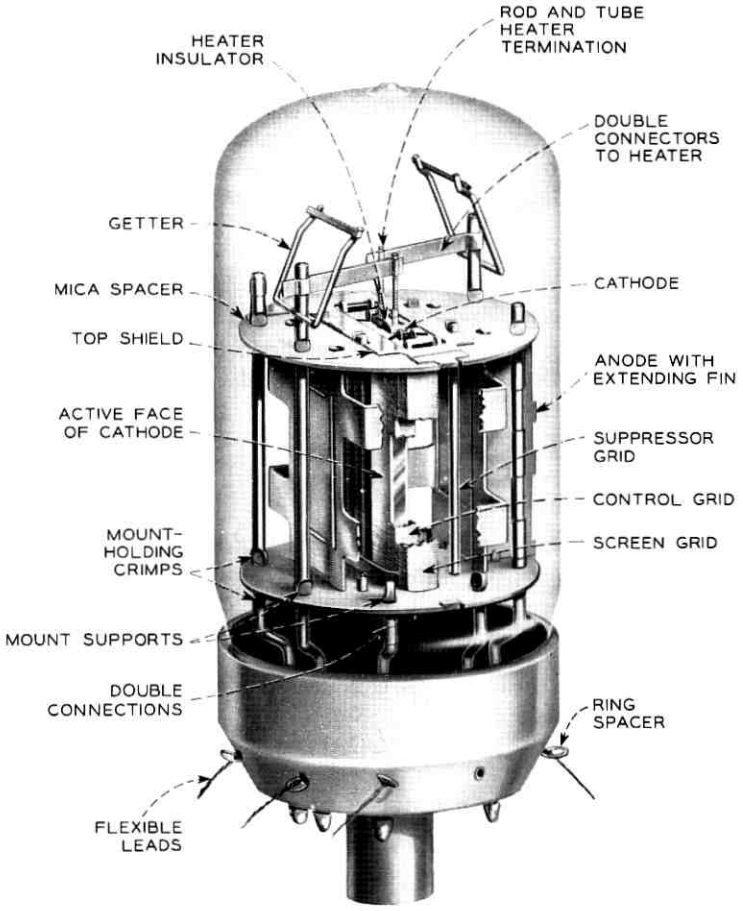


Fig. 1 — Structural details of the 455A-F electron tube.

whose inner nickel portions form the supports for the element assembly. The entire mount is supported from the stem leads. It was so designed to eliminate the mica insulator to bulb contact present in the earlier tube, with its resultant production of small mica particles.

The element "cage" assembly consists of the molybdenum anode, the oxide-coated cathode sleeve, frame-type control and screen grids, four rods that act as a suppressor grid, and two end shields. All of these are held together in proper relationship by two magnesium oxide-coated mica insulators. The "cage" assembly slides over the support leads, which are then crimped to position the cage. The coated heater and insulator

assembly slides into the cathode sleeve from the top. The attachment of the heater connectors and getters completes the mount. The judicious placement of standoffs, tabs and lead crimps creates a truss-like structure that permits the tube to withstand a 500-g, 2-msec shock in any direction without damage. (This is ten times the predicted maximum shock to which the tube might be subjected in cable laying.)

In a repeater the based tube is supported on rubber cushions in a methacrylate housing as shown in Fig. 2. The flexible leads pass through the housing to form circuit connections and to provide for relative motion between the circuit and the tube. The small rings that pass through the flexible leads control the length of lead in the various regions.

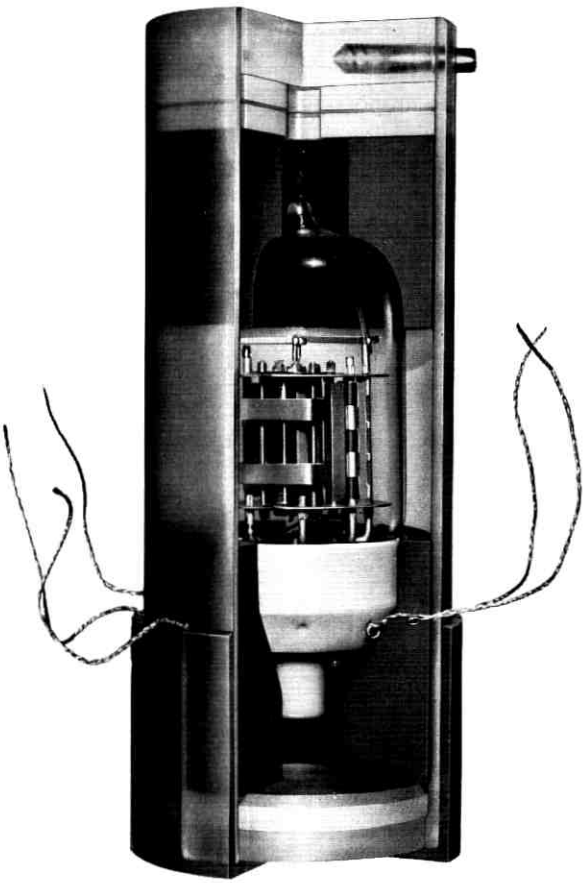


Fig. 2 — 455A-F tube cushioned in a methacrylate housing.

3.2.3 *Lead Welding and Crimping*

Again referring to the particle problem, a hazardous condition is created by splattering from the spot welds made during mount fabrication. Particles of 0.005 inch can be seen quite readily when viewed under a microscope on a flat surface. They are, however, extremely difficult to detect in a mount structure. This problem has been minimized by designing the tube so that the number of spot weld positions is roughly one-third those in the early flexible repeater tube. The tube design is such, however, that even with the reduction in the number of welds, every active tube element is redundantly connected so that should a single weld fail, the tube would still function.

Another design feature to minimize particle contamination is the use of crimped leads to support the mount structure. The older submarine cable tubes used metal eyelets crimped to the micas and welded to the stem leads as the mount supporting arrangement. Experience had shown that eyelets so used provided excellent structural support but, in a close-spaced tube, had the disadvantages of increasing the weld splash problem, generating mica particles, and catching and retaining all kinds of particulate matter and concealing it until late in the tube processing.

3.2.4 *The Anode*

The anode is a one-piece detail with a side fin. An earlier design, utilizing a two-piece plate with no side fin, optimized the interelectrode capacitances and provided sufficient mount ruggedness in the direction parallel to the grid plane, but was somewhat weak in the transverse direction. Ruggedness for tubes intended for the quiescent environment of the ocean floor may seem incongruous but is needed to protect against shocks during cable handling and laying. The anode structure, as can be seen in Fig. 1, has had all nonfunctional portions eliminated to provide the most open structure possible. This was done to permit microscopic examination of the cathode surface and the critical 0.0055-inch grid-to-cathode region to insure that no foreign matter has been entrapped.

3.2.5 *The Grids*

The control grid and screen grid are both made with fine tungsten wire wound at high tension on frames blanked and formed from sheet molybdenum (see Fig. 3). The frames are made in two sections, lapped to very precise dimensions, and paired to close tolerances. After the wires are wound on the paired frames, a single furnace operation brazes the frame

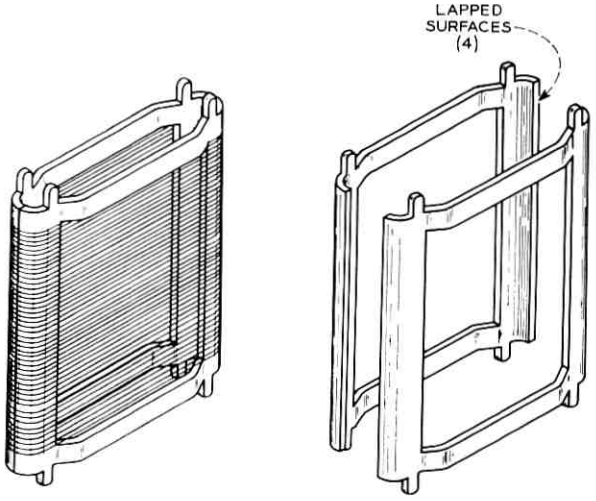


Fig. 3 — Wound grid and a pair of grid frames for the 455A-F tube.

sections together, brazes the fine wire to the frame and causes molten gold to cover the fine wire. While other types of frame grid construction were available,[2,3] this type, patterned after the original frame grid,[4] was selected because of fewer welds, lower grid mass, and the elimination of frame embrittlement.

3.2.6 *The Heater*

Heater reliability is of major concern, as this element is probably the most susceptible to catastrophic failure in that tungsten is subject to possible recrystallization and embrittlement. The heater in the 455A-F tube is a coiled tungsten wire formed into a precisely dimensioned M-shaped heater, spray coated with alumina, and slipped into a formed alumina block. Details of the heater and connectors are shown in Fig. 4. Connection to the tungsten heater is made in a manner similar to that which proved successful for the 175HQ tubes[1] with a single modification, namely, the crimping of the nickel sleeves to obtain a more intimate heater-to-connector contact. The 1090°C heater temperature is consistent with the conservative approach to design and operation.

To insure a large supply of uniform quality material, a procurement program was worked out with the supplier, the Westinghouse Electric Company. Special ingots of tungsten were made specifically for this use. The ingots were reduced to the wire form and wound onto many small

spools, each one related by code marking to each other and to the original ingot. The spools were sampled in a statistical manner by making heaters and running regular and accelerated life tests on them. The accelerated tests included higher temperature operation as well as ON and OFF cycling. Use of a spool of wire was contingent upon obtaining no failures on these tests.

The insulator block shown in Fig. 4 is novel[5] and worthy of discussion. It is essentially a four-bore alumina block with the wall broken through between the two center holes, giving this opening a dumbbell shape. This permits the heater in its final M shape to be inserted into the block from one end. In addition to permitting an easier assembly operation, the dumbbell insulator makes it possible to insert a coated heater that has been fired. Use of a coated heater minimizes the transfer of tungsten to the insulator walls and insures high insulation resistance between the heater and cathode.

3.2.7 *The Cathode*

3.2.7.1 *Study of Cathode Materials.* There was evidence that impurities in commercial cathode nickel could, in time, cause performance degrada-

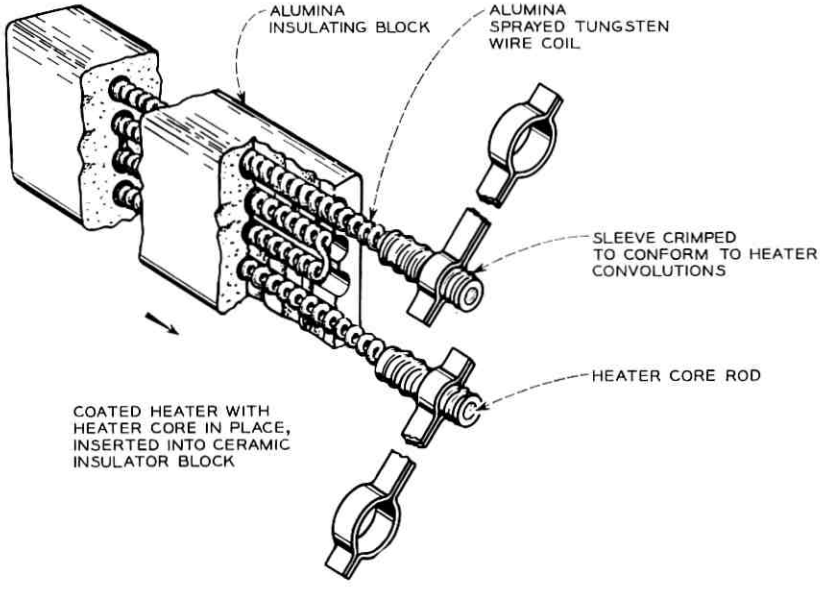


Fig. 4 — Heater and insulator assembly, and heater connector arrangement for 455A-F tube.

tion either through sublimation with consequent leakage between elements or with the development of cathode interface impedance. A survey showed that there was no source of supply for cathode nickels of the purity needed for the newer long-life high-transconductance tubes with their more stringent requirements.

Accordingly, a development program was undertaken to (1) provide special materials for the tube development work and (2) establish means for procuring production quantities in cooperation with the Western Electric Company. The first part of this program made available small billets of high-purity nickel and nickel with various additives[6] in which no single impurity exceeded 0.005 per cent by weight. Studies of the fundamental properties of these materials and a comprehensive evaluation of their suitability for use in long-life electron tubes were conducted.[7,8]

3.2.7.2 *Cathode Fabrication.* The cathodes are made of rectangular tubing with the emissive coating on the two broad surfaces. Initially they were fabricated from two channel-shaped pieces of nickel joined together by a diffusion welding process using sapphire tooling to avoid contamination of the nickel.

Approximately 700 tubes were fabricated and put on life test, using as a test vehicle a modification of the Western Electric 435A electron tube. This pentode was used because its close electrode spacing and high transconductance would accentuate the effects being studied. The tests included nine different cathode melts run at nine different operating conditions. Throughout the development program complete analytical tests were made at various points to insure the cleanliness and purity of the cathodes.

The making of cathodes from sheet material had provided ultra-pure sleeves for all of the cathode program. The fabrication method had been developed to an extent that would permit its use in production, and some 230 tubes of preliminary and refined designs for the 455 A-F amplifier tube were on life test. Since there were some undesirable aspects to the welded cathode, a program to produce seamless cathode sleeves from raw material supplied by Bell Laboratories was worked out with the Superior Tube Company of Norristown, Pennsylvania. Special measures were applied at their plant to insure that no contaminants were picked up and that each individual cathode sleeve was identified with respect to its starting billet. Analytical tests were applied to insure the purity of the material at critical points in the production. No changes could be observed from the original data. Tubes made with the new seamless cathode sleeves were put on life tests along with suitable control lots.

3.2.7.3 *Choice of Cathode Material.* As the development neared completion, a pilot production run of 300 tubes was made using the final design with all fabrication techniques and parts processing oriented toward a scaled-up production run. The life data from the cathode programs had, by this time, indicated two melts with superior characteristics for long life. These were (1) a single-additive, nickel plus 2.0 per cent tungsten, and (2) a double-additive, nickel plus 2.0 per cent tungsten plus 0.02 per cent magnesium. Because ample supplies of these two kinds of cathode materials had been stockpiled, it was possible to delay the final choice until production was ready to start, thus accumulating more evidence to support the final decision. The melt ultimately chosen for production was the double-additive, since the immediate effectiveness of the magnesium as a reducing agent permits a relatively short age-in period. The tungsten, with its slower diffusion and reaction rates, would not by itself make available tubes of adequate uniformity with 5000 hours of aging.

3.3 *Processing and Cleaning Controls*

A basic principle of reliability is that while inspections are important they are usually after-the-fact, and hence reliability must be built into the product. This is done by meticulous workmanship and by a system of multiple checking on parts and processing.

The system that was evolved recognized that it is virtually impossible at the time of tube fabrication to predict what specific information would be important when the tube is to be evaluated 5000 hours later, or if the tube is on life test, perhaps years later. Each tube was given a unique serial number, and all records were arranged so that from the serial number it would be possible to:

(a) trace every item in the tube back to its raw material lot,
(b) know who treated each part at every step in the tube fabrication,
(c) know when each part was treated, and
(d) know how each part was treated.

This was accomplished by tying together with proper records:

(a) the tube serial number,
(b) a serial number on the cathode, (duplicated on the plate)
(c) a serial number on the control grid,
(d) a serial number on the screen grid,
(e) lot numbers on each group of parts, and
(f) lot numbers on all material batches.

The record keeping consumed a large number of man-hours, even though maximum use was made of modern machine aids. This detailed

information on individual tubes was essential in analyzing anomalies in test results and in rationalizing the acceptability of processes and tubes.

The cleaning procedures used in tube fabrication are of utmost importance in producing long-life, highly reliable tubes. The basic cleaning procedure adopted was that developed at Bell Laboratories.[9] This is:

(a) removal of grease by solvents,
(b) removal of physical contaminants by ultrasonic agitation,
(c) rinse in deionized water in a cascade-type cleaner,
(d) light oxidation to remove residual organic materials,
(e) reduction in hydrogen to outgas parts,
(f) testing for surface contaminants by the water wettability (atomizer) test, and
(g) storage in "atomizer clean" containers with strict limitations on duration of storage, i.e., 96 hours maximum for coated cathodes to ten days maximum for shields.

These basic processes were used wherever applicable at each step in the fabrication. Early contaminant elimination and low carryover was thus achieved. The resulting fabrication and processing procedures consisted of some 350 separate operations, a formidable number under any circumstances. However, the associated test and life data proved beyond question that a uniformly high-quality tube was being turned out on a production basis.

3.4 *Electrical Characteristics and Life*

3.4.1 *Operating Characteristics*

The arrangement of the electron tubes in the SD system repeater is shown in the simplified circuit schematic of Fig. 5. In the amplifier, stages A, B, C and D, E, F comprise the two parallel amplification paths mentioned earlier in this article. Table II lists the operating voltages and principal electrical characteristics of the 455A-F tubes for input and output conditions of amplifier operation. Grid noise is important in stages A and D, and significant in B and E, while output capability is most important in stages C and F. The typical equivalent grid noise figure is 825 ohms compared to the theoretical value of 750 ohms.

A family of plate current versus plate voltage curves for a typical tube over the approximate region of operation is shown in Fig. 6. While the curves have the general appearance of a pentode family, the sharp breaks at the knee region are more characteristic of a tetrode. Actually, the tube is a cross between the two since the suppressor grid consists of only four rods and there is considerable dependence on space charge for suppres-

transmission of current to the plate. The double shoulders in the curves are attributable to lack of perfect symmetry.

3.4.2 *Thermionic Life*

The thermionic performance of tubes during life is of major importance to a cable system. In the discussion of reliability, it was assumed that deterioration due to decreasing thermionic emission would be negligible. This was based largely on the performance of 175HQ tubes in SB cable systems and backed up by the life data accumulated on development models of 455A-F tubes. During the development program many tubes, representing a variety of structural features, many different cathode melts and several sets of operating conditions, chiefly cathode temperatures, were put on life test.

Curve (a) of Fig. 7 shows the life pattern for the oldest test of development model 455A-F tubes. This is for a representative group of tubes with cathodes of grade 220 nickel operating at 700°C.* Transconductance has been normalized to the median value at the 5000-hour test. Per cent of this reference transconductance is plotted versus total aging time in hours and years. Median values and ranges are shown at the various test points. The decrease in transconductance with time is attributable to the development of cathode interface impedance which, at the latest test point, had a median value of about six ohms.

Life performance of a group of tubes using cathodes with the tungsten and magnesium additives (final type) are shown in curve (b). At nearly 40,000 hours life the median transconductance exceeds the 5000-hour reference value by about one per cent. There was no measurable interface impedance in these tubes at the latest test. Again the cathode operating temperature was 700°C. Comparative tests subsequently showed 670°C to give results comparable to those at 700°. Consequently, in the interest of conservative operation, the lower temperature was selected for the final design.

Curve (c) in Fig. 7 is for tubes typical of Western Electric Company production operating at 670°C. While the total age is shorter for these tubes, it is gratifying to note that the trend is following closely the pattern set by the development tubes at comparable age.

The activity data of Fig. 8 present an indication of future trend in transconductance for the same groups shown in Fig. 7. Here are plotted the changes in transconductance (per cent ΔG_m) which accompany a 20 per cent decrease in tube heater current. It is noted that the activity picture is favorable for development and production tubes. The wider

* All cathode temperatures referred to in this paper are "true" temperatures.

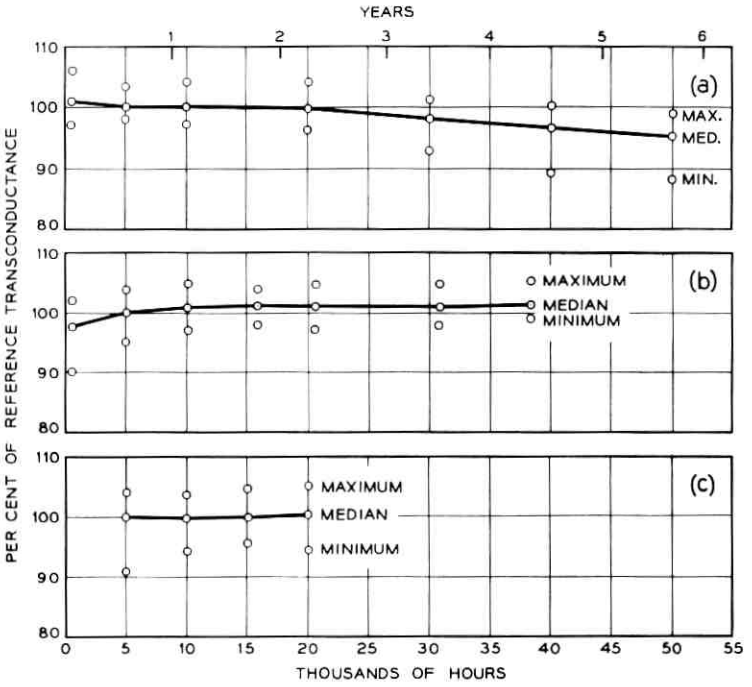


Fig. 7 — Transconductance as a function of life for 455A-F tubes.

spread for the Western Electric product is consistent with the larger sample on test. From the stability exhibited in curves (b) and (c) of Figs. 7 and 8 it appears that the extrapolation to the higher current density, the use of the special cathode alloy and operation at 670°C are all justified.

3.5 *Tube Production*

Production of tubes has been underway for nearly four years at the Allentown Works of the Western Electric Company. The philosophy of building reliability into the tubes was mentioned above, and it was pointed out that only tested and proved materials would be used. This was successful during the development and has been carried over into production. All fully certified parts and materials are given lot numbers, certain piece parts are individually numbered and each tube is given a unique serial number. To expedite the handling of product a "kit" system was adopted with six tubes in a kit. Six sets of parts and the records pertinent to them are given to an assembly operator. At the completion of the assembly, the operator passes on six tube mounts

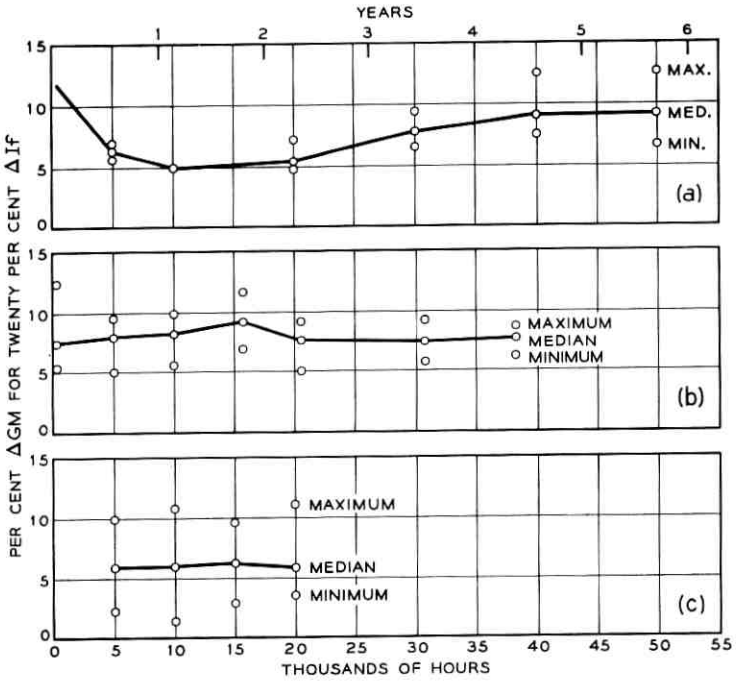


Fig. 8 — Cathode activity as a function of life for 455A-F tubes.

(assemblies) and the pertinent papers. Some operations, for example pumping, can handle two or more kits at a time. The kit system also permits examination of data for "batch effects." As the name implies, these are deviations or trends in data that can be studied in related groups of tubes as well as in individuals.

In the course of fabrication all tubes are given ten inspections, six of which include thorough searches for particles. Also, at specific times within the 5000 hours of aging the general electrical characteristics are tested five times. Supplementary tests are also made of special electrical parameters such as power output, modulation, equivalent grid noise, etc.

3.6 *Tube Selection Procedure*

3.6.1 *Review by Committee*

The selection of tubes for sea-bottom service goes beyond the mere matching of test data against limits. There are certain attributes or second-order responses of a tube to which it is difficult to assign numbers.

These attributes are assessed by the selection committee much as if the candidate were a person. Before a tube is submitted to the committee, the processing and inspection records are checked for consistency and completeness.

The selection committee reviews all of the data in detail, studying trends and minor variations occurring within the limits of acceptability. As examples: the age-in characteristics are compared to the characteristics at the 5000 hours acceptance point. The residual gas test and plate current age-in curves are reviewed. The departure of any data from the normal pattern is critically analyzed, and only those tubes which have exhibited normal behavior are accepted for sea-bottom use.

3.6.2 *Grouping for Repeater Use*

The tubes accepted for sea bottom are permanently grouped by sixes for use in an individual repeater. This grouping has no relation to the "kit" of six used for production control. The grouping takes into account the sum of the heater voltages at a fixed current, the slope of the plate current aging-in curve, and the products of the transconductances for the tubes in each of the two amplification paths (tubes A $\times$ B $\times$ C compared to tubes D $\times$ E $\times$ F). Tubes with low noise figure are assigned to input stages while tubes of high power output are used in output stages.

As a final check on the static characteristics and on the grouping of the tubes, they are given an electrical test in a circuit simulating the dc circuit of the amplifier, energized at rated cable current. Individual tube plate and screen grid currents are measured in this circuit, and then a comparison is made between tube performance in the working circuit and in the tube testing equipment. The tubes are now ready for shipment to the repeater assembly factory, where they are immediately restested in an amplifier simulating circuit identical to the one at the tube manufacturing plant. With satisfactory agreement between these two tests, the group of tubes is acceptable for assembly into a repeater.

IV. THE GAS-FILLED PROTECTIVE TUBES

4.1 *Circuit Function*

In those portions of the cable system operating at more than about 1500 volts to ground, a fault resulting in the grounding of the center conductor produces severe electrical transients. These transients propagate through several repeaters to either side of the fault before being

attenuated to a safe level. If no protection were provided, damage to repeater components would be probable, and repeater failure would be possible.

Two types of gas tubes were developed to provide the desired protection. Both types are electrically symmetrical diodes, designed to conduct current in either direction, and both are of the cold cathode variety, requiring no power in the standby condition. Cutaway views of the two tubes are shown in Fig. 9.

4.2 *The 456A Power Bypass Gas Tube*

4.2.1 *Operating Requirements*

The 456A tube is bridged across the six series-connected amplifier tube heaters as shown in Fig. 5 and has two functions: (1) in the event of a fault on the cable causing an abnormal current to flow through the repeater, the tube will fire and conduct the current, preventing damage to the tube heaters and to other parallel low-voltage components; (2) if an amplifier tube heater opens, the rising voltage will fire the gas tube, holding the heater circuit voltage at a safe level. The cable voltage is then

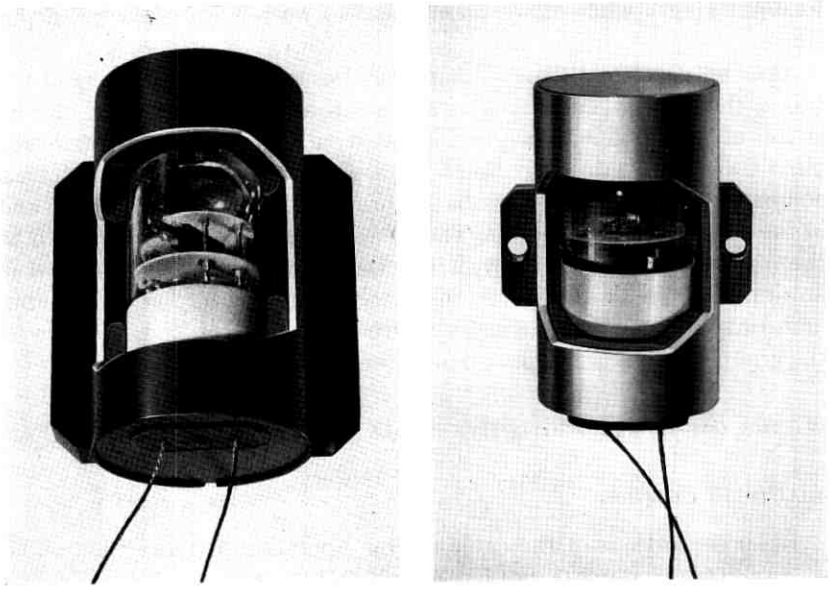


Fig. 9 — 456A (left) and 458A (right) electron tubes cushioned in aluminum housings.

turned down and the procedure carried out for energizing the heat coil, which in turn melts the fusible link to bypass the defective heater.[10] By means of a center-tap connection, only half of the amplifier tubes (ABC or DEF) will be turned off. The minimum firing voltage of the gas tube is set slightly above the voltage required by the heater circuit for this operation so that the tube does not refire. Should a heater in the second string subsequently fail, the gas tube will again fire and maintain dc continuity during trouble location tests.

For surge protection service, the tube is designed to pass a charge of 0.6 coulomb at a peak current of 75 amperes in either direction. This provides a reasonable margin for the maximum reverse surge which would occur in a near-shore repeater with a fault on the shore side, and for the maximum forward surge which would occur at the one-half voltage to ground point with a fault on the seaward side of that repeater. For a system of maximum length, the charge passed in either case is approximately 0.5 coulomb (the charge stored in a 100-mf capacitor at 5 kv). The magnitude of the peak current in the heater curcuit is less than 50 amperes under either condition and well within the capability of the tube.

The tube has a nominal breakdown voltage of 190 volts, and being located inside the power separation filters where the rate of rise of the transient voltage is relatively slow, the voltage rises only a few tens of volts above breakdown before the tube fires. Glow conduction is established within five microseconds at a tube voltage drop of about 70 volts. In less than one-half millisecond the cathode is heated sufficiently by ion bombardment to cause a transition to arc conduction, giving a tube voltage drop of about 11 volts. In this mode, as an ionically heated cathode device, the tube can conduct the large surge transient or the normal cable current as required. The power dissipation in the tube at normal cable current is approximately 5 watts. Tube life in this condition is more than 1000 hours, providing ample margin over the 100 hours estimated maximum time required to locate a cable fault. As a surge protection device the tube can conduct more than 50 maximum-energy surges without going out of firing voltage limits.

The fundamental characteristics and ratings are given in Table III.

4.2.2 *Design Details*

The detailed design of the working parts of the 456A is shown in Fig. 10. The two cathanodes are each mounted on three leads which are in turn strengthened, and made to move as a unit under shock, by means of the ceramic baffle. The baffle also prevents the arc discharge

TABLE III — 456A COLD CATHODE, GAS-FILLED ELECTRON TUBE

Maximum Ratings	
Average cathode current	450 ma
Surge cathode current	75 a
Coulombic charge	0.6 coulomb
Starts (firings)	100 max.
Surges	50 max.
Shelf life	20 years min.
Conducting life at 450 ma	100 hr min.
Ambient temperature	−10 to +50°C
Shock — 5 msec	50 g
Electrical Data	
Breakdown (firing) voltage	160 v min. 230 v max.
Breakdown time — to glow	20 μsec max.
Transition time — to arc	500 μsec max.
Sustaining voltage — glow	70 v
Sustaining voltage — arc	11 v

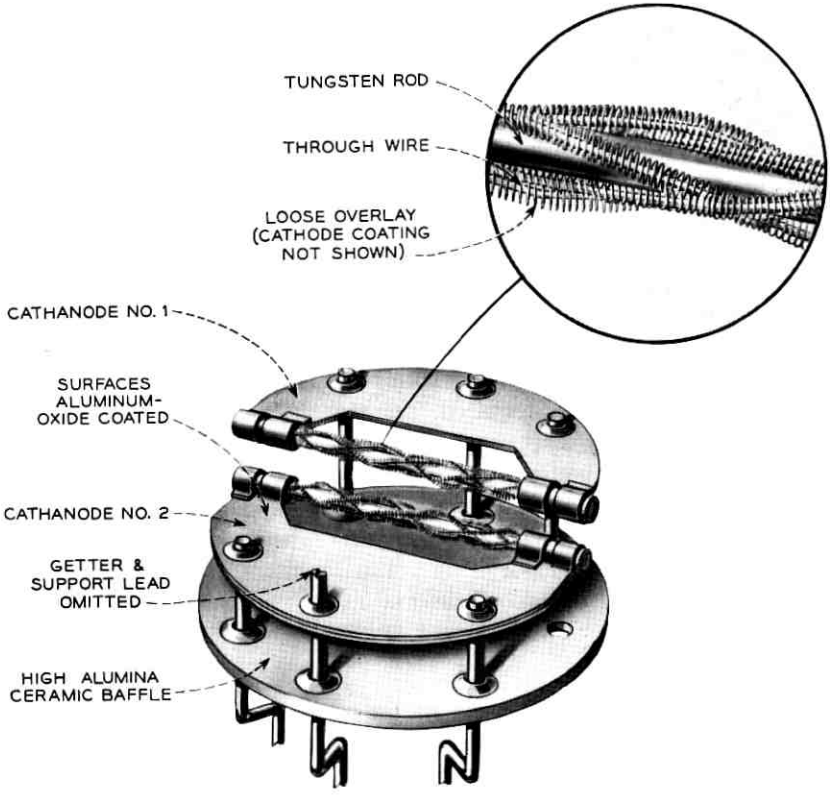


Fig. 10 — Structural features of the 456A tube.

from forming on the dumet seals. Each cathanode is formed from two identical "C" shaped molybdenum plates which clamp the active tungsten cathode.

The cathode is a composite structure made up of a 0.030-inch tungsten rod over which are braided six strands of 0.0045-inch tungsten wire, each strand carrying a loose over-winding of 0.001-inch tungsten wire. The crossed structure of braided wire and over-winding is filled with the emissive oxides of barium and strontium, providing small emission zones that have a loose thermal coupling with the rod. In establishing conduction, any one of these zones will be heated by ion bombardment, at a fraction of normal cable current, to a temperature sufficient for thermionic emission and operation as an ionically heated cathode. The zones are sufficiently short, however, that the heat energy resulting from high-current surges is conducted to the rod at the crossover points, preventing major damage to the finer wires.

The width and thickness of the molybdenum "C" plates are so proportioned that during the pumping process each part of the cathanode structure comes to the proper processing temperature simultaneously when heated by high-frequency induction. The two cathanodes are spaced approximately 0.100 inch apart and the gas filling is 18 torr of reagent-grade argon. A barium getter (not shown) is used to aid clean-up of impurity gases. The tube has a priming of one microcurie of radium bromide to insure fast breakdown in a dark environment.

4.3 *The 458A Signal Path Protective Tube*

4.3.1 *Operating Requirements*

The 458A tube is bridged across the transmission path at both the input and output of the repeater, just inside of the power separation filters (see Fig. 5). The most severe voltage surge the tube is required to handle is that caused by a short-circuit fault in the adjacent cable section. In the higher-voltage portions of the system this surge voltage may rise to a value of more than four kv in approximately one μsec. Since it is desirable to limit the voltage on many of the transmission path components to less than 1500 v, a very fast tube is required. The signal path tube is designed to fire in from 0.2 to 0.3 μsec on a 4 kv per μsec transient, limiting the surge to the transmission path to a maximum of 1200 v.

The charge shunted by the gas tube is a substantial portion of the charge stored in the high-voltage capacitors of the repeater, and may be as much as 1.5 millicoulombs. The discharge is oscillatory in nature and lasts about 10 μsec. The peak current through the tube on the first swing

may be as high as 1200 a. These high currents are carried by the tube in the metallic arc mode of conduction at a tube drop in the order of 15 v.

The ability of the tube to pass such surges is tested in a circuit equivalent to that in a repeater. The size of the capacitors is doubled, however, to insure an adequate testing margin. In this test each tube is surged ten times in each direction with a total integrated charge of 4.5 millicoulombs and a peak current of 1800 a. The tube is conservatively rated to pass 50 maximum cable surges.

In use in the cable the tube is not required to carry continuous current. There is a secondary use of the tube in the power supply for equipment protection in which approximately 2 ma dc is conducted until the trouble is corrected. For this use the tube is given a 5-ma average current rating.

The fundamental ratings and characteristics are given in Table IV.

4.3.2 *Mechanical Features*

The structural details of the 458A are shown in Fig. 11. The two identical cathanodes are mounted on a high-alumina ceramic disk, one on either side, with their support tabs at 90°. The cathanodes face each other through an aperture in the support disk.

Each cathanode is a square nickel cup with integral mounting tabs. The facing surfaces are coated with a thin layer of emissive oxides of barium and strontium, activated during tube processing by means of a high-frequency discharge to develop super-emissive cold cathodes. The "nutmeg grater" shaped perforations perform two functions: (1) the

TABLE — IV — 458A COLD CATHODE, GAS-FILLED ELECTRON TUBE

Maximum ratings	
average cathode current	5 ma
surge cathode current	1500 a
coulombic charge	1.5 millicoulombs
surges	50 max.
shelf life	20 years (min.)
conducting life at 5 ma	200 hours (min.)
ambient temperature	−10 to +50°C
shock — 5 msec	50 g
Electrical data	
breakdown voltage	60 v
sustaining voltage — 5 ma	50 v
breakdown time — 500 v	5 μsec max.
breakdown time — 4 kv/μsec	0.7 μsec max.

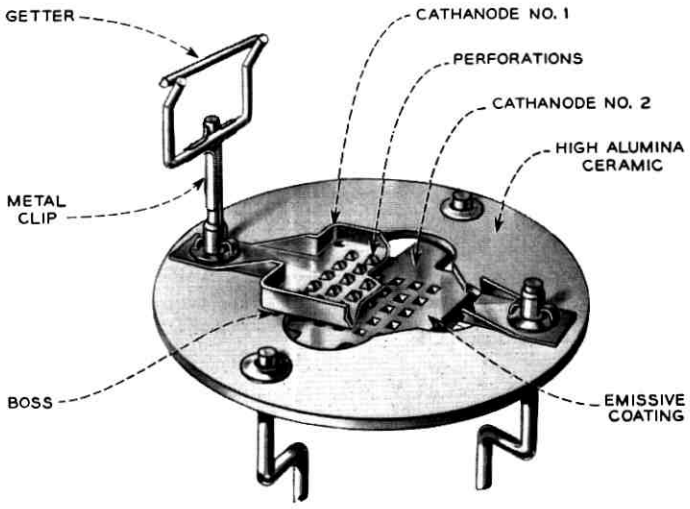


Fig. 11 — Structural features of the 458A tube.

hollow cathode effect of the small depressions increases the emission efficiency and life; (2) they allow a visual observation of the glow over the cathode surface to determine the uniformity of emission and cathode coverage. A small boss is provided at each corner of the cathanode to limit and position the contact area on the ceramic disk. This provides a leakage path of greater than 1000 megohms between the elements, even after the sputtering of the cathode material due to high-current arcs. The barium getter is the same as that used in the 456A tube.

The gas filling in the 458A tube is 1 per cent argon and 99 per cent neon at 60 torr pressure. The plane-parallel electrode geometry at a 0.030-inch spacing gives a nominal breakdown voltage of 60 v and a sustaining voltage at 5 ma of 50 v. One microcurie of radium bromide is used as a priming to insure sufficient initial ionization for high-speed operation in the absence of light.

4.4 *General Considerations*

In the design of the two gas-filled tubes the glassware, stem, bulb, and basing follow the pattern of the 455A-F amplifier tube. Both gas tubes are mounted between rubber cushions in aluminum housings (See Fig. 9). The housing for the power bypass tube has a black anodized finish and is secured to the power separation filter chassis, which it uses as a heat

sink. This insures a bulb temperature of less than 130°C when carrying full cable current. The signal path tube has no heat problem and is in a bright housing. This tube and lead-out braid, however, require insulation from the housing for 6-kv operation.

In addition to the basic twenty-year standby life requirement, the tubes must also not reduce the system reliability through their own failure. Shorting of the elements has been made virtually impossible by generous spacings, rugged structural design and the multiple securement of parts. Both types can withstand some five times the shock levels expected in the present cable systems. The tubes operate on the high-pressure side of the Paschen minimum, and gas leak-in cannot reduce the firing voltages to unsafe values.

Raw material control through strict specifications, adequate testing procedures, and lot prove-in follow the pattern of the amplifier tube. Similarly, the parts production, handling, processing and cleaning, quality tests and inspections have also followed approved procedures.

The tubes are carried through a comprehensive aging and operating procedure simulating operating conditions in the cable. Elaborate testing schedules evaluate the performance at each stage in the processing, giving the detailed behavior of each tube. The tests include high-energy surge tests at or above maximum ratings and a two-hour thermal pulse at 125°C to evaluate over-all cleanliness. The tubes are checked for stability over a minimum period of three months and for operation in the dark to insure adequate radium priming. The power bypass tube is given a 30-day test in the dark at 4°C with cable voltage applied.

V. CONCLUSION

When electron tubes were first used in deep-sea repeaters for the SB systems, it was recognized that the undertaking was ambitious and perhaps even audacious. The faith exhibited in proceeding with this application of electron tubes appears well justified with more than 80 million amplifier tube hours of operation on sea-bottom with no failures. Strict attention to details of processing, fabrication, aging, testing, and selection constitute the background for this achievement. By modification and extension of this formula, it is expected that a similar record will be achieved for the electron tubes in the SD systems.

The development of the three codes of tubes for use in the SD system was a project staffed by many members of Bell Laboratories, who contributed a variety of knowledge and skills. No attempt will be made to single out individuals for special mention. It was a team effort.

APPENDIX

Discussion of Reliability

The reliability objective for the electron tubes of the SD system was stated as: "The probability of system failure due to a tube failure shall not exceed 50 per cent for a twenty-year service period for a 3000-mile system."

The probability of no system failures

$$P_{f(0)} = e^{-\lambda}$$

where λ is the expected number of tube failures for the period. For a 50 per cent probability $e^{-\lambda} = 0.5$ and $\lambda = 0.69$.

The mean time between system failures is $20/\lambda$ or twenty-nine years.

For a 3000-mile system with a twenty-mile repeater spacing there will be 150 repeaters. For defects not minimized by circuit redundancy, there are six tubes per repeater. Then, for a twenty-year period

$$\lambda = (150 \times 6 \times 20)/T$$

where T is the mean time (in years) between random failures. For $\lambda = 0.69$, $T = 26{,}100$ years.

If 0.69 failure is experienced among the 900 tubes for a twenty-year period, this corresponds to a failure rate of $0.69/900 = 0.00077$ or 0.08 per cent for the twenty years, or to 4.4×10^{-9} failure per tube per hour (4.4 fits).

For the defects which, because of redundancy, would cause failure only if both circuits were affected:

$$\lambda = 150[(3 \times 20/T)^2] = 0.69 \text{ (approximately)}$$

and

$$T = 885 \text{ years mean time between failures.}$$

Since with redundancy the required tube life is about $\frac{1}{30}$th that required without it, the corresponding failure rate would be 2.4 per cent and the number of fits 132.

This analysis has assumed that no "wear-out" mechanism is involved in the twenty-year service period. While mechanisms such as depletion of cathode coating or reducing agents are known, the rates are such that they have no significant effect in this period.

REFERENCES

1. McNally, J. O., Metson, G. H., Veazie, E. A., and Holmes, M. F., Electron Tubes for Transatlantic Cable System, B.S.T.J., **36,** Jan., 1957, p. 163.
2. Ford, G. T., The 404A — A Broadband Amplifier Tube, Bell Laboratories Record, **27,** Feb., 1949, p. 59.
3. Ford, G. T., and Walsh, E. J., The Development of Electron Tubes for a New Coaxial Transmission System, B.S.T.J., **30,** Oct., 1951, p. 1103.
4. Walsh, E. J., Fine Wire Type Vacuum Tube Grid, Bell Laboratories Record, **28,** Apr., 1950, p. 165.
5. U. S. Patent No. 2,936,385, awarded to R. L. Kaufman of Western Electric Co.
6. Olsen, K. M., High-Purity Nickel, Bell Laboratories Record, **38,** Feb., 1960, p. 54.
7. Kern, H. E., Research on Oxide-Coated Cathodes, Bell Laboratories Record, **38,** Dec., 1960, p. 451.
8. Frost, H. B., High-Purity Nickel Cathodes: Performance Studies, Bell Laboratories Record, **39,** Jan., 1961, p. 18.
9. Biondi, F. J., et al., A.S.T.M. Spec. Tech. Publ. No. 246, 1958; reprinted as Bell System Monograph 3143.
10. Brewer, S. T., Dickinson, F. R., and von Roesgen, C. A., Repeaters and Equalizers for the SD Submarine Cable System, B.S.T.J., this issue, p. 1243.

Cable Power Facility

By J. D. BISHOP and S. MOTTEL

(Manuscript received March 19, 1964)

Submarine cable system power supplies must be designed to provide an uninterrupted flow of precisely controlled, transient-free operating power for a time interval equal to the service life of the cable system. Heavy reliance on redundancy along with conservative application of semiconductor and magnetic circuit elements form the basis for the circuit design. The equipment design emphasizes the techniques required to withstand high voltages, provide ease of maintenance and insure personnel safety.

Saturable reactors operated in the constrained mode serve as the basic power control element. AC power for the saturable reactors is obtained from a transistor inverter fed from a continuously floated storage battery. Metering-type magnetic amplifiers are used to isolate the regulation, metering and alarm circuits from the high-voltage output. Transistor amplifiers and temperature-compensated silicon reference diodes complete the regulating loop.

I. INTRODUCTION

The power supplies for the SD submarine cable system[1] must be designed to provide an uninterrupted flow of precisely controlled, transient-free operating power if the long life and stable transmission characteristics inherent in the submerged repeater design are to be obtained. Ideally, the power system should be capable of providing this power without interruption for the life of the cable system. Equally important is the need for reliable automatic alarm and shutdown circuits. These signal the presence of abnormal cable currents and voltages and turn down the power if hazardous levels are reached, whether the cause of the trouble is within or external to the power plant.

II. DESIGN REQUIREMENTS

Fig. 1 illustrates the basic power feeding arrangement used. Nominal current and voltage for a 3600-nautical mile (nm) system are shown.

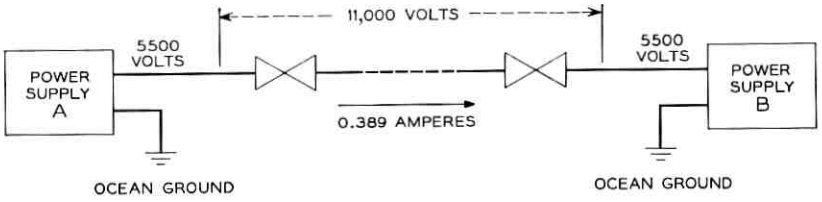


Fig. 1 — Basic power feed arrangement.

Power is normally applied from both ends to minimize the applied voltage stress. For systems shorter than 1800 nm, power may be applied optionally from both ends or one end only.

Complete electrical design requirements are given in Table I. These values apply for supplying power at one end of the cable.

TABLE I — ELECTRICAL DESIGN REQUIREMENTS

Normal voltage	5500 v
Normal current	0.389 a
Ripple	1 v rms maximum
Noise	−120 dbm maximum (0.1 mc–1.1 mc) referred to transmission input
Normal accuracy of current control	±0.4% absolute
Load current regulation	
1000-v change	1.0%
Rectifier failure	−0.5%
Output resistance above 6500 v	2500 ohms

External alarms:

Power plant output

Minor alarm	Major alarm
current ±2%	current ±5%
voltage ±5%	voltage ±10%

Rectifier output

Minor alarm	Major alarm
current −5%	current +5%
less than 500 v	more than 7500 v

Protective shutdowns:

Rectifier

5% over-current
over 7500 v

Plant

over 9000 v

III. DESIGN PHILOSOPHY

Redundancy was considered the key to achieving the required degree of reliability. To fully exploit this approach, three basic principles were felt to be essential: minimum interdependence of one functional element upon another, use of static components only, and minimization of the number of components which could not be made redundant directly connected to the power plant output. Fig. 2 illustrates the basic power system. Two identical power plants are used, independent of each other except for the sharing of the ac input power source and cable terminating equipment.

Power for the battery charging rectifiers is normally obtained from commercial sources. When commercial power failure occurs, these rectifiers are fed from the emergency power system of the station. The batteries are normally maintained on continuous charge with their outputs fed directly to the inverters. The inverters are designed to operate over the normal battery voltage range without end-cell switching or other voltage adjustment means.

Regulation of the cable current is done by the high-voltage rectifiers. These, along with their associated inverters, are designed to have full

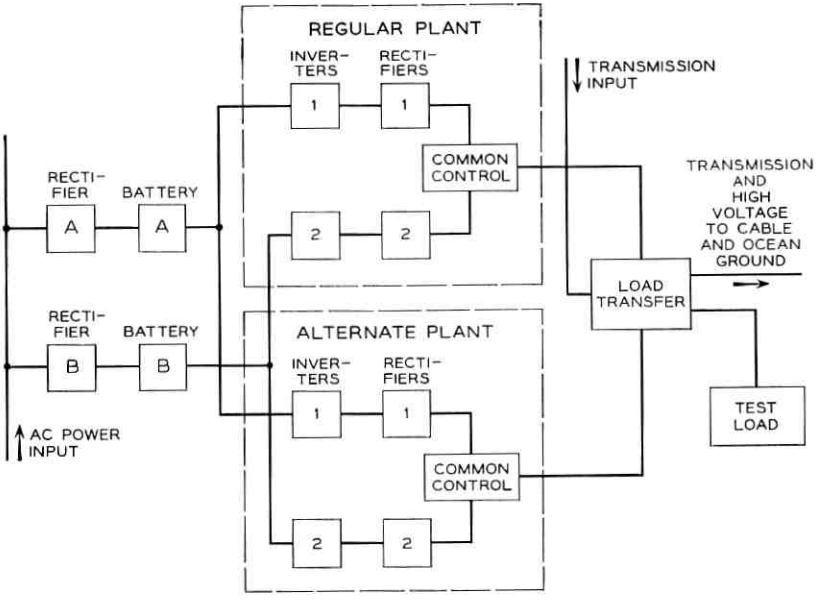


Fig. 2 — Power system functional block diagram.

load capability. In normal operation they are connected in series and adjusted to share the load equally. If the output of one of the rectifiers fails, the remaining rectifier assumes the cable load. When this occurs, it is normal practice to transfer the power feed to the standby power plant. The load transfer is designed to permit such a transfer without disturbance of the cable current.

An adjustable test load is provided to permit testing of both individual rectifiers or a complete power plant. A precision calibrating circuit is included as an integral part of the test load to permit periodic calibration of the cable current meters.

IV. EQUIPMENT DESIGN OBJECTIVES

The equipment design objectives were as follows:

(a) to provide an equipment arrangement capable of withstanding the high voltages present without dielectric breakdowns or corona for extremely long service periods,

(b) to give easy access to apparatus for repair or maintenance operations

(c) to guarantee personnel safety and proper operation of the plant by design features,

(d) to provide mechanical integrity to withstand 3-g shock loads due to atomic blasts when installed in hardened sites,

(e) to require a minimum of installation effort, and

(f) to provide a distinctive, pleasant appearance.

V. GENERAL DESCRIPTION

5.1 *Equipment Design*

The power plant is composed of a number of seven-foot-high cabinets, the exact number and arrangement being a function of the specific application. Among the considerations involved are the length of cable being powered, the type of installation (hardened or nonhardened), and the degree of redundancy desired. An arrangement of cabinets for a regular supply is shown in Fig. 3. An alternate power plant does not include a load transfer. A test load cabinet is furnished with each regular supply.

The power plant equipment cabinets are 7 feet high, with split rear doors. Two widths are available. The cabinets are of welded aluminum construction. A different type of cabinet is used for the test load and is available with wheels or for direct mounting to the floor. A number of slide-mounted equipment units of a generally similar mechanical configuration are incorporated in the cabinets. Typical units are shown in Fig. 4.

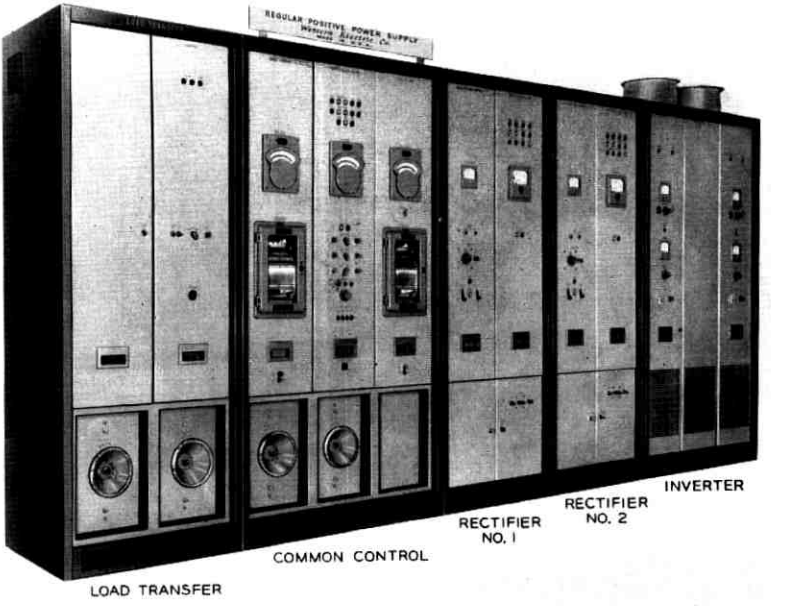


Fig. 3 — Terminal plant — regular power supply.

5.2 *High-Voltage Rectifier Circuit Design*

A series-connected saturable reactor,[2] operated well into the constrained mode, is used as the main power-control element in the rectifiers. This device electrically approximates an ideal current source by virtue of its ability to maintain an output current whose magnitude is proportional to the control current but is virtually independent of supply and load voltage variations. When appropriately designed with a control circuit having a long time constant, it behaves as a current source under dynamic as well as static operating conditions.[3]

Fig. 5 shows the circuit schematically with typical wave shapes. The rectangular wave shape of the load current is independent of supply and load voltage wave shapes in the linear operating range, as shown. Typical saturable reactors used in this system exhibit only a 2 per cent change in load current for a 25 per cent change in supply voltage and a load voltage change equivalent to 6500 v.

The principal reasons for the use of the saturable reactor were:

(1) it is inherently a current source, requiring only a nominal amount

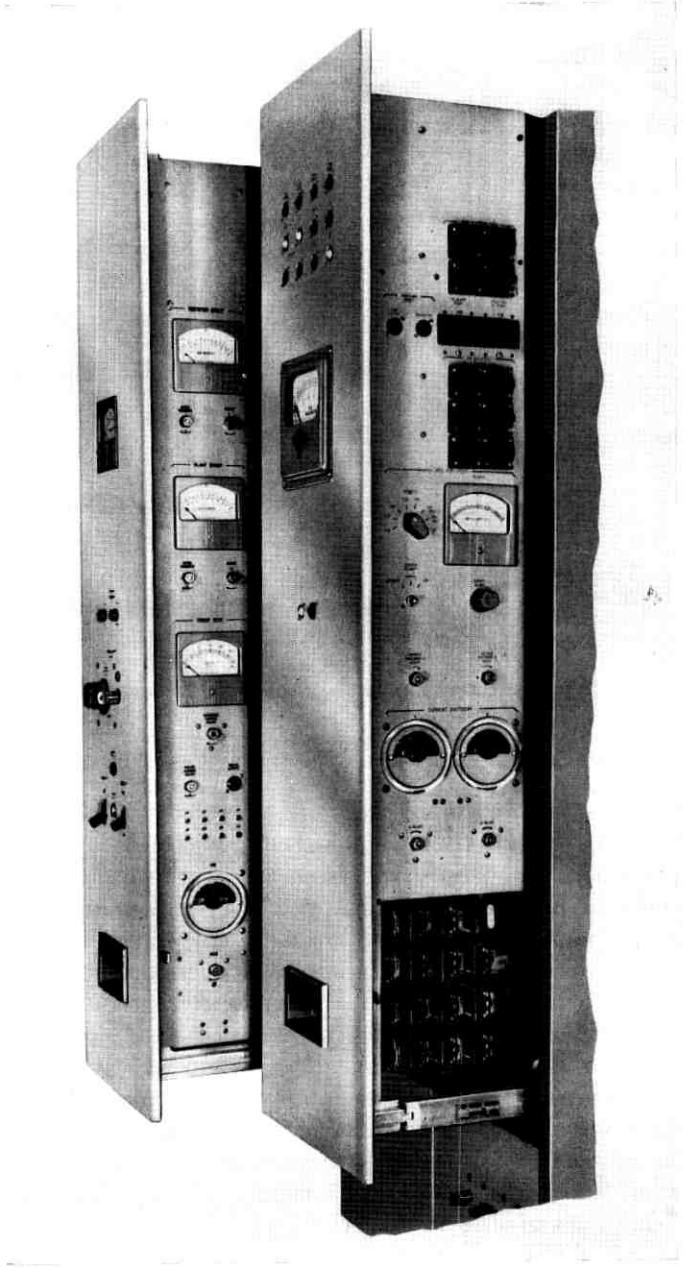


Fig. 4 — Typical slide-mounted units.

of over-all regulating circuit loop gain to meet both static and dynamic performance requirements;

(2) the rectangular load current wave shape is ideal for rectification, giving a small percentage output ripple;

(3) a long control circuit time constant prevents sudden and rapid change in the rectifier output, providing ample time for detection and turndown of a runaway rectifier before significant changes in cable current can occur; and

(4) insensitivity to supply voltage variations permits use of an unregulated ac power source.

The practical problems associated with designing a saturable reactor circuit for operation at these power levels virtually rule out the use of commercial power frequencies. A 60-cycle design would be impracticably large in volume and weight. The attractiveness of the saturable reactor

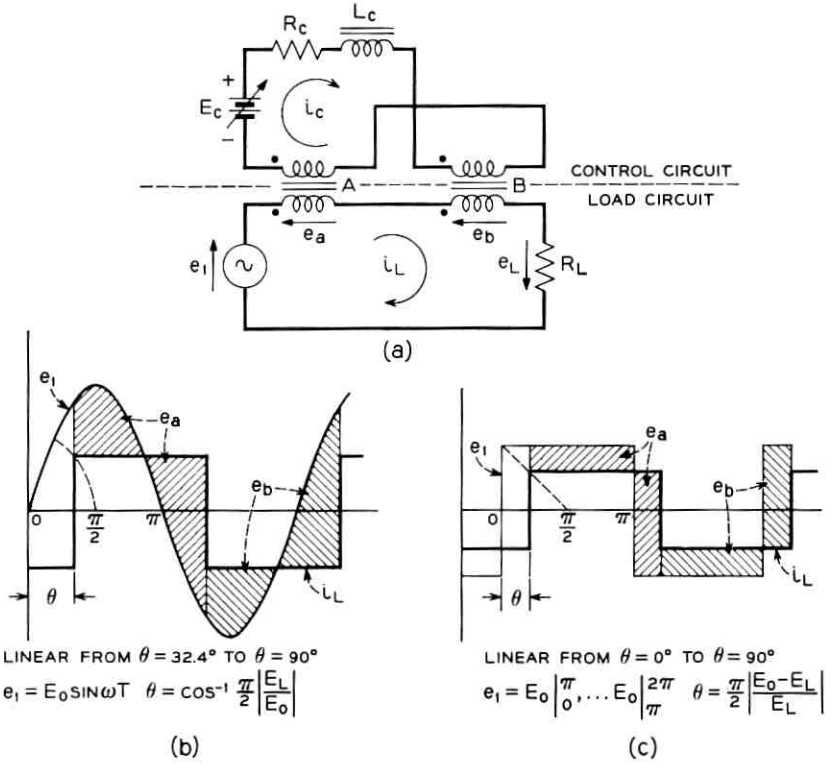


Fig. 5 — Saturable reactor regulator circuit: (a) schematic, (b) waveforms with sinusoidal excitation, (c) waveforms with square-wave excitation.

approach, however, was sufficiently great to justify the use of higher supply frequencies.

Since a reliable source of ac operating power is essential to fulfillment of the design objectives, careful consideration must be given to its selection. The availability of power transistors having collector dissipation ratings of 250 watts makes possible static inverter designs capable of meeting the power requirements for this system. These inverters, in their simplest form, generate rectangular output voltage wave shapes (although sinusoidal outputs can be obtained by appropriate techniques). Saturable reactors as used here are inherently insensitive to supply voltage wave shape, within broad limitations, and a single design can be made to give satisfactory performance with either wave shape. The rectangular wave shape offers the advantage of the widest possible linear operating range for a given reactor, as may be seen by reference to Fig. 5, permitting the most economical design for a given set of requirements. Where operation with both wave shapes is required, this economy cannot be obtained.

Proven schemes for generation of reliable ac power based on the use of rotating machines are widely used. Because of the availability of reliable 400-cycle rotating machines and the broad experience gained with this frequency in military systems, a decision to use nominal 400-cycle power was made (although even higher frequencies would have been more suitable from a size standpoint).

The advantages of the static inverter, including freedom from routine maintenance, make it the preferred ac power source for this system. Although a suitable proven design did not exist at the outset of this project, the availability of rotating machines as a back-up approach permitted ample time to prove in an inverter design. This was accomplished, and static inverters are used at all terminal installations.

To obtain the full potential of saturable reactor regulation, it is necessary to connect the outputs of the rectifiers in series. This stems largely from the desire to maintain a constant cable current under all operating conditions. If the rectifier outputs were connected in parallel, for example, loss of output from one of the rectifiers would require a sudden increase in the output current of the remaining rectifier if a significant dip in cable current were to be avoided. The relatively long time constant of the saturable reactor circuit precludes rapid adjustment of rectifier output current. A second disadvantage would be the need for a greatly increased regulator loop gain in order to provide the required gross changes in control current without a significant change in cable current. While this approach represents a departure from past practices, it is not entirely with-

out precedents. A dual of this method may be found in the parallel operation of voltage sources in the commercial power industry. This arrangement permits addition or removal of supplementary sources without significant disturbance to the load.

Fig. 6 shows the basic interconnections within a plant. Shunt diodes are connected across the output of each rectifier. Under normal operating conditions these diodes are reverse biased and conduct a negligible current. If the output current of a rectifier should fail, its voltage will drop to zero and attempt to reverse polarity as a result of current flow from the remaining rectifier. The shunt diode, becoming forward biased, provides an alternate path for the current, permitting the remaining rectifier to supply the load. Only two rectifiers are shown in the illustration, but the technique can be extended to any number of series-connected rectifiers.

To insure stable load voltage sharing between rectifiers, two steps must be taken. The first is a reduction of the nearly infinite rectifier output resistance to a finite stable value. The second is provision of a highly stable regulating circuit.

While any desired amount of rectifier output resistance could have been obtained by a combination of both current and voltage feedback, the complexity of this approach was felt to be excessive when contrasted with the simplicity of an equivalent physical resistance connected across the output. To limit the change in cable current to less than 0.5 per cent following failure of a rectifier, this resistance must have a value of 1.5 megohms. The less than 3-watt power consumption is insignificant.

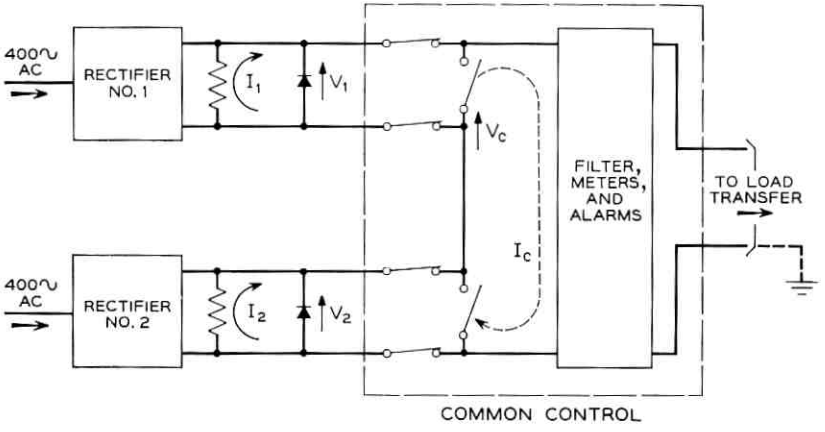


Fig. 6 — Power plant interconnections.

If the drift in voltage between two rectifiers is to be held to less than 75 volts for long periods of time, the drift in rectifier output current must be held to less than about 0.025 per cent for equivalent periods of time. Achievement of this high level of stability requires careful regulator circuit design. The availability of temperature-compensated silicon reference diodes and matched transistors permitted meeting this requirement with direct-coupled transistor amplifier circuits.

No special precautions were found necessary other than compliance with accepted feedback stability criteria to insure freedom from oscillation and hunting. The rectifiers are designed to be stable for both open- and short-circuit conditions.

5.3 *Common Control Circuit*

By virtue of its relationship to the over-all system, the common control is one of the more critical parts of the system. Either terminal-to-terminal short circuits of shunt-connected apparatus or a grounding failure of apparatus connected to the high-voltage conductor will result in immediate loss of cable power. Great care was therefore taken to minimize the likelihood of such failures. This was done largely by keeping the number of such components to a minimum and by paying special attention to the design and manufacture of these components to insure the highest possible quality.

While an open-circuit failure of series-connected components will also result in immediate loss of cable power, series-connected components are limited to magnetic amplifiers and filter inductors. With suitable design and manufacture the probability of an open-circuit failure in these can be made virtually zero.

VI. DETAILED DESCRIPTION

6.1 *Inverter*

The design of the inverter was based on the use of 250-watt silicon power transistors developed by Westinghouse Electric Corporation.

The inverter is designed to operate from a 42 to 52-vdc source. Two outputs are provided. The main output, rated at 4.5 kva, is unregulated and supplies power to the saturable reactor in the rectifier. A second output, rated at 0.3 kva, supplies regulated operating power to the rectifier feedback amplifier and the metering magnetic amplifiers in both the rectifier and common control. The regulated output voltage is held statically to within ± 1 per cent for combined input voltage and load

current changes. No special speed of response requirements was necessary, however, because of the type of regulating circuit used in the rectifier.

Special attention was given to the design of the main power stage to insure reliable operation with the saturable reactor load. The inverter must perform satisfactorily with a load current whose magnitude and phase angle describe the region bounded by 0 to full-load current and 0 to unity power factor. For example, a rectifier delivering rated current into a short circuit will draw full-load kva at nearly 0 power factor from the inverter.

With the exception of the oscillator, each inverter stage is of the bridge configuration. The oscillator uses a center-tap circuit to permit simple application of series LC timing.

Provision is included in the oscillator circuit for frequency synchronization of all inverters in a power plant. The synchronizing circuit is loosely coupled, so that faults on the synchronizing bus will result only in loss of synchronization. Synchronization is provided to eliminate possible beat frequency interference between the combined outputs of the rectifiers in a plant.

6.1.1 *Reliability*

Conservative transistor operation was considered essential in obtaining reliable inverter performance. To achieve this, not only must the average and peak transistor power dissipation be held within conservative limits, but also the locus of the product of instantaneous collector current and voltage must be maintained within the reliable operating region for the transistors used.

The peak collector power dissipation of the transistors has been held to less than their rated average power dissipation. With the exception of the power stage, average collector power dissipation has been held far below ratings. The heat sinks provided are capable of maintaining collector junction temperatures below 100°C under all operating conditions.

Economic considerations required that the transistors in the power stage be operated much closer to their ratings, in order that the total number of devices be held within reasonable limits. The power stage design permits failure of up to one transistor in each of the four circuit legs before catastrophic inverter failure becomes imminent. This is accomplished by the use of 8 parallel-connected individually-fused transistors in each leg of the power stage. The fuses are of the fast acting current limiting type. Sufficient overload current capacity is available in

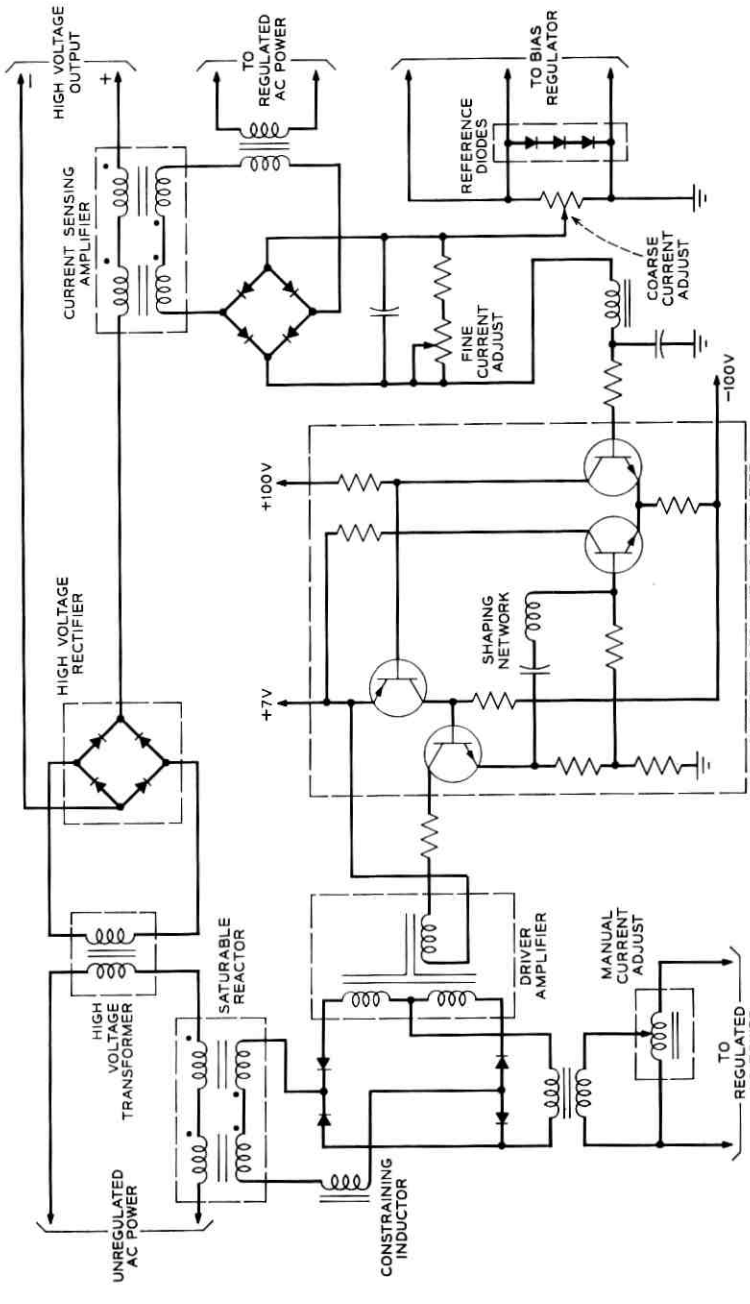


Fig. 7 — Current regulating feedback circuit.

the circuit to blow a fuse following a transistor failure without serious overload of the remaining transistors. The circuit has been designed to operate at rated load for at least 8 hours with only 7 transistors per leg.

An automatic fuse failure alarm is provided to signal the presence of a blown fuse. A manually operated metering circuit is provided to permit in-service measurement of individual transistor currents, permitting detection of low-gain or open-circuited transistors.

6.2 *Rectifier*

It was shown earlier that output failure of one of the rectifiers—resulting from a terminal-to-terminal short circuit, for example—while undesirable, does not result in loss of cable power. Grounding failures within the rectifier directly connected to the high-voltage power plant output conductor will result in immediate loss of cable power. As in the common control, it again becomes important to minimize the likelihood of grounding failures.

Metering-type magnetic amplifiers, which functionally exhibit at dc all the attributes of a high-quality ac current transformer, are used to isolate the metering, regulating and alarm circuits from the high-voltage part of the circuit. This approach offers the dual advantage of materially reducing the number of components directly subjected to high voltage and permits use of conventional devices and circuit techniques for metering, regulation and alarms. Only those components which by their basic function must be connected to the high-voltage conductor are so connected.

6.2.1 *Current Regulation*

The current regulating circuit is shown in Fig. 7. Negative feedback is provided around the 3-stage transistor amplifier to reduce the effects of unit-to-unit variations in transistor parameters and to provide a convenient point for applying the shaping network required to insure adequate gain and phase stability margins for the regulation system.

A reference voltage of about 20 v obtained from temperature-compensated silicon diodes is required to obtain the required regulator dc stability. Bias current for the reference diodes is obtained from a single-stage transistor regulator.

The output of the transistor amplifier controls a self-saturating magnetic amplifier (driver amplifier) which in turn drives the main regulating saturable reactor.

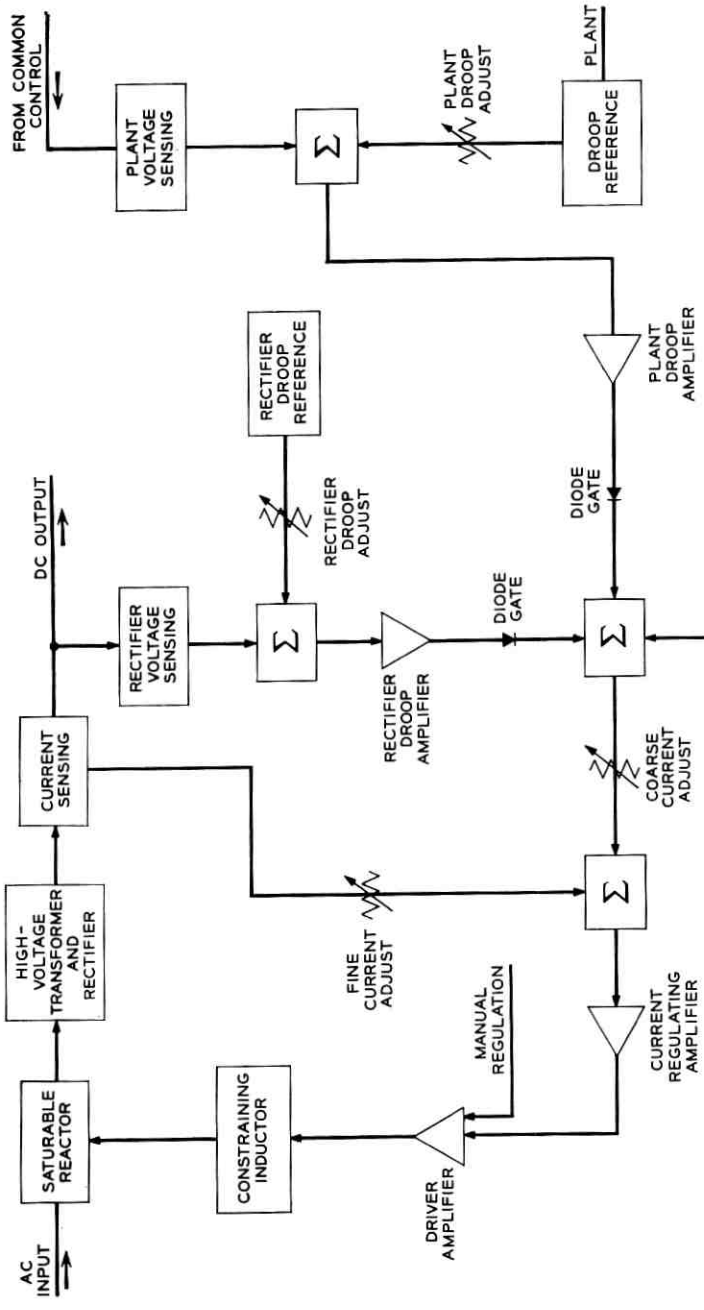


Fig. 8 — Rectifier feedback regulator functional block diagram.

6.2.2 *Current Droop*

Voltage limitations in the repeater power separation filter capacitors make it preferable, when the applied cable voltage exceeds 6500 volts (as might occur with abnormally high earth potentials), to reduce cable current rather than allow further voltage increase. A transition from current to voltage regulation is obtained by the addition of a second feedback path to the regulator, responsive to plant output voltage. This feedback path is coupled into the first path at the current regulator reference. The circuit is so designed that in the current droop region the magnitude of the reference voltage is reduced in direct proportion to the increase in output voltage. Since the current regulator acts to maintain the rectifier output current accurately proportional to the magnitude of its reference (voltage), effective control of current is possible.

The gain of the droop circuit is adjusted to produce a 100 per cent reduction in cable current for a 1000-v increase in cable voltage (normally from 6500 to 7500 v). A 3-stage single-ended transistor feedback amplifier provides the controlled gain and power required to modulate the current regulator reference voltage. A functional block diagram of the over-all rectifier feedback regulating circuit is shown in Fig. 8.

Two separate droop circuits are provided with each rectifier. The first, as described above, responds to plant output voltage. The second, functioning in a similar manner, responds to rectifier output voltage and serves as back-up protection. It is normally adjusted to act at the same voltage as the plant droop circuit. Since the voltage of an individual rectifier is normally about one-half the plant voltage, the rectifier droop circuit will function only if one of the plant droop circuits fails.

6.2.3 *Droop Circuit Test*

Under normal circumstances, the occurrence of large earth potentials is infrequent, but it is important that the droop circuits function properly when required. Confidence in these circuits can be obtained if an in-service test can be performed. To be meaningful, this test should involve as much of the droop circuit as possible.

A second control winding (test winding) is provided on each of the metering magnetic amplifiers. The output of these amplifiers is proportional to their net control ampere turns. By passing a dc current through the test winding, it is possible to either increase or decrease the amplifier output while in no way affecting normal system output.

The input and output of each droop amplifier is metered during test. The input meter indicates either the plant or rectifier voltage; the output

meter indicates the voltage difference between the droop amplifier output and the current regulator reference voltage. This voltage indicates the amount of margin existing before droop action takes place. By adjustment of the test winding current, this margin can be brought to zero. The indication on the input meter then corresponds to the voltage at which droop will occur.

Selector switches are provided to permit application of an adjustable current to each of the test windings. Normal droop action of the circuit is inhibited during test.

Fig. 9 illustrates the voltage-current characteristic of a power plant when the effects of the current droop circuit are included.

6.2.4 *Manual Regulation*

The main regulating saturable reactor is sufficiently stable over time intervals of 10–15 minutes to permit usable rectifier operation on a manually controlled basis. This provides a back-up operating mode. Operation in this mode involves only the driver magnetic amplifier and a continually adjustable autotransformer, the latter serving as the manual current adjust control. In this mode of operation, control current is removed from the driver magnetic amplifier. It then behaves purely as a passive element, delivering a dc output voltage proportional to input voltage obtained from the manual current adjust control. Input

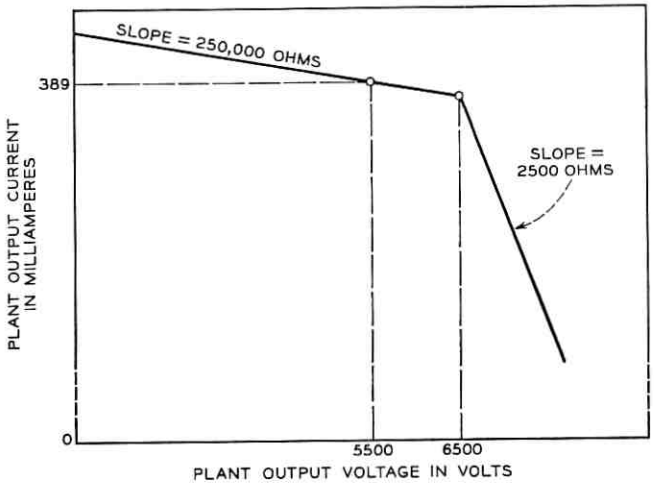


Fig. 9 — Power plant voltage-current characteristic curve.

power for the driver amplifier is taken from the regulated ac output of the inverter.

6.2.5 *Rectifier Alarms and Shutdown Circuits*

To meet the over-all design objectives, the rectifiers must provide automatic alarm and shutdown protection when abnormal currents or voltages exist. These abnormal conditions can occur as a result of malfunction of a rectifier, operator error, excessive earth potentials or faults on the cable system.

Two classes of alarms are provided. Those designated "minor" imply the presence of an abnormal but tolerable current or voltage. Those designated "major" imply the presence of hazardous current or voltage and require shutdown of the offender.

Two independent sensing circuits feeding separate, permanently adjusted, meter-type relays provide the current alarms. Each relay responds to both high- and low-current conditions. When the high-current contact on either relay makes, shutdown as well as an alarm indication takes place. Shutdown is produced by release of the main ac input contactor. A second fast-acting electronic shutdown circuit, fed along with one of the meter relays from one of the sensing magnetic amplifiers, is provided. This circuit produces shutdown by interrupting the dc input to the inverter oscillator. Power flow from the inverter ceases within about 15 msec after the shutdown level is reached. The shutdown circuits are all adjusted to operate at the same current. The electronic circuit, being faster, operates first. The relays provide back-up protection.

A single sensing circuit feeding both an adjustable contact indicating meter relay and a fast-acting electronic circuit is provided for the voltage alarm. Shutdown is accomplished in the same manner as for the current alarms, with both circuits adjusted to operate at the same voltage. The meter relay serves also as the rectifier output voltmeter.

While great care has been taken in the design of the rectifier and inverter, a malfunction is likely to result in abnormal output. Where low current occurs, the shunt diode automatically bypasses the output. When high output occurs, the alarms will provide immediate warning, and if the output is over 5 per cent above normal will result in immediate shutdown. When sudden failure of the feedback regulator occurs, resulting in maximum output, the long (0.3-second) time constant of the main saturable reactor control circuit provides ample time for detection and shutdown before the current can climb appreciably above the shutdown level.

Alarm lamps are provided for each specific alarm function. These lamp indications persist for any alarm that signals the office alarm. An approximately 2-second time delay is provided on all minor alarms. If the alarm comes in and clears before the delay has elapsed, the alarm clears automatically and no office alarm is transmitted.

6.2.6 *Alarm Testing Circuits*

Under normal conditions, the frequency of alarm indications is very low. It is important, however, as with the droop circuits, that these circuits function properly when abnormal conditions occur. Additional positions are provided on the magnetic amplifier test selector switches to permit in-service testing of these circuits. Where the function to be tested includes shutdown, shutdown is inhibited. All other aspects of the alarm and shutdown circuits are unaffected by the inhibiting action. Through application of the test signal to the magnetic amplifier, confidence in the integrity of the alarm circuits is maintained, since the complete alarm path is tested. Inhibiting involves bridging of normally closed relay contacts.

The design of the alarm test circuit does not permit accurate adjustment of the alarm operating points, but rather is intended solely as a functional check. The design of the meter-type relays insures that they will hold their calibration over long periods of time. Accurate calibration of these relays is possible by taking the equipment out of service and operating it with the test load. Fig. 10 shows the test circuit used.

6.2.7 *Spark Gap*

A protective spark gap is connected across the dc output of the rectifier to limit the maximum instantaneous voltage which can be developed. This is done to protect components in the rectifier against damage resulting from excessively high voltage. A resistor and a current sensing relay are connected in series with the spark gap. Operation of this relay results in rectifier shutdown. The resistance is provided to limit the peak current in the protector circuit.

6.3 *Common Control*

The basic functions of the common control are:

(1) connect and disconnect means for the individual rectifiers in a plant,

(2) voltage and cable current metering,

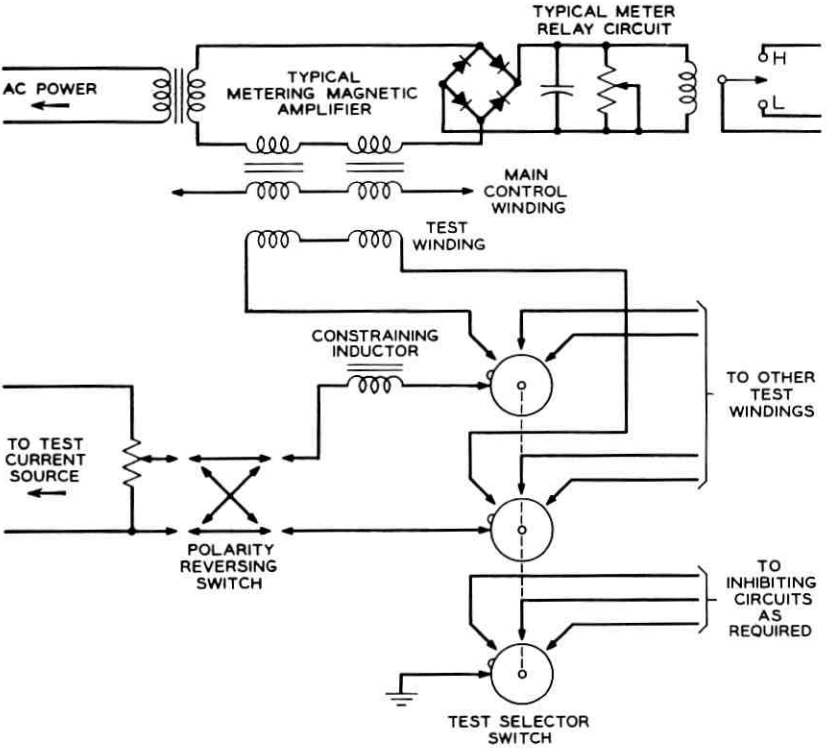


Fig. 10 — Magnetic amplifier test circuit.

(3) monitoring of cable voltage and cable current by means of alarms,
(4) final power supply output ripple filtering, and
(5) means for surge voltage limiting.

As in the rectifier, metering magnetic amplifiers are used to isolate the alarm circuits from the high voltage circuit. Since one side of the plant output is at ground potential, indicating voltmeters are operated directly with high-resistance multipliers.

A suppressed-zero meter (300 to 450 ma), operating from a metering magnetic amplifier having exceptional stability, is used for precision cable current metering. The magnetic amplifier is connected in the high-voltage conductor at the output of the common control. AC power for these magnetic amplifiers is selectable by means of a selector switch so that power can be obtained from any of the rectifiers.

Since none of these alarms results in shutdown, their failure results only in loss of alarm capability. Alarm failure resulting from loss of ac

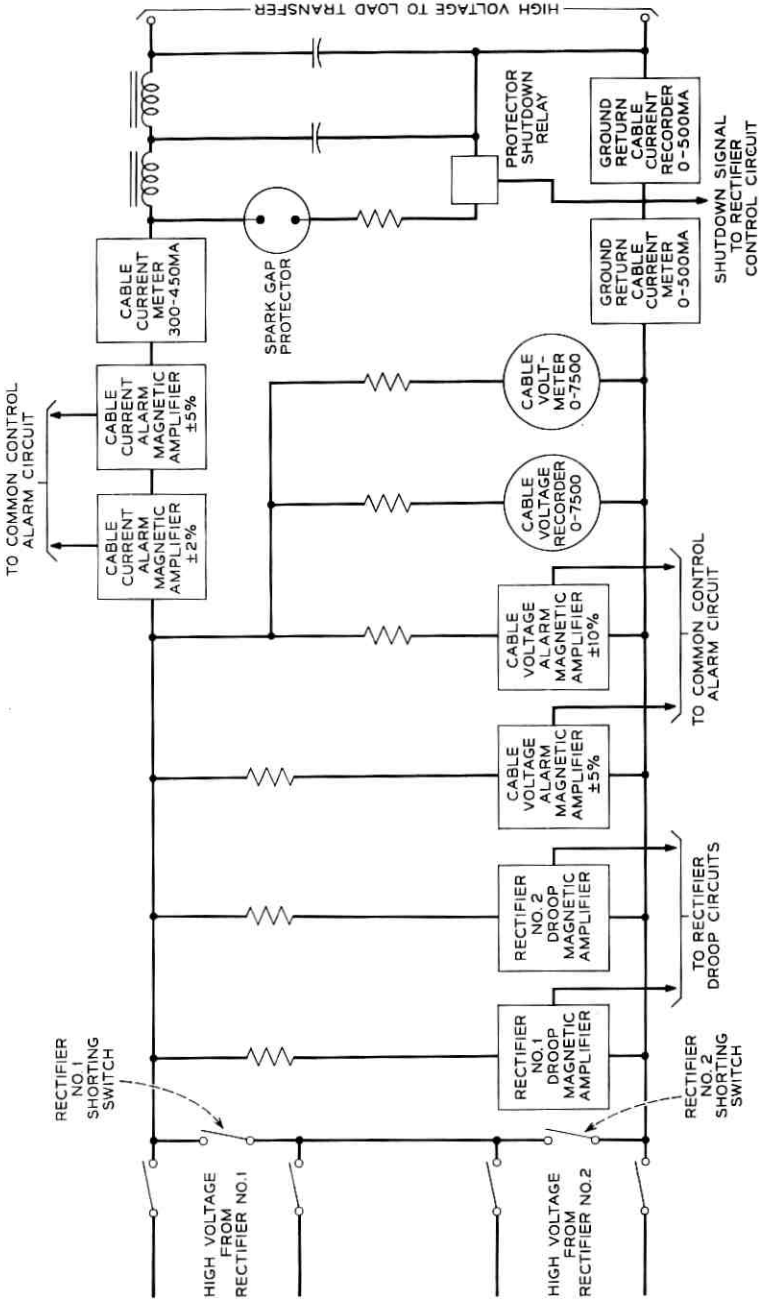


Fig. 11 — Common control functional circuit.

power produces immediate transmission of low-limit major alarms to the office alarm system.

An alarm test circuit similar to that used with the rectifiers is provided for in-service testing of all alarms. Fig. 11 illustrates the common control functional circuit.

6.3.1 *Surge Voltage Protection*

The series connection of the rectifiers results in an asymptotic open-circuit voltage greater than 15,000 v. Application of such a voltage to filter capacitors within both the power plant and the repeaters will severely shorten their life. Because of the constant-current nature of the power plant, such voltages are possible following an open-circuit cable break. The rate of rise of voltage can be as high as 120 v per msec depending on the location of the break. This rate of rise is too fast to control by direct overvoltage sensing and turndown circuits. Instead, a spark gap, designed for fast triggering and capable of carrying the full plant output current indefinitely, is used for surge voltage protection.

A current limiting resistance and current sensing relay is included in the spark gap circuit to limit the magnitude of the surge current and to produce shutdown of the power plant.

6.4 *Load Transfer*

The purpose of the load transfer is to permit transfer of power from the working power plant to the standby power plant without interruption or disturbance of the cable current. Fig. 12 illustrates the circuit used. The transfer operation consists of energizing the standby plant through the connected path (A) into the test load; then simultaneously adjusting the standby plant and test load until the current and voltage equal that of the working plant; then closing the presently open contacts (B) and opening the formerly closed contacts (A). It may be seen at this point that the former working plant is now feeding the test load and the standby plant is now feeding the cable.

Alarm transfer relays synchronized with the high-voltage contacts connect the office alarm to the working plant.

6.5 *Test Load*

The test load is designed to dissipate the full power output of either an individual rectifier or a complete plant. Load voltage is adjustable to permit operation at any voltage from 0 to 7500 v. This flexibility is

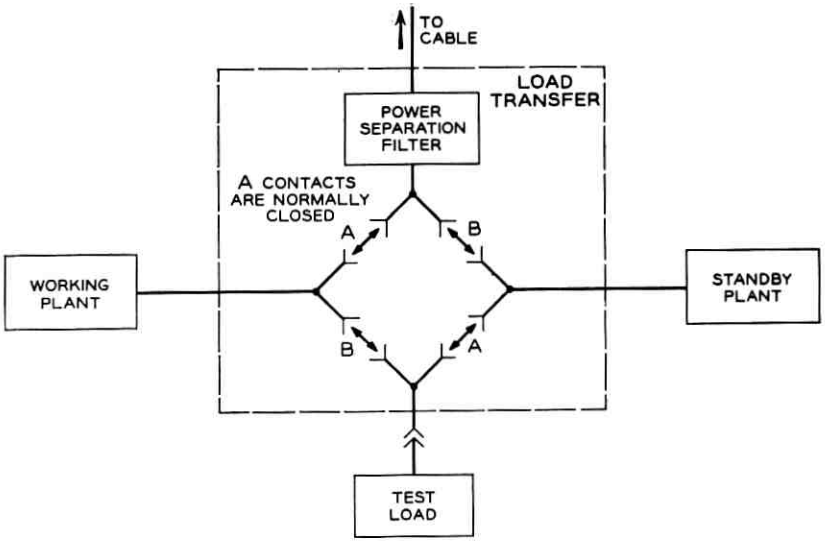


Fig. 12 — Load transfer circuit.

required to accommodate all possible lengths of cable systems. Meters accurate to 0.25 per cent are provided for current and voltage measurement.

6.5.1 *Ripple Measurement*

An attenuator is provided to permit rectifier or plant output ripple voltage to be measured at a safe potential with respect to ground.

6.5.2 *Precision Current Meter Calibrator*

Two sets of shunts and standard cells are provided to permit an accurate calibration of the power plant cable current meters. The shunts have a resistance which gives a voltage equal to the nominal standard cell voltage at 0.389 ampere of current. The voltage drop across the shunt is compared to the voltage of the standard cell by means of a sensitive galvanometer. Switches are provided to permit comparison of the voltage drop across each shunt to the voltage of each standard cell and permit cross comparison of calibrating circuit elements to reveal a drift or aging in any of these elements.

VII. DETAILED EQUIPMENT DESCRIPTION

7.1 *Appearance*

In the equipment development the aim was to design a distinctive, pleasant-appearing, and functional set of equipment. A distinctive blue color was selected for the exterior of the cabinets and the rear doors. Light gray panels are used for ease of legibility, readability and maintainability. Panels which are adjacent to one another have rounded edge contours to minimize the effects of assembly tolerances. An effort was made to provide an up-dated appearance for components such as meters recorders, etc. The number of cables, conduits and other appurtenances to the equipment bay tops is minimized in order to achieve a clean, uncluttered look.

7.2 *Voltage Isolation, Personnel Protection and Key Interlock System*

Hazardous voltages over 600 volts within the power supply are isolated so that access to components with these potentials is restricted when the equipment is energized. These voltages are isolated in locked units or in locked compartments. Fig. 13 demonstrates the isolation within the equipments. The high- and low-voltage sections within the cabinets are defined by vertical or horizontal partitions. The tops of the cabinets are enclosed with sheet-metal covers. The foregoing is generally applicable to all equipment cabinets except the inverter cabinet, which has no potentials above 600 volts and is therefore unlocked.

Operating procedures designed to minimize any interruptions to service or injury or damage to equipment were developed to allow access to high-voltage portions of the equipment only when the potentials had been removed. Each step in the procedure requires the operation of a key or number of keys wherein the key is released, captured or exchanged.

The key interlocks are used to control the operation of switches, the opening of doors, the operation of variable autotransformers, the operation of the patch panels, the operation of patching facilities for testing, etc.

7.3 *Maintenance*

A number of maintenance features have been included. The major feature is the provision of regular and alternate power supplies which are in either the working or standby condition. This permits testing and main-

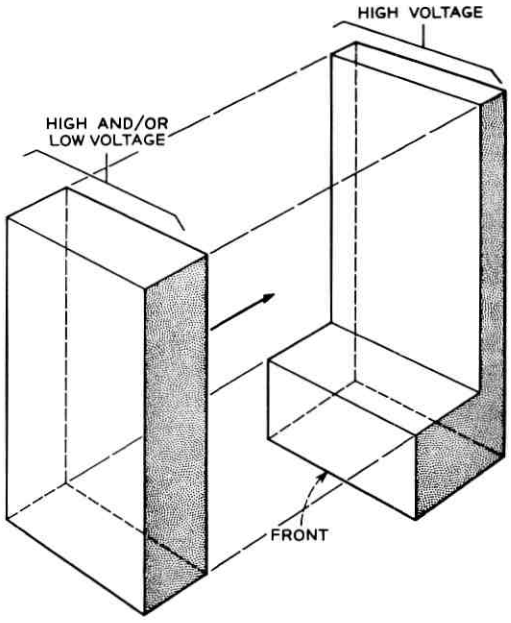


Fig. 13 — Voltage isolation in equipment.

taining the standby plant without concern, since there is no effect on the submarine cable system.

Recording ammeters and voltmeters monitor the output voltage and current of the regular and alternate power supplies. These recorders serve as an invaluable aid in maintenance procedures in the event of an unscheduled shutdown of the power supply. Alarm lamps and indicators have been furnished to assist trouble shooting or adjustment. Test jacks have been provided for in-service checks of low-potential circuits. A major maintenance tool is the test load.

Human engineering principles have been applied wherever possible in the location of controls, meters, switches, etc.

Equipment units on pull-out ball bearing slides are used for better packaging and also to improve maintenance access. Units with only low-voltage circuitry have an interior secondary control area for controls and apparatus required for maintenance adjustments. These are exposed when the slide unit is extended. A typical arrangement is shown in Fig. 4. Indicators of system performance are front panel mounted and are visible at all times. This approach eliminates the vast clutter of instrumentation controls, switches, etc. which might otherwise appear on the

front of the equipment panels, and has simplified maintenance and training.

7.4 *Corona and Dielectric Tests*

For system reliability the power plant must withstand corona and dielectric voltage test requirements. Portions of the equipment must pass 19,000-vac or vdc dielectric test requirements and 10,000-vdc corona tests. Workmanship items which include elimination of sharp projections, loose strands of wire and the achievement of smooth, round solder joints are important in passing these dielectric and corona requirements. Ceramic standoff insulators required baked silicone varnish surface treatments to pass corona requirements.

7.5 *Apparatus*

7.5.1 *Patch Panels*

Since the power plant includes redundant rectifiers and provision for regular and alternate supplies, it was necessary to incorporate a suitable disconnect and voltage isolation means whose mechanical operation could be visually verified. The term "patch panel" has been applied to this piece of equipment, which is seen in Fig. 3 in the lower part of the common control cabinet.

The patch panel is essentially a mechanically operated jumper cable. When a handwheel is turned clockwise, specially designed male and female molded high-voltage cable assemblies are connected. The patch panels are locked in the connected and in the disconnected position by means of key interlocks.

7.5.2 *Vacuum Relays*

Vacuum relays are used as shorting switches and control relays. Their operation is associated with the operation of the key interlock system. Vacuum relays were selected rather than mechanical switches because of their superior voltage breakdown characteristics and small size. Relays are normally used in pairs with paralleled contacts. The relays are designed to fail-safe in the event that the coil power is lost.

7.5.3 *Semiconductor Devices*

Semiconductor devices have been extensively employed in the equipment. No tubes, other than gas tubes which will be described later, have

been used. In the high-voltage portion of the rectifiers, encapsulated silicon diode sticks with capabilities far in excess of operating conditions have been specified. The diode-encapsulated sticks are mounted using ceramic standoffs to further improve their voltage capability to ground.

7.5.4 *Gas Tubes*

In order to further assure personnel safety and plant reliability, gas tubes have been used which are counterparts of the gas tubes used in the submarine cable repeaters. These gas tubes protect power supply high-voltage meters and recorders if their ground side opens. The tubes will fire and clamp the instruments to ground. Another gas tube, which can carry full cable current, is shunted around the current recorder and the direct-reading ground current return meter to protect against an open in the meter shunts.

7.6 *Hardening*

For protection of communications facilities, most submarine cable terminal stations are installed at hardened sites. In order to achieve a minimum usage of floor space for the power supply, a design with the equipment cabinets arranged in two rows, facing each other, was developed as shown in Fig. 14. The equipments are placed on a platform mounted on rubber shock mounts.

VIII. SPECIAL EQUIPMENT

8.1 *Ground Supply*

The ground supply bay is primarily designed to act as a ground and transmission termination for a cable system powered from one end. A power separation filter is included to permit insertion and removal of transmission circuits. The ground supply bay can be connected to a second ground supply bay to serve as the through connection at an intermediate point on the submarine cable system. It is also possible, on an emergency basis, to patch in a high-voltage power plant to the ground supply to power a part of a submarine cable system.

8.2 *Shipboard Power Plant*

During the laying and repair operation in the submarine cable system it is necessary to provide power to the cable from both land and ship or possibly from one source alone. For these situations a shipboard power

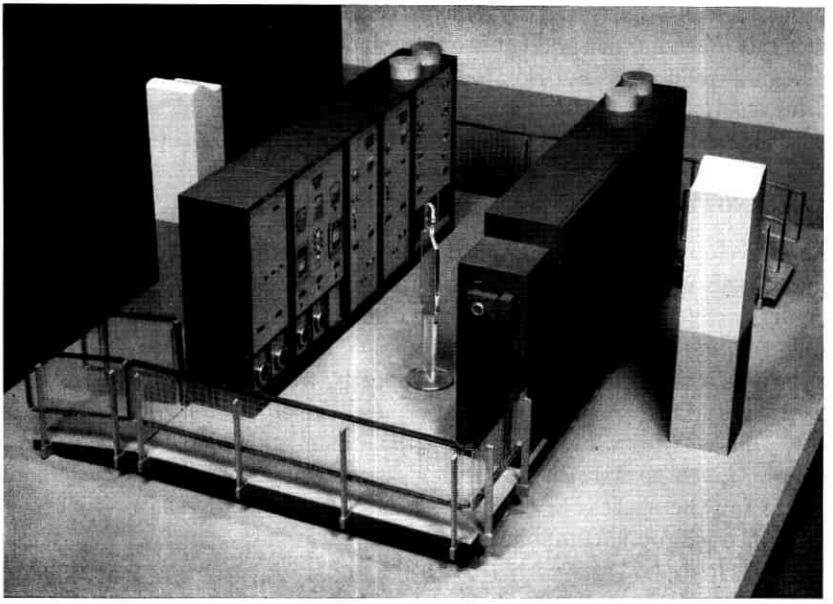


Fig. 14 — Hardened site installation.

supply has been designed and is installed on C. S. *Long Lines*. The shipboard power supply is available for cable laying operations and powering the cable stored on a ship. It is also possible during cable laying operations to power from the shore end and apply a ground to the cable aboard the ship. The shipboard power supply is designed to operate on 400-cycle ac power supplied by a pair of motor alternators. Its design includes several options permitting use with existing foreign or domestic submarine cable systems.

8.3 *Power Plant for Short Cable Systems*

A small power plant similar in design to one of the main power plants has been developed to supply operating power for distances less than 300 nm with double-end power feed or 150 nm with single-end power feed. The major differences are the use of only a single two-rectifier power plant, 1200 cycles ac power, and elimination of key interlocks.

IX. ELECTRICAL NOISE

Some difficulty has been found with electrical noise interference, both from sources within and external to the power plant equipment. This

problem was aggravated by the presence of three distinctly different grounds (ocean, office and cable outer conductor) which must be common at signal frequencies but isolated at dc. A satisfactory solution was found after on-site investigations and involved minor equipment wiring changes and the addition of capacitors at appropriate points in the circuit.

X. ACKNOWLEDGMENTS

A development of this complexity obviously required the work of many individuals in order to bring the project to successful completion. Special credit goes to Mr. B. H. Hamilton, who formulated the basic approach followed and supervised the circuit design; to Mr. V. B. Boros, who contributed to the concept and was responsible for a portion of the circuit design; to Mr. R. R. Gay, who supervised the equipment design; and to Mr. D. E. Trucksess, who managed the over-all project. Credit should also be given to Mr. A. D. Hasley's magnetic apparatus development group for the successful design of a large number of reliable high-performance magnetic components.

REFERENCES

1. Ehrbar, R. D., Fraser, J. M., Kelley, R. A., Morris, L. H., Mottram, E. T., and Rounds, P. W., The SD Submarine Cable System, B.S.T.J., this issue, p. 1155.
2. Storm, H. F., Series-Connected Saturable Reactor with Control Source of Comparatively High Impedance, Trans. AIEE, **69**, 1950, Pt. II, pp. 1299–1309.
3. Goldstein, H. L., Observation of Transients in the Series Connected Saturable Reactor with High Impedance Control Source, Proc. Nonlinear Magnetics and Magnetic Amplifier Conference, August 6–8, 1958, pp. 47–61.

A Cable Laying Facility

By R. D. EHRBAR

(Manuscript received March 2, 1964)

In order fully to utilize new broadband submarine cable systems it has been necessary to consider new techniques for placing the system on the ocean bottom. This article describes the general development plan that resulted in a new cable ship capable of efficiently handling any new system. The article thus serves as an introduction to the remaining articles of this special issue.

I. INTRODUCTION

The need for new and more complex submarine cable systems to handle a growing transoceanic communications business produced many new features in the design of cable and repeaters and auxiliary electronic equipment.[1,2,3] It also made necessary new solutions to the problem of placing the cable system on the ocean floor. This task appeared formidable to the early telegraph cable engineers, and many failures preceded the final development of a satisfactory mode of operation. In terms of modern systems their task was simple, since they were interested only in a single-conductor circuit that would remain a low-resistance path well insulated from the sea.

Modern broadband cable systems present a large number of new problems. The cable is now a complex transmission line whose characteristics up to high frequencies must be predictable and stable after the laying process. Repeaters and equalizers requiring large rigid containers are now inserted in the cable at frequent intervals. Both cable and repeaters are expensive and must not be wasted by the lack of precise payout control.

When development of a repeatered telephone system[4] was initiated by the Bell System, it was recognized that the laying of such a system in deep water was a sizable problem. Consequently, effort was concentrated on a flexible repeater that could be handled like cable with only minor modifications to existing cable ships.

At the same time it was realized that this restriction would unduly

constrain future developments of broadband systems, and a program was started in 1953 to develop more versatile and more precise techniques and machinery for cable laying that could be integrated into a working cable ship. The objective was a modern cable laying facility that would permit the efficient installation of any type of cable system that might be developed in succeeding years. The many parts of the facility are described in companion articles.[5,6,7] These were developed as an integrated system under a common development plan described below.

As a first step, analytical and simple model work was started on the fundamental mechanics of cable laying and recovery. Some of the questions to be answered were:

(1) What is the configuration of a cable on its way to the sea bottom or during recovery? What are the forces involved?

(2) What is the nature of the ocean bottom environment?

(3) How must the cable be controlled for proper distribution on the bottom?

(4) What are the effects of wind, currents, and wave motion on the laying process?

Many of these questions have not yet been completely answered and are the subject of continuing exploration and study. However, a sufficient background of fundamental knowledge was assembled to guide the development of cable laying techniques and equipment. Some of this work has been reported.[8] Additional information will be published as studies of general interest are completed.

II. OBJECTIVES AND REQUIREMENTS

The objectives of a perfect cable laying technique can be stated quite simply:

1. The cable should just follow the bottom contour — laying wasteful excess cable must be avoided without increasing the chance of failure during or after laying. Cables suspended above the bottom in areas of varying depth are subject to damage by chafe, abrasion, and man-made hazards.

2. In the trip from shipboard to its place on the bottom no part of the cable or repeaters should be subjected to excessive mechanical deformation or shock. This will insure that the reliability and transmission stability of the system are not affected by the laying process.

With these objectives in mind, analytical work continued. Basic ideas were studied with table-top models and $\frac{1}{4}$-scale working models. Finally the better ideas were converted to full-scale models. At the same

time naval architects were brought in as consultants to consider the proposals in the terms of a marine environment and to incorporate them in preliminary arrangements of a new cable ship. Other persons engaged in cable laying were briefed on the new ideas and their comments solicited to take advantage of prior experience.

From these studies came four fundamental requirements to guide the final development:

(1) The cable laying process should be continuous at speeds of 4–8 knots to avoid unpredictable transients and ship control difficulties.

(2) The cable, when under high tension, should be in a straight line and should be handled by machinery that would treat it as gently as possible to eliminate transmission variations and possible damage.

(3) Personnel requirements should be kept to a minimum so that long systems with close-spaced repeaters would not require oversized, highly trained ship crews.

(4) Methods should be safe and simple, avoiding the need for close timing or coordination, since a ship at sea can be a very unstable platform.

III. NEW DEVELOPMENTS

The exploratory work suggested that the cable ship be considered in terms of two independent functions: (1) high-speed continuous laying of rigid repeater systems, and (2) repair of cable systems of any type or slow-speed laying of a few repeaters. The first function was assigned to the stern of the ship and the second function to the bow.

For stern laying three new developments were needed:

(1) a new cable engine which would hold cable without damage and simultaneously accommodate rigid repeaters without change in speed,

(2) a method to control the cable and repeater during overboarding, and

(3) a system for handling and stowing cable and repeaters in the ship that would permit continuous and essentially unattended payout.

For the functions at the bow, conventional cable engines using large-diameter drums were adaptable. Many new features were desirable, but of most importance was controlled movement of the cable line in the horizontal plane to permit repeaters to be handled without overriding cable on the drum.

Development of methods and equipment to meet these specific needs was undertaken in early 1957. These did not by themselves make a cable handling facility, but had to be integrated into the design of the cable ship. In general, they required large spaces and involved heavy

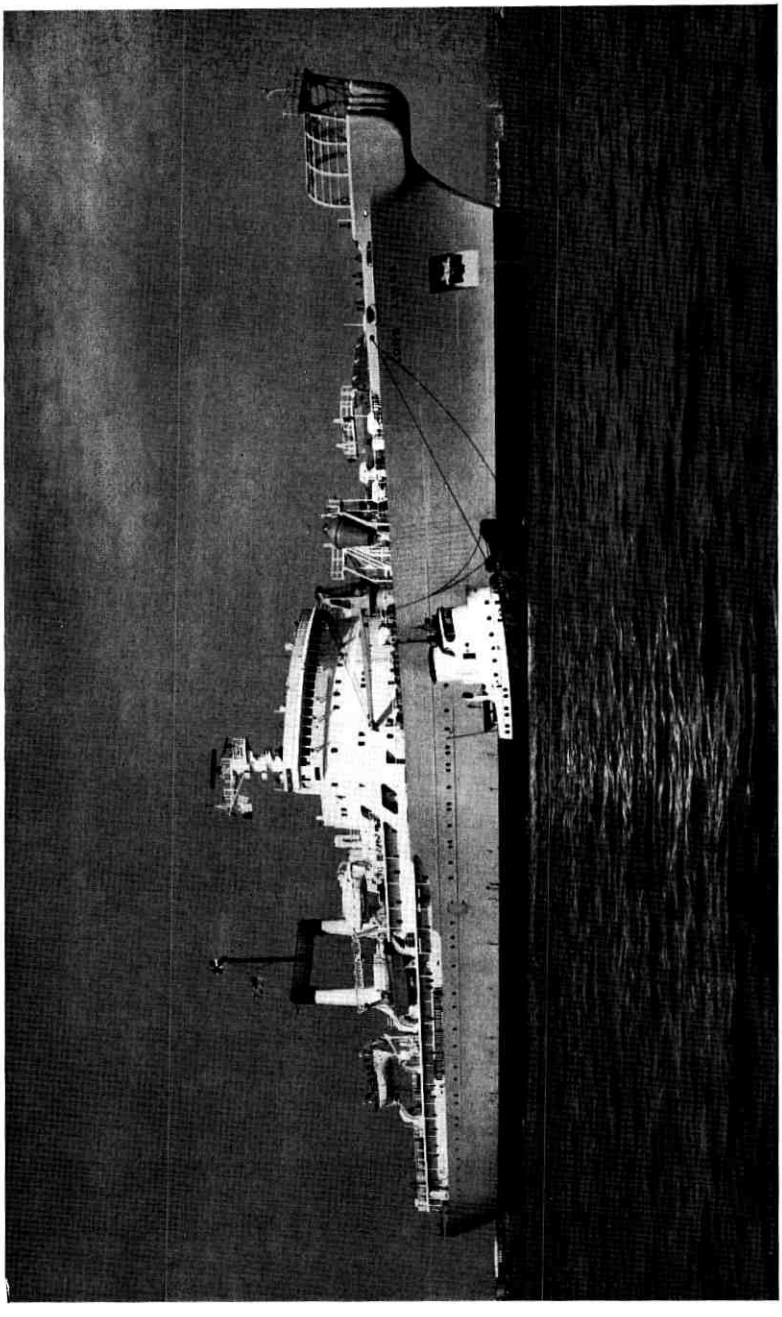


Fig. 1 – C. S. *Long Lines*.

machinery, so provision had to be made in the basic structure and in the spatial arrangements of the ship. It proved to be essential that the development of the laying system and the design of the ship be coordinated under an integrated system concept.

Then the ship with its cable gear had to be put to work in a dynamic sense. This involved operation of the cable gear and the ship by personnel in accordance with requirements dictated by the cable system. To accomplish this most effectively, it was necessary to consider navigation and ship control, control of the laying process, transmission testing and adjustment, cable splicing and loading, communications and many other factors to insure that the cable and repeaters could be placed safely and precisely along the desired route.

IV. CABLE SHIP *LONG LINES*

The new Bell System cable ship, *Long Lines*, shown in Fig. 1, represents the first example of cable ship design following the basic philosophy of an integrated system design. It was not sufficient that the ship be able to lay the newest system, but was necessary also that its equipment be flexible enough to lay other systems which might follow in the natural evolution of undersea transmission systems. The papers that follow in this special issue describe some of the more important developments and present the story of this new ship. Throughout the descriptions the need for coordination of all the parts should become apparent.

This new cable handling facility has proved, in its first important test of laying a new transatlantic cable, that the thorough development program was successful. Cable and repeaters were payed out continuously at 8 knots with a ship's complement about 25 per cent less than that used on other cable laying ships of comparable size. A facility is available that can be adapted to handle any transoceanic system that might be considered in the next decade. The basic design is such that new features can be added as our knowledge increases and new techniques are developed.

Many organizations and individuals were responsible for the success of this undertaking. During the development the following organizations outside the Bell System were important members of the development team:

Bergen Research and Engineering Company assisted in much of the quarter-scale model design, construction and testing.

Western Gear Corporation was responsible for final design and manufacture of the cable engines.

Gibbs & Cox, Inc. provided the necessary marine background and contributed to many of the design concepts.

The final design and construction of the cable ship was the responsibility of the Long Lines Department of the American Telephone and Telegraph Company. Gibbs & Cox, Inc. acted as its design agent in this undertaking. A contract for construction of the ship was signed with Schlieker Werft, Hamburg, Germany on October 27, 1960. The launching occurred on September 24, 1961, and the ship arrived in the United States on April 13, 1963. It sailed for its first cable laying job on July 23, 1963.

The ship is owned by Transoceanic Cable Ship Company, a wholly owned subsidiary of A.T.&T. Co. It is operated by Isthmian Lines, who handle the cable laying and repair work with the technical assistance of Bell System personnel.

REFERENCES

1. Ehrbar, R. D., Fraser, J. M., Kelley, R. A., Morris, L. H., Mottram, E. T., and Rounds, P. W., The SD Submarine Cable System, B.S.T.J., this issue, p. 1155.
2. Bowker, M. W., Nutt, W. G., and Riley, R. M., Design of Armorless Ocean Cable, B.S.T.J., this issue, p. 1185.
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., this issue, p. 1243.
4. Special transatlantic cable issue, B.S.T.J., **36**, January, 1957.
5. Grismore, O. D., Cable and Repeater Handling System, B.S.T.J., this issue, p. 1373.
6. Gretter, R. W., Cable Payout System, B.S.T.J., this issue, p. 1395.
7. Butler, J. H., Altenburg, C. J., McSweeney, R. J., and Sutton, L. E., Design and Powering of Cable Ship *Long Lines*, B.S.T.J., this issue, p. 1435.
8. Zajac, E. E., Dynamics and Kinematics of the Laying and Recovery of Submarine Cable, B.S.T.J., **36**, Sept., 1957, p. 1129.

Cable and Repeater Handling System

By O. D. GRISMORE

(Manuscript received March 24, 1964)

This paper describes the development of methods and equipment for handling cable and repeaters in laying submarine cable systems and the application of these developments to C.S. "Long Lines." Both the laying process and the associated operation of assembling the system components in preparation for laying are described.

I. INTRODUCTION

The underwater portion of a modern submarine cable system is composed of three basic components; cable, repeaters and equalizers. These are delivered to dockside as individual items, and the handling of these units from this point to the time when they leave the ship on the way to the ocean bottom is the work of the cable and repeater handling system. The planning of arrangements and procedures, and the design of equipment and details to implement the following operations are included in the handling system:

(1) movement from dockside to appropriate stowage positions aboard ship,

(2) connection of components into a transmission system,

(3) cable test during loading,

(4) system test during laying,

(5) preparation of repeaters and equalizers for launching,

(6) control of cable and repeater movement during laying, and

(7) cable jointing for laying and repair.

For the performance of these operations, the following requirements were set for equipment and arrangements:

(1) cable to be handled in a fashion to avoid damage from bending, twist, abrasion and tension;

(2) repeaters stowed and handled to avoid excessive shock and extremes of temperature;

(3) stowage of cable simplified to avoid possibility of cable fouling

and permit easy checking of clearances and provide flexibility in sequence of tanking;

(4) cable and repeaters to be payed out in a steady-state continuous operation at speeds to 8 knots;

(5) minimum manpower requirements during the laying operation.

II. GENERAL SCHEME OF OPERATION

From a knowledge of the route over which the cable is to be laid, a loading plan is prepared which assigns cable sections, repeaters and equalizers to particular places in the system and particular positions on the ship. The components are then brought aboard ship in the proper order for stowage to permit laying in the correct sequence.

The cable and repeaters, which comprise a shipload, are connected together aboard ship to form a single complete circuit. This may consist of as much as 2000 nautical miles (nm) of cable, 100 repeaters and 10 equalizers if the ship is loaded to design capacity. The circuit thus formed consists of a series of "ocean blocks" 192 nm in length, each containing 10 repeaters. The ocean blocks are connected together through ocean-block equalizers which are adjusted aboard ship just prior to overboarding.

After the shipload of cable and repeaters is assembled into a complete circuit, the ends are connected to the shipboard power supply through power separation filters and the circuit is powered from end to end. Transmission measurements are then made on one block at a time from one equalizer to the next adjacent. Through measurements from end to end of the system are not possible when the cable is aboard ship, because the temperature of the cable in the ship's tanks is much higher than it is at sea bottom, and the excess loss of a complete shipload due to temperature is in the order of 200 db.

Following the shipboard transmission check and upon arrival at the cable ground, the outboard end of the cable is connected to the shore station or to the end of cable previously laid. As soon as the connection is complete, power is turned up on the system, ship-to-shore communication is established and transmission measurements are made. Cable laying is then started, and from this time until the bitter end of the cable is reached there is no interruption in power or signal transmission.

The amount of cable payed out is slightly in excess of the distance traveled over the ground, in order that there be sufficient cable length to fill bottom irregularities, the excess length being known as "slack." To measure the distance over the ground a continuous length of piano wire, called the "taut wire", is payed out. This wire is anchored to the

ocean bottom and is payed out under constant tension. The speed of taut wire payout is measured and the cable payout speed adjusted to provide the slack appropriate for the bottom conditions at the point of cable touchdown. Bottom contours are measured by echo sounders, the results of the measurements being plotted on a precision depth recorder. In laying cable over a flat ocean bed or over constant, moderate slopes of less than about 4°, the cable and repeaters are payed out at top speed of 8 knots, weather and sea conditions permitting. Over rough bottom with steep slopes, ship's speed is altered as necessary to get proper coverage of the bottom.[1]

As cable is payed out, the next repeater or equalizer which is to be launched is moved from a stowage position to a launch position, where it is free to move in line with the cable along the working deck, through the cable engine and into the water. At this time a parachute is attached to the repeater body. When in the water, the parachute opens to slow the repeater descent to more nearly match the sinking rate of the cable.

As soon as a repeater is launched the next repeater is prepared for overboarding. The preparation for launching takes about 10 minutes and the time interval between repeaters or between repeaters and equalizers varies from about $\frac{3}{4}$ hour to 5 hours, depending upon cable laying speed.

While laying the system, transmission measurements are being made and communications are being carried on over the cable between ship and shore. The transmission and equalizer adjustment connections to the cable are made at the equalizer, and measurements are made between shore and the next equalizer to be launched. Approximately 3 hours before the equalizer launching a decision is made as to the optimum setting of the equalizer. It is then adjusted by means of the stepping switch. After rechecking the transmission characteristic, the transmission and stepping lead connections are removed, the lead ends are sealed and the equalizer is prepared for launching.

The process of cable laying and repeater and equalizer launching continues section-by-section and block-by-block until the end of the load is reached. The cable end is then joined to the shore end cable or is streamed on the bottom to be picked up later after the ship returns with another load of cable.

III. CABLE LAYING SYSTEM

The most significant single contribution to improved cable laying was the design and development of the linear cable engine,[2] which made it possible to pay out cable and large, heavy, rigid repeater casings with

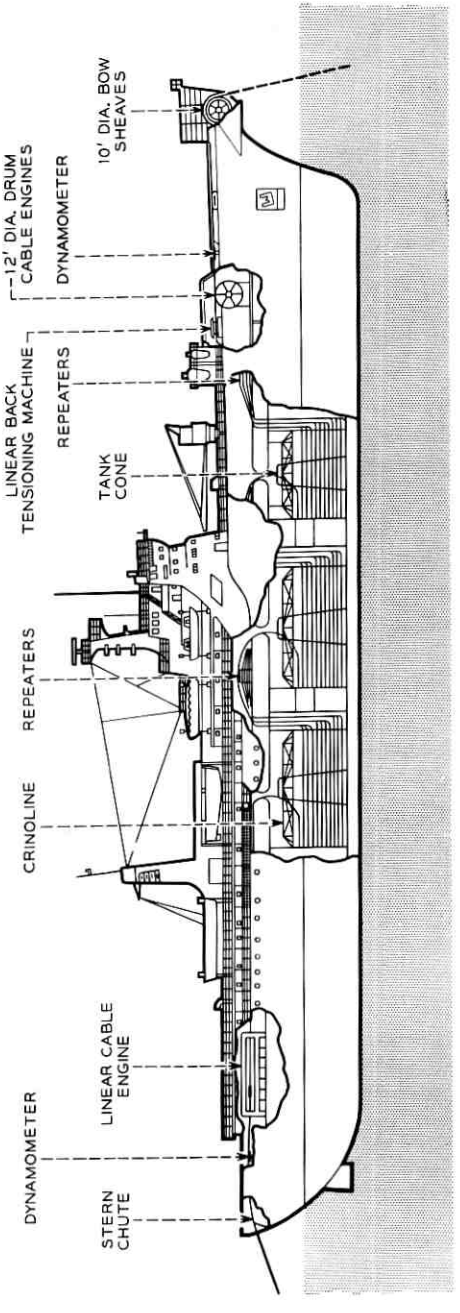


Fig. 1 — General layout of the cable laying system.

equal ease. The cable handling system exploits to the fullest extent the capabilities of the linear cable engine and provides for continuous substantially automatic, payout of cable and repeater.

The general layout of the cable laying system is shown in Fig. 1 and an individual tank arrangement is shown in Fig. 2. Cable is loaded into the cylindrical tanks in a reverse sequence from which it is to be laid. The cable is coiled as shown, in layers called "flakes," around the central core of the tank, which has the form of a truncated cone. There are three main cable tanks along the center line of the ship which hold the cable to be laid. In general, all of the cable in one tank is payed out before transferring to another; the preferred order is tanks 3–2–1. The arrangements are completely flexible, however, and practically any order can be used at a sacrifice in simplicity of stowage.

Cable is payed out from the center of the top of the tank over the smooth faired surface to the cable working deck. It moves aft along this deck to enter the repeater trough and passes through the trough to the cable engine. On passing through the cable engine it reenters the trough, continues over the dynamometer, through the stern overboarding chute and into the water. The guiding principles have been that the moving cable and repeaters travel over faired surfaces, smooth deck areas or through open troughs. Closed guides, rollers and gates are avoided, and it is unnecessary for personnel to handle or guide the moving cable or repeaters at any point in the payout process. Once payout is started the only action required by personnel is to move the repeaters from the stowed to launch position and to dress the following cable on the deck forward of the repeater in the launch position.

In Fig. 2, tank 3 is shown with a repeater in the launch position and the repeater stowage area filled with a complete complement of repeaters. Payout of the cable is aft, and a bight of cable is shown rising through the deck slot just before reaching the working deck. As the bight of cable reaches the working deck the bight is straightened out, the cable is pulled taut, and the repeater is accelerated to payout speed.

IV. CABLE TANKS

Of the seven cable tanks in the ship only the three main tanks which carry the cable to be laid are of particular interest. The other four are small tanks which hold short lengths of spare and repair cable; cable is not payed out directly from these.

The three main cable tanks are similar, differing only in diameter. Tank 1 forward is 42.5 feet in diameter, with a capacity of 31,000 cu. ft., while tanks 2 and 3 are 55 feet in diameter and have a capacity of

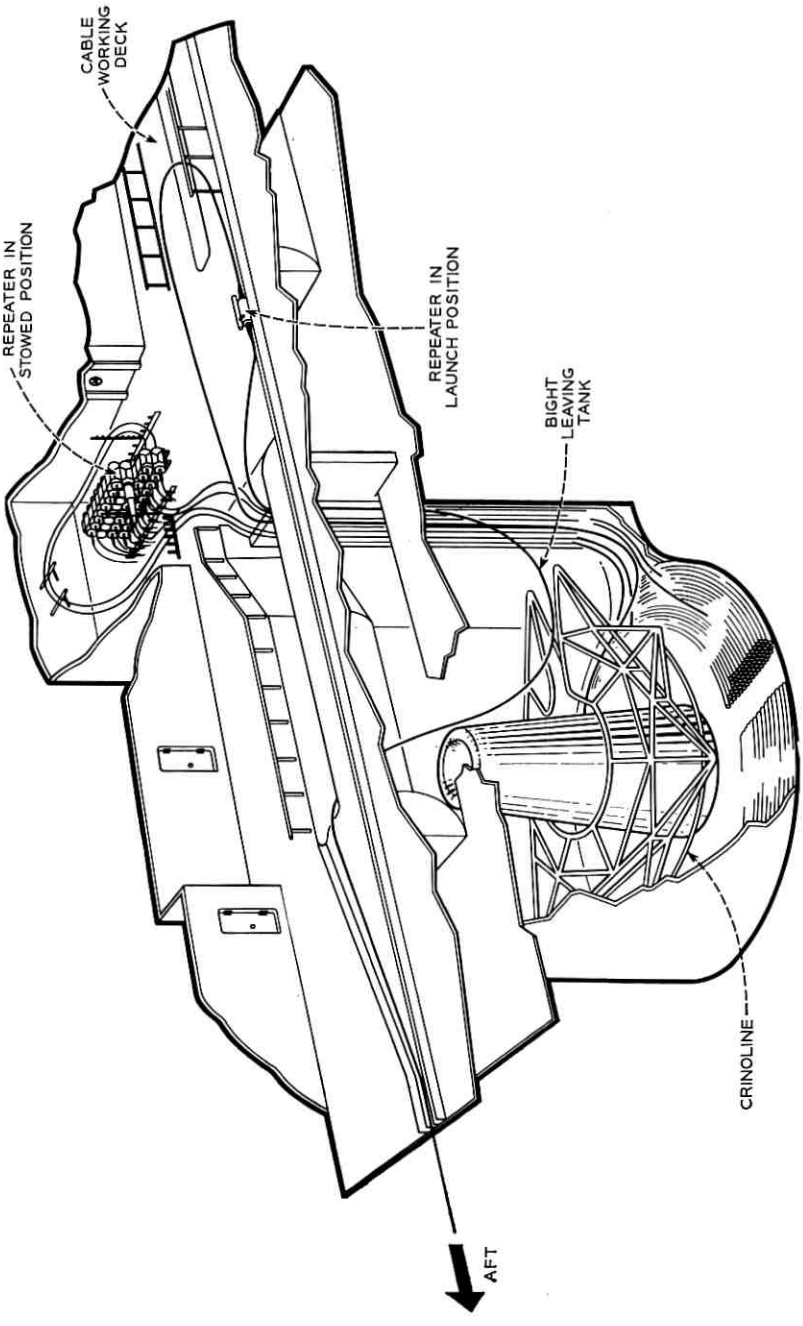


Fig. 2 — Cable and repeater stowage.

54,000 cu. ft. each. The central cones are 15 feet in diameter at the base and taper to 10 feet at the top. The coiling height of 24 feet is the same for all tanks. The total tank capacity is sufficient to contain 2000 nm of $1\frac{1}{4}$-inch diameter deep-sea armorless cable.

The cable tanks differ from previous designs in several respects, all of them pointed toward automating the cable payout. The usual tank opening, consisting of a relatively small round or rectangular central hatch, has been replaced by the long narrow slot shown, extending from the center of the tank forward to the tank wall. This slot permits the cable bights to rise freely from the tanks without danger of twisting or kinking. A nylon net is stretched taut across the entire slot opening a few inches above the deck surface, since it was recognized that an unprotected open slot would be a hazard to personnel. Cable and repeaters move freely under the net during payout. The net also serves to restrain the rising bight of cable as it is pulled out of the tank and into the trough.

The forward wall of the tank has been opened to provide a full-length vertical slot to contain the cables running from the tank to the repeater stowage area. As many as 88 cables are stowed in holders or restrainers mounted in these slots. The slots also contain small elevators or man hoists used for raising and lowering personnel and materials into the tank and as working platforms for personnel placing cables in the cable holders.

The cable holders mounted on the forward wall of the vertical slot are shown in Fig. 3. These are a series of narrow slots deep enough to hold four separate cables and wide enough to accommodate single armored cable (1.88-inch). After the cables are stowed in the holders, the slots are closed with flat strips of rubber which prevent the cable from falling out and control the rate at which the cable leaves the slot. This is particularly important to the cable entering the tank, since if too loosely held it may fall freely and form a kink, and if too heavily restrained it may be bent at too small a radius as it is pulled free. The choice of material as well as its thickness and amount of closure were all critical in a single design to control the several different cables under varying temperature conditions.

The crinoline, seen as the slotted circular pipe framework within the tank in Fig. 2, represents an important development in the automation of bight payout. The purpose of the crinoline is to restrain and control the cable as it moves around the tank during payout at high speeds. The open slot extending from the center ring to the tank wall prevents the cable from running out into the slot during normal payout but permits

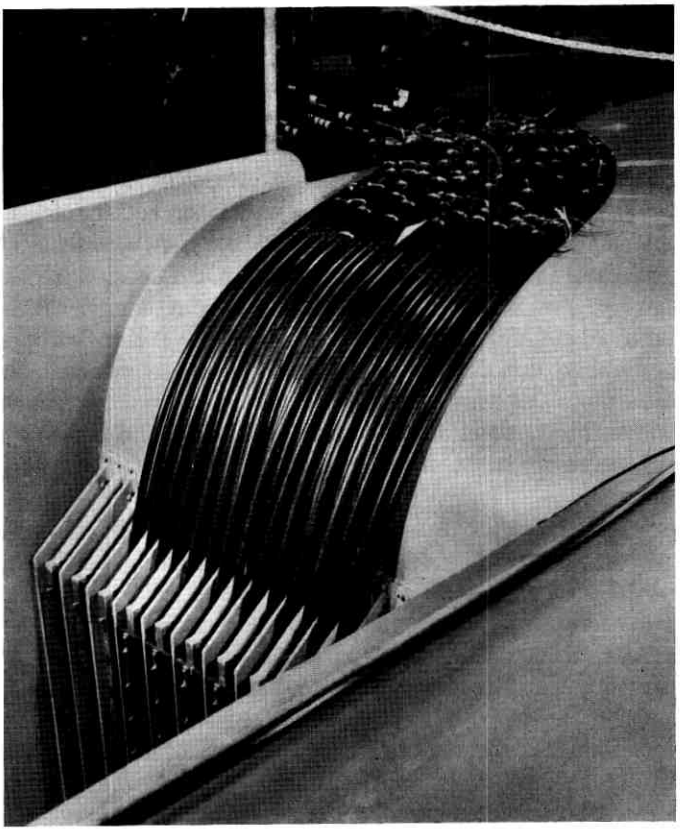


Fig. 3 — Cable holders on forward wall of vertical slot.

the bight to rise freely through the slot to the cable working deck at cable speeds as high as 8 knots. Although the crinoline working height above the top flake is not critical, a height of 6–7 feet gives the best results. Two auxiliary rings, having a diameter (17 feet) equal to the inner ring of the main crinoline, are provided above the main crinoline so that unrestrained cable lengths need not exceed 7–8 feet. The main crinoline and auxiliary rings are positioned at appropriate heights above the top cable flake by means of electrically operated hoists. The hoists are controlled from crinoline control stations at the main deck level. These stations are manned continuously during payout by personnel who can observe the cable action in the tank, adjust the crinoline height as necessary, and report any observed irregularities during the procedure.

V. REPEATER STOWAGE

The repeater stowage areas are located on the port side, forward of the tanks in which the cable is stowed. This position was chosen because it simplified the cable arrangement and the repeater handling process. The repeaters are clamped in individual rack frameworks shown in Fig. 4. These racks, which are demountable, are then bolted one to another and to a deck foundation to form a stack with as many as 44 repeaters arranged 11 wide and 4 high. Since these areas are not air-conditioned, forced-air ventilation for repeater cooling is provided through ducts opening under the repeater stacks. A nearly complete stack of repeaters is shown in Fig. 5.

At the outboard end of the repeater rack foundation a vertical frame is placed to serve as a lateral support for the repeater stack. This also serves as a mounting for power separation filters which terminate the cable and for terminations of transmission and test leads from the transmission testroom. Flexible leads connect the equalizers to measure transmission and to step the equalizer switch to adjust loss during the equalization process.

A station for auxiliary services is located in each of the repeater stowage areas. This provides electrical power, compressed air and fresh water for equipment used in the cable-to-repeater assembly procedure.

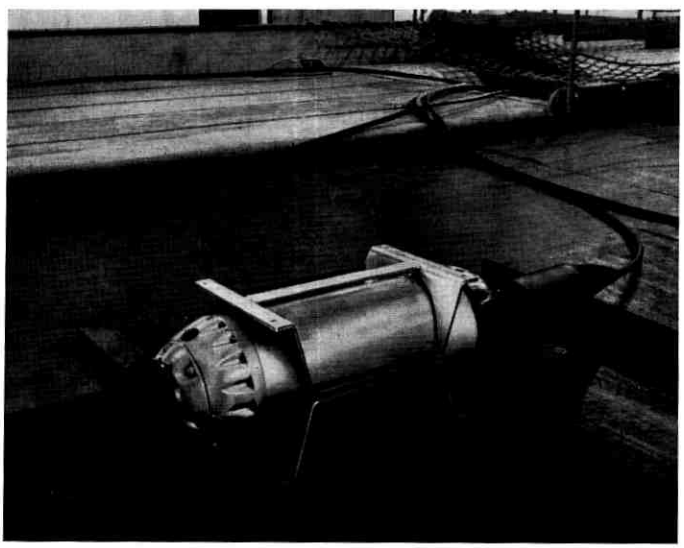


Fig. 4 — Repeater clamped in individual rack framework.

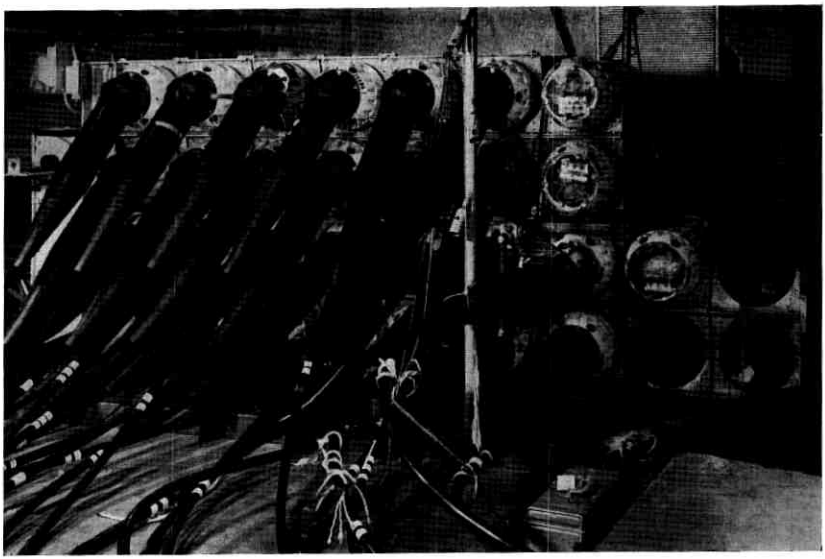


Fig. 5 — Repeater stack.

Repeaters are moved into stowage position and from the stowage area to the launch position by means of an electrically operated hoist. This hoist is driven for both rise and traverse and can reach any of the working areas at the upper deck level.

Because of the large numbers of cables which must be stowed in a repeater bay and because of the need to prevent bending below a minimum radius of 3 feet, the cable runs were carefully worked out, and supports and guides were devised to assure that the proper paths would be followed and that the cable would not be subjected to damaging conditions.

VI. CABLE WORKING AREA, REPEATER TROUGH AND OVERBOARDING CHUTE

Fig. 6 shows the cable working deck looking aft from forward of tank 2 with a central area called the "highway" raised approximately 6 inches above the deck surface. The cable during payout moves across this highway surface until it enters the repeater trough section shown to the left of the highway in the illustration. This deck has been built without sheer in order that cable lines will lie flat on the deck.

The repeater trough is rectangular in cross-section over most of its length but becomes "vee" shaped a short distance forward of the cable engine and between the cable engine and the stern chute, where it

Fig. 6 — Cable working deck looking aft from forward of cable tank 2.

again becomes rectangular. The transition from rectangle to vee forward of the cable engine is designed to rotate the parachute attached to the repeater through an angle of 25° so that it is outside the gripping surfaces as the repeater moves through the engine. The vee shape continues aft of the engine past the dynamometer (see Fig. 1), where the

transition from vee to rectangular shape rotates the parachute back to its original orientation to clear the side of the stern chute.

Since the sensitivity of the dynamometer is a function of the cable angle at the dynamometer bearing surface, it is essential that this angle be constant. This is accomplished by providing a cable hold-down device in the stern chute which restrains the cable and prevents it rising off the trough bottom when the ship pitches. This hold-down is raised automatically before the repeater reaches this position.

When cable moves across broad, flat deck areas and over large-radius faired surfaces at low tension, the stiffness of the cable is sufficient to cause it to move laterally, and surface wear at these surfaces is low although no lubrication is provided. This situation exists on all surfaces over which the cable passes forward of the cable engine.

In the cable trough aft of the cable engine the cable is in contact with the trough bottom at one point and with the dynamometer at another. At these points the cable is under tension (600–7000 lbs.) and is restrained from lateral motion by the trough shape. Tests showed that under these conditions a $\frac{1}{4}$-inch thick plate of mild steel would have a life well over 10,000 nm. To further increase the life and simplify replacement, hardened steel surfaces in the form of bolted inserts are installed at these points, and cooling is provided by water flooding over the wear surfaces.

The cable and repeaters leave the ship through a stern chute instead of over a sheave as is common practice on most cable ships. The chute is advantageous in that the large radius needed (>3.5 feet) can be obtained in much less space and at considerably lower cost than by the use of sheaves. Several proposals were considered before the final design, shown in Fig. 7, was reached in which the ship's hull was shaped to form the sides and bottom of the chute. The structural and shaping problems were not difficult, since the intersecting surfaces are all cylindrical sections with no compound curves required. Wear in the chute after 3000 nm has been so small that it has been difficult to detect.

VII. PARACHUTE DESIGN AND OPERATION

There are two reasons for controlling the rate of descent of the repeater. The first of these has to do with the laying of bottom slack and the second concerns the protection of the repeater from excessive shock on bottoming.

Under steady-state laying conditions the configuration of the cable between the ship and ocean bottom is a straight line; it is desirable to

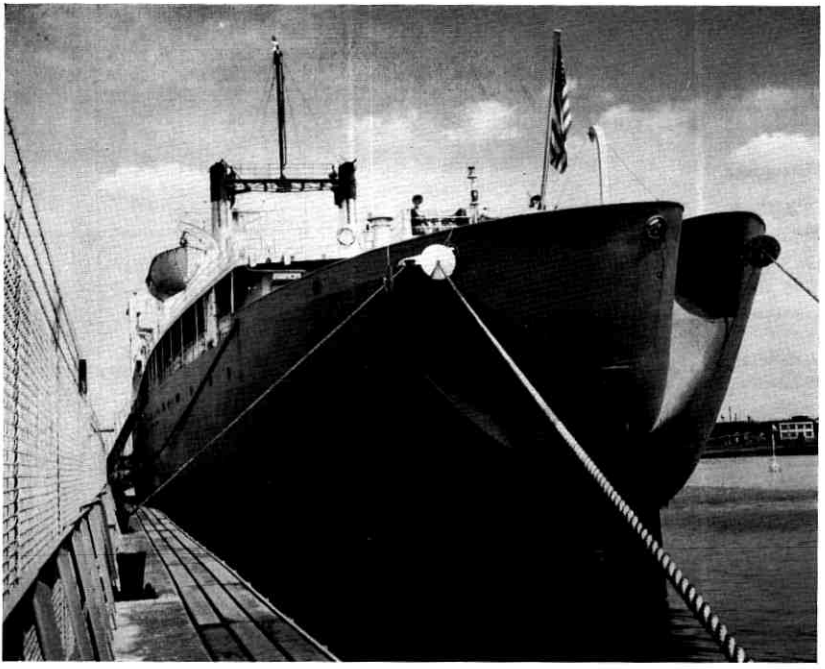


Fig. 7 — View of stern chute of C.S. *Long Lines*.

maintain this configuration to properly control the distribution of slack. The presence of the large mass of the repeater in the suspended cable destroys this condition and adversely affects slack control since the repeater sinking rate is several times that of the cable (8 k vs $\frac{3}{4}$ k).

If the cable were laid in still water the problem would be fairly simple, and substantially perfect match could be obtained between cable and repeater sinking rates. This condition never exists, however, because of the presence of ocean currents of varying and unpredictable velocities which affect the parachute drag. A compromise size was chosen which will not present excessive drag but limits the repeater bottoming speed to a safe magnitude of 2 knots.

The nylon parachute is of simple circular design, seven feet in diameter. The complete parachute mechanism consists of four parts: parachute, container, flotation bag and harness. Fig. 8 shows a repeater ready for launching with a parachute attached.

The parachute is packed in a zipper-closed pocket at the end of a cylindrical air-filled flotation bag. This flotation bag is placed in a fabric

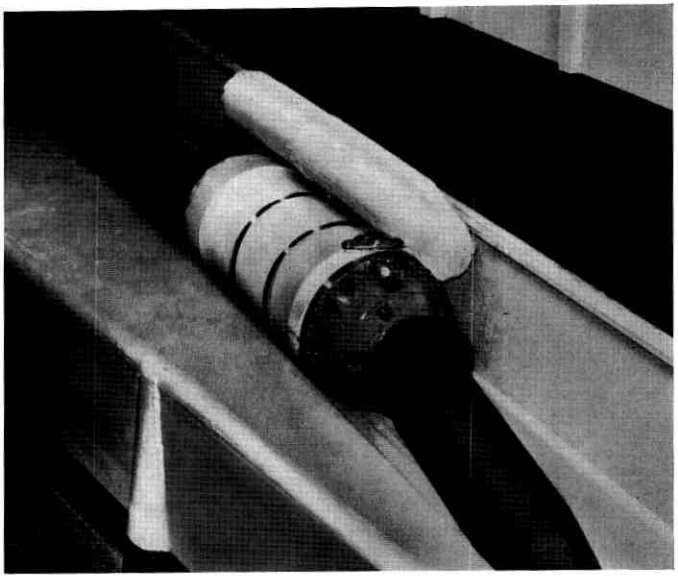


Fig. 8 — Repeater ready for launching with attached parachute.

container which is attached to the harness. The harness serves the dual purpose of fastening the parachute to the repeater and, in connection with the trough shape, positions the parachute properly during payout.

The operation is entirely automatic. When the repeater enters the water the flotation bag pops out and floats to the surface, releasing the parachute to deploy in the water. A corrosive link between the parachute shrouds and the repeater releases the parachute after about 24 hours of exposure to sea water. Several hundred parachutes have been used thus far and there has been no failure in operation.

VIII. FORE DECK

The fore deck at the boat deck level (shown in Fig. 9) is the location for all repair and recovery and buoy handling operations. It is also expected that a small amount of cable laying, particularly of shore ends, will be done over the bow.

The three large-diameter bow sheaves are housed in the faired surfaces shown in Fig. 10. The three sheaves are mounted coaxially, the outer two being grooved and the center one flat surfaced. Cable, rope and chain are ordinarily carried over the outer sheaves, and heavy, bulky gear such as cable mooring anchors and grapnels is handled over the

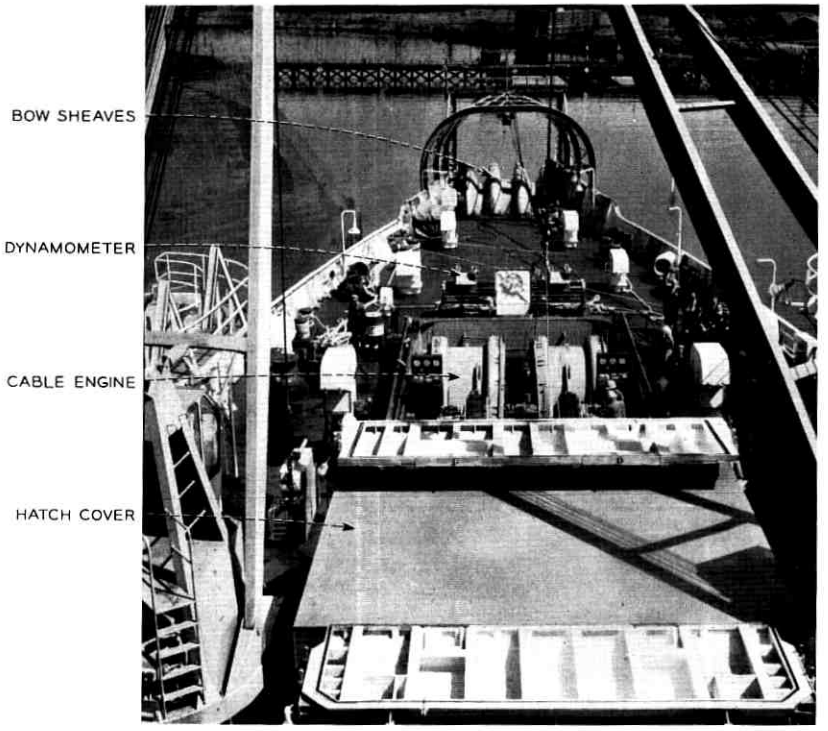


Fig. 9 — Foredeck layout of C.S. *Long Lines*.

center sheave. Two cable handling sheaves are required because during repair or splicing operations two cables are handled simultaneously. The gantry frame shown provides support for an electrically driven hoist used for moving heavy grapnels and mooring anchors over the sheaves.

A ship control station is located just aft of the bow sheaves, and the ship's operation can be handled directly from this position. Cable payout information is also repeated at the foredeck control. This includes cable tension and amount of cable picked up or payed out through the bow engines.

There are two large rectangular hatches located well forward on this deck. The first of these provides access to the two drum-type cable engines located below this hatch at the upper deck level. The after hatch opening gives access to the upper deck, and cable, rope and chain are carried through this hatch to the upper deck.

The covers for these cable working hatches are unique on a cable ship. The after cover telescopes under the forward cover, and the covers can

Fig. 10 — Bow sheaves of C.S. *Long Lines*.

be moved individually on longitudinal rails to open either or both of the hatches. By this means and by opening the ends of the covers, almost any degree of protection can be obtained for personnel and machinery during foul weather operations.

IX. CHESTER LABORATORY DEVELOPMENT WORK

The use of full-scale models as exemplified in the operation of "C.S. Fantastic" was an essential part of the successful development of the cable handling facilities. While scale models were useful for purposes of visualization, they were of little practical value in predicting operating characteristics, since it was impossible to scale the cable dynamically.

All of the details for handling cable and repeaters were built in prototype and subjected to exhaustive tests. Since ship's motion (pitch, roll and yaw) could not be simulated, all of the payout processes were required to operate successfully at speeds well in excess of the design objective of 8 knots to allow for possible adverse effects from this motion.

Cable payout, repeater launching, parachute positioning, crinoline operation and transfer of payout from one tank to another were accomplished successfully at speeds as high as 11.5 knots, the maximum capability of the Chester facility.

Measurements were also made of the shocks to which the repeater was subjected during acceleration from standstill to payout speeds in the launching operation. The validity of these tests was later confirmed by similar measurements aboard ship; even at payout speeds of 9.5 knots, accelerations do not exceed 19 g, well below design requirements.

It is interesting to note that many of the innovations developed at Chester have been used in cable ships other than C.S. *Long Lines*.

X. CABLE LOADING AND REPEATER JOINING

The arrangements and details for loading, although somewhat less exacting than those required for laying, are of considerable importance. It is here that the possibility of confusion and error is greatest, and extreme care must be taken to prevent locked turns and crosses which could make it impossible to get the cable out without cutting and splicing.

Insofar as possible, the cable is loaded in a single continuous length and the mechanical continuity is never broken. This is possible only to a limited extent, since two load lines are normally in operation simultaneously. It is possible, however, to load each tank with a single mechanically continuous length, and this practice is followed in general on C.S. *Long Lines*.

The individual cable sections are taken in pans to the loading dock area and placed as required at the end of the loading line. The cable is then pulled aboard ship through the load line by a linear transporter located at the forward end of the cable tank slot and fed down into the tank for coiling. On the straight runs and around curves in a vertical plane the cable moves through open troughs or over faired surfaces, but for turns in a horizontal plane, sheaves having a minimum diameter of 6 feet are used.

Coiling of cable is in a clockwise direction, beginning at the outer tank wall and working into the central cone; then it continues working from the cone outward. In the large tanks each flake contains approximately 3 nm of cable and there are about 6 flakes per section.

This is the first time that out-in, in-out coiling of cable has been successfully used. With conventional armored cable the practice has been to coil from the tank wall in to the cone and then lay out to the tank wall and resume coiling again toward the cone. Coiling from the cone outward

was not successful with the armored cable because in the reverse operation of paying out there was a possibility of turns sticking, resulting in lifting an adjacent turn with high probability of forming a kink. With the out-in, in-out coiling, load factors approximately 92 per cent of theoretical (57.5 cu. ft./nm) for perfect pack have been achieved.

When the last end of a cable section is reached, it is joined to the first end of the next succeeding section by means of a coupling connector usually referred to as a "dummy repeater." The dummy repeater shown in Fig. 11 consists of two coupling covers connected together by a tie bar; the combination is equal in length to a repeater or equalizer. This dummy serves as a convenient device for positioning the cable section ends in the repeater stack while the cables to and from the tank are being

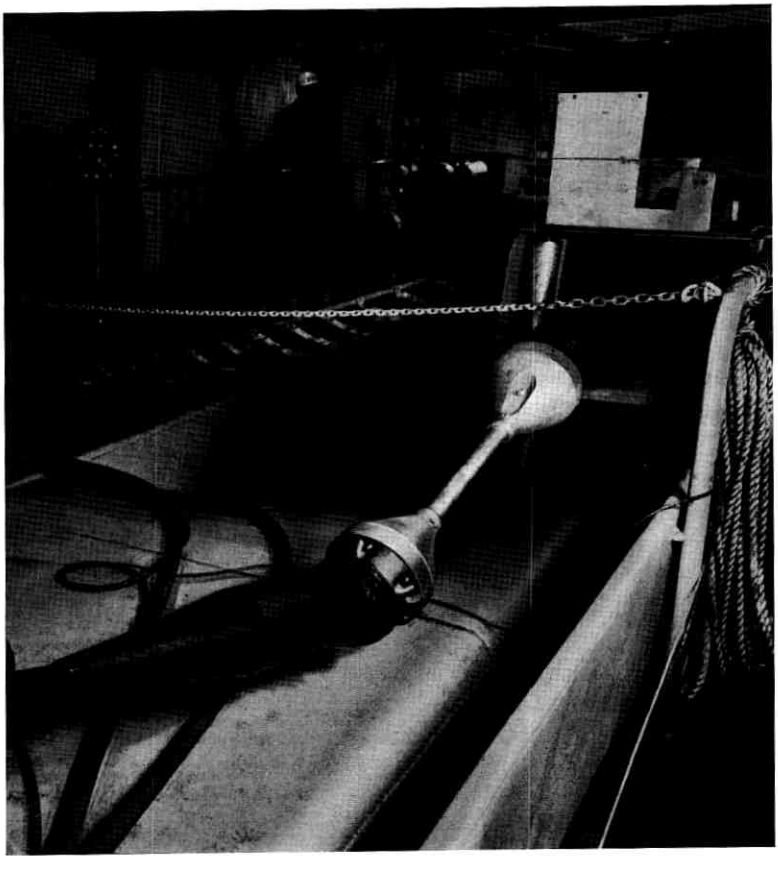


Fig. 11 — Coupling connector ("dummy repeater").

dressed into position and fastened down. In addition to furnishing a means of mechanical connection to the cable section, the coupling covers protect the coupling pigtail and provide mountings for electrical connection to the center conductor for electrical test during loading. Before the couplings are disconnected from the dummy for joining to a repeater, the cable ends are joined together by a rope tie. This tie is not removed until the process of joining the cables to the repeaters is finished and the lead position and continuity have been checked.

Coiling in the tanks is manual and crews of about 25 men per tank are used to perform this work. Loading speeds as high as 4 knots per line can be achieved, and, with allowance for down time to bring the cables from the tank to the repeater stowage area and dress them into position, an average speed of 2.5 knots per line is possible.

As soon as the first cable couplings have been placed in position at the repeater rack, a repeater is moved into place and the work of joining cable to repeaters is started. The space for the electrical and mechanical connection of cable to repeater is restricted, and the equipment designs and processes had to be carefully coordinated to work in the space available.

The joining process requires exact positioning of the coupling with respect to the repeater, preparation of pigtail ends, brazing of center conductor, overmold of conductor joint, X-ray examination of the mold, assembly of ground leads, and bolting of coupling to repeater. Mounting jigs like that shown in Fig. 12 hold the coupling in alignment with the repeater body during the joining procedure, and the jig serves in turn as a mounting for the brazer, molding machine and X-ray unit and as a bolting guide for final assembly. The illustration shows the molding machine in position.

XI. CABLE AND SYSTEM TESTS

Electrical tests of the cable during loading are rather rudimentary but will indicate any serious cable faults as soon as possible. The tests consist of dc resistance measurements made before, during and after loading; insulation resistance measurements made after loading; and pulse echo tests made after loading.

The system tests are much more comprehensive and are a measure of system transmission performance. The transmission characteristic of each block is measured and the crystal peak gain of each repeater is measured. In addition, each of the equalizers in the system is checked at every one of the 32 positions of the stepping switch.

Since the cable loss — and hence the system characteristic — varies

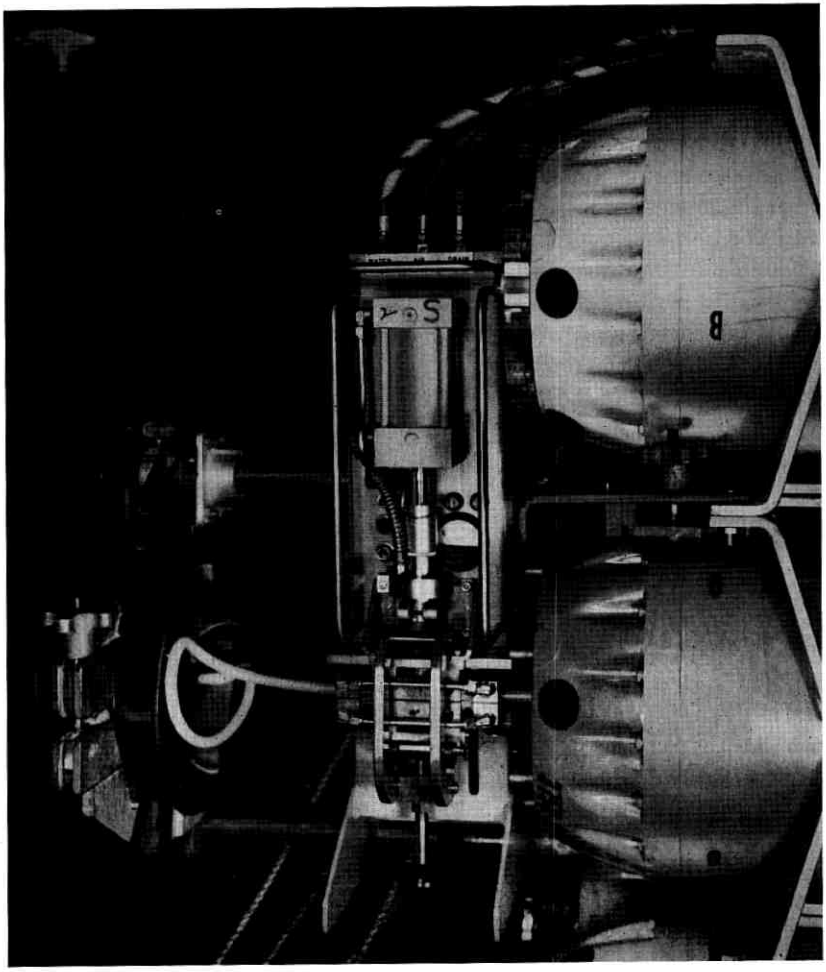


Fig. 12 — Mounting jig with molding machine in position.

with temperature, it is necessary to know the temperature of the cable at the time the transmission measurements are made. Temperature measurements are made at 15 positions in each of the repeater tanks. Three thermocouples are located in the wall of each central cone, and 12 thermocouples are placed in designated positions within the cable coils of each tank. The measurement of temperature is semiautomatic; to determine the temperature at any of the 45 positions, it is necessary only

to press the appropriate button and read out the temperature directly on the meter scale.

At the conclusion of these tests, power is turned down and no further measurements are made until the initial splice is made and laying is ready to start.

XII. SPACE LOCATIONS AND ARRANGEMENTS

Efforts were made to assign spaces for cable operations for the greatest convenience and utility. Consideration was given not only to immediate requirements for equipment areas, working space and servicing needs but also to probable future requirements.

Floor plans were worked out for all of the areas directly or indirectly connected with the cable laying operation. Lighting, air-conditioning and ventilation requirements were set for all these areas.

Special attention was given to the layouts in the drum room, transmission testroom, jointing room and taut wire room, since these are the active operating areas during cable laying, and convenience and efficiency are especially important.

In the transmission testroom all equipment is furnished in duplicate so that the failure of any unit does not interrupt system operation and test. Cable payout information which is of assistance to the transmission engineers is repeated in the test room. This information consists of cable payout speed, mileage and miles to the next repeater. Cable temperature measuring equipment is also located in this room.

Cable engine control during laying is from the cable operations center, known as the "drum room," and the layout centers around the cable engine control console. All of the payout control equipment is located here. This includes cable engine controls (both bow and stern), mileage counters of cable and taut wire, slack computers, depth sounders, precision depth recorder and ship's speed indicator. Slack computers and cable and taut wire mileage counters are furnished in duplicate.

Cable jointing, whenever possible, is carried on in the jointing room. The location on the upper deck opposite cable tank 1 was chosen, since it is the space most readily accessible for cable ends brought over the bow and through the cable hatch. The layout permits splicing operations on all types of cable without exceeding the minimum bending radius requirements.

The taut wire room houses not only the taut wire machine itself but provides for stowage of full and empty taut wire reels. Racks were designed and layout arranged so that the bulky reels, containing 140 nm of

wire weighing approximately 2000 pounds, could be safely handled under rough sea conditions.

XIII. ELECTRICAL SERVICES AND COMMUNICATIONS

The factors of safety, reliability, stability and convenience determined the choice of electrical service to the working areas. Alternate power sources supply the regular and alternate test and power equipment in the transmission testroom for maximum reliability, and both steady-state and transient voltage and frequency requirements were set for the prime power sources.

Fusing and switching of all of the equipment was specified so that these were compatible with the services furnished. In areas such as the transmission testroom and electronics maintenance room where portable equipment may be used or equipment serviced, the convenience outlets are fed through isolation transformers to provide single-phase grounded service in place of the usual shipboard single-phase balanced-to-ground supply.

Coaxial circuits are installed from the switchboard bay in the transmission testroom to all of the cable working areas and, by appropriate patching at the switchboard, test and communication circuits can be set up between any of these locations.

A cable operations communication circuit connects all of the cable working areas. Operating stations consisting of microphone and loudspeaker are located at strategic locations where there is active cable movement or cable control, and announcements over this circuit are repeated at all locations.

For the purpose of safety to personnel and to cable, "safety alarm" stations are also provided at strategic locations. These stations consist of distinctively colored button switches connected to all of the cable engines and to an audible alarm. The operation of any of these switches immediately stops any operating cable engine and cable payout or pickup comes to an immediate halt. Needless to say, this circuit is used only in extreme emergencies, to stop operations when personnel are endangered or an equipment failure requires immediate or drastic action.

REFERENCES

1. Zajac, E. E., Dynamics and Kinematics of the Laying and Recovery of Submarine Cable, B.S.T.J., **36**, Sept., 1957, p. 1129.
2. Gretter, R. W., Cable Payout System, B.S.T.J., this issue, p. 1395.

Cable Payout System

By R. W. GRETTER

(Manuscript received March 24, 1964)

This paper describes the development of cable machinery and control equipment capable of laying modern submarine cable systems. The necessity for reliable continuous operation is established, and the problems involved in gripping armorless cable are pointed out. The concept of a linear cable engine is introduced, and the evolution of a track design is followed through several model stages.

The various test programs employed to establish feasibility, prove-in component reliability, and check performance of the final machine are discussed. A control philosophy is presented and the development of a control system is described. Final design, construction, and testing of the cable machinery and its control equipment are covered briefly.

I. INTRODUCTION

In essence, a cable payout system consists of stowage for the cable, machinery for exerting a braking action on the cable to regulate the payout rate and instrumentation for determining the correct payout rate. In the first attempt to lay a telegraph cable across the English Channel, the stowage and braking functions of a payout system were provided by a reel mounted on the after deck of a tug and fitted with a brake. For cable systems more than a very few miles in length, the size of the reel soon became unreasonable and less obvious approaches had to be taken.

Stowage of cable in cylindrical holds or tanks was employed early in the development of the ocean cable laying art, and seems to have been a natural evolution from the common nautical practice of tiering chain in chain lockers or of coiling whale line in tubs. Similarly, the drum-type cable engines employed for many years as braking devices in handling telegraph cables appear to have been an early adaptation of the anchor windlass.

While a combination of tank stowage with drum machinery was adequate for telegraph systems, it leads to problems when systems with

rigid repeaters are laid, since the ship must be stopped and turns taken off the drum while each rigid repeater is passed overboard. Analytical work in the field of cable mechanics has shown that adequate control of slack requires continuity of operation. It was therefore necessary to develop continuous methods of handling rigid repeaters. Drum-type cable engines can be fitted with auxiliary devices for handling rigid repeaters continuously at slow or moderate speeds, and cable engines consisting of multiple V-sheaves have been designed for handling rigid repeaters at slow speeds by means of bypass ropes. For continuous handling of rigid repeaters at full cable laying speeds, however, a linear or straight line cable engine is essential.

II. LINEAR CABLE ENGINE DESIGN REQUIREMENTS

Modern submarine cable systems, in addition to having large rigid repeaters, differ from previous systems in another aspect which is important from the cable machinery viewpoint, namely that the cable is of the so-called armorless type with the strength member in the center.[1] The strength member is composed of very high-strength steel, and is surrounded by the delicate transmission structure consisting of polyethylene and thin sheet copper.

The cable tension which results from the weight of the length of cable suspended between a ship and the ocean bottom must be transferred to the ship by frictional shear forces exerted on the cable surface. In the case of drum engines the normal forces on the surface required for the development of the frictional forces result from the wrapping of the cable around the drum. In the linear cable engine, the curved surface is absent and the cable tension does not result in any normal force on the cable's surface. The cable must therefore be squeezed between the tracks of the cable engine.

The cable engine design requirements, from the viewpoint of the cable system, can be summarized as follows: the cable engine must handle cable and repeaters continuously at a steady speed, in a straight line, gripping the cable so as to transfer the tension to the ship without damaging the transmission structure and passing repeaters without excessive shock.

III. DESIGN APPROACH

The development approach adopted for the linear cable engine was strongly influenced by two very firmly held principles. First, the cable engine and associated equipment was to be designed to handle a mechani-

cally optimum cable and repeater system; in other words, the cable system designers would not be asked to compromise their designs in order to ease machinery problems. Secondly, the entire design would be based on achieving the highest possible reliability so that there would be a very high probability of completing a lay (say 2000 nm) in one continuous operation without any interruption due to cable engine failure. Because of the first of these requirements, it was natural to design the cable engine from the inside out; that is, to commence by considering the problem of gripping a straight cable in such a way as to transfer the highest possible cable tension to the ship without damage to the transmission structure. The moving track which accomplished the gripping was then to be designed to accommodate repeaters. A system of sprockets and shafting to support and drive the track would be added next, and addition of a framework and a drive and control system would complete the design. At each stage in its evolution the design was evaluated for reliability, and an extensive testing program was planned to prove-in each element.

3.1 *Cable Gripping*

The squeezing forces required to develop cable tension must be applied in a symmetrical manner to avoid any tendency to separate the dielectric from the center conductor. Analysis, described in Appendix A, showed that the cable would have to be gripped along at least four lines, equally spaced around its circumference, if regions of negative pressure on the center conductor interface were to be avoided. The double V-block design which is shown in Fig. 1 gave the desired cable gripping configuration.

Because of the nature of the resulting pressure distribution in the dielectric, the frictional shear forces which can be transmitted across the various unbonded interfaces within the cable are directly proportional to the radii of the interfaces. This means that the stretching of the center conductor which occurs under high tension cannot be accommodated by slip of the entire cable, because slip will occur between the center conductor and the dielectric before the outer jacket slips in the grooved blocks of the cable engine. For this reason, a linear cable engine which is to handle cable without interface bonds must have a so-called "shear-limiting" feature which limits the shear force applied to a unit length of the cable. The allowable value of shear force per unit length depends in turn on the squeeze applied to the cable.

Experimental work on cable samples showed that when the cable was gripped between the double V-blocks no permanent distortion of the

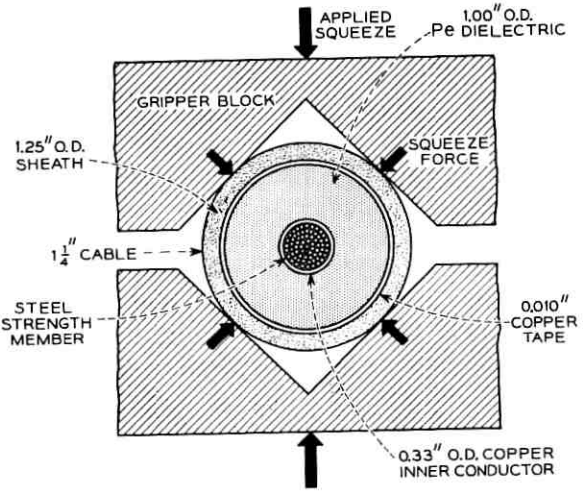


Fig. 1 — Double V-block cable gripping.

cross section would occur if the squeeze force did not exceed the value which gave a contact force of 500 pounds per inch along each contact. A working value of one-half the maximum or 250 pounds was used for design purposes. With this value of squeeze force and a coefficient of friction between copper and polyethylene of approximately 0.16, a tension decay rate of 40 pounds per inch of cable could be obtained without causing slip on the center conductor interface, which had a circumference of approximately one inch. Since the nominal tensile strength of the cable was to be approximately 16,000 pounds, this meant that the cable would have to be gripped over a length of 400 inches, or approximately 33 feet, if it was to be held up to breaking tensions without any internal slip. Thus the length of the gripping region of the cable engine was set by cable characteristics.

3.2 *Quarter-Scale Model*

The development of the linear cable engine involved several phases. In the first phase a one-quarter scale model was built to test the basic concept. At this stage in the development the cable gripping problem was not completely understood, and there was considerable optimism regarding the development of a cable having bonds on the various interfaces to transmit shear stresses. It was also thought that the proposed tracks would have to be forced apart by some sort of cam to allow the repeater to pass through. For these reasons the initial quarter-scale

model did not include the shear limiting feature, but was fitted with a pair of traveling cams which were clutched into the main drive system so as to spread the tracks apart, leaving a pocket to accommodate the repeater. The tracks which gripped the cable were pressed together by rollers mounted on saddles, the saddles in turn being mounted on compliant air bags. The quarter-scale model, shown in Fig. 2, was operated for several months in carrying out tests of various stowage arrangements and overboarding devices. Tests with this machine indicated that model repeaters would pass through with no difficulty and that the cam system was not necessary.

3.3 *Full-Scale Track Tests*

During the evaluation of the quarter-scale engine it was decided that shear limiting would be provided in the final engine, since it was certain that reliable chemically bonded interfaces could not be produced in time.

Before proceeding with the design of a mchine incorporating shear limiting, the shear limiting concept was checked out on a bench-type test rig which gripped a length of cable several feet long between blocks mounted on rollers between squeezing units. Axial motion in the direction of applied cable tension was controlled by the shear limiting units. Various spring configurations were tried and an ordinary coil spring

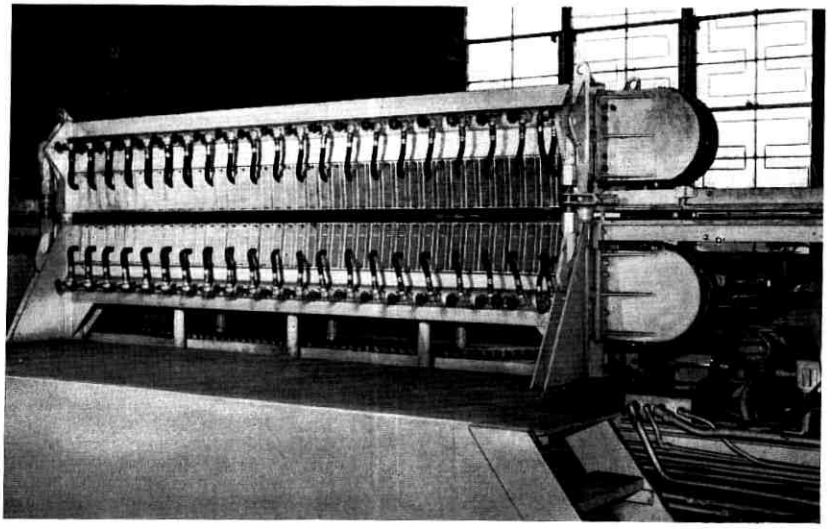


Fig. 2 — Quarter-scale model.

mounted with a preload in a cartridge was found to be the most satisfactory. The shear limiters function as shown schematically in Fig. 3. The observed deflection is essentially parabolic, as predicted by theory.

In order to allow for shear limiting motion, it was necessary to leave gaps between the individual members of the track. The track also had to articulate for envelopment of the repeater. Because the track moves with the cable, it was necessary for the track to move at high speed with respect to the elements which applied the squeezing force. After an investigation of possible sliding contacts, it was decided that rollers on the track elements would be necessary to handle the large squeezing forces and high speeds.

A track test rig was designed to check out the track as conceived at that stage of its development. Fig. 4 shows this machine, which was full size as regards width of track, but was only a few feet in gripping length. In addition, the test rig involved only one track, the opposing track being simulated by a series of simple roller carriages traveling on a rigid base. The moving track was backed up by an articulated belt, and squeeze was applied by air springs as shown in Fig. 5. This test rig was run for many hundreds of miles and at intervals a longitudinally split half repeater was passed through.

3.4 *Design and Construction Contractor*

At this point Western Gear Corporation was selected to do final design and to manufacture the linear cable engine. Its representatives joined with engineers of Bell Laboratories and Bergen Research Engineering

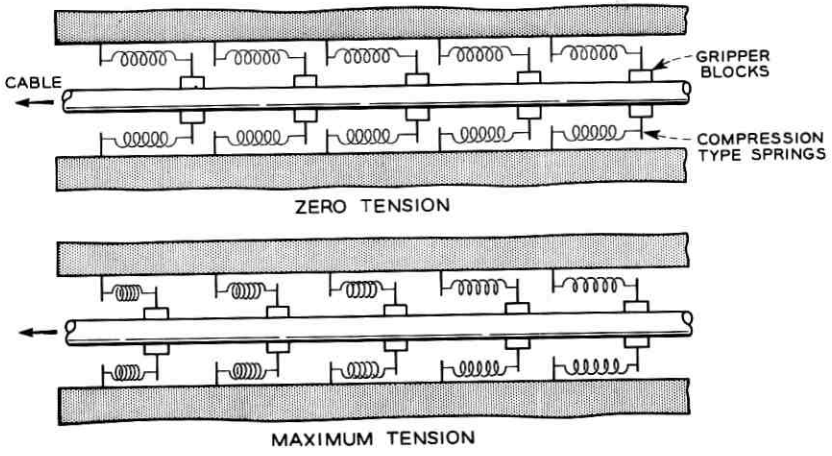


Fig. 3 — Shear limiting.

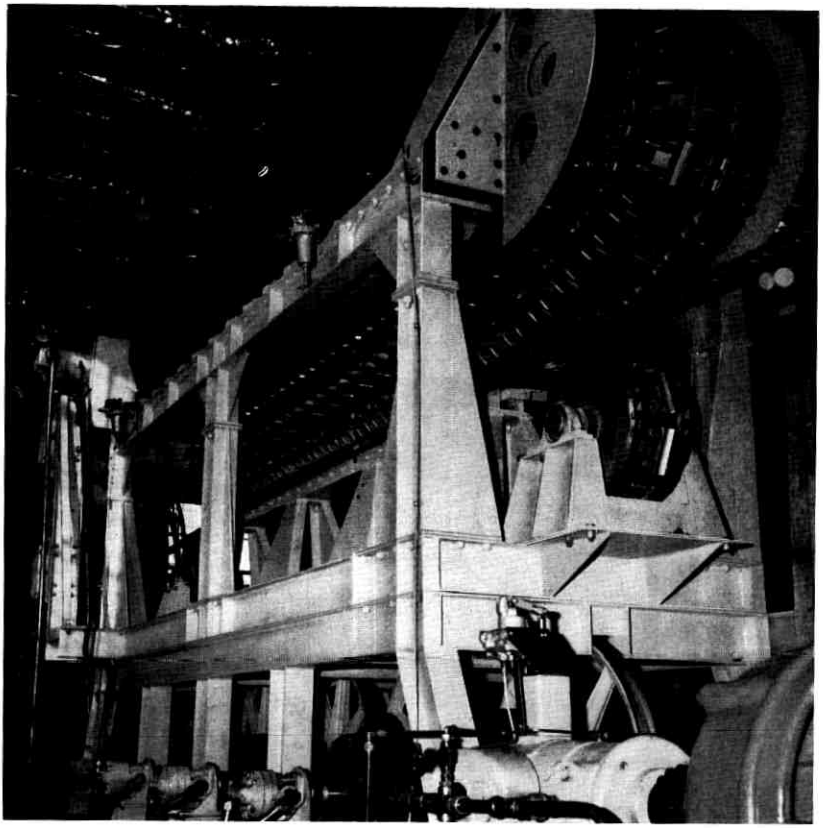


Fig. 4 — Track test apparatus.

Corporation in planning and evaluation of the test program for the track test rig. After observing the tests and studying the test results, the three groups concurred in the opinion that the rubber rollers and their bearings would be the limiting factors in designing a squeeze-type engine to have sufficient reliability. It was also concluded that the air bags which provided the compliant backing for the belts were causing an intolerable pressure increase when repeaters passed through the machine and that the solution of this problem by increasing the number of air bags in each stack would lead to a stability problem.

3.5 *Linear Cable Engine Design Studies*

Upon completion of the test program and associated studies, preliminary design work on a prototype linear cable engine was started. Initially,

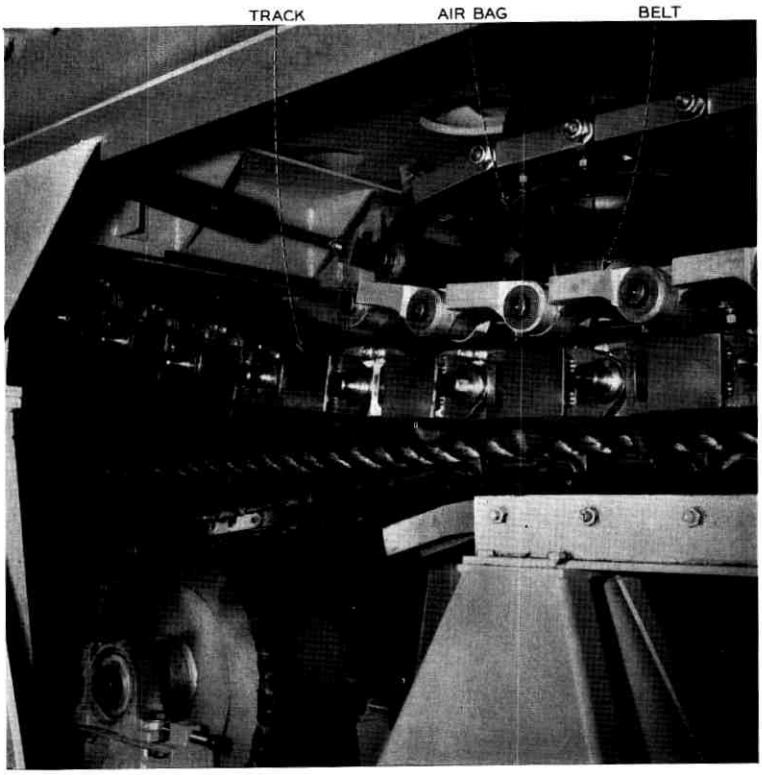


Fig. 5 — Track test apparatus; detail showing articulated belt and air springs.

effort was directed along two parallel lines: first, attempts to resolve the problems inherent in the squeeze-type linear cable engine and second, paper and model studies of alternative linear cable engine concepts proposed by Bell Laboratories.*

A nitrogen-accumulator-backed oil-hydraulic squeeze system, proposed as a substitute for the air bags, was studied. These studies showed that a combination of hydraulic cylinders with linkages could be designed to maintain the squeeze pressure essentially constant as the repeater entered the machine.

The question of transients incidental to repeater passage at high speeds was resolved by means of a test rig based on a large lathe. A cam mounted on the face plate subjected a cylinder and linkage assembly to the displacements and velocities incidental to the passage of a repeater

* These involved other gripping methods which would avoid the problem of applying high squeezing forces to a moving track.

at speeds up to 10 knots. After many hours of operation of this test rig it was concluded that the hydraulic cylinders with their linkages and nitrogen accumulators would be a satisfactory substitute for the air bags used in the previous models.

Discussion of the rubber roller problem with many rubber manufacturers led to the conclusion that the loads and speeds involved in the proposed linear cable engine would be marginal for available rubber compounds. Because our particular operating conditions and duty cycle were unique, a test program was required to establish reliability for the rubber rollers. The rubber compounds which appeared to be most suitable were made up in sample rollers, and these were run for several thousands of hours on rotating drums. The rollers were pressed against the drums intermittently to simulate the duty cycle of rollers passing through the squeeze zone on the linear cable engine. On the basis of the results of these tests, it was concluded that the rollers could be depended upon to function reliably for at least 2000 nm and possibly for several times this figure.

Initial design studies of track configurations capable of passing around sprockets of reasonable size and enveloping repeaters indicated that approximately 3000 antifriction bearings would be required to support the necessary rubber rollers. Because of the space restrictions and the load and speed requirements, the B_{10} life* of the bearings could not be made high enough to give an acceptably large probability of no bearing failures in 2000 miles of operation. Thus it was necessary to assure ourselves that a bearing which had failed by the usual bearing industry criterion would not lead to a catastrophic failure.

The Timken Roller Bearing Company had recommended a bearing identical with that used in the front wheels of compact automobiles for this application. Its engineers felt that because of the heat treatment given to the bearing races an initial fatigue failure would not propagate rapidly and thus lead to an early catastrophic failure. In order to test this hypothesis a roller bearing test program was planned. The Timken Company produced initial fatigue failues in a group of 20 bearings in their laboratory. Prototype rubber rollers were mounted on these bearings and run at prototype loads, speeds, and duty cycles for thousands of hours. The rollers were then disassembled and the bearings photographed. The photographs were compared with photographs taken at the time of initial fatigue failure. In most cases there was no discernible increase in the spalled area of the initial failure. In one or two cases the area did increase noticeably, but there was no indication that catastrophic failure would

* Life exceeded by 90 per cent of a population of bearings.

have occurred within many thousands of hours. Details of the failure propagation study are given in Appendix B.

3.6 *Track Side Chains*

The proposed track was to consist of a series of roller carriages mounted on roller chains which were in turn carried on sprockets. From exploratory work it was concluded that it was not feasible to lubricate modern roller chains for high-speed operation in a salt water atmosphere without exposing the cable to contamination by the lubricating medium. A special precision pintle-type chain, having oil-impregnated sintered-metal bearings, was therefore developed. Since there are no rotating rollers in this type of chain and each link merely oscillates through a small angle with respect to its neighbor, it was possible to seal in the lubricating medium by means of rubber shear seals which could tolerate the small angle of rotation. Because of the lack of rollers, there is of course a sliding motion of the links with respect to the sprocket. It was possible to provide lubrication for this motion by means of a very viscous type of grease which would not be thrown off the sprocket onto the cable. A test section of the new type of chain was manufactured and was found to operate satisfactorily under the prototype loads and speeds.

3.7 *Roller Carriage Design*

The roller carriages were to be connected to the side, or tension, chains by means of shear limiters which would allow essentially free axial motion of the roller carriage with respect to the chain under a certain axial load, thus limiting the shear force per unit length applied to the cable. In addition, it was necessary to design the roller carriages so that the squeezing forces applied to the cable through the rollers were not transmitted to the chains, since this would put large frictional side loads on the shear limiters. It was also desirable that the two chains be connected transversely by a rigid member. The roller carriage design which met all of these requirements was, of course, somewhat complicated. Fig. 6 is a photograph of a wooden model of the design. The model included three links of the pintle chain on each end of the roller carriage. Fig. 7 shows the relative orientations of the roller carriages, the side chains, and the articulated metal belts which apply the squeezing forces to the roller sides of the carriages.

3.8 *Final Design of Linear Cable Engine*

Upon completion of the development testing which proved-in the bearings, rollers, shear limiter, pintle chain, and the hydraulic squeezing

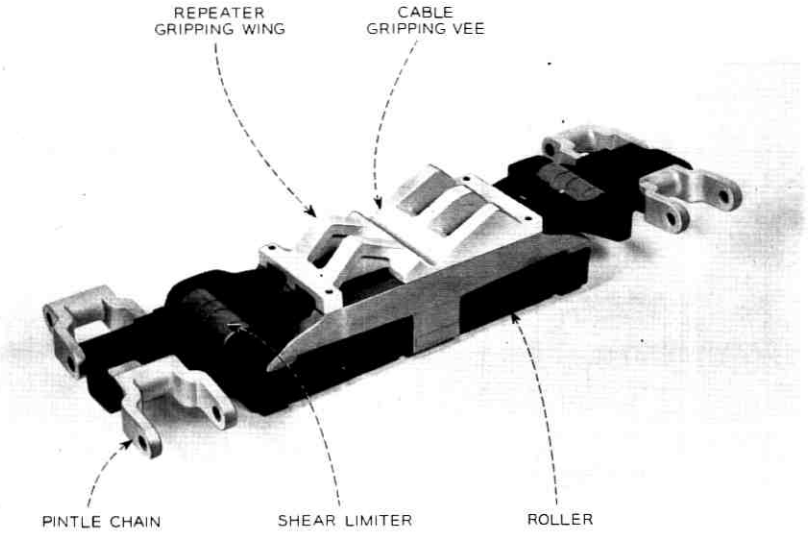


Fig. 6 — Model of roller carriage.

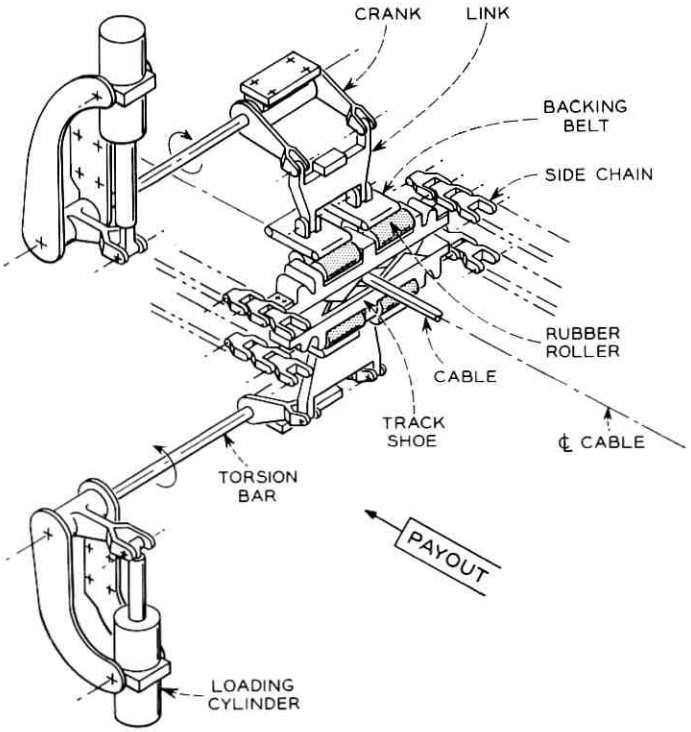
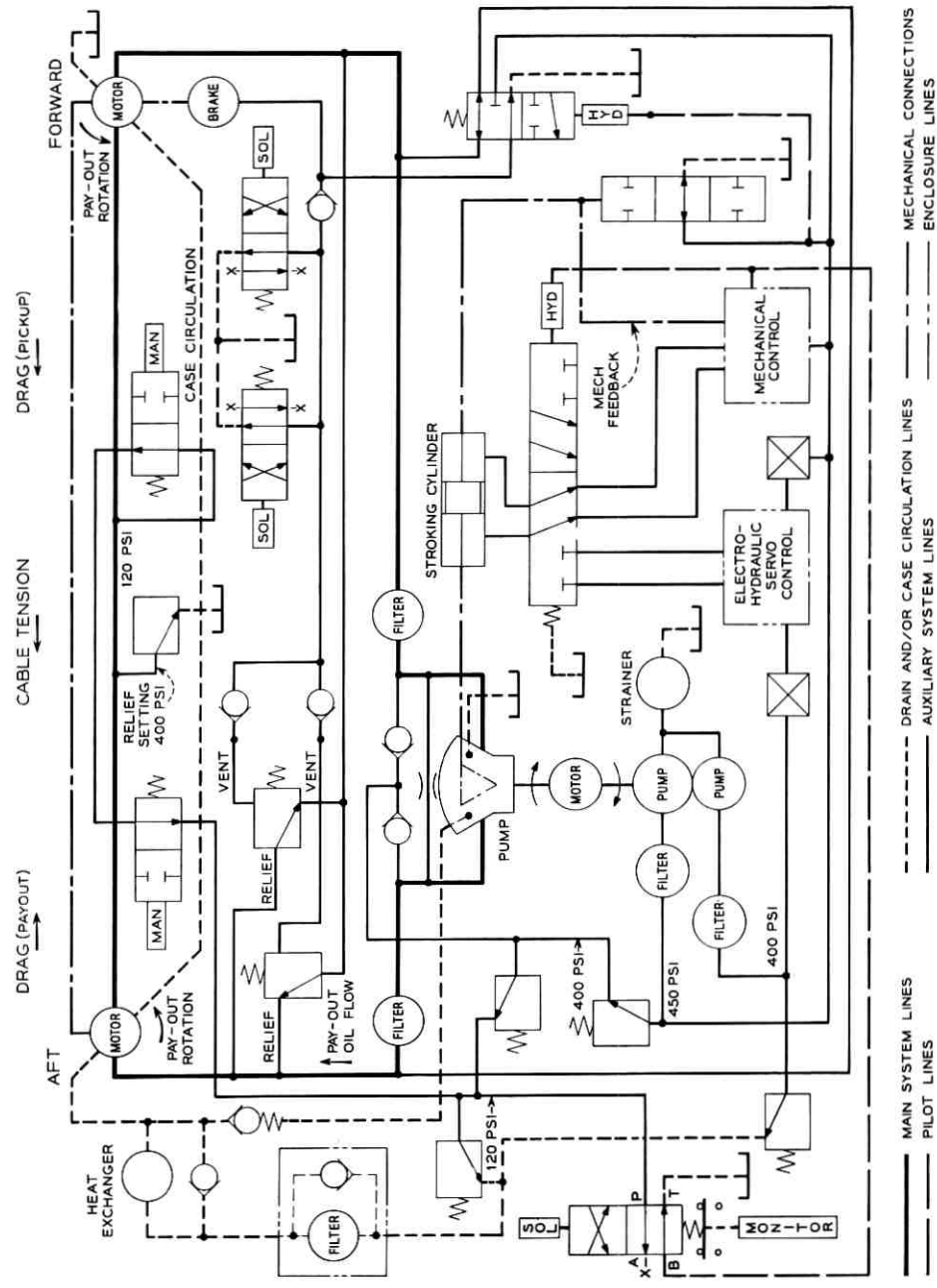


Fig. 7 — Roller carriage assembly detail.


FORWARD
DRAG (PICKUP)
CABLE TENSION
DRAG (PAYOUT)
AFT
MOTOR
BRAKE
PAY-OUT ROTATION
SOL
MAN
CASE CIRCULATION
120 PSI
RELIEF SETTING 400 PSI
VENT
RELIEF
PAY-OUT OIL FLOW
FILTER
STROKING CYLINDER
HYD
MECH FEEDBACK
MECHANICAL CONTROL
ELECTRO-HYDRAULIC SERVO CONTROL
STRAINER
PUMP
450 PSI
400 PSI
HEAT EXCHANGER
MONITOR
MAIN SYSTEM LINES
PILOT LINES
DRAIN AND/OR CASE CIRCULATION LINES
AUXILIARY SYSTEM LINES
MECHANICAL CONNECTIONS
ENCLOSURE LINES

units, the entire design was reviewed for comparison with some of the other types of linear cable engines under consideration. Although it was felt that at least one of the other approaches was probably capable of yielding a much lighter, smaller and cheaper machine, the schedule did not allow for the additional study and testing work needed to prove-in the alternate design. The development schedule, in conjunction with the encouraging results of the endurance tests on the bearings and rollers, therefore led to a decision to drop further study of alternative designs and to proceed with the final design of a squeeze-type linear cable engine.

Functionally, the linear cable engine was to consist of two major subsystems, the drive system and the squeeze system. The drive system included the track which grips the cable and moves with it, the sprockets, sprocket shafts, gear boxes, and a hydraulic system for driving the track. The squeeze system was comprised of the articulated belts between which the tracks are squeezed, the hydraulic units for applying the squeeze force to the track and the associated piping, accumulators, and hydraulic pumps for supplying oil pressure and nitrogen back-up pressure to the squeezing units.

3.8.1 *Drive System*

Because of rolling resistance of the rubber rollers and drag of the bearing seals, it was anticipated that there would be several thousand pounds of track drag. This meant that at low cable tensions it would be necessary to drive the cable engine in the payout direction, while at high cable tensions, on the other hand, it would be necessary to exert a braking effort on the track. To avoid pulling the slack, required to envelope the repeater, out of the return side of the chain and introducing it into the squeezed portion, it was necessary to drive the outboard sprocket when pushing cable out and to drive the inboard sprocket when braking. This called for a drive system consisting of a hydraulic pump with two motors in series, one attached to the sprocket shafts at each end of the machine. The replenishing oil required to prevent cavitation of the pump was introduced into the loop in such a way as to cause the two motors to pull against each other and thus keep the track under tension in the squeezing zone. A hydraulic schematic of the system is shown in Fig. 8.

3.8.2 *Squeeze System*

Fig. 9 is an artist's rendering of the linear cable engine as envisioned at this stage in its development. The machine is shown built into the deck of a ship, with a repeater enveloped by the track. The tracks in turn are backed up by the articulated "belts" which appear light-colored

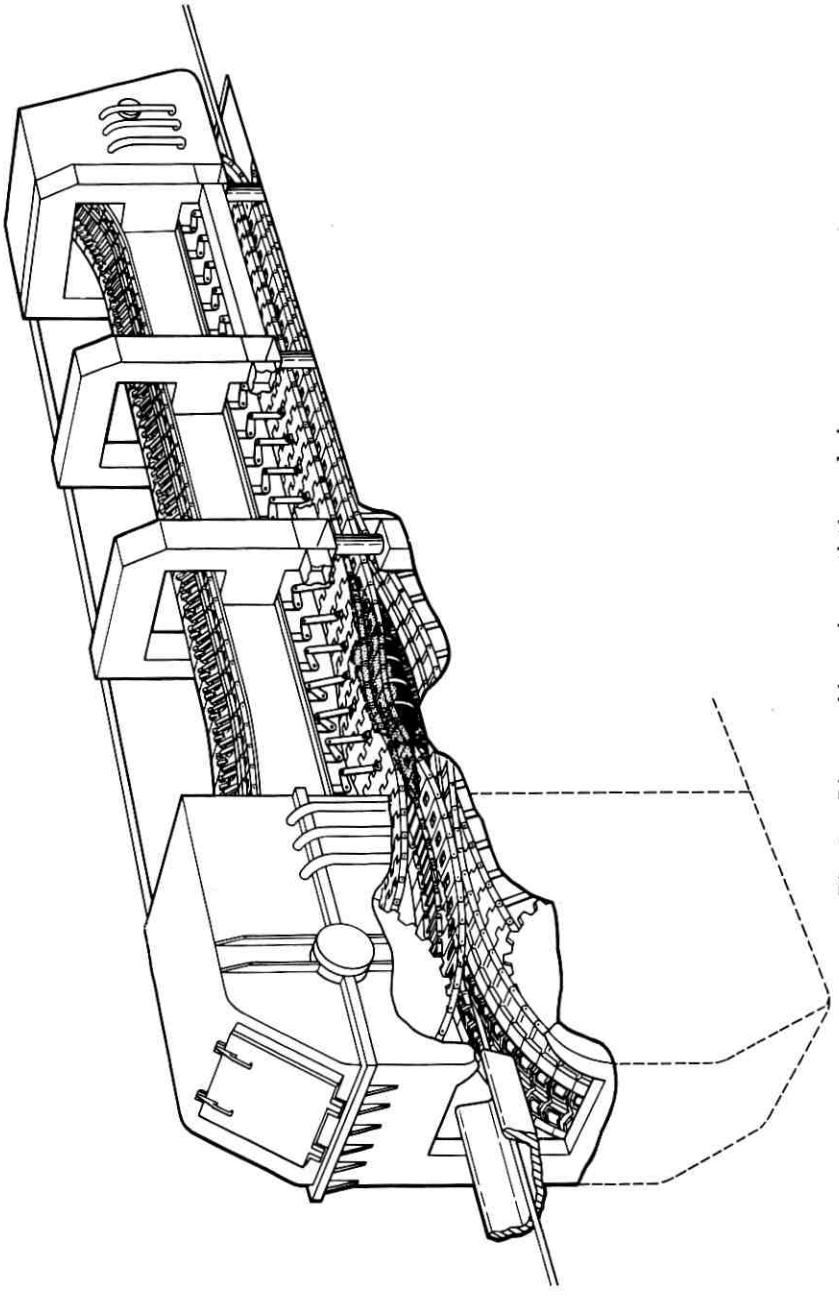


Fig. 9 — Linear cable engine, artist's rendering.

in the figure. Details of the proposed hydraulic squeezing systems are given in Figs. 7 and 10. For reliability, the gripping region of the cable engine was to be divided into seven zones, each of which had its own system of hydraulic valves and accumulator. The seven sections were to be supplied from a dual hydraulic power supply, as shown in the schematic.

IV. CONSTRUCTION

During the final stages of the design a contract was negotiated between Transoceanic Cable Ship Company, a subsidiary of the American Telephone and Telegraph Company, and Western Gear Corporation for the construction of the linear cable engine and other machinery to be described below. This contract called for manufacture in the Seattle, Washington, area with assembly at a newly acquired facility at Everett, Washington.

The completed linear cable engine is shown in Figs. 11 and 12. Fig. 13 is a close-up of the tracks and belts.

V. TEST PROGRAM

In addition to the large amount of development testing which was used to prove-in various components and subsystems, two other major testing programs were carried out to insure reliability of the linear cable engine. The first of these was in the area of production testing. As a supplement to a 100 per cent inspection of all machined parts, extensive use was made of X-rays for castings and Magnaflux for the large number of aluminum forgings in the track. Many special production tests were made on subassemblies. For example, a special test rig was made and calibrated for checking the bonding of the rubber rollers to their shafts. Another example was the production testing of the roller carriages. After assembly of the rollers with the Timken tapered roller bearings, each carriage was put on a test rig which pressed it under the full prototype load against a drum which rotated with a peripheral speed of 800 feet per minute. Temperature measurements were made and each carriage was run until the temperatures at several points stabilized. This check was intended to detect bearings with excess axial preload and faulty seals.

The final category of testing involved the completed machine. This program in turn consisted of two categories: first, performance testing in which it was determined that the machine met all of the specification requirements as to speed, tension, etc., and second, an endurance run of approximately 500 miles.

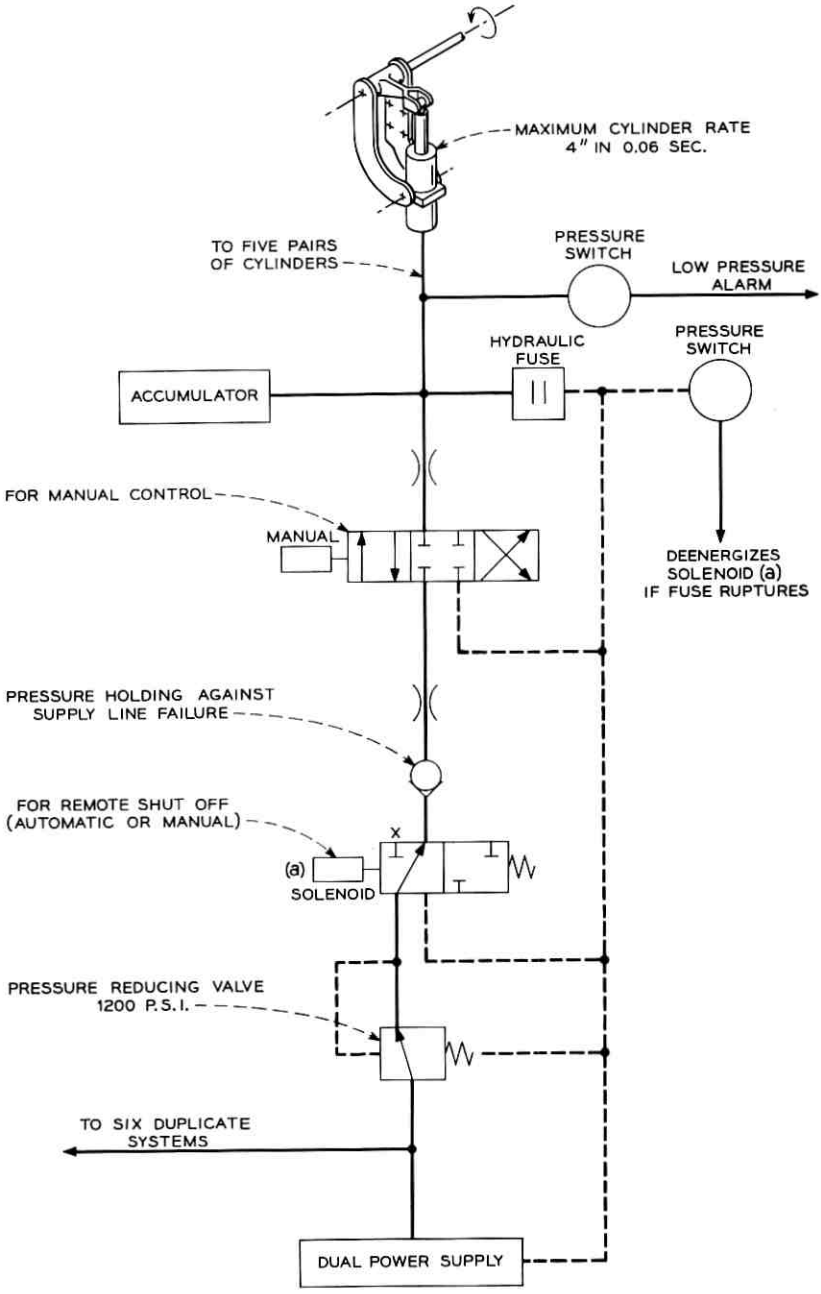


Fig. 10 — Schematic of hydraulic squeezing system.

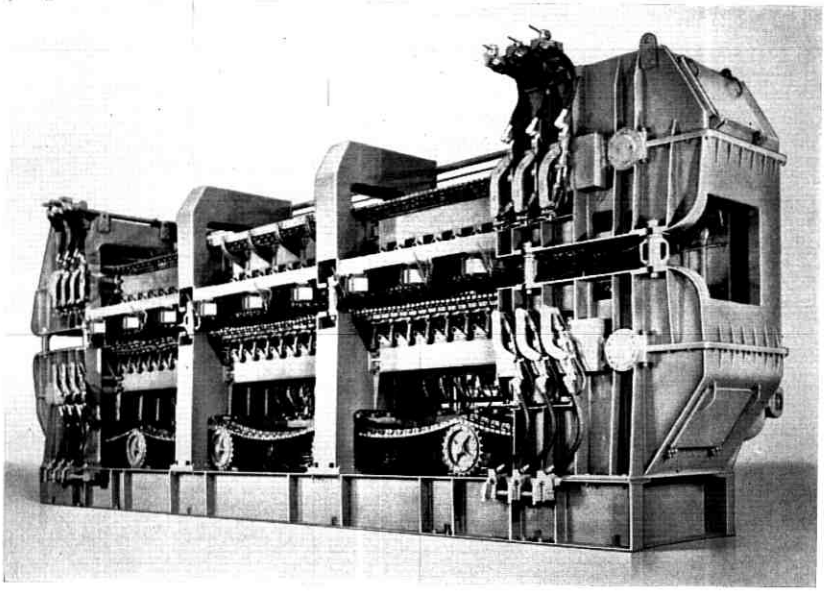


Fig. 11 — Linear cable engine.

Fig. 12 — Linear cable engine.

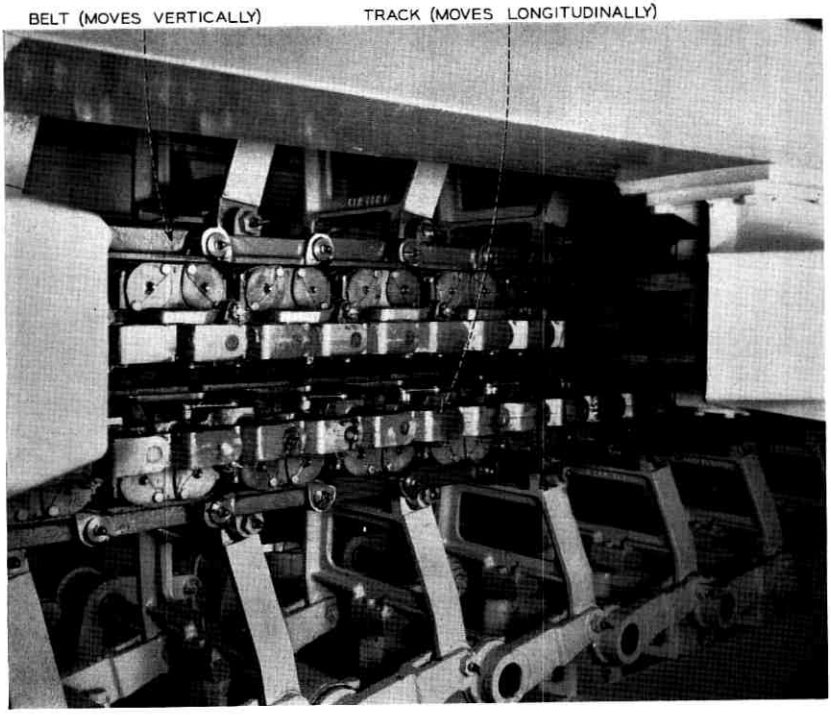


Fig. 13 — Linear cable engine; detail showing tracks and belts.

For the performance evaluation and the endurance test, the completed cable machinery was set up in such a manner that the linear cable engine could be operated against one of the drums of the bow cable machinery which will be described below. A 1000-foot length of $1\frac{1}{2}$-inch diameter "Spring Lay" mooring line was spliced into a closed loop and was run through the linear engine and over one of the drums. Large horizontal sheaves were used to complete the loop. The machine was then run at various speeds and tensions up to a maximum of 8 knots speed and 8000 pounds tension, and a wooden repeater having the size and shape of the prototype repeater was introduced into the inboard end of the machine and run through to check the articulation of the track.

The linear cable engine was designed to pull up to tensions of 16,000 pounds at low speed in the pickup direction. Pickup performance was checked by picking up at high tension with the machine in low gear. Next the brakes were set and a 40,000-pound static pull applied to the linear cable engine without slip or damage.

After successful completion of the performance checkout, the linear cable engine was run continuously day and night for 500 miles at 8 knots with the wooden repeater being sent through at intervals corresponding to the 20-mile repeater spacing. In the final test the linear cable engine was mounted on angle blocks and run inclined at an angle of $22\frac{1}{2}°$. Principal characteristics of the linear cable engine are summarized in Table I.

VI. OTHER CABLE HANDLING EQUIPMENT

In addition to the linear cable engine which was designed for steady-state high-speed laying, it was necessary to provide drum-type cable engines for handling grapnel rope, buoy moorings, and various kinds of cable encountered in repair work on telephone, telegraph, and other cable systems. Drum cable engines are essentially large-diameter winches in which tensions are built up in several turns of cable. In the case of an ordinary deck winch the necessary back tension is applied by a seaman who pulls on the hawser on the low-tension side of the winch. For a drum-type cable engine, a so-called "drawoff-holdback gear" serves this function.

The double-drum bow cable engine was also designed to handle rigid repeaters at low speeds during repair operations. This requirement called for large-diameter, wide drums and special auxiliary equipment. The 12-foot diameter specified for the drums allowed some margin over the minimum called for by the repeater rigid length, gimbal angle, and cable bending radius; the three-foot width allowed for five turns of heavy shallow-water cable on either side of a repeater.

As each turn of cable comes onto a drum, it tends to fall alongside the

TABLE I — LINEAR CABLE ENGINE CAPABILITIES

Cable diameter	
armorless cable (double-vee gripping)	1.20″ min.–1.80″ max.
armored cable	1.20″ min.*
Repeater dimensions	
diameter	14″ max.
length	12′ max.
Speed	
high gear	9 knots max.
low gear	2 knots max.
Tension	
high gear	8,000 lbs max.
low gear	16,000 lbs max.

* Diameters greater than 1.80″ contacted by "wings" on gripper blocks.

preceding turn; in other words, the cable "spools" itself across the drum. Drum cable engines are therefore normally fitted with so-called "fleeting knives" which plow the cable aside one cable diameter for each rotation of the drum to maintain the several turns of cable in an axially stationary position on the drum surface. To avoid rough treatment of the cable by a stationary knife, the cable engine was fitted with fleeting rings which rotate freely about axes inclined slightly with the axis of each drum. The rings are so oriented that they are further apart at the top of the drum, where the cable comes on, and closer together at the bottom so that, whether the machine is paying out or picking up, the oncoming cable falls freely onto the drum surface and then is pushed axially by contact with the inclined ring as the drum rotates. The several turns of cable, therefore, lie against each other and against one fleeting ring or the other, depending on whether cable is being picked up or payed out.

The two drums are carried on antifriction bearings on stationary shafts which are cantilevered toward the center of the ship from opposite sides of a closed rectangular box frame. This type of mounting has two distinct advantages. First, the end of the "dead" shaft carries the stationary mounting for one fleeting ring. Since this structure is closed in by expanded metal panels, there are no moving parts which can be contacted by a seaman walking between the drums. Second, the fact that the drums face each other facilitates transfer of cable or rope from one drum to another and decreases the size of the deck opening required.

Each drum is driven by an independent direct-current electric drive system consisting of a heavy duty mill-type motor, a separately excited generator driven by one of the main propulsion turbines, and the necessary amplidynes, excitation generators and control equipment. In the payout gear ratio, the engine is capable of paying out or picking up cable at tensions up to 16,000 pounds at speeds up to 8 knots. In two other gear ratios, tensions of up to 100,000 pounds can be handled at lower speeds. The drum speed and direction of rotation can be controlled manually from a console near the engine or remotely from the cable control center. In addition to the dynamic braking inherent in the electrical drive, the drums have mechanical shoe brakes which can be applied hydraulically or by hand.

In order to avoid bending the cable to a small radius and subjecting it to concentrated loads, a linear drawoff-holdback gear was designed. The linear design also facilitates handling of rigid repeaters and of rigid connectors in grapnel rope. A hydraulic drive system is used to allow operation at constant tension when functioning as drawoff or holdback gear and at constant speed if it should be necessary to use the gear as a cable hauler when loading cable.

Cable tension is measured by means of a dynamometer which deflects the cable and weighs the resulting component of cable tension. Earlier dynamometers consisted of large-diameter wheels or sheaves with load cells or other devices for indicating tension. The dynamometers for C.S. *Long Lines* substitute sliding of the cable on a curved "table" for the rotation of a large sheave. The load is weighed by a conventional load cell arrangement.

To allow handling of repeaters without causing the overriding turns on the drum which would result if an attempt were made to fleet large-diameter repeaters and cable together, the drawoff-holdback gear and dynamometer for each drum were arranged to traverse. Auxiliary traversable fleeting knives were also provided. When a repeater approaches the drum, the several turns of cable are fleeted away from the active fleeting ring by the traversable fleeting knife, which is then retracted. The drawoff-holdback gear or dynamometer is then traversed to direct the oncoming lead of cable onto the drum in such a manner that the repeater clears the cable on the drum. Several turns of oncoming cable are then wound onto the drum between the repeater and the active fleeting ring as the turns on the other side come off. Further rotation of the drum causes the turns to "spool" along the drum until they again contact the fleeting ring to complete the cycle.

Fig. 14 is an artist's rendering of one drum of the bow cable engine

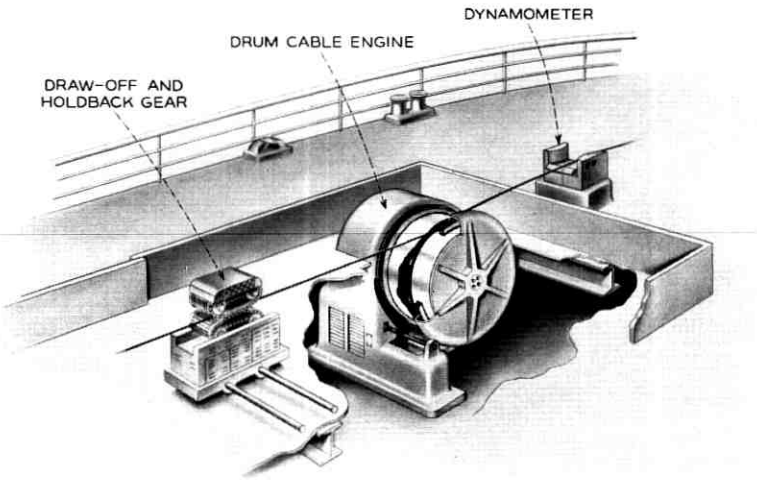


Fig. 14 — Cable engine drum with dynamometer and drawoff-holdback gear.

with its associated dynamometer and drawoff-holdback gear. Fig. 15 is a photograph of the actual drums as set up for test at the plant of the manufacturer. The fleeting rings can be seen in Fig. 16, which shows the starboard drum as installed in C.S. *Long Lines.*

6.1 *Cable Control Center*

On previous cable ships it has been the practice to control each cable engine from a nearby station so that the operator could maintain a watch on the engine and make speed adjustments by hand. Information needed for calculation of the required payout speed was collected from various sources throughout the ship. The required payout speed was then calculated, generally on the bridge, and payout speed or tension orders were then relayed to the engine operator. To integrate the cable laying operations and to permit a degree of automation, the concept of a "cable control center" has been introduced. This center was to be a room, located beneath the ship's bridge, containing all the instrumentation relating to the cable laying operation. It was also to be a control station from which any of the cable engines could be operated. Local

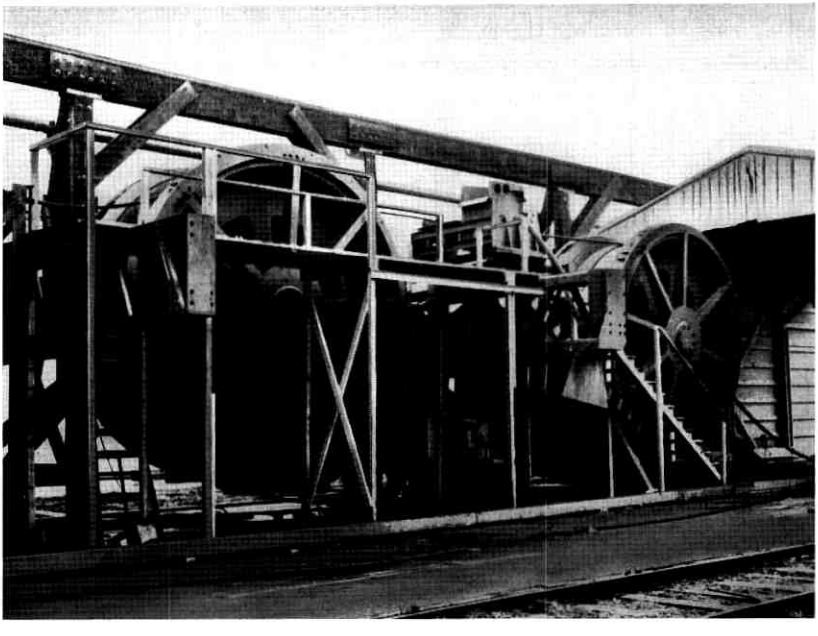


Fig. 15 — Cable engine drums set up for manufacturer's tests.

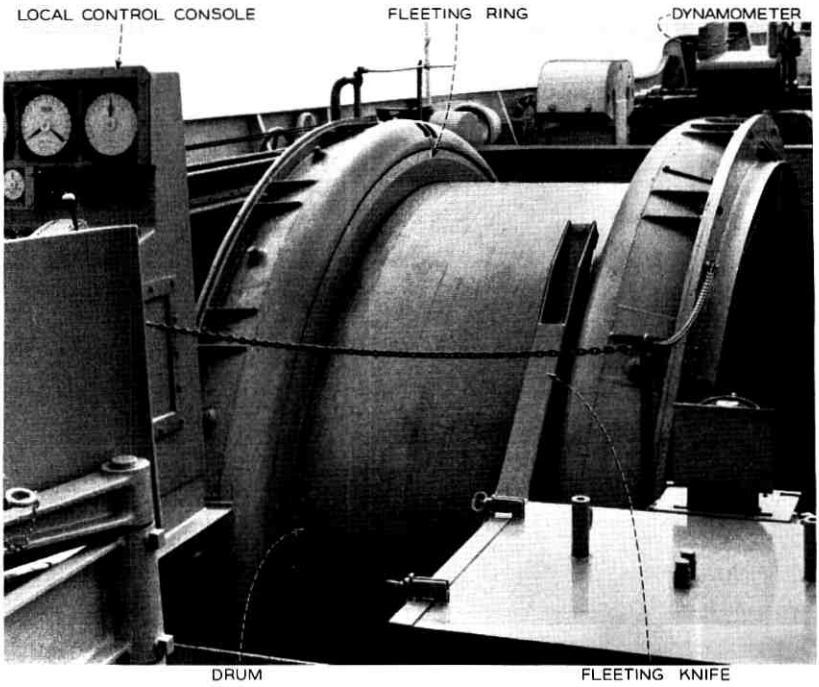


Fig. 16 — Cable engine drum installed in C.S. *Long Lines*.

control stations at each engine were still to be retained for periods when engine observation was necessary, such as during start-up or for cable repair operations. Fig. 17 shows the cable payout control console in the cable control center on C.S. *Long Lines*.

6.2 *Cable Engine Payout Control — Control Modes*

Submarine cable should be laid with just sufficient slack to enable conformance with the ocean bed, and to avoid residual tension in the cable after laying. Under these ideal conditions the cable tension at the ship during steady-state laying conditions would be essentially equal to what is usually referred to as the *wh* factor — that is, the product of the weight per unit length of the cable in sea water and the depth of the water. The tension at the ship will increase above this value if residual tension exists in the cable on the sea bed and conversely will be less than *wh* if sufficient excess slack is payed out to result in a significant longitudinal velocity of the cable relative to the water. Severe ship's motion

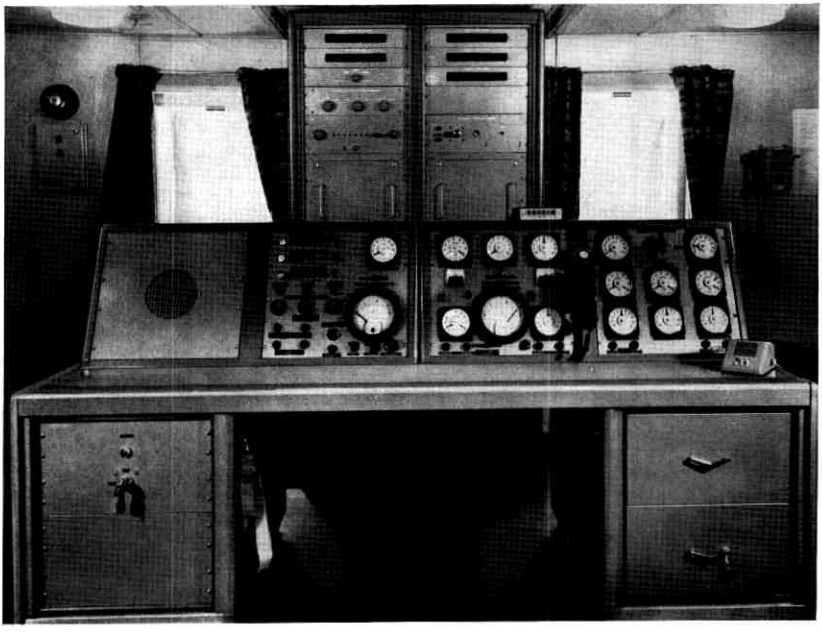


Fig. 17 — Cable payout control console.

will introduce transients into the steady-state laying conditions and will cause undesirable tension surges if the cable is payed out at a fixed speed. These tension surges can be eliminated by paying out at a fixed tension.

Because of the relationship among cable tension, payout rate, ship's speed, depth of water, and cable characteristics, a particular value of tension will yield a slack percentage which depends on the values of the other parameters. Because of uncertainties in some of these values and because of the sensitivity of payout rate to small changes in tension, it is not feasible to calculate the required value of tension from measured values of the other variables. A practical method, however, is to set up the steady-state laying conditions on the basis of cable engine speed as a result of ship speed and desired slack, to observe the resultant tension at that particular depth of water, and then to use this value of tension if ship's motion becomes excessive. Under tension control, the cable mileage payed out would be compared with distance steamed and the tension adjusted to increase or decrease the payout rate as required to produce the desired percentage of slack on the bottom.

From these considerations the following basic modes of operation were selected:

(a) Manual speed control, to obtain a desired amount of slack to cover the contour of the sea bed. Because the linear cable engine can actually push cable out, it is possible to lay slack even in very shallow water where the tensions approach zero.

(b) Manual tension control, to adjust the tension to give the average speed of the engine which results in the desired amount of slack. The instantaneous speed of the engine varies to reduce short-term tension variations.

(c) Automatic slack control, in which the amount of required slack is calculated and set into a computer which controls the cable engine average speed to maintain the required amount of slack despite changes in ship speed. The fast-acting constant tension mode can still be retained to operate within the slow acting automatic slack control, and hence the effect of ship motion on the cable tension can be reduced.

The control system design was based on these principles and also incorporated a fail-safe feature which was intended to ensure that if a component or power source should fail during the operation, control would revert without interruption to one of the simpler modes. The primary control of the engine was therefore mechanical speed control operating from a lever connected by a mechanical linkage to the stroke of the hydraulic pump which drives the engine.

Improving on this to enable a more precise adjustment of speed is the closed-loop speed control system in which a handwheel, operating through clutches and gear reductions, drives a dc-excited potentiometer, the output of which is compared with a tachometer signal from the cable engine to yield an error signal capable of operating the pump stroke control via a servo valve and linkage. The basic tension control was constructed in a similar manner, but in this case the feedback signal is obtained from a load cell arranged to measure cable tension. Again the error signal adjusts engine speed via the servo valve. Fig. 18 is a block diagram of the cable engine control system.

The manual inputs for these three control arrangements are all in close proximity to the cable engine, and remote operation of the latter two systems is obtained by means of control motors activated by push buttons on the console in the cable control center. When remote operation is desired, the motors are clutched into the systems in place of the handwheels. All necessary information regarding the behavior of the engine or cable is transmitted to the cable control center by selsyn signals, or as dc voltages, and displayed on the cable payout control console as shown in Fig. 17.

Once control of the engine, in either the speed or tension modes, is

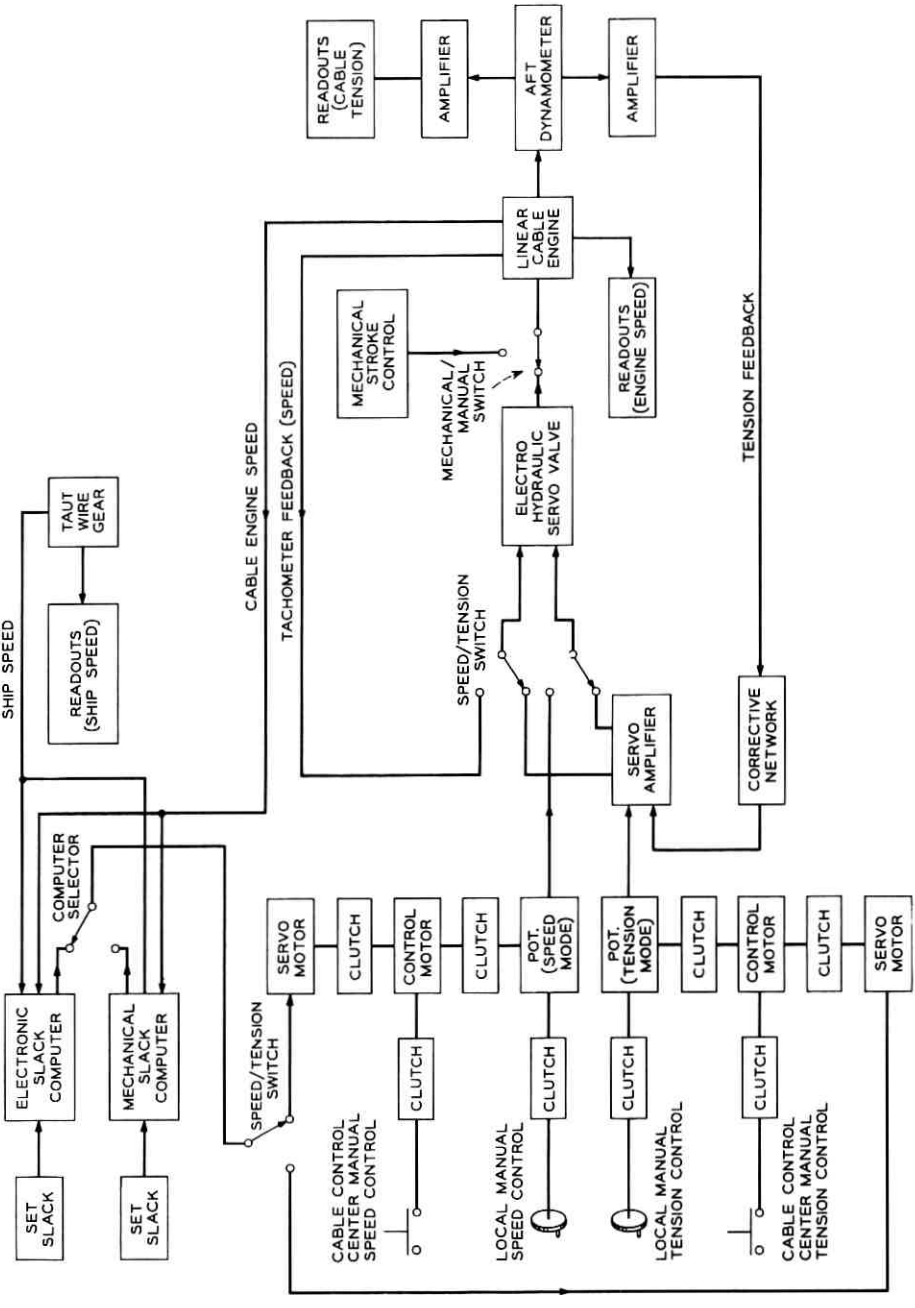


Fig. 18 — Cable engine control system.

transferred to the cable control center, the automatic slack mode can be superimposed on the system. Now the operator is merely required to set the required slack as the input to a computer, which then operates the speed or tension control potentiometers via a further series of servo motors, clutches and gear reductions.

The computer is, for reasons of simplicity and reliability, merely a refinement of the cone-and-cylinder type of mechanical computer used for many years on cable ships and miniaturized for convenience. As a standby a proprietary electronic computer is also available in the cable control center, and control could be readily transferred to this if necessary.

A separate unit was provided in the cable control center for the sole function of providing a continuous indication of cable mileage and taut wire mileage. This unit, which is shown in Fig. 17, operates independently of the other systems and is actuated by small pulse generators attached to each engine. The pulse generating units transmit one pulse for each fathom of cable or wire to an assembly of stepping switches. The accumulated count is displayed on "Nixie" readouts. Remote repeaters with identical readouts are also provided in the test room, on the bridge, and on the forecastle. This type of presentation was selected to avoid the noise of counting relays in the test room and on the bridge, and to allow all units to be zero-set or reset simultaneously from one location.

VII. CONCLUSION

The cable engines and control system were checked out during several short crew-training cruises soon after delivery of the ship. Cable and repeaters were payed out by the linear cable engine at speeds up to 8 knots in all control modes. The drum engines successfully handled armored cable, armorless cable, repeaters, grapnel rope, chains, shackles and other hardware.

Note Added in Proof

As of May 15, 1964, the linear cable engine has payed out approximately 8200 nm of SD cable systems.

APPENDIX A

Cable Gripping

Consider an armorless coaxial submarine cable having a high-strength steel central member surrounded by a coaxial transmission structure con-

sisting of thin copper sheet and polyethylene dielectric. An axial tensile load on the strength member can only be dissipated by axial shear forces applied to the external surface and transmitted through the transmission structure to the central strength member. The magnitude of the maximum allowable axial shear force per unit length will be limited because the shear stresses in the copper and polyethylene, and particularly on the interfaces, must be kept to a minimum. For a given allowable maximum shear force per unit length, the linear cable engine which makes most efficient use of gripping length is that which applies a uniform shear force per unit length to the cable.

Assume that a linear cable engine can be designed so that the tension in the cable is dissipated uniformly in a gripping length λ. The tension distribution will then be as shown in Fig. 19. As a basis for the design of a cable engine which would achieve this tension decay, it is of interest to calculate the displacement of the cross section of the center conductor at the high-tension end of the machine.

The axial strain in the center conductor is given by

$$\epsilon = T/EA \tag{1}$$

where

T = local tension

E = Young's modulus

A = cross-sectional area of the strand.

If the axial strain is integrated from the point in the cable at which the tension is zero, we find that the cross section at the high-tension end will move, with respect to the assumed stationary cross section at the zero tension point, an amount given by

$$u = \tfrac{1}{2}\epsilon_{max}\lambda \tag{2}$$

or

$$u = T_o\lambda/2AE. \tag{3}$$

With a knowledge of the cable properties the maximum displacement could be calculated from (3). To avoid becoming involved in the details of the cable design it is more convenient to work directly with (2) and the fact that the strength member will be composed of "improved plow steel." Since this type of steel has a tensile strength of 300,000 psi and its Young's modulus is 30 million psi, the maximum strain in the strength member will reach 1 per cent under the ultimate load. For lower tensile

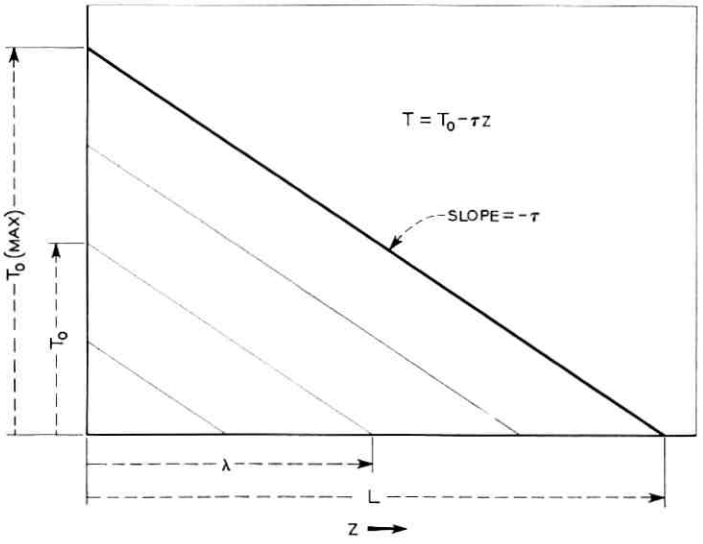


Fig. 19 — Uniformly dissipated cable tension distribution.

loads the maximum strain in the cable will be given, approximately, by

$$\epsilon_{max} = \frac{T_o}{T_{o(ultimate)}} \cdot \epsilon_{(max)(ultimate)} . \tag{4}$$

Equation (1) can now be put into the form

$$u = \frac{1}{2} \frac{T_o}{T_{o(ultimate)}} \cdot \epsilon_{(max)(ultimate)} \cdot \lambda \tag{5}$$

and the total displacement can be evaluated for a given operating tension, ultimate tension, and gripping length. The gripping length, in turn, is determined by the allowable shear force per unit length.

Fig. 20 is a plot of normal stress on the center conductor of an armorless cable having bonded interfaces and subjected to a tension T on a V-groove sheave of radius R and V-groove angle φ. In the special case where the V-groove angle equals 180°, the two normal reaction forces on the cable surface coalesce into one and we have an ordinary cylindrical drum. The calculations were actually made for this case, and the 90° situation was obtained by superposition.

For both the 90° V-groove and the plain cylindrical drum there are negative normal stresses on the center conductor in certain regions. Unless the cable has a bonded interface with sufficient strength to tolerate

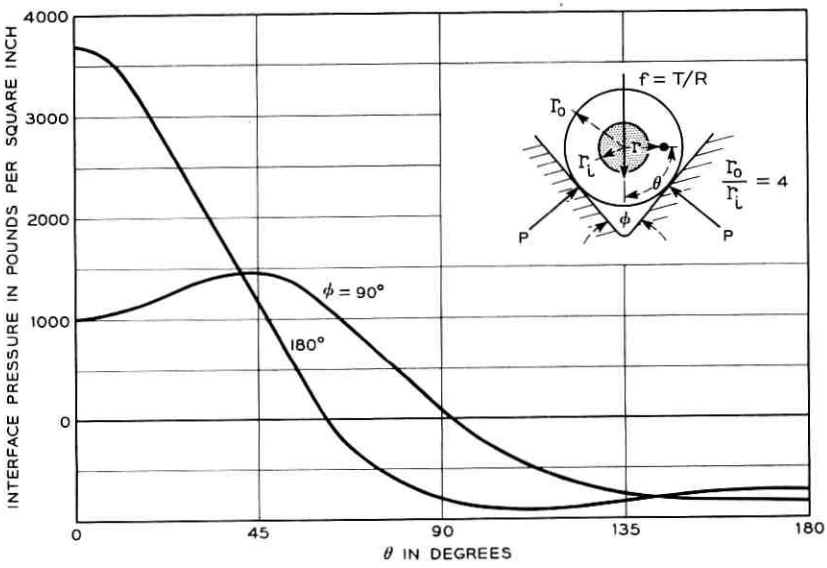


Fig. 20 — Center conductor normal stress — one V-block.

the tensile stresses or there is a shrink pressure on this interface equal to the maximum tensile stress, there will be separation and of course the calculated stress distributions will no longer apply.

Fig. 21 depicts the center conductor normal stress distribution for the case of two symmetrically located V-blocks which represent the gripper blocks of a linear cable engine. In this case the radius of the equivalent V-groove sheave has gone to infinity and the finite cable tension can give no normal component of force. The two blocks must therefore be squeezed together. It is to be noted that the normal stress on the center conductor interface does not become tensile anywhere around the circumference. This type of gripping will therefore not cause radial separation of an armorless cable with unbonded interfaces, a polyethylene dielectric and a ratio of outside radius to center conductor radius of four.

The cable gripping forces which are shown as concentrated in Fig. 21 will actually cause slight flat spots to develop on the cable and the contact stress distribution will be semielliptical, as shown in Fig. 22. The pressure distribution around the outer periphery of the cable can be represented by a Fourier series which will have a constant term given by the average pressure distribution which, in turn, is simply the sum of the four contact forces divided by the outside circumference. In addition there will be a series of sinusoidal terms. If the center conductor is as-

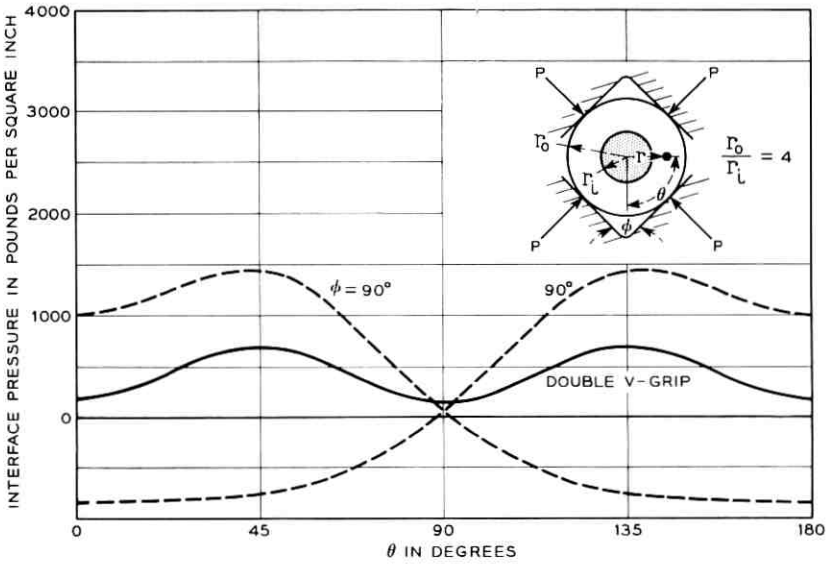


Fig. 21 — Center conductor normal stress — two V blocks.

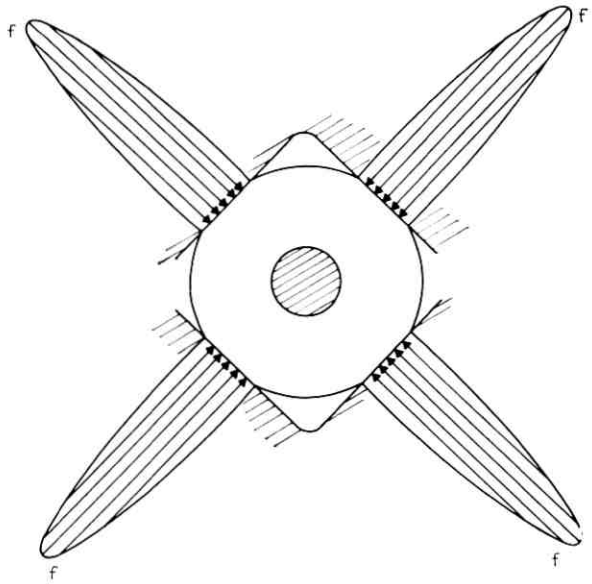


Fig. 22 — Contact stress distribution — two V-blocks.

sumed to be infinitely rigid and the polyethylene incompressible, the average pressure term will be transmitted undiminished to the center conductor. The various sinusoidal terms will be attenuated in some fashion, but in any case their contribution to the average pressure on the center conductor interface will be zero. The average pressure on the center conductor interface will therefore be equal to the average pressure on the external circumference of the cable.

The frictional shear stress which can be developed on any cylindrical interface will be given by the product of the coefficient of friction upon that interface and the normal pressure. The shear force per unit length which can be transmitted across the interface can be obtained by integrating the frictional shear stress around the circumference. If the coefficient of friction is independent of pressure, the frictional shear force at the condition of impending slippage will be given by the product of the coefficient of friction, the circumference, and the average interface pressure. For the center conductor interface we have

$$\tau_i = 2\pi r_i \mu_i P_{i(\text{av})} \tag{6}$$

and for the outer surface

$$\tau_o = 4\mu_o f \tag{7}$$

where

$$P_{i(\text{av})} = P_{o(\text{av})} \tag{8}$$

$$P_{o(\text{av})} = 4f/2\pi r_o \tag{9}$$

τ = slip value of frictional shear force per unit length

r = interface radius

μ = coefficient of friction

P = pressure

and

f = compressive force per unit length for the area of contact with each of the four plane faces of the gripper

and the subscripts "i" and "o" refer to inner and outer, respectively. Combining (6), (7), (8), and (9) yields

$$\tau_i/\tau_o = \mu_i r_i/\mu_o r_o \,. \tag{10}$$

Equation (10) indicates that slip will occur on the center conductor in-

terface for values of shear force per unit length which will not cause the outer jacket to slip in the V-blocks unless we have

$$\mu_0/\mu_i \leqq r_i/r_o \tag{11}$$

Since the coefficient of friction on the center conductor interface will be approximately 0.16 and the ratio of outside to inside radius will be approximately four, inequality (11) calls for an external friction coefficient of less than 0.04. It was not considered feasible to maintain this value of coefficient of friction between the outer surface of the cable and the gripping elements of the cable engine. A shear limiting feature, which allowed axial motion of the gripper block at a shear force per unit length which would not cause slip on the center conductor interface, was therefore included in the track design.

Equations (6), (8), and (9) can be combined to give

$$\tau_i = 4\mu_i(r_i/r_o)f \tag{12}$$

or, for our case

$$\tau_i = \mu_i f \tag{13}$$

since the outside-to-inside radius ratio is four.

Experimental work on gripping of cable in double V-blocks has established 500 pounds per inch as the maximum allowable contact force. A design value of 250 pounds per inch was used in setting the allowable shear force per unit length. Substitution of this value and of the previously mentioned coefficient of friction of 0.16 into (13) gives

$$\tau_{(\text{shear limit})} \leqq \tau_{i(\max)} = 40 \text{ lb/in.} \tag{14}$$

The τ notation is used for the shear limiter setting because the effect of an ideal shear limiter is equivalent to slip of the jacket in a gripper block. The active gripping length is then

$$\lambda = \frac{T_o}{\tau_{(\text{shear limit})}} \tag{15}$$

and (3) and (5) can be written

$$u = \frac{T_o^2}{2AE\tau_{(\text{shear limit})}} \tag{16}$$

and

$$u = \frac{T_o^2 \epsilon_{(\max)(\text{ultimate})}}{2T_{o(\text{ultimate})}\,\tau_{(\text{shear limit})}} \tag{17}$$

respectively. For the cable in question we have

$$T_{o(\text{ultimate})} = 16{,}000 \text{ lb}$$

$$\epsilon_{(\text{max}) \ (\text{ultimate})} = 0.01$$

and

$$\tau_{(\text{shear limit})} = 40 \text{ lb/in.}$$

whence

$$u = \frac{T_o^2}{128 \times 10^6} \text{ in.} \tag{18}$$

for T_o in pounds.

If this cable were gripped and tensioned in such a manner as to achieve the linear decay of Fig. 19 for all tensions up to the nominal ultimate, (18) would give

$$u_{(T_o=16{,}000 \text{ lb})} = 2 \text{ in.}$$

For smaller working tensions the required shear limiter travel decreases parabolically and

$$u_{(T_o=8{,}000 \text{ lb})} = \tfrac{1}{2} \text{ in.}$$

APPENDIX B

*Bearing Reliability**

The bearing industry defines the life of bearings as that period of service which is limited by fatigue phenomena. The fatigue-limited life terminates in what will be referred to as the "initial failure" of a bearing. It is characterized by fatigue of the material due to repeated stress under rotation resulting in flaking and ultimately in spalled areas on the rollers and/or races of the bearing. This initial failure will not make a bearing inoperable for cable engine application but will merely tend to increase the noise level and possibly the rotating torque to a small degree.

When a bearing is operated after initial failure the spalled areas increase in size until ultimately the bearing is incapable of performing its design function. It may "lock-up" so that the cup and cone cannot rotate with respect to each other, or the rollers and cage may disintegrate, allowing large radial displacements of cup with respect to cone. In either of these cases the bearing will be said to have failed catastrophically.

Since castrophic failure is a result of running an initially failed bearing

* Furnished by G. J. Levenbach and T. N. Grogean.

beyond initial failure, it is necessary to discuss the fatigue life of bearings as a basis for discussion of the subsequent catastrophic failure.

B.1 *Initial Fatigue Failure*

The following discussion is based on the assumption that the bearings are run under ideal conditions. We assume proper lubrication, the absence of foreign materials, isolation from the atmosphere, and proper installation and handling.

With these assumed running conditions, the probability that no initial fatigue failure will occur in a bearing running for a given time is contained in a formula given by Palmgren:[2]

$$\text{Log}_{10} \frac{1}{S} = (L_s)^b \log_{10} \frac{1}{0.9} \tag{19}$$

where

S = probability of a bearing exceeding a life L_s

L_s = life measured as a fraction of the life exceeded by 90 per cent of population of bearings (so-called B_{10} life)

b = shape constant in the Weibull distribution.

Palmgren's formula is based on extensive experimental evidence that the fatigue failure time distribution can be described by a Weibull distribution.

For a sample of n bearings the probability, S_n , that all of them exceed L_s is derived from (19) and takes the form

$$S_n = \exp\left[-(L_s)^b \cdot \frac{n}{a}\right] \tag{20}$$

where a is a conversion constant given by

$$a = \frac{\log_{10} e}{\log_{10} (0.9)^{-1}} = 9.5$$

and e is the base of the Naperian logarithm.

The shape constant b is a property of a specific type of bearing.[3] Palmgren notes that for commonly used bearing steels b has values in the range of 1.1 to 1.5.

Data obtained from 30 bearings fatigue-failed by Timken Roller Bearing Company can be used to estimate the shape factor b. When the data listed in Table II are plotted on Weibull probability paper the sample distribution should appear as a straight line.[4] The slope of the line

TABLE II — INITIAL FAILURE PATTERN OF TEST LOT OF 30 LM11700 SERIES TIMKEN BEARINGS (TIMKEN TEST NUMBER 151.3-G)

Number of bearings exceeding life L_s	29	23	16	10	3
L_s	0.64	2.80	5.00	9.0	9.90

gives an estimate of $b = 1.45$. Using (20), this can be used to calculate the probability that no bearing failure will occur in h hours of operation. This probability is plotted in Fig. 23.

The expected number of failures (n') in h hours can be determined by multiplying the probability of a bearing failure p by the total number of bearings n and noting that the probability of a bearing fatigue-failing is $p = 1 - S$.

$$n' = pn = (1 - S)n = \left[1 - \exp\left\{-(L_s)^b \cdot \frac{1}{a}\right\}\right] n. \qquad (21)$$

Equation (21) is plotted in Fig. 24 showing the expected number of failures vs hours of operation.

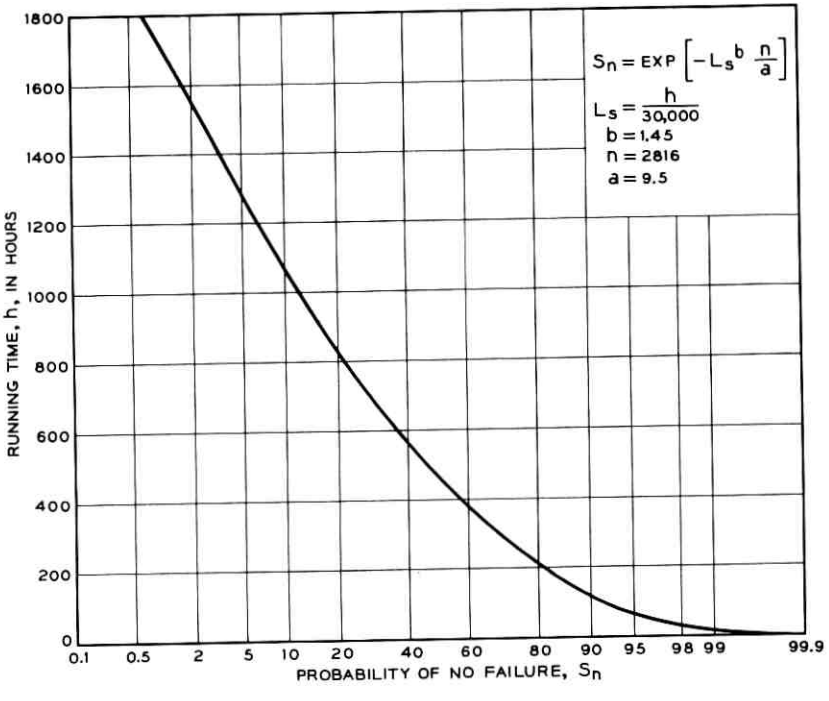


Fig. 23 — Probability of no bearing failure in l hours.

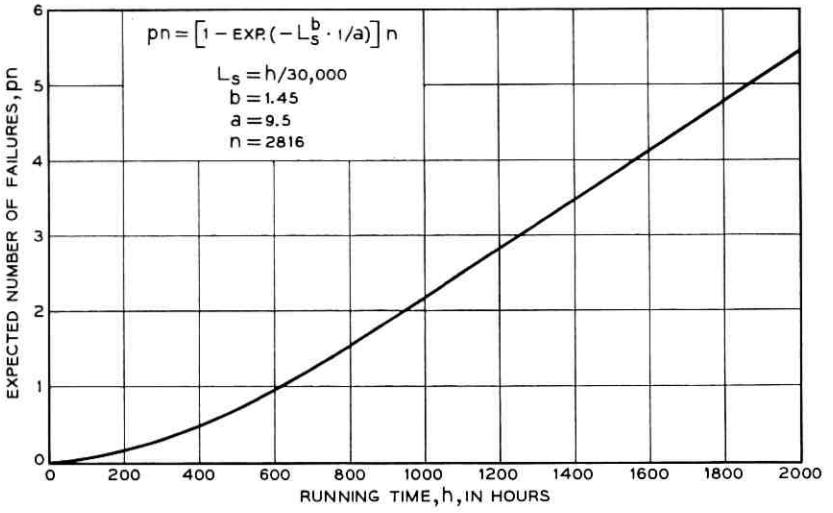


Fig. 24 — Expected number of failures in h hours.

Another useful result is the probability that there will be N or more bearing failures in a given number of hours. Basically this can be described by a binomial process. The binomial formula gives the probability of N failures of a total of n bearings. For easy computation, this can be approximated by using the pn in the previous discussion as the parameter in the Poisson distribution. The resulting relation shows

$$P = e^{-pn} \sum_{i=N}^{n} \frac{(pn)^i}{i!} \tag{22}$$

where P is the probability that there will be N or more failures in a total number n with pn the expected number of failures. Equation (22) is plotted in Fig. 25 for P vs N for various values of h.

It is seen from the previous discussion that there is a high probability that there will be initial failures in relatively short operating times. For example, in 300 hours, which is roughly the time for one ocean crossing, Fig. 23 shows that there is only 70 per cent probability that there will be no bearing failures; Fig. 24 shows that the probable number of failures is approximately 1; and Fig. 25 shows that there is a 5 per cent probability that there will be 2 or more.

With these facts known for bearings run under ideal conditions, it is necessary to accept the fact that there will be initially failed bearings in the cable engine during extended running periods. The possibility of propagation to catastrophic failure is therefore of interest.

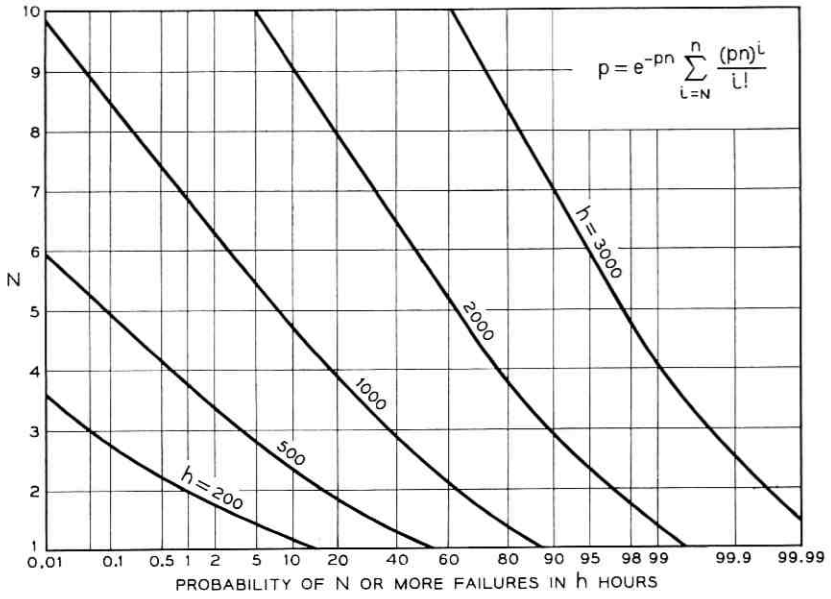


Fig. 25 — Probability of N or more failures for various hours of operation.

B.2 *Failure Propagation*

The bearing industry has not been greatly concerned with catastrophic failure; consequently, little information is available. At the request of Bell Laboratories, Timken Roller Bearing Company ran accelerated life tests on four bearings of the type considered here. The four bearings were run on their laboratory fatigue test machines at high speed under heavy load with oil lubrication. Two of the four bearings experienced what might be considered catastrophic failure at the end of 3.4 and 4.0 B_{10} life. The other two were still operable at the end of 4.7 B_{10} life.

In addition to this accelerated test result, Timken supplied Bell Laboratories with twenty bearings which had been run to the point of initial fatigue failure with oil lubrication. Eight of the twenty bearings were run under prototype loads at prototype speed by Western Gear Corporation and lubricated every 400 hours with Shell Darina A-X. The test was terminated after 3020 hours with little change in running properties. Pictures taken before and after the test indicate that the failure did not propagate to any considerable degree.

The remaining twelve fatigue-failed bearings were tested at Bergen Research Engineering Corporation. In this case they were initially

greased with Shell Darina A-X and run continuously without further lubrication. At 5500 hours the driving torque on one roller had increased significantly, and the test rig was shut down. Inspection indicated that the grease had become dry and hard in both bearings of the offending roller. The remaining rollers were regreased and the test resumed.

At 5850 hours, it was observed that several seals had been forced out against the ends of the rollers, apparently because dried grease had blocked the vent fittings. The test was therefore terminated at 5850 hours and all twelve bearings were returned to the Timken Roller Bearing Company for examination. Photographs taken before and after the 5850-hour failure propagation test were compared. Although there had been a considerable increase in spalled areas in some cases none of the bearings could be said to have failed catastrophically.

B.3 *Interpretation of Results*

Because of the small amount of failure propagation data available, it is not possible to develop a failure distribution theory similar to that which applies to the initial fatigue failure. It might, however, be of interest to look at the results under the somewhat tenuous assumption that the times to catastrophic failures are exponentially distributed.

(a) Assuming the accelerated life test of the four bearings terminated at 4.0 B_{10}, which is conservative, we estimate the mean life for catastrophic failure to be[5]

$$\theta_a = 231{,}000 \text{ hours} = 7.7\ B_{10}.$$

(b) Assuming one failure in the eight-bearing test to have occurred at 3020 hours provides an estimate of

$$\theta_b = 24{,}000 \text{ hours} = 0.8\ B_{10}.$$

(c) Assuming one failure in the 12-bearing test to have occurred at 5500 hours provides an estimate of

$$\theta_c = 66{,}000 \text{ hours} = 2.2\ B_{10}.$$

Taking the lowest estimate at its face value would yield a probability that an undetected initial fatigue failure would not propagate in the next 300 hours to the point of catastrophic failure as

$$\exp(-t/\theta) = \exp(-300/24{,}000) = 0.99.$$

This result is conservative in that there was actually no catastrophic failure in the eight-bearing test.

REFERENCES

1. Bowker, M. W., Nutt, W. G., and Riley, R. M., Design of Armorless Ocean Cable, B.S.T.J., this issue, p. 1185.
2. Palmgren, A. *Ball and Roller Bearing Engineering*, third ed., Philadelphia, SKF Ind. Corp., 1959.
3. Lieblein, G., and Zelen, M., Statistical Investigation of the Fatigue Life of Deep-Groove Ball Bearings, J. Res. N.B.S., **57**, Nov., 1956, pp. 273–316.
4. Kao, J. H. K., A Graphical Estimation of Mixed Weibull Parameters in Life Testing of Electron Tubes, Technometrics, **1**, 1959, pp. 389–407.
5. Epstein, B., and Sobel, M., Life Testing, J. Am. Stat. Assoc., **48**, 1953, pp. 486–582.

Design and Powering of Cable Ship "Long Lines"

By J. H. BUTLER, C. J. ALTENBURG, R. J. McSWEENEY and L. E. SUTTON*

(Manuscript received March 31, 1964)

The C.S. "Long Lines" was designed both to lay the new SD submarine cable system and to repair new and existing undersea cables. Within the limits imposed by this dual purpose, an entirely new cable ship design has been evolved which gives maximum working efficiency, maneuverability, and stability. While most of the special features of the ship — wide-range power plant, cable storage and handling facilities, etc. — are the result of its special function, a great deal of attention has been paid to providing optimum working conditions for, and safety of, the ship's complement.

I. INTRODUCTION

1.1 *Factors Affecting Ship Design*

The primary purpose of the new Cable Ship *Long Lines* was to lay a new type of cable system. Its secondary, but scarcely less important, purpose was to be capable of repairing all existing and new cable systems, both telephone and telegraph. The ship then, a tool in itself, had to carry other tools or hardware to do the job for which it was designed. First in importance were the cable tanks, the number and size depending on the amount and type of cable to be carried, then the means of payout for laying, and last the type of pickup for repair work.[1,2]

Speed and range were developed when the owner specified the extent of the ship's run, its longest service steaming time, and its time on the so-called "cable grounds."

Since the *Long Lines* was to be at sea for long periods of time and carry a multimillion-dollar cargo, great emphasis was put on the reliability of the ship and its component parts. All machinery and equipment used were designed to operate satisfactorily with a momentary roll of 30 degrees to either side or a pitch of 10 degrees, a permanent list of

* All of Gibbs & Cox, Inc., New York, N.Y.

15 degrees to either side, and a permanent inclination of 5 degrees in either direction fore and aft.

Various agencies whose regulations had to be considered in the design, construction, and installation of all its parts, equipment, and outfit are listed in Appendix A. This listing shows that, in addition to the owner's specifications, regulations had to be followed with regard to the construction of the ship itself, the machinery, auxiliaries and outfit, the safety of the ship in the event of damage by flooding or fire, and the safety of the crew and those engaged in loading the ship.

1.2 *Contract Specifications and Plans*

When the cable and repeater laying techniques had been developed and test facilities and other requirements determined, authorization was given by the Long Lines Department of the American Telephone and Telegraph Company to proceed with the preparation of bidding data for a new cable ship. These data included the preparation of detailed ship specifications incorporating the requirements for the ship itself, its detailed characteristics, full machinery and electrical requirements, outfit, accommodations, messing, and decoration.

In addition to these specifications, approximately 30 contract plans and numerous other detail plans were prepared as further guidance to the prospective contractor. These included ship construction, ship arrangements, and plans peculiar to a cable ship, viz., bow sheaves, electronic spaces, jointing room, drum room, etc.

1.3 *Opening of Bids and Contract Award*

All of the foregoing bidding data were incorporated as a package unit and transmitted to shipyards in several countries who had expressed a desire to bid on the ship. A period of eight weeks was allowed for the preparation and submittal of the bids to the owner. The contract was awarded to Schlieker Werft of Hamburg, West Germany, who submitted the lowest bid.

Owing to the unfortunate circumstances of bankruptcy proceedings against Schlieker Werft when the ship was about ninety per cent completed, it was taken out of this yard and completed at Deutsche Werft, Hamburg, West Germany.

II. DESIGN

The design of a special-purpose ship such as a cable ship is complex and requires consideration of many factors not usually encountered in

cargo ship design. The discussion that follows has been limited to areas of particular importance in cable operations.

2.1 *Selection of Hull Form*

A cable ship must be a stable working platform and have good sea-keeping qualities. It must operate at a variety of speeds and with varying loads. Consequently, considerable attention must be paid to the hull form.

A ship that is to be operated at high speeds is designed with relatively fine lines, while a ship operated at lower speed can have much fuller lines. This is because at low speeds the frictional resistance, which depends on the area of the ship below the water line, is most important, but at higher speeds the wave-making resistance increases in importance; it is desirable to reduce this as much as possible by fining down the ends of the ship. The fineness of a ship is indicated by the block coefficient,[3] which is the volume up to the designed water line divided by the product of the length, beam and draft of the circumscribing block, thus:

$$C_b = \frac{V}{L \times B \times d}$$

where

V = volume of displacement in cubic feet
L = length of ship at load waterline in feet
B = extreme immersed breadth in feet
d = draft of water in feet
C_b = block coefficient (or coefficient of fineness).

The block coefficient is selected fine enough to give a reasonably low resistance when the ship is propelled at the design speed, but high enough to avoid an excessively large and costly ship. In selecting the block coefficient, the speed of the ship considered is not the actual speed but a speed in relation to the length of the ship, namely the design speed divided by the square root of the length. A longer ship can therefore be made fuller than a shorter ship for the same speed.

The heavy weights of the bow sheaves and the cable engines located at the ends of the vessel increase the moment of inertia and have a tendency to increase the pitching motion. This was considered in the development of the lines in order to reduce this adverse effect as much as possible.

2.2 *Displacement and Light Ship Weight*[3]

The displacement of a ship is the total weight of the water displaced by the ship, and is equal to the weight of the ship itself (light ship weight) and all weights carried on board. The full load displacement is the light ship weight plus the deadweight. The deadweight is the total of the variable weights on board.

The light ship weight can be divided up into the weight of the hull structure, equipment and outfit, and machinery. In the case of the *Long Lines*, these weights and the deadweight are subdivided as follows:

(1) hull structure, the weight of the steel in the hull;

(2) equipment and outfit, the weight of all joiner work and furniture, all machinery for handling the cable, steering gear, anchors and cable gear, boats and all other equipment needed for operating the ship, including the electric generating plants for the ship's use, and the approximately 62 miles of electric cable in the ship;

(3) machinery, the weight of all machinery and auxiliaries needed for the propulsion of the ship at the required speed; and

(4) deadweight, the weight of the cable, water in cable tanks, spare cable, cable gear (such as buoys, grapnel, rope, etc), crew, stores, fresh water for ship's use, reserve feed water for boilers, and fuel oil for the specified cruising range, part at full speed, part at cable laying speed.

With the exception of the first few items of deadweight that are the owner's requirements, all other weights are interrelated and depend on the size of the vessel. The procedure, therefore, is to select a family of ships of different displacements with proportions and coefficients based on data from other ships. Preliminary rough estimates of the weight of hull, equipment and outfit, horsepower required and weight of machinery and fuel oil for each of these ships are then made. Based on these estimates, dimensions of a hull able to carry the required payload are selected for further consideration. Arrangements are then drawn up to check that sufficient space is available for cable tanks, cable handling, machinery, crew accommodation, etc. Based on these arrangements, additional checks are made of the weight, and adjustments are then made of the dimensions.

2.3 *Stability*

In selecting the dimensions of the ship, the operational stability of the ship has to be considered. The stability depends on both the vertical location of the center of gravity of the loaded ship and the beam of the vessel. With a given center of gravity, the beam must be wide enough

to avoid excessive heel due to high wind abeam or excessive angles of roll in a seaway. On the other hand the beam must not be too wide, since this may result in an excessively fast period of roll and violent motions when rolling, making it hazardous to move around on the deck.

When the ship is floating upright as shown in Fig. 1(a), the weight of the ship and its contents, W, acting downwards through its center of gravity, G, is opposed by an equal buoyance force acting upwards through the center of buoyancy, B, and the ship is in equilibrium. The buoyancy is the sum of the vertical components of the water pressure acting on the outside of the hull, and is equal to the weight of the water displaced by the ship. The center of buoyancy is the center of gravity of the underwater volume of the ship.

If the ship is heeled over by an external moment, the buoyancy provided by the wedge A in the upright position is replaced by the buoyancy provided by the wedge A_1. As a result the center of buoyancy moves from B to B_1 and the weight and the buoyancy are no longer in line but provide a couple, $W \cdot \overline{GZ}$, tending to bring the ship back to the upright position, as shown in Fig. 1(b).

The force acting through the center of buoyancy intersects the centerline of the ship at a point M. This point is called the metacenter. It is constant for moderate angles of heel at a particular water line.

The location of the metacenter for different drafts depends only on the width and properties of the water line and the displacement at that draft, and can readily be calculated.

Since the stability of the ship depends on the location of the center of gravity in the different loading conditions, careful calculations of the center of gravity of the light ship and all items of deadweight are made in

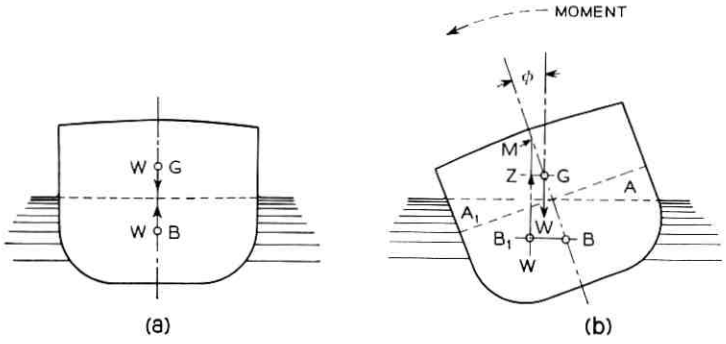


Fig. 1 — Factors affecting stability: (a) equilibrium state; (b) heeling.

order to be able to calculate the center of gravity and the stability for each loading condition.

The freeboard, or the distance between the water line and the deck at side, is also an important factor in the safety of the ship, since it determines the maximum angle of heel the ship can take without losing stability and increasing the chance of capsizing.

The righting moment is $W \cdot \overline{GZ}$ where $\overline{GZ} = \overline{GM} \sin \phi$. $\overline{GM}$ is referred to as the metacentric height, and is a measure of the stability of the ship.

2.4 *Model Testing*

When the dimensions and coefficients of the ship have been selected, a preliminary set of lines defining the shape of the hull is drawn, and a scale model of the ship's hull is built for testing.

This model is towed in a tank to determine the resistances of the hull at different speeds. The results of these tests are compared with the results of other model tests to determine that the resistance is reasonably low. Often several sets of lines are tested in order to obtain the best possible lines.

After completion of this towing test for bare hull resistance, the hull is tested with appendages added such as bilge keels* and rudders, and finally a self-propelled test is conducted to determine the interaction of the hull and propeller, the propulsion coefficient, and the actual horsepower required to drive the ship at different speeds.

In order to insure that the flow around the hull is as smooth as possible, especially near the propellers, where an unsatisfactory condition could lead to hull vibrations, the models are frequently tested in a circulating water tank. In this tank the water is moved in a uniform stream at different speeds by large pumps and the model remains stationary. Tufts are attached directly to the hull or to small pins a short distance from the hull. Windows are provided in the side and bottom of the tank, permitting observation of the direction and steadiness of the flow as indicated by the tufts. In addition, small quantities of dye are injected in the water at different points along the hull to indicate the flow. The model is observed both with and without the propeller running at the correct rpm for the relative speed of the ship and water, in order to observe the effect of the propeller on the flow around the model.

For *Long Lines*, preliminary tests for bare hull and appendaged hull

* A bilge keel is a long fin running fore and aft and fitted to the hull, port and starboard sides at the turn of the bilge, to reduce rolling. In the case of the *Long Lines* these bilge keels were designed with a deeper than usual profile and a shorter longitudinal extent in order to minimize resistance and increase damping to effect a steady cable laying platform.

(with bilge keels attached to the hull, bossings and rudders) resistance were conducted at Stevens Institute of Technology using a scale model of wood, approximately five feet long.

In addition to these still water tests, the Stevens Institute Towing Tank conducted tests on the 5-foot model to evaluate the behavior of the ship in waves. Two types of long-crested irregular "seas" were chosen, a moderate sea with a significant height* of 19 feet generated by a 28-knot wind, and a steep sea with a significant height of 28 feet generated by a 33-knot wind. Test conditions were chosen to simulate full-scale operation when laying cable over the stern at slow speeds and repairing cable over the bow at zero speed. For each sea type, ship speeds were varied from zero to 14.6 knots for head seas, zero to 11.6 knots for following seas and zero speed for bow, beam and quartering seas. The model demonstrated superior seakeeping qualities under all conditions, and it was concluded that cable operations could be conducted under all conditions except steep head seas for ship speeds of about 10 knots and above.

On the basis of these test results, final ship lines were chosen for testing at the David Taylor Model Basin, Washington, D.C. Using a larger scale model, about twenty feet long, made of special wax, bare hull and appendaged hull resistance of the ship were tested. Then, with a suitably chosen set of stock propellers, the self-propelled tests were run. The deep water basin where these tests were conducted is 2775 feet long, 51 feet wide and 22 feet deep.

Lines of flow for bilge keel locations were also determined at the David Taylor Model Basin. After completion of the self-propelled test series, the fully appendaged model was checked in the circulating water channel at several displacements and under various operating conditions.

Final model tests were conducted by Schlieker Werft at the Hamburg Model Basin, Hamburg, Germany. These tests were all conducted at the full load displacement of the ship, and comprised an appendaged resistance test with several self-propelled tests. The self-propelled tests employed a series of propellers and formed the basis for the selection of important characteristics of the propellers designed for the ship.

2.5 *Inclining Experiment*

When the ship is essentially completed, the calculated light ship weight and the center of gravity of the ship are checked by the shipbuilder by means of an inclining experiment.

* The significant height of a wave system is the average of the highest one-third of the waves which pass during an interval.

The ship is moored in such a way that it floats freely with minimum restraint. The draft of the ship at the draft marks is measured carefully from open boats simultaneously on both sides, forward, aft and midship, and the specific gravity of the water is obtained from samples at several locations along the ship. From this information and the shape of the hull the displacement of the ship in the inclined condition is very accurately determined.

The ship is inclined by moving one or more weights from a location on the centerline of the ship, Fig. 2(a).

The weights are moved to several positions away from the centerline of the ship, first on one side, then on the other, and finally on the centerline. The distance of the weights from the initial position is carefully measured after each movement and the resulting heeling moment $w \cdot x$ (a moment tending to incline the ship away from the upright position) is calculated.[3]

The angles of heel ϕ for each position of the weights are obtained by measuring the deviation of three long pendulums in different locations in the ship. The tangents of these angles are plotted against the corresponding heeling moments to keep a check on the consistency of the readings, since theoretically all observed points should fall in a straight line through the origin of coordinates.

If the displacement of the ship is W, and the heeling moment is $w \cdot x$, then, from Fig. 2(b)

$$\overline{GG_1} = \frac{wx}{W} = \overline{GM} \tan \phi$$

$$\overline{GM} = \left[W \left(\frac{\tan \phi}{wx} \right) \right]^{-1}$$

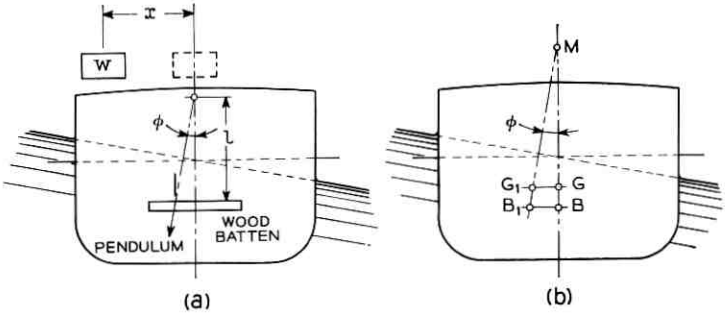


Fig. 2 — Inclining experiments: (a) placement of weights; (b) forces affecting height of center of gravity during heeling.

Taking the average value of (tan ϕ/wx) from the plot, the distance GM is calculated. Since the location of the metacenter M for the inclined condition is known, the location of the center of gravity of the ship in the inclined condition can be determined.

At the time of the inclining experiment the ship is completely surveyed, and the weight and location of all items needed to complete the ship is noted. All items aboard the ship not belonging to the light ship weight, such as inclining weights, liquids in tanks, tools, etc., are also recorded. By addition or deduction of the weight and moments of these items from the weight and moments of the ship as inclined, the light ship weight and center of gravity can be calculated.

Based on the inclining experiment, a stability book is prepared giving all information to the Master to enable him to load the ship to insure sufficient stability in all conditions of loading.

2.6 *Flooding and Damaged Stability*

The ship is divided into compartments by transverse watertight bulkheads spaced so that if any two adjacent compartments are flooded the ship will not sink. The ship loses stability when flooded, and it is possible that even if the ship would not sink, it could capsize due to insufficient stability or excessive heeling moments due to unsymmetrical flooding. The effect of the flooding of adjacent compartments on the stability of the ship is therefore calculated for different drafts, and the minimum initial stability that the ship requires before damage occurs is obtained. The stability required to prevent capsizing in case of damage of the most critical adjacent compartments for different drafts is tabulated in the stability book, and the ship must meet this requirement in all conditions of loading.

III. MANEUVERABILITY

Early in the conceptual period of *Long Lines* it was recognized that a high order of maneuverability and precise control of propulsion power were required to permit the full exploitation of the advanced cable handling techniques.

A cable ship should be capable of precise, rapid maneuvering at low speeds, such as is desirable in grappling and coming-up on a buoy, and should be able to maintain accurate station-keeping in adverse sea and weather conditions when standing on a cable. These requirements demand consideration of maneuverability aids and a propulsion system which is easily and rapidly controllable and can provide high propeller

torques at low speeds for rapid accelerations. Further, it is desirable that the propulsion plant be readily adaptable to remote control, in order to permit maneuvering the ship for cable handling from the most advantageous position. This approach eliminates the time lag and opportunity for misunderstanding inherent in systems in which orders are communicated through several people to the engine room.

3.1 *Maneuverability Aids*

3.1.1 *Number of Screws*

Twin screws versus a single screw were considered and selection was made in favor of twin screws.

3.1.2 *Rudders*

Twin spade rudders were selected and are located in the propeller race. They have an area 50 per cent larger than normally used for merchant vessels, and rudder angles of 45 degrees instead of the usual 35 degrees, thus providing sharper turning angles.

3.1.3 *Bow Jet*

A propeller-driven bow jet was decided upon as an effective aid in maneuvering with the vessel operating at very low speed. The propeller, 5 feet in diameter, is mounted in an athwartship tunnel near the bow and is driven by a reversible dc electric motor through a reduction gear. It is controlled from the wheelhouse and from forward and aft control stations.

IV. PROPULSION STUDIES

Various propulsion schemes were studied in the light of the ship requirements, and it quickly became apparent that a variable-voltage direct current system provided precisely the requirements desired. Its primary advantages are as follows:

(1) Very large torques (limited by shaft strength to 175 per cent of full load torque) are available at any propeller speed from zero to full speed. This is essential to provide the rapid propeller accelerations and reversals demanded by the precise ship maneuvering requirements. Propeller reversals from full ahead can be accomplished in less than two seconds.

(2) Remote control is easily accomplished by the provision of a simple rheostat and appropriate instrumentation at the remote location.

(3) Propeller speed is smoothly and continuously variable from zero to full in either direction.

(4) The ship will spend a great deal of its life at reduced powers during cable operations; electric drive permits available propulsion prime mover power during these periods to be used for other purposes, such as supplying cable engines and some of the ship's service electric power.

(5) Electric propulsion is inherently flexible, permitting shutdown of unused units during periods of low demand and easy cross-connection in the event of a casualty.

The major disadvantages of dc electric propulsion are its relatively high weight and somewhat low efficiency. The high weight is acceptable in return for the desired operating characteristics. The low efficiency is not very significant in this ship because the low propulsion power of the ship, combined with the high hotel and auxiliary power demand created by the working nature of the vessel, results in the propulsion power being a much smaller percentage of the total ship's power requirements than in most conventional ships. A large portion of the ship's operating time will be at very low propulsion power for cable laying (about one-eighth power), decreasing the importance of high propulsion efficiency.

4.1 *Selection of Prime Mover*

Having selected dc electric drive, there remained the question of the type of prime mover to power the generators. After preliminary studies diesel engine and steam turbine prime movers were selected for further study. Their respective advantages are as follows:

(1) Advantages of the diesel engine
   (a) lower first cost,
   (b) lower weight of fuel required,
   (c) lower machinery weight,
   (d) lower amount of fresh water required, and
   (e) smaller machinery spaces.

(2) Advantages of the steam turbine:
   (a) lower operating fuel costs,
   (b) lower maintenance and repair costs,
   (c) increased reliability,
   (d) less noise and vibration, and
   (e) the large number of seagoing licensed marine steam engineers available.

4.2 *Final Selection of Propulsion Plant Drive*

After consideration of all the facts, such as the heavy work schedule for the new cable ship, limited availability of licensed marine diesel engineers, reliability and minimum time out for maintenance and repair, and the problem of noise and vibration in the engine room and crew quarters, it was concluded that the propulsion plant drive should be of the steam turbine type.

TABLE I — PRINCIPAL CHARACTERISTICS OF C.S. "LONG LINES"

Length over-all	511′-6″
Beam	69′-6″
Load draft	26′-9″
Horsepower normal, total	7700 shp
Sustained sea speed	15 knots
Cable capacity ($1\frac{1}{4}$″ dia. cable)	2000 nm

V. OVER-ALL DESCRIPTION OF SHIP

The principal characteristics of the *Long Lines* are given in Table I. The hull is all steel and is designed with transverse framing except for

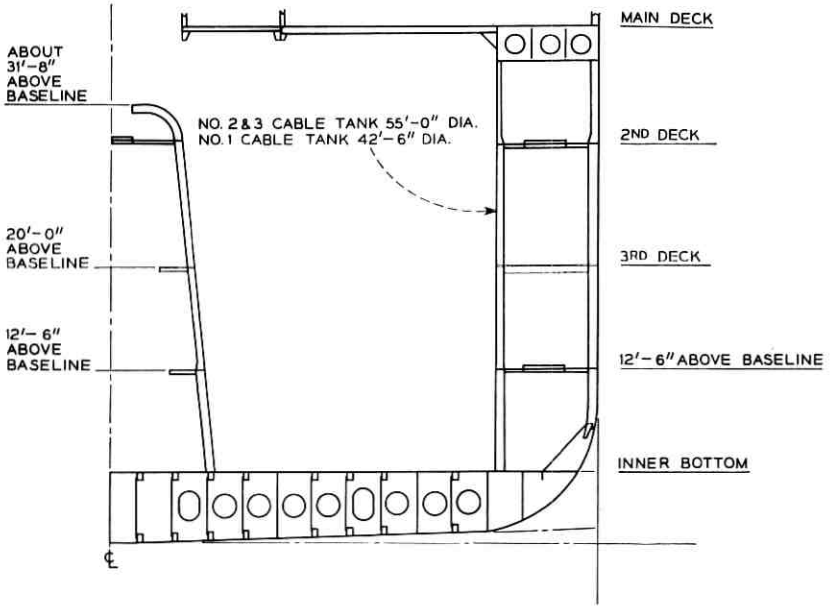


Fig. 3 — Typical cross section in way of cable tanks.

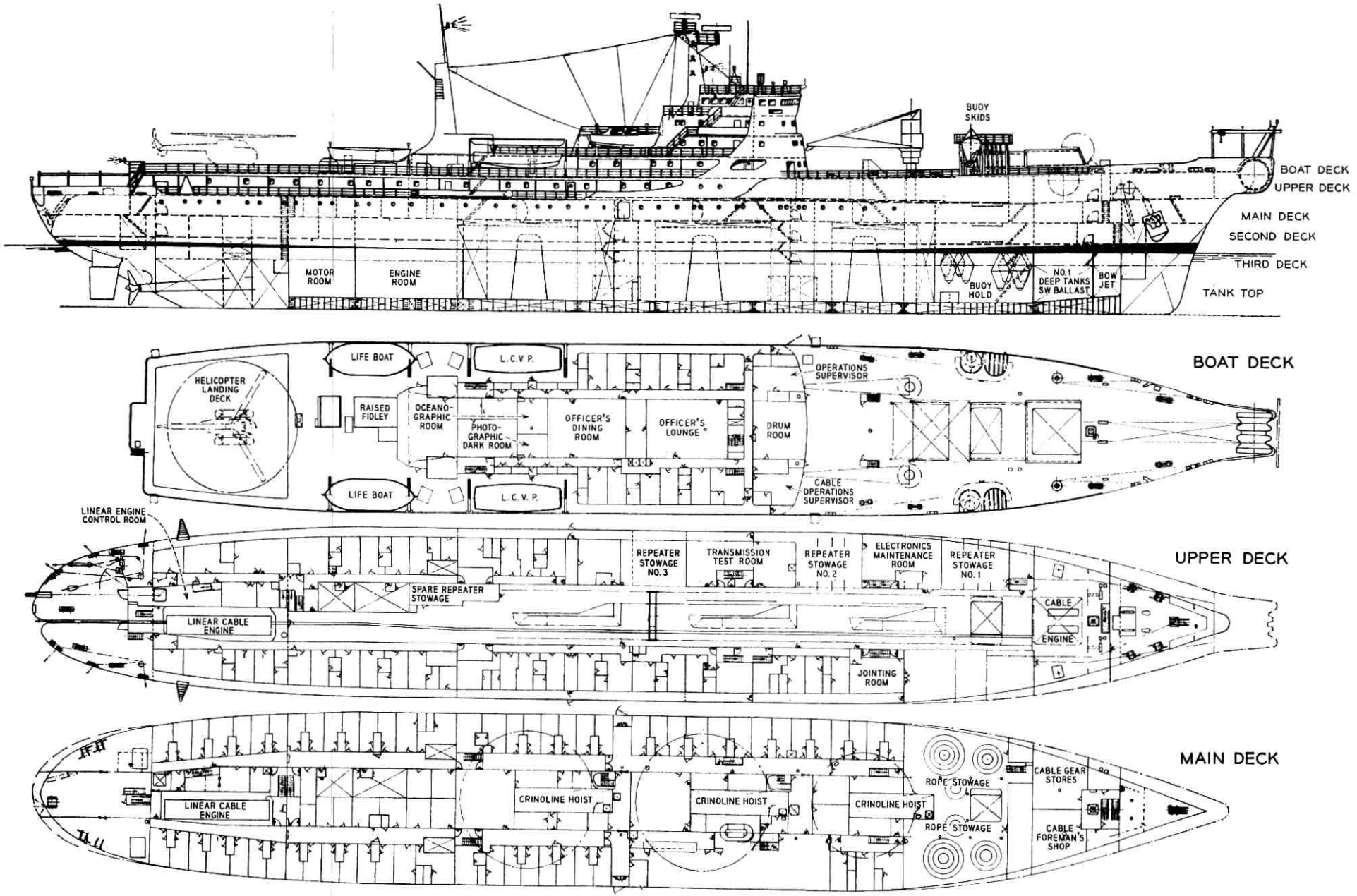


Fig. 4 — General arrangement of decks.

the inner bottom and bottom shell, which are framed longitudinally. It is of welded construction throughout except for certain shell seams and deck edge connections which are riveted. The shell is strengthened for navigation in ice. The propeller shafts also include an allowance for ice strengthening.

The three large-diameter cable tanks are located forward of the machinery spaces and provide approximately 138,500 cubic feet of space for cable storage. They are fitted with watertight cones and also with recesses for cable bights[1] at the forward end which extend for the full depth of the tanks. Fig. 3 shows the cable tank construction.

Ship service electric power and power for the linear cable engine are furnished by four ac generators, two driven by the propulsion turbines and two driven by independent steam turbines. Emergency service electric power is furnished by a diesel engine-driven ac generator set. Electric power for the bow cable engines is furnished by two dc generators driven by the propulsion turbines.

A helicopter deck, with its own lighting and with special safety rails around its periphery, is located at the aft end of the houses.

All joiner work throughout the ship, linings, bulkheads and ceilings, is of incombustible construction. Incombustible furniture and fire-resistant draperies, portieres, bedspreads, etc., are installed throughout as far as possible. All living accommodations, offices, lounges, wheelhouse, chartroom, radio room, and the principal working spaces are air-conditioned for personnel comfort and to provide a good environment for the large amount of electronic equipment.

A general arrangement of the decks is shown in Fig. 4.

VI. PROPULSION PLANT

The propulsion machinery is of the twin screw, steam turbine, direct current electric drive type.

6.1 *Design Conditions*

Design conditions for the propulsion plant are given in Table II. The layout of the main machinery spaces is shown in Fig. 5.

6.2 *Generator Train and Propulsion Motors*

Each generator train is driven through a single reduction gear by a 5000-hp steam turbine. Each train consists of the following, all driven in tandem at a constant 900 rpm by the propulsion turbine (see also Fig. 6):

TABLE II — PROPULSION MACHINERY DESIGN CONDITIONS

Shaft horsepower — normal, each shaft	3,850
Shaft horsepower — normal, total	7,700
Shaft horsepower — maximum, each shaft	4,250
Shaft horsepower — maximum, total	8,500
Sustained sea speed (at 80 per cent of normal power at full load draft)	15 knots
Steam conditions at boiler superheater outlet	600 psig — 850°F
Full fuel capacity (half cruising at full power 15 knots and half cruising 7.5 knots)	14,350 nm

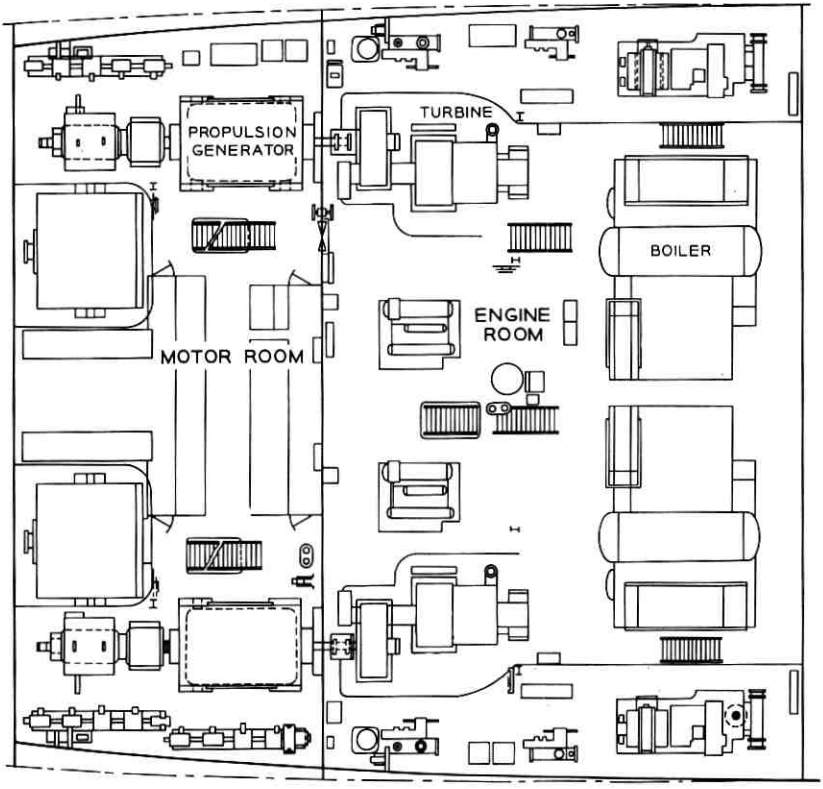


Fig. 5 — Layout of main machinery spaces.

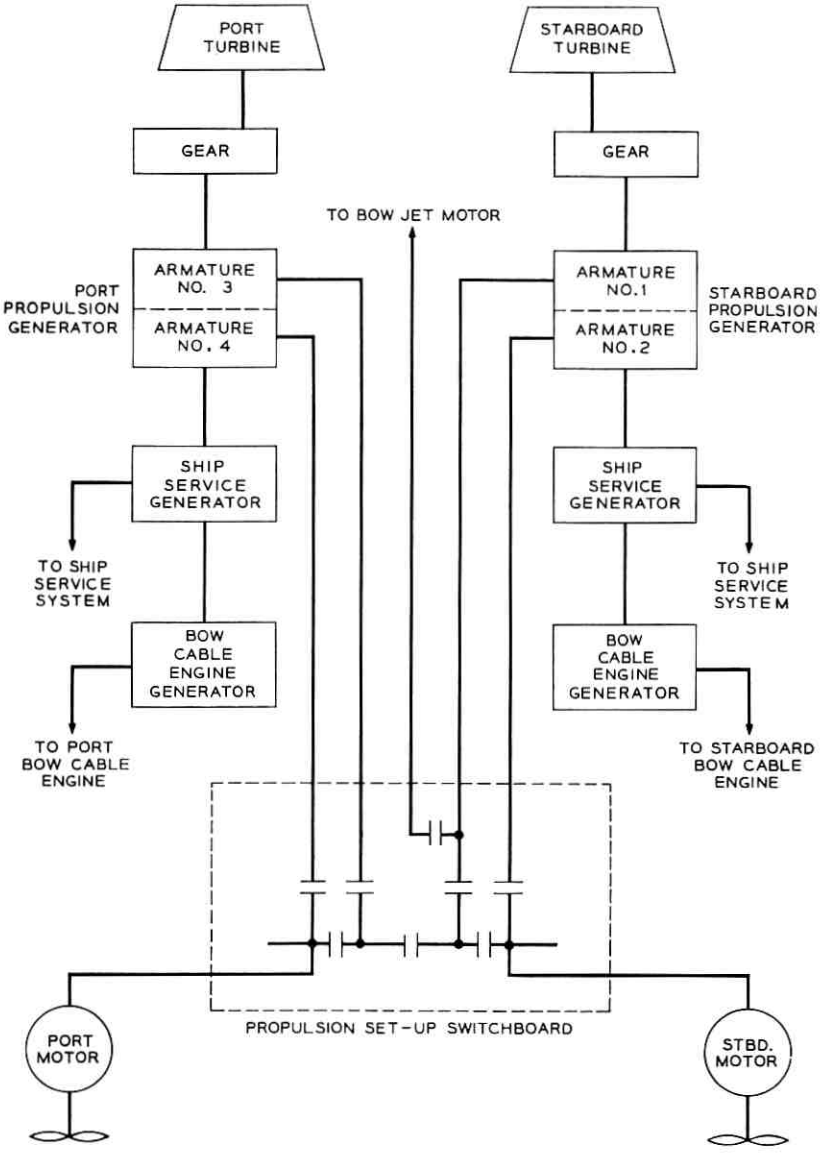


Fig. 6 — Generator train.

6.2.1 *DC Main Propulsion Generators*

The two armatures in each generator are completely independent electrically. The generator is totally enclosed, water cooled and force ventilated by a separately driven cooling blower.

6.2.2 *Ship Service Generators — 3-phase ac*

The machine is open and air cooled. This generator can be utilized to supply some of the ship's service power during periods when the ship is not operating at full propulsion power.

6.2.3 *Bow Cable Engine Generators*

This generator, rated 480 kw, 800 amperes at 600 volts, is totally enclosed, water cooled, and force ventilated by a separately powered 3-hp blower. Because of the infrequent service it is also provided with an externally operable brush-lifting device to lift the brushes off the commutator when the generator is not in use.

6.2.4 *Propulsion Motors*

Each propulsion motor is rated 4250 horsepower at 135 rpm. It is separately excited, with field weakening to reach 147 rpm in order to utilize full available power in the light ship condition. The motors are totally enclosed, water cooled and force ventilated by a separately powered 30-hp blower.

6.2.5 *Cooling of Units*

All propulsion motors and generators are kept slightly pressurized by means of a small make-up air fan. This fan supplies clean, electrostatically filtered air to the inside of the machines and maintains a slight positive pressure so that any leakage will be out of the machine. Thus the machine interiors are protected from engine room air contaminants.

6.2.6 *Control and Switching*

The basic principle of control is motor speed control by variable armature voltage with constant motor field. The manually operated speed control rheostat works in a control field of an amplidyne, the output of which supplies generator excitation. Since generator speed is held constant, this varies armature voltage, thus controlling motor speed. The system is speed regulated by means of generator output voltage

feedback, and current and power limiting controls are incorporated to prevent overstressing or overloading the mechanical portions of the system. In addition, the necessary switching, interlocks and safety devices are incorporated in the control to take advantage of the flexibility available, yet prevent malfunction. One of the switching arrangements provides the ability to connect one of the propulsion armatures to the bow jet motor, thus providing 750 hp to the bow jet without the need for a separate generator. The control components and switchgear are contained in a propulsion set-up switchboard and a propulsion control console, both located in the motor room. In addition, there are propulsion control consoles in the bow of the ship, near the bow sheaves; in the wheelhouse, port, centerline and starboard; and in the stern of the ship, in view of the overboarding chute.

6.3 *Main Turbines*

The main turbines are of the high-speed, single-casing type and are connected to the main reduction gears through a flexible mechanical dental-type coupling. The port and starboard turbines are identical and have the same rotation. The turbines are designed to deliver the normal rated horsepower at the point of best economy with initial steam condition at the turbine throttle of 585 psig and 840°F total temperature and a vacuum of 28.5 inches Hg at the exhaust. The turbines are also designed for uncontrolled bleed steam for feedwater heating, boiler combustion air heating and other purposes.

6.4 *Reduction Gears*

The reduction gears are of the single-reduction, double-helical type, arranged to allow the main condensers to be mounted athwartship under the turbine.

6.5 *Shafting*

The shafts are solid forged steel, American Bureau of Shipping Grade No. 2 material.

VII. ENGINE AND MOTOR ROOM AUXILIARIES

7.1 *Distilling Plant*

Since the ship will be at sea for long periods of time, all the fresh water required on board is produced by two complete sea water distilling plants

of the flash type. Each distilling plant is designed to meet the fresh water purity requirements of the United States Public Health Service, and each unit has a capacity of 12,000 gallons per day of fresh water of potable quality. One distilling plant can provide all the fresh water requirements for drinking water, hot and cold washing water, hot and cold fresh water for galley and pantry service (cooking and dishwashing), laundries, and make-up water for the boilers.

7.2 *Boilers*

Two natural circulation water tube boilers are installed, and each boiler is of the bent tube, oil-fired, single-furnace, single-uptake, air-encased type with water-walled furnaces. Each boiler is fitted with double-cavity, walk-in type superheaters, economizers, steam air heater and submerged internal type desuperheaters installed in the steam drum. These boilers produce superheated steam at 600 psig, 850°F at the superheater outlet.

The boilers are fitted with steam-mechanical atomizing fuel oil burners suitable for use with the automatic combustion control system. These burners combine the features of steam and mechanical fuel oil atomization. The system is unique in that the burners are used in two ways. They provide straight mechanical atomization (no steam assistance) of the fuel burned when the ship is operating at full power. In addition, they function as a steam atomizing burner (steam assisting in the atomization of the fuel oil) at low steaming rates, as during periods of reduced propulsion steam demands (cable laying and cable repair). This second function results in reduced fuel consumption when the straight mechanical atomizing efficiency falls. The primary advantage of the steam-mechanical atomizer is that good combustion efficiency over a wide range of operation is obtained with minimum steam consumption and a low fuel oil system operating pressure. The importance of wide-range burners may be appreciated when considering that the ship will operate for extended periods at light steaming loads, as during cable laying, cable repairing, maneuvering and in-port operation. These burners also have the advantage that the furnace can be kept clean under low steam demand operation of the vessel, because more efficient steam atomization of the oil results in more complete combustion.

The boilers are fitted with a pneumatic-type automatic combustion control system designed to maintain constant superheater outlet pressure. The combustion control system attains maximum fuel economy by maintaining the boiler steam pressure and temperature at the highest

design conditions by controlling the quantity of oil and air being delivered to the boiler in the proper ratio to maintain maximum combustion efficiency. Maintaining the proper steam pressure and temperature insures maximum turbine performance. The control system will follow the fluctuations in the boiler loads due to changes in the ship's speed or electrical load changes, at the same time maintaining maximum fuel economy with a minimum of smoke.

VIII. SHIP'S SERVICE REFRIGERATION MACHINERY

A ship's service refrigeration system is provided and designed for the direct expansion of Freon 12 gas. Since the ship will be at sea for long periods of time, approximately 8000 cubic feet of refrigerated stores have been provided.

IX. PIPING

All piping installed in the ship was designed to meet the highest marine piping standards used in the U.S.A. Corrosion-resistant materials were used to reduce maintenance and repair costs and to obtain long life in service. All piping carrying sea water is of 90-10 copper-nickel alloy. This alloy is very resistant to the corrosion-erosion effect of sea water. Stainless steel has been used in many of the other piping systems. There are approximately 12 miles of piping installed in the *Long Lines*.

X. SAFETY AIDS

10.1 *Watertight Doors, Firescreen Doors*

Five watertight horizontal sliding doors have been installed at vantage points in some bulkheads to maintain the watertight integrity of the ship. These doors, normally open, are electrically operated and can be controlled collectively or individually from the wheelhouse. There are local door controls on both sides of each door.

Approximately 48 firescreen doors, normally held open by magnet control, can be operated from the wheelhouse and locally; about 30 more firescreen doors, normally closed, are operated locally. All of these are installed in the firescreen bulkheads for protection against fire.

10.2 *Fire Detection System*

A thermostatic detection system is provided to sound an alarm in the event of a fire. Detectors are installed in all public spaces, lockers, holds

and similar unattended spaces. Manually operated alarm boxes are provided in all fire zones in readily accessible locations. An annunciator panel, which indicates the area originating an alarm, is installed in the wheelhouse, and alarm bells are provided in the wheelhouse, engine room and quarters of the firefighting crew.

10.3 *Firefighting Systems — General*

The vessel is equipped with the following firefighting systems:

(1) sea water extinguishing system for use in fighting fires in ordinary combustible materials, such as dunnage, canvas, wood, etc,

(2) carbon dioxide extinguishing systems and liquid foam extinguishing systems for use in fighting fires in substances such as gasoline, oils, etc., and fires in live electrical equipment, and

(3) steam smothering extinguishing system for limited use in fighting fires in substances such as fuel and diesel oil.

XI. BOW SHEAVES

An assembly of three roller-bearing bow sheaves is mounted in the overhang of the stem and enclosed within a supporting structure fairing into the forward body of the ship. The center sheave has a diameter of 11 feet, and the two outboard sheaves are each 10 feet in diameter at the root. The ring and hub of each sheave is of cast steel, and the sheave body is made up of steel weldments. The shafting is in three sections of different diameters to permit easy installation and removal. Automatic lubrication for the bearings has been provided.

For handling forward, a monorail and hoist have been provided extending over the bow sheaves. This is an impressive looking structure, since it includes a frame of heavy scantlings for overboarding repeaters at the bow. The frame can be fitted with a vinyl-coated nylon fabric covering for protection in severe weather.

XII. STERN CHUTE

Instead of a conventional sheave at the stern, a stern chute has been employed. This chute is located to starboard of the centerline of the ship and is formed by a 16-inch wide indent in the upper deck plating at the extreme aft end. This indent gradually falls away and fairs into the shell plating at the stern.

XIII. ALUMINUM ROLLING HATCH

A power operated aluminum rolling hatch about 50 feet long and constructed of two nesting sections is located over the bow cable engines. This construction coupled with rolling end flaps, also power operated, permits continuous bow cable operations in adverse weather.

XIV. SPECIAL STOWAGE

Stowages for various types of apparatus and equipment peculiar to the functions of cable laying have been provided and are located at vantage points throughout the ship.

Buoys are stowed in a large hold, two decks high, located forward of the cable tanks. This stowage is conveniently arranged so that the buoys can be transferred by a minimum amount of handling with a monorail and hoist device into the center of the hold; the hoist device picks up each buoy from its stowage position and, through a self-releasing hook, transfers the buoy to the deck crane hoist and from there to its "ready" stowage on the port or starboard buoy skids, or directly overboard. These buoy skids have been especially designed with overhead platforms to facilitate control in launching and positioning the buoys.

Stowages for grapnels, mushroom anchors and similar heavy equipment, together with chains and buoy rope, are located in the same hold as the buoys, but in the two tween decks directly above the buoy stowage. The chain is stowed in bins having a raised floor in portable sections, all to permit easy maintenance. Chain pipes are installed port and starboard, extending from the upper deck to the stowage on the third deck. By means of swiveling devices at the lower ends of these chain pipes, the chain can be run direct into each of the seven chain stowage compartments.

A total of four tanks, each 20 feet in diameter by 10.5 feet high, for the stowage of spare cable, are installed on the tank top level, port and starboard, between the main cable tanks. Twelve circular steel bins, all 4 feet high but of diameters varying from 4 to 16 feet, are provided for grapnel rope stowage. New rope is stowed on one side of the ship and used rope on the other side.

XV. NAVIGATION EQUIPMENT

In considering the nature of the mission of the *Long Lines*, it becomes apparent that it is of utmost importance that navigational facilities be

provided that will determine the ship's location with the maximum possible accuracy at all times during cable laying or pick-up and repair operations.

Standard radio direction finding equipment, as required for certification by the United States Coast Guard, has been provided for taking fixes on land-based telegraph stations and beacons and on other ships at sea.

Loran equipment installed is of the latest type, known as Loran "C." This system provides greater range and greater accuracy than the standard Loran "A". It also has Loran "A" capabilities built-in for use in areas where Loran "C" signals are not available.

Decca Navigator equipment is installed in the chartroom. This equipment, operating in an area of Decca signal coverage, provides a continuous instantaneous indication that will give the ship's location by reference to Decca charts.

Two Decca true-motion radar sets are installed with indicators in the wheelhouse. True-motion presentation shows a fixed geographic area on the indicator with all moving objects, including own ship, moving across the displayed area, making it easier to plot vessel speeds, approach paths, etc., than is possible on the standard relative motion display. One radar, operating on a 10-cm wavelength, is superior at long range and in penetrating adverse weather conditions. The other radar is a 3-cm set that exhibits superior qualities in close navigation, as in rivers, harbors, buoyed channels and in close proximity to other vessels.

In order to provide correlation with charted depths and to furnish accurate depth information regarding a cable lay, two echo depth sounders are installed. Each sounder is provided with alternate transducers, one set forward and one aft. Kelvin-Hughes sounding equipment is provided in the chartroom for navigational purposes, and a duplicate indicator is installed in the drum room. The second echo sounder, located in the drum room, is an EDO instrument. In addition to the two echo depth sounders, there is a Westrex Model XV Precision Depth Recorder that is capable of extremely accurate depth recordings in conjunction with either the Kelvin-Hughes or EDO equipment in either the drum room or chartroom.

In addition to the foregoing, there is a standard Sperry Gyro Compass system. Steering control may be manual or by gyro pilot, using either of the dual controls in the wheelhouse, or manual from the bow or stern ship control console or, in emergency, from the trick wheels in the steering gear room.

A Pitometer Log is provided for showing the ship speed in the chart-

room and wheelhouse and the draft of the vessel in the wheelhouse. A Chernikeeff Log System is provided to show ship speed in the drum room and to furnish input of ship speed to the true-motion radars.

XVI. EXTERIOR COMMUNICATION FACILITIES

A standard radiotelegraph installation has been installed for ship-to-shore and ship-to-ship communications, as is customary on all vessels of this size.

In addition, to provide optimum communications between the ship and the shore stations during her long periods at sea, a 5-kw single-sideband transmitter has been provided. As a back-up for this unit, and for use when maximum power is not required, a 1-kw single-sideband transmitter is provided. Compatible receiving equipment is installed, and interconnecting facilities are provided to permit use of the radiotelephone equipment from selected stations of the ship's automatic telephone system.

Low-power radiotelephone equipment is provided for use in harbor communications and VHF frequency-modulated units are provided for communication with the cable work boats or for communication with similarly equipped stations on other ships or on shore.

XVII. INTERIOR COMMUNICATION FACILITIES

Communication between key stations for ship operation is provided by sound-powered telephone, since the Coast Guard requires that such communication be independent of any other system on the ship. Normal communications facilities are provided by an automatic dial telephone system of latest design which incorporates private tie-lines between selected points in addition to dialing facilities.

Two public address systems are provided, one for general ship's use and the other for communication between functional stations of the cable operations system.

XVIII. SHIP'S ILLUMINATION

Conventional marine illumination standards, as set forth in the Illuminating Engineering Society's "Recommended Practice for Marine Lighting," [4] were applied to all interior living spaces, work spaces, machinery compartments and exterior areas. The standards were modified to provide special illumination treatment to suit the function of each area, since it was important that good lighting always be available.

Recreation, berthing, messing and office spaces were illuminated to a standard employed for first-class passenger ships.

The floodlight installation on the forward open deck was designed to facilitate cable handling operations at night and includes special fixtures to illuminate the exposed cable from the deck to the water surface.

XIX. SEA TRIALS

On March 6, 1963, C.S. *Long Lines* sailed from Hamburg to the North Sea for sea trials, and returned on March 12, 1963.

These trials included the usual proving of all systems: speed and economy trials; the adequacy of the control system under all conditions of operation; steering and maneuvering tests, including ahead and astern steering, Z maneuvers and turning circles; and special steering trials to determine the maneuverability of the ship when laying and repairing cable.

All the trials and tests were successfully conducted and the ship was remarkably steady and free from vibration.

APPENDIX A

Various agencies whose regulations were required to be met in the construction and installation of all parts, equipment and outfit of C.S. *Long Lines*, include the following:

(1) American Bureau of Shipping to the highest class,
(2) United States Coast Guard,
(3) United States Public Health Service,
(4) Federal Communications Commission,
(5) United States Customs,
(6) Panama Canal Company,
(7) Suez Canal Authorities, and
(8) International Convention for the Safety of Life at Sea, London, 1948.

In addition to the foregoing, it was required that the recommended methods, materials and practices contained in the latest issue of the following published standards be followed in the construction of the ship:

(1) American Institute of Electrical Engineers, Standard No. 45,
(2) American Standard Safety Code for Elevators, Dumb-waiters and Escalators,
(3) American Bureau of Shipping recommendations for reducing stress concentration at keyways in propeller shafts,

(4) Code of Recommended Precautions Against Accidents connected with the loading and unloading of Merchant Vessels,

(5) American Society of Refrigerating Engineers Recommended Practice for Shipboard Installation, and

(6) Society of Naval Architects and Marine Engineers Ship Trials and Test Codes.

REFERENCES

1. Grismore, O. D., Cable and Repeater Handling System, B.S.T.J., this issue, p. 1373.
2. Gretter, R. W., Cable Payout System, B.S.T.J., this issue, p. 1395.
3. *Principles of Naval Architecture*, **1**, The Society of Naval Architects and Marine Engineers.
4. *Recommended Practices for Marine Lighting*, AIA File No. 31.

Contributors to This Issue

C. J. Altenburg, M.E., 1934, Stevens Institute of Technology, Gibbs and Cox, Inc., 1936—. He is a senior engineer and assistant department head in the engineering department responsible for design of the C.S. *Long Lines*. Member, Society of Naval Architects and Engineers, American Society of Naval Engineers, National Association of Corrosion Engineers, American Welding Society and A.S.T.M.

J. D. Bishop, B.S. in E.E., Ohio University, 1953; Bell Telephone Laboratories 1953—. He has been engaged in circuit design of a broad range of power supplies and power systems for specific applications. These include microwave systems, carrier systems, electronic switching systems, data systems and military projects. He currently is in charge of a power supply circuit design group. Member, IEEE, Audio Engineering Society and Tau Beta Pi.

Miles W. Bowker, Bell Telephone Laboratories, 1940—. Mr. Bowker's early work was concerned with the development of outside plant apparatus, construction methods, and maintenance methods. This work included the development of the first solderless splice for polyethylene insulated coaxial cable and the original feasibility studies on the use of gas pressure in the exchange plant. From 1956 to 1960, he contributed to the design of, and the manufacturing feasibility studies for, the armorless ocean cable. From 1960 to 1963, he was concerned with the fabrication and placing methods for experimental circular waveguide. Currently, he is involved in analytical and experimental engineering mechanics studies covering a variety of problems, including infrared transmission, heat transfer, servomechanisms, statics and dynamics, and nuclear weapon effects.

S. Theodore Brewer, B.S. in E.E., 1937, M.S. in E., 1938, Purdue University, Bell Telephone Laboratories, 1937—. In his early assignments, he contributed to the development of broadband coaxial systems and video feedback amplifiers, including the design of measuring equip-

ment associated with these developments. Later, he was concerned with electronically controlled automatic switching. He holds patents on control and feedback systems, switching networks, and repeater circuits. In the SD system development, he was in charge of a group responsible for the electrical design of the undersea repeater. Currently, he is in charge of the repeater circuit design for a high-capacity transistor submarine cable system. During World War II, he served as radar staff officer with the 62nd Fighter Wing. Member, IEEE, Eta Kappa Nu, Tau Beta Pi and Sigma Xi.

JOHN H. BUTLER, Gibbs and Cox, Inc., 1933—. His work with this firm has included participation in the design of the C.S. *Long Lines*. Member, Society of Naval Architects and Marine Engineers.

FRANK R. DICKINSON, B.S.E.E., 1927, Union College; Bell Telephone Laboratories, 1931—. His early work at Bell Laboratories was in the engineering of trial installations of new telephone equipment, followed by analysis work on current engineering problems. From 1936 until 1954 he was engaged in mechanical design of carrier equipment units for C, J, and L systems, except for the period during World War II when he was involved in the mechanical design of airborne radar bombsight units. In 1954 he became supervisor of a group responsible for the mechanical design of repeaters for ocean cables and releated problems, and is currently involved in design of a new broader-band repeater. Member, Eta Kappa Nu.

R. D. EHRBAR, B.S.E.E., 1937, Johns Hopkins University; Bell Telephone Laboratories, 1937—. He first worked on K and L-type carrier equipment. During World War II he participated in the development of various airborne radar systems. After the war he worked on development of the L3 coaxial cable system. In 1955 he became Head, Submarine Cable Systems Department, and has since been associated with various submarine cable projects. Senior member, IEEE.

JOHN M. FRASER, B.S. in E.E., 1945, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1934—. His early work dealt with the design and evaluation of communication systems for both the military and the Bell System. Since then he has been engaged in transmission system engineering on a variety of carrier systems, including submarine cable systems. He is currently responsible for a group working on a new

TASI, the SD submarine cable, and new terminals to improve high-frequency overseas radio transmission. Senior member, IEEE; member, Sigma Xi, Tau Beta Pi and Eta Kappa Nu.

R. W. Gretter, S.B. in M.E., 1950, S.M. in M.E. 1951, Mech. E. 1953, and Sc.D. in M.E. 1956, Massachusetts Institute of Technology; Bell Telephone Laboratories 1955—. Initially Mr. Gretter did analytical work in cable mechanics. This was followed by participation in the development of cable machinery for C.S. *Long Lines*. He now supervises a group which is responsible for cable handling equipment and for some aspects of the submarine cable burier. He is a licensed professional engineer and a member of Pi Tau Sigma, Tau Beta Pi and Sigma Xi.

O. D. Grismore, B.S.E.E., 1927, Purdue University; American Telephone and Telegraph Co., 1927–1934; Bell Telephone Laboratories, 1934—. He first worked on field measurements of inductive interference from power systems and electrical railways; following this he became concerned with field trial installations and equalization studies of coaxial carrier systems. During World War II he worked on weapons, high-altitude radar bombing, and field studies of radar target identification. After the war he resumed work on coaxial carrier systems. His work with submarine cable systems started with the transmission aspects of cable manufacture and has continued in cable handling operations on cable ships. Registered professional engineer; member, Eta Kappa Nu and Tau Beta Pi.

V. Lyman Holdaway, B.S., 1929, M.S., 1930, California Institute of Technology; Bell Telephone Laboratories, 1930—. He was first engaged in the development of a family of mercury vapor rectifier tubes for radio transmitters and public address systems. This was followed by work on the development of a series of thyratron tubes for telephone power plants. During World War II, he worked on medium-voltage temperature-free thyratrons for use in gun directors and military communications systems and on ruggedized miniature cold cathode tubes for use in magnetic mines laid from aircraft. Later work was concerned with cold cathode tubes for telephone plants. These included multi-element tubes for selective ringing, voltage reference and regulator tubes, and the talking-path diode for the first electronic switching system (Morris, Ill.). He was also active in the renovation of the safety organization of Bell Laboratories. More recently his work was centered on gas tubes for the protection of submarine cable systems, and his latest assign-

ment is in the field of gas masers. He holds nine Bell System patents, is a member of IEEE and Sigma Xi, and is a registered professional engineer of the State of New York.

SVEN G. JOHANSSON, B.S. in E.E. 1924, Tekniska Gymnasiet, Orebro, Sweden. Mr. Johansson joined the Western Electric Company Engineer of Manufacture Organization at Kearny, New Jersey, in 1929. After various assignments in the physical electrical laboratory, design of electrical testing equipment, and crystal unit manufacturing planning, he was promoted to Department Chief in charge of Microwave Electron Tube Engineering at Western Electric, Allentown Works, in 1947. In 1953 he was assigned to Hillside, New Jersey, as Engineering Department Chief on flexible submarine cable repeaters. In 1959 he was promoted to his present position, Assistant Superintendent in charge of engineering organizations for manufacture of flexible and rigid submarine cable repeaters at the Western Electric Hillside and Clark, N.J., shops. Member, IEEE.

R. A. KELLEY, B.S.E.E., 1947, and M.S.E.E., 1948, Purdue University; Eastman Kodak 1948–1950; Bell Telephone Laboratories, 1950—. After working on the L3 coaxial system, Mr. Kelley was concerned with the development of broadband submarine cable systems, including the system design and repeater circuitry for the SD System. He is now Director, Digital Transmission Laboratory. Member, IEEE, Tau Beta Pi, Eta Kappa Nu and Sigma Xi.

BROOKE W. LERCH, B.S. in M.E., University of Michigan, 1931; Western Electric Company, Inc., 1937—. Mr. Lerch's first assignments at the Baltimore Works of Western Electric Co. were textile applications and rubber extrusion for drop and station wires. During World War II he was the engineering department chief at the Western Electric Scranton, Pa. shops, engaged in design and production of cords and field cables for the armed forces. Returning to Baltimore Works, Mr. Lerch's assignments were concerned with outside plant wires and cords. Later, he became Assistant Superintendent, Toll and PIC Exchange Cable. In 1961 his duties included ocean cable engineering and inspection. Member, Wire Association.

ROBERT J. MCSWEENEY, B.S., Stevens Institute of Technology, 1954, Gibbs & Cox, 1954—. He was originally involved in analysis of propulsion systems for naval vessels and was responsible for purchase specifica-

tion preparation and subsequent technical evaluation of resultant proposals. He is at present responsible, with others, for the design of complete shipboard electric power systems. In connection with C.S. *Long Lines*, he was responsible for the preliminary and detailed design of the electric propulsion system, including coordination of the drive characteristics with the dynamics of the ship-propeller-water system and the boiler-steam-turbine cycle. He maintained liaison with the shipyard during and after construction, and organized, directed, and analyzed the main propulsion system sea trials.

L. H. Morris, B.S.E.E., 1935, City College of New York; Bell Telephone Laboratories, 1928—. He has worked on coaxial cable systems from the early one-megacycle trial systems through the L3 system, and on the TH microwave radio relay system. Since 1959 he has headed a department responsible for system and terminal aspects of various submarine cable projects.

Samuel Mottel, B.S.M.E., 1950, City College of New York, Bell Telephone Laboratories, 1952—. Mr. Mottel has been concerned with power equipment and systems development since joining the Laboratories. He has worked on power for carrier systems, microwave systems, Bell System and military submarine cable systems, data systems and various military communications systems. Since 1963 he has supervised a group working in systems power equipment development areas. Member, A.S.M.E.

Elliott T. Mottram, B.S.M.E., 1928, Columbia University; Bell Telephone Laboratories 1928—. His early work was in development of disc recording and reproducing equipment and techniques. Later he was concerned with development related to recording and reproducing sound on film and on tape. During, and for a short time after, World War II (1939–50) he was engaged in development of airborne radio and radar equipment, electronic computers and bomb sights, and airborne homing missiles. In 1950 he became Director of Transmission Systems Development with responsibility for development of television, wire and military communication systems. This included submarine cable development, which has occupied an increasing proportion of his interest as new systems have been developed and laid. Member, IEEE.

Wendell G. Nutt, B.S.M.E., 1949, Texas Technological College; Southwestern Bell Telephone Co. 1949–1953; Bell Telephone Labora-

tories, 1953—. At Bell Laboratories he first worked on gas pressure maintenance and splicing of multipair cables. Next he was concerned with the development of armorless ocean cable and with waveguide for the circular electric mode. He is currently supervisor of a group concerned with the development of broadband multicoaxial cables for land routes. Member, Tau Beta Pi, Kappa Mu Epsilon and Alpha Chi.

James W. Phelps, B.S.E.E., Iowa State University, 1951; Bell Telephone Laboratories, 1951—. His early work was in the design and specification of electrical protection systems. In 1956 he transferred to a group responsible for the design and development of armorless ocean cable. He headed the group responsible for the operation of the experimental cable laboratory in Cambridge, Massachusetts, where the early cable samples were produced. At present he supervises a group working on a new cable structure for broadband use. Member, Tau Beta Pi and Eta Kappa Nu.

R. M. Riley, A.B., 1943, Park College; M.S., 1948, University of Minnesota; Instructor, University of Minnesota, 1946–1948; Instructor, Iowa State University, 1948–1949; Bell Telephone Laboratories, 1949–1953; Chief Engineer, Visioneering Company, 1954–1955; Bell Telephone Laboratories, 1955—. He was first engaged in studies for the outside plant department. He has more recently been engaged in the development of ocean cable and is presently Assistant Director, Outside Plant Laboratory. Member, IEEE and Mathematical Association of America.

Philip W. Rounds, A.B., 1929, Harvard University; Bell Telephone Laboratories, 1929—. Prior to World War II, Mr. Rounds was concerned with the development of transmission networks for toll telephone, telephoto and program systems. During the war he developed computing networks for antiaircraft gun directors and bombsights, as well as transmission networks for sonar systems. He has since worked on the development of transmission networks for television systems and more recently the development of submarine cable systems. Member, IEEE.

Laurus E. Sutton, III, M.E., 1948, Stevens Institute of Technology; graduate study, Stevens Institute of Technology, Columbia University and Massachusetts Institute of Technology; Gibbs and Cox, Inc., 1948—. He has worked on the design of various experimental naval craft, including hydrofoil research craft, and commercial ship design. He is presently head of the scientific section of the electrical

division of Gibbs and Cox, Inc. Member, IEEE and American Ordnance Association.

WILBUR VAN HASTE, B.S.E.E., 1936, New York University; Bell Telephone Laboratories, 1928—. He has been concerned with the design and development of grid-controlled electron tubes for use in a wide variety of Bell System communication projects. These include tubes for open wire, coaxial cable, microwave, and submarine cable systems. He presently supervises a group responsible for the final design and acceptability of tubes for SD submarine cable systems. Senior member, IEEE; member, Iota Alpha.

CHARLES A. VON ROESGEN, Dipl. Ing., 1952, Swiss Federal Institute of Technology, Zurich (Switzerland); Bell Telephone Laboratories 1953—. He first worked on the development of cables for submarine use; following that he engaged in automatic transmission test equipment design. In 1960 he became a supervisor of a group concerned with the development of the SD repeater. At present, Mr. von Roesgen is responsible for the design of new submarine cable terminal equipment. Member, IEEE.

EDWARD J. WALSH, Bell Telephone Laboratories, 1928—. He has chiefly been engaged in the mechanical design of electron tube structures and enclosures, including the design during World War II of proximity fuse tubes and the thermally tuned klystron, and later of the frame-grid tubes used in microwave radio relay systems. More recently he has supervised a group responsible for mechanical design of electron tubes for the SD submarine cable system; he is presently in charge of a group working on structures for the gaseous optical maser, photomultipliers, and other electron devices.

B. S. T. J. BRIEFS

A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases

By D. W. BERREMAN

(Manuscript received May 26, 1964)

Development of systems for long-distance communications using laser beams is of considerable current interest. A major problem is to avoid losses in transmission systems for light. Losses by scattering and absorption in solid transmission schemes are likely to be large. By using gases as the media that guide the light beam, such losses should be minimized. A number of schemes in which gases can be made to guide a light beam have recently been investigated.

Tests on models of one type of gas lens or light guide utilizing refractive index gradients caused by temperature gradients are to be described in this paper. These particular models consisted of a suitably shaped continuous heating element or train of separate elements supported by insulating material inside a cooler pipe.

Heating elements composed either of a series of closely spaced, doughnut-shaped rings or toruses, or of a single, continuous helix of suitable dimensions were used in the experiments to be described. The beam travels through the doughnut holes or down the axis of the helix. Helices were used in the more detailed tests to be reported here because they are much easier to construct and to support in the cooler outer jacket. They focus at least as strongly and with as little aberration as toruses of approximately equivalent dimensions.

Lens trains of the types described focus to some extent by the alternating gradient (AG) focusing mechanism.*[1,2] However, if the temperature distribution is suitably distorted by moving the gas past the heating element, a great enhancement of focusing can be obtained. The enhancement is due to the ordinary focusing effect of a radially decreasing average refractive index, which is absent if the gas is static. Convection alone caused the gas motion in the lenses described here.

The best of several experimental models of convective thermal gas

* A. R. Hutson first proposed use of AG focusing in gas lenses to the author in private discussion.

lenses, in terms of specific convergence* vs power consumption by the heating element as well as absence of aberrations, is illustrated in Fig. 1. The model contains a helix about 0.75 meters long and is closed at the ends with flat glass plates to confine the gas. In tests of power consumption the helix was warmed electrically, but for detailed tests for convergence vs temperature difference between helix and outer jacket, it proved more convenient to run warm water through the helix, which was made of copper tubing. Temperature differences were measured with a thermocouple (not shown), fastened to the helix and to the inside of the outer jacket. Measurements of focusing strength and aberration were made with considerable precision using a modification of the Foucault knife-edge test.[3] A very fine pinhole light source, two 80-cm telescope objective lenses (one for making the incident light parallel and the other for enhancing its reconvergence), and both horizontal and vertical knife edges were used in the tests. The two knife-edge orientations enabled us to measure not only focal length, but also spherical aberration, astigmatism, and a type of aberration that might be characterized as "S-shaped" or "sagging lens" aberration. The latter was revealed by the *horizontal* knife-edge test as an inequality in focal length between the upper and lower halves of the clear aperture. One should expect astigmatism and sagging lens aberration in a convective system.

The amount of astigmatism is surprisingly small, and the amounts of spherical and "sagging lens" aberration were too small to detect with the gas lens shown in Fig. 1, using air, CO_2 or propane gas at temperature differences that gave focal lengths of more than about 5 meters or specific convergences less than about 0.3 meters^{-2}. At higher temperature differences, both astigmatism and sagging lens aberration began to appear, but were relatively small perturbations on the convergence. Fig. 2 shows a detailed study of the astigmatism and convergence ($1/F$) versus temperature difference (helix temperature minus jacket temperature) with the jacket at approximately room temperature using CO_2 at one atmosphere pressure.

When helices with considerably larger apertures for the light beam

* The periodic light guides described here had the property that the focal length of any one cycle of the guide was very much longer than the length of one cycle. Consequently, AG focusing was very weak and the focusing was due largely to the fact that the mean refractive index along the axis was higher than along parallel lines off the axis. The mean refractive index at radius r from the axis can be described by the equation $n = n_0(1 - \frac{1}{2}Cr^2)$ when aberrations are absent. When AG focusing is negligible, the quantity C is equal to average focusing power, or convergence, of a segment of unit length out of the continuous "lens train." Hence, it is appropriate to call C the *specific convergence* of such a light guide, a term suggested by W. L. Bond. The maximum allowable curvature of such a light guide is proportional to its specific convergence and to the diameter of its aperture.

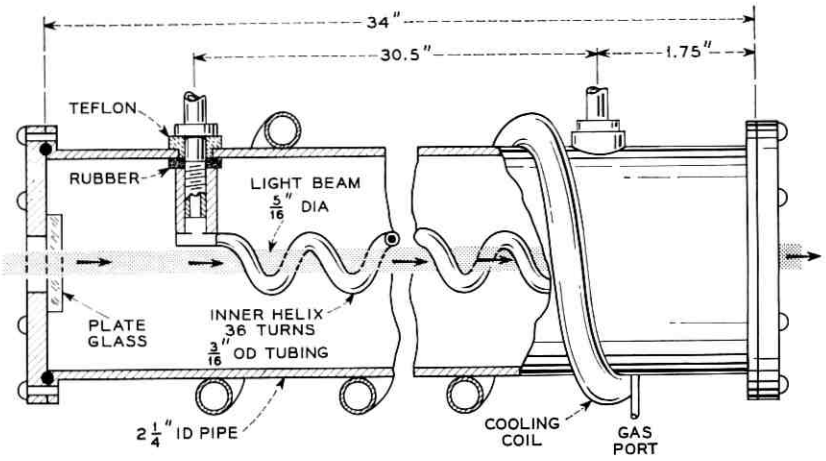


Fig. 1 — Convective gas lens showing the best of several warming elements, a helix, inserted in such a way as to be heated with warm water; the jacket is also kept at a fixed lower temperature with water.

were tested, there was a large amount of spherical aberration when pure CO_2 was used, but not when a mixture of half CO_2 and half helium or when pure argon was used. Likewise, the helix shown in Fig. 1 gave measurable spherical aberration when C_4F_{10} was used but not with any of the lighter gases tested. A helix with an even smaller aperture would probably greatly reduce the spherical aberration with C_4F_{10} . These facts are consequences of the differences in relative rates of convective and conductive heat flow in the gases.

Another effect that depends on these factors, and also on the specific refractive index of the gases, is the specific convergence when the power per unit length consumed by the heating element is fixed, or what may be called the *efficiency* of the lens. Efficiency has the dimensions of meters^{-1} watts^{-1}, or diopters per watt. The efficiency of the lens shown in Fig. 1 was found to be almost independent of power consumption in the range where aberrations were small. Measurements at different pressures with CO_2 showed that efficiency is approximately proportional to pressure. When one watt of electrical power was supplied continuously to the helix with the jacket at room temperature, the temperature differences, efficiencies and values of specific convergence shown in Table I were measured with the gases listed.

The amount of AG focusing with a helical heating element was calculated and shown to be entirely negligible when the helix turns are as closely spaced as in the lens shown in Fig. 1. This is confirmed by the

TABLE I

Gas	ΔT	Efficiency	Specific Convergence
air	4.0°C	0.045 $m^{-1}w^{-1}$	0.06 m^{-2}
CO_2	3.9°C	0.091 $m^{-1}w^{-1}$	0.12 m^{-2}
C_3H_8	3.4°C	0.12 $m^{-1}w^{-1}$	0.16 m^{-2}
C_4F_{10}*	2.6°C	0.15 $m^{-1}w^{-1}$	0.20 m^{-2}

* This gas was suggested by K. B. McAfee. The focal length is only approximate because of spherical aberration.

symmetry of Fig. 2, since AG focusing would be independent of the sign of the temperature difference. Fig. 3(a) is a three-dimensional plot showing how the refractive index, n, would vary as a function of position in a plane normal to the optic axis if there were negligible convection. In that case, both the temperature and the refractive index would very nearly obey Laplace's equation. Suppose one considers two orthogonal cross sections of the plot parallel to and containing the optic axis. Near the axis, a cross section that cuts through the helix has curvature $\partial^2 n/\partial y^2$ that is equal in magnitude but opposite in sign to the curvature $\partial^2 n/\partial x^2$ in the orthogonal cross section. (Cf. lines YOY' and XOX' on Fig. 3a.) This is a direct consequence of the Laplace equation and the symmetry

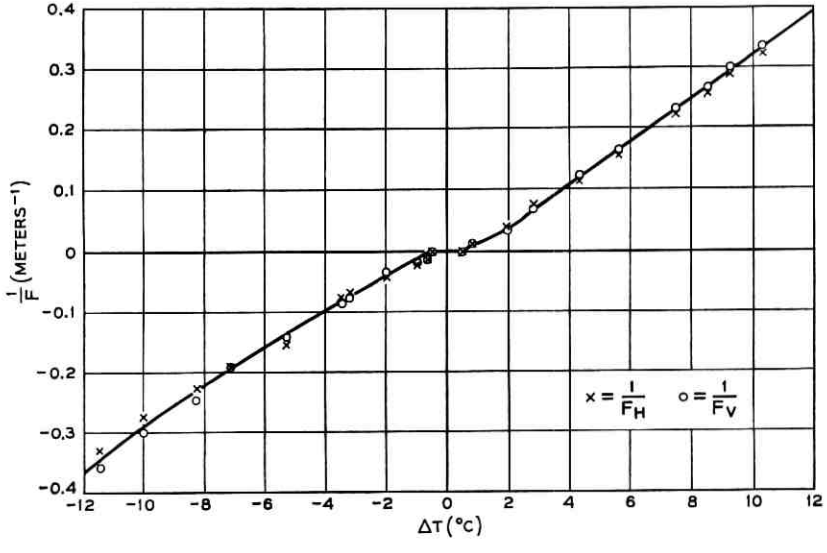


Fig. 2 — Reciprocal focal length, or convergence in vertical plane ($1/F_V$) and in horizontal plane ($1/F_H$), vs temperature of inner helix minus temperature of outer tube for the lens shown in Fig. 1 with CO_2 at one atmosphere pressure. The difference between horizontal and vertical convergence is a measure of astigmatism.

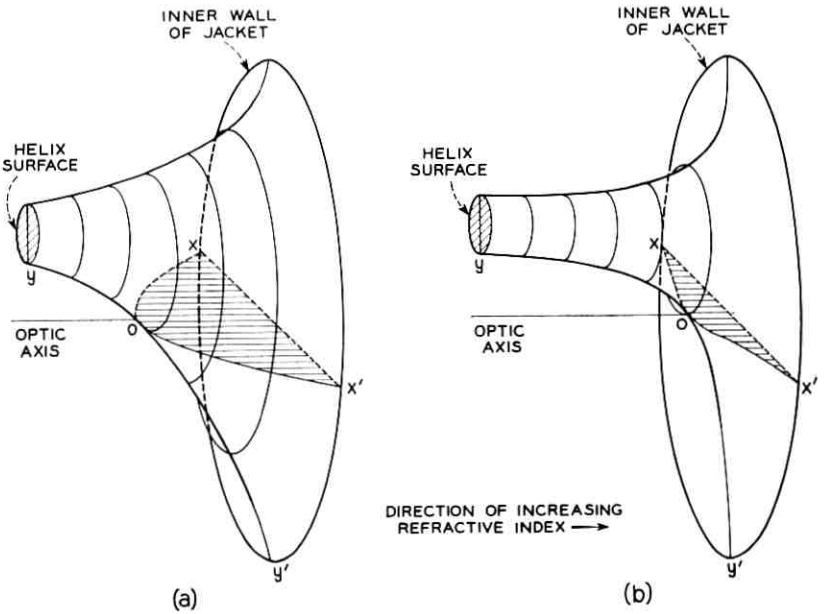


Fig. 3 —(a) Refractive index vs two coordinates of position in a plane normal to the optic axis in the gas lens. The refractive index is highest at the jacket wall, represented by the large circle seen in perspective on the right side, and lowest at the warm helix wall, represented in cross section by the small circle on the left. This figure represents the hypothetical situation in the absence of convection. (b) Similar to (a) but with convection. Note difference in magnitudes of curvature of the plot in cross sections through the optic axis normal (XOX') and parallel (YOY') to the plane of the drawing. Such difference is zero in (a). Details would vary somewhat with azimuth of helix at the plane of the plot.

of the helix. No ordinary focusing occurs in this case, because there is no change in *average* refractive index with radial distance from the axis on lines parallel to the axis. A more rigorous analysis shows that no ordinary focusing occurs anywhere inside the aperture of any system having either helical or ordinary axial symmetry if n obeys Laplace's equation. Fig. 3(b) is the same sort of plot as Fig. 3(a), except that the contours are changed by convection. The convection produces a sharper temperature gradient near the helix surface and the temperature is more nearly uniform elsewhere. Consequently, the plot of refractive index curves more sharply, near the axis, than before in the cross section through the axis and the helical tube, and less sharply in the orthogonal direction. The result is that the *average* refractive index is then higher along lines near the axis than along lines farther away, so that there is a net positive ordinary focusing effect.

A detailed mathematical analysis of such convective flow has not been

made to the author's knowledge, but experiments on temperatures and flow in gases between concentric cylinders of unequal temperature have been reported in Ref. 4. That paper shows temperature distribution alteration by convection in a quantitative way and also presents the relation between convection and the relative size of inner and outer cylinders and the viscosity, specific heat, density and heat conductivity of the convective fluid.

Although any element of the helix is far from straight and is not very near the center in any cross-sectional element, the convective flow in any plane is qualitatively very similar to that described for small flow rates in Ref. 4. The flow patterns in several planes were observed and photographed in a glass-walled replica of the lens using tobacco smoke in air. The flow was perfectly steady and nonturbulent, even with a temperature

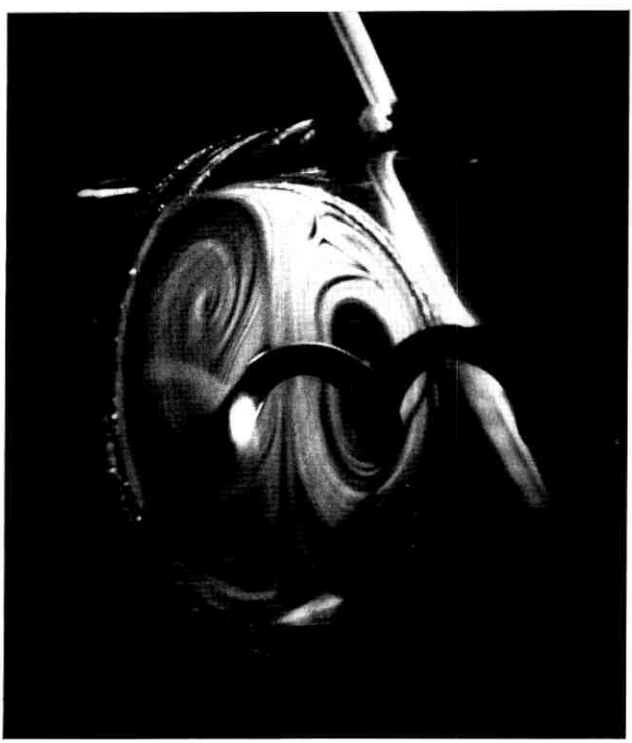


Fig. 4 — Smoke moving with air in glass-walled gas lens. The camera is somewhat below and to one side of the lens. Smoke is flowing upward around the helix and downward on both sides, inside the jacket. Helix is 10°C warmer than the jacket. Illumination is from the upper right through a narrow slit normal to the axis. The illuminated region cuts through the helix at a point level with the axis.

difference of over 10°C, and always showed a two-kidney shaped pattern similar to that obtained with a concentric cylinders.[4] Fig. 4 is one of these photographs, taken in a plane where the helical tube is on a horizontal line from the axis, where one might expect greatest departure from the results with concentric cylinders.

It is evident, especially from Ref. 4, that temperature gradients are not axially symmetric even when averaged over a complete loop of the helix, because of the relation between the convective flow patterns and the direction of gravity. It is therefore surprising, but none the less true, that nearly perfect, aberration-free focusing can be achieved with the convective gas lens if a uniform helix of the proper cross-sectional dimensions is chosen for the warming element. There seems to be no reason why a continuous helix and tubular jacket of similar cross section but of great length might not have the same local focusing properties as the short segment that was tested.

Acknowledgments. The author wishes to acknowledge the assistance of D. E. Collins in the construction and testing of the lenses and taking the smoke photographs. A. R. Hutson, K. B. McAfee and J. A. Lewis have been especially helpful in discussions of the gas lens.

REFERENCES

1. Courant, E. D., Livingston, M. S., and Snyder, H. S., The Strong Focusing Synchrotron — A New High-Energy Accelerator, Phys. Rev., **88**, 1952, pp. 1190–1196.
2. Tien, P. K., Focusing of a Long Cylindrical Electron Stream, J. Appl. Phys. **25**, 1954, pp. 1281–1288. (P. K. Tien pointed out in private communications that this analysis of AG focusing of electrons can be used with only minor changes for the present case.)
3. L. Foucault, Description of Various Processes Made Use of for Finding Out the Configuration of Optical Surfaces, Monthly Notices of the Astronomical Society of London **19**, 1859, pp. 283–287 and Annales de l'Observatoire de Paris, **5**, 1859, pp. 197–237.
4. Liu, Chen-Yi, Mueller, W. K., and Landis, F., Natural Convection Heat Transfer in Long Horizontal Cylindrical Annuli, Int. Dev. in Heat Transfer, Proc. 1961–62 Heat Transfer Conf., Boulder, Colo., publ. A.S.M.E., 1963, pp. 976–984.

A Gas Lens Using Unlike, Counter-Flowing Gases

By D. W. BERREMAN

(Manuscript received May 26, 1964)

Because of the current interest in gas lenses for possibe use in long-distance laser beam transmission, it seems worthwhile to report on the successful operation of an example of another class of gas lenses. This class of lenses utilizes the difference of specific refractive index of different gases to achieve focusing in a region where there is a variation in relative concentration of different gases. In the device to be described, two gases flow together continuously from opposing tubes into a mixing chamber. The chamber is designed so that the effective interface where the gases meet is an axially symmetric curved surface, which acts as a lens for light passing down the axis of the opposing tubes.

In order to minimize distortion due to gravitational force, it is desirable to use two gases of about the same density. It may be most economical to separate and recycle the gases after they have run together. In principle, continuous separation with a semipermeable membrane would consume only a moderate amount of power. Except for a computation of the theoretical minimum of power consumption, no further discussion of the problem of gas separation will be presented in this paper.

A working model of a single element of a counter-flowing gas lens is shown in Fig. 1. Rather elaborate precautions were taken to avoid

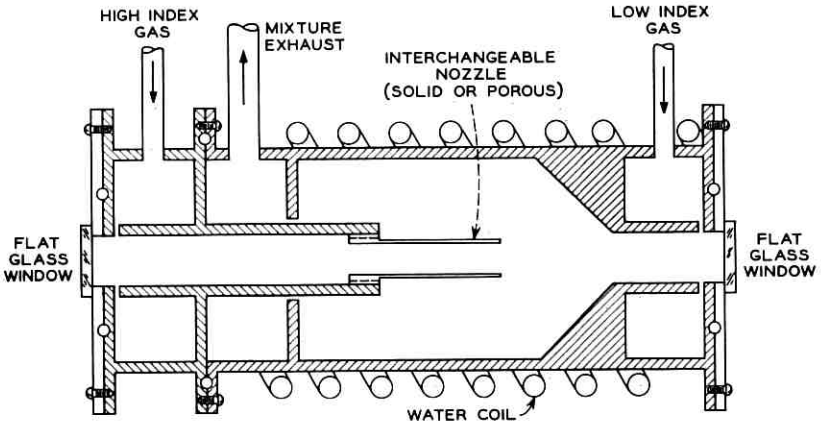


Fig. 1 — Experimental model of counter-flowing, unlike gas lens; over-all length is 8 inches.

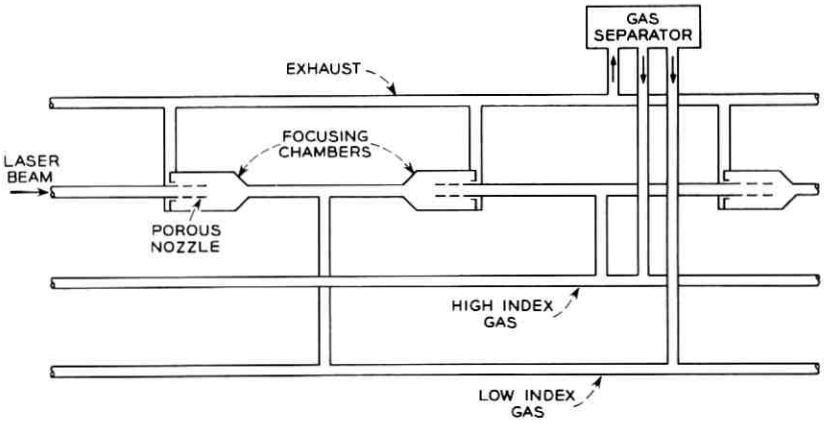


Fig. 2 — Schematic diagram of a part of a continuous series of counter-flowing, unlike gas lenses connected to a separator for recycling gases.

turbulence or asymmetry that might be introduced if the gas entered directly into the tubes through holes in the walls. Such precautions could probably be dropped if greater length of tubes existed between the points at which the gases enter and the regions in which they mix (see Fig. 2). For a continuous series of such lenses, one might have the actual lenses or "mixing chambers" a few feet apart, with gases of alternate types entering midway between the mixing chambers, as shown in Fig. 2.

The focal length and aberration of the lens as a function of flow rates were measured using a modification of the Foucault knife-edge test mentioned in the accompanying B.S.T.J. Brief on thermal gas lenses.[1] The two gases were run through long parallel copper tubes, which had been soldered together, before they entered the gas lens. Water was circulated through a tube soldered to the copper gas tubes and through a tubular coil wrapped around the gas lens (see Fig. 1). This precaution insured the absence of any focusing due to thermal differences. At moderate flow rates no appreciable aberration was observed, but some spherical aberration appeared before turbulence set in.

Fig. 3 is a graph of the variation of focal length with rates of flow of argon and CO_2. The lower group of curves shows such variation in the simplest model, in which CO_2 flowed out the end of a $\frac{3}{8}$-inch ID tube or nozzle into a relatively large cavity surrounding the end of the nozzle. The cavity was supplied with argon from the opening opposite the nozzle. The upper group of curves shows the same parameters when the solid nozzle was replaced with a cylindrical tube of 50-mesh copper screen. The screen allowed mixing of the CO_2 with argon over a longer distance with greater radial variation in concentration. The result is

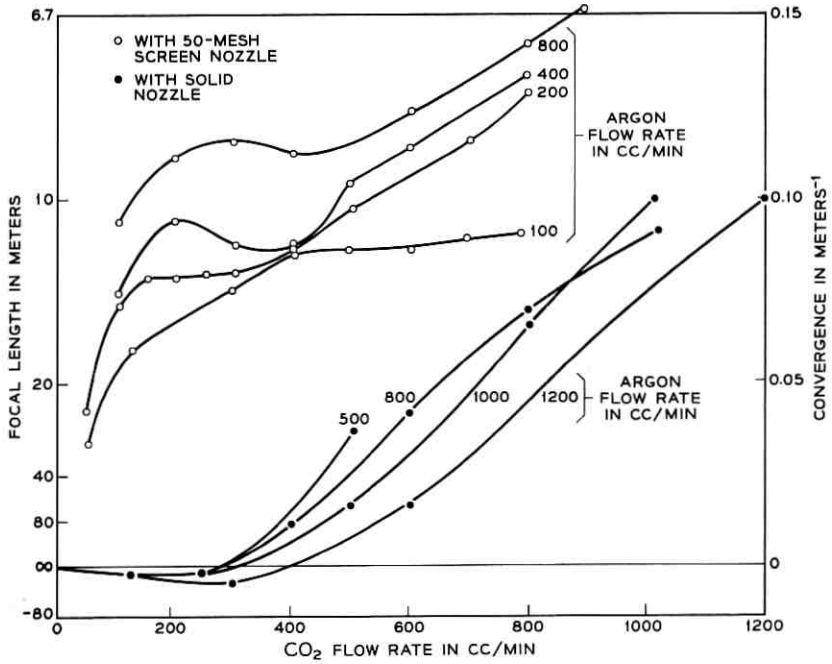


Fig. 3 — Graph showing variation of focal length and convergence with gas flow rates. Note increased efficiency using screen nozzle.

much greater curvature of the effective interface between the CO_2 and the argon, which gave a much shorter focal length at a given flow rate. Diffusion is certainly very important in the enhancement of focusing strength when the screen is used. A. R. Hutson was the first to suggest gas lenses utilizing refractive index gradients controlled by diffusion of gases of unequal refractive index into one another (private discussions with the author).

The curves stop at flow rates of about 1 liter per second of either gas because turbulence appeared at higher flow rates. Flow appeared to be completely free of turbulence at lesser flow rates.

The minimum power consumption, if the gases must be recycled, can be obtained using a simple thermodynamic argument due to K. B. McAfee.[2] If it were possible to find two semipermeable membranes, one for one gas and one for the other, the power required to separate the gases reversibly, neglecting small departures from ideal gas properties, would be

$$P = -fRT\left[\ln X + \frac{1 - X}{X}\ln (1 - X)\right]$$

where f is the flow rate of one of the gases in moles per second, R is the gas constant in joules/mole °K, T is the absolute temperature, and X is the mole fraction of that gas in the mixture. If the flow rate of each gas is 200 cc/minute, so that $X = 0.5$, the formula gives a power consumption

$$P = 0.50 \text{ watts.}$$

Tests showed that the focal length at such a flow rate was about 15 meters in the lens with a screen nozzle, using CO_2 and argon. The theoretical maximum "efficiency" of the lens is thus 0.13 diopters per watt at this flow rate. It may be of interest to note that this efficiency is the same order of magnitude as that of the convective helical gas lens reported in Ref. 2. However, it should be emphasized that such low power consumption is only a theoretical limit for a counter-flowing gas lens of these particular dimensions.

Acknowledgments. It is a pleasure to acknowledge many helpful discussions and much encouragement given by A. R. Hutson in the course of this work. D. E. Collins has been of great help in setting up and performing the experiments.

REFERENCES

1. Berreman, D. W., A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases, B.S.T.J., this issue, p. 1469.
2. McAfee, K. B., Diffusion Separation, *Encyclopedia of Chemical Technology*, Suppl. 2, Interscience, New York, 1960, p. 297ff [the formula presented here comes from Eq. (13)].

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLIII JULY 1964 NUMBER 4, PART 2

Timing Errors in a Chain of Regenerative Repeaters, III

By B. K. KINARIWALA

(Manuscript received March 2, 1964)

We present here a general and rigorous theory of the jitter accumulation in a chain of regenerative repeaters. The sources of jitter are assumed to be the signal-dependent sources, as distinguished from purely random sources independent of the signal.

Our results show that while the absolute jitter and its dispersion grow without bound with the number of repeaters, the spacing and the alignment jitter remain bounded. In particular, the spacing jitter bounds are quite optimistic for most practical situations, viz., no greater than twice the absolute jitter injected at a single repeater. This result is of importance in that it ensures proper decoding of the binary signal. Its further importance is that it does ensure, in most cases, the validity of the basic model and thus the validity of other results obtained by that model. One such result shows that the alignment jitter is slowly-varying for repeaters farther along the chain. We also include some results which would be of use in computations, together with a simple example.

I. INTRODUCTION

1.1 *Purpose*

Pulse regeneration is an attractive feature of digital communication systems. A regenerator or a regenerative repeater must: (*i*) detect the presence or absence of a pulse at certain time instants which are, ideally,

multiples of the basic pulse repetition period, (*ii*) regenerate the pulse shapes, and (*iii*) retime these pulses so they occur at appropriate instants of time in the outgoing signal. In practice, errors in detection due to noise, distortion, etc. in the system and errors in retiming the signal impose limitations on the operation of such systems. Except for noise which may lead to the detection of a pulse where no pulse exists in the original signal and vice versa, the imperfections in the system show up as a jittering of the pulse positions in the outgoing signal. In self-timing repeaters,[1] the jitter from all sources except random noise is also dependent on the signal itself.[2]

Several workers have investigated many different aspects of the timing jitter problem.[1-9] We study here the signal-dependent jitter due to repeated regeneration. This article leans quite heavily on our previous discussion, and in fact it supplements as well as complements the previous results referred to in this article as Parts I and II.[3]

1.2 *Background*

In our previous discussion, we showed how the timing jitter in a pulse train accumulates as the pulse train is repeatedly regenerated. The timing information was assumed to be extracted from the incoming signal itself, and the timing extractor was assumed to be a tuned circuit. Mistuning in the tuned circuit was assumed, as a convenience, to be the major source of jitter. However, the accumulation properties of timing jitter are not dependent on a particular source of jitter. We shall attempt to clarify this point here.

Our previous results showed that the displacements of the pulse positions from their original positions (or the "absolute jitter") increase indefinitely with the number of repeaters. The major component in the absolute jitter was found to be flat delay (i.e., the same displacement at every pulse position). A natural question follows: How does the absolute jitter behave if the flat delay is removed? This is the "dispersion" or the absolute jitter measured against the reference clock delayed by an appropriate amount. It was shown that the dispersion also increases without bound except when the pulse trains are severely constrained (e.g., periodic, finite, etc.). These results are valid even under the constraint that there exists at least one pulse in a predetermined number of the basic periods or "time slots."

It is worth noting that the absolute jitter and the dispersion have as counterparts the average and the variance of the random variable represented by the pulse displacement at any pulse position. We wish to em-

phasize, however, that our results are independent of any a priori statistics concerning the pulse train ensemble.

1.3 *Results*

In the present article, our most significant result concerns a more important parameter, viz., the "spacing jitter," which is the variation in the spacing between adjacent pulses. We show that the spacing jitter is bounded for an indefinitely long chain of repeaters, and the bound is directly related to the minimum pulse density. Such bounds may be precisely evaluated both for the infinite as well as the finite chain of repeaters. The importance of these results lies in the precise evaluation of the bounds, the means to control these bounds, and in our ability to relate these bounds either with resulting errors in the decoding of the signal,[2] or with distortion in the analog signal,[4] depending on whether the bounds are large or small. As will be seen later, these results also determine the validity or otherwise of all the other previous results on the timing problem, since the present results have a direct bearing on the validity of the model used by most people.

We also present a rather thorough discussion of the computational aspects of the problem in the Appendix. Special situations such as periodic patterns, truncations, pattern transitions, etc. are included in our discussion.

The case of nonidentical repeaters is examined briefly. It appears that the bounds on jitter are not appreciably different if the repeaters are not appreciably different. The "misalignment," or the jitter introduced by a single repeater in an already jittered pulse train, is also examined briefly. We show that the misalignment is slowly-varying for repeaters further along in the chain.

Let us emphasize, in conclusion, that our results are not dependent on any a priori statistics. Our analysis is quite rigorous once the basic model is derived. The basic model is essentially the same as that of other investigators, and the major assumption in the model asserts that a single repeater introduces only slowly-varying jitter in a jitter-free pulse train. Such an assumption is quite reasonable for any practical repeater.

1.4 *Organization*

We start with a mathematical statement of the problem. Our formulation shows that the input and the output jitter sequences are related to each other by a linear operator which maps the space of bounded sequences into itself. The dimensionality of the space is determined by the

memory of the system, which is infinite for an infinitely long chain of repeaters. We are thus led to a discussion of the operator in a Banach space of infinite dimensions. The spectral properties of the operator determine the behavior of the absolute jitter and the dispersion. Next, we consider the spacing jitter obtained by a difference operation on the absolute jitter. The brief discussions on the misalignment, unequal repeaters, etc. follow. Finally, we present several results to facilitate computations.

II. STATEMENT OF PROBLEM

The basic model of the repeater is represented by a tuned circuit excited by a train of pulses. The natural frequency of the tuned circuit is assumed to be very close to the pulse repetition frequency. At upward (or downward) zero crossings of the response of the tuned circuit, timing pulses are generated which determine the instants of outgoing pulses; the presence or absence of a particular pulse in the output signal is determined by the presence or absence of a pulse in the input signal.

Let $\{\cdots, -2\tau, -\tau, 0, \tau, \cdots\}$ represent the instants of occurrence of pulses for the ideal pulse train. These would be the centers of the corresponding time slots. The occurrence or nonoccurrence of a pulse at $t = -n\tau$ is determined by the value of the random binary variable $\mathcal{A}_n$. A pulse is present when $\mathcal{A}_n = 1$ and no pulse is present when $\mathcal{A}_n = 0$.

At this point it is convenient to assume that the pulse train consists of impulses located at the actual pulse positions defined above. The actual pulse shapes modify the zero crossings of the response of the tuned circuit, and a term representing such a correction can be added separately.

Finally, let $\xi_k^{\,l}$ be the displacement of the kth pulse (originally located at $t = -k\tau$) at the output of the lth repeater in a chain of repeaters. At the input of the lth repeater the timing displacement is given by $\xi_k^{\,l-1}$, which is the jitter value at the output of the $(l-1)$th repeater. The displacement (or jitter) is measured in radians, where 2π corresponds to the pulse interval τ.

In order to determine $\xi_k^{\,l}$, we merely find the appropriate zero crossing of the response of the tuned circuit excited by a train of pulses. The exciting pulse train is itself jittered and this fact is represented by the values of $\{\xi_k^{\,l-1}\}$. It turns out that the $\xi_k^{\,l}$ is actually a nonlinear function of the set $\{\xi_{k+n}^{\,l-1}\}$ with $n = 0, 1, 2, \cdots$. Furthermore, it also depends on the original signal represented by the set $\{\mathcal{A}_n\}$ and, of course, the Q of the resonant circuit. If, however, $\{\xi_{k+n}^{\,l-1}\}$ satisfy certain conditions, it is possible to represent the $\xi_k^{\,l}$ as a linear function of the variables

$\{\xi_{k+n}{}^{l-1}\}$. The function is still dependent on $\{\mathcal{A}_n\}$ and the Q of the circuit. This is the fundamental relation between the input and output jitter and, if there were no jitter introduced by the repeater, it would take the following form:[2]

$$\xi_k{}^{l} = \frac{\sum_{n=0}^{\infty} \mathcal{A}_{n+k}\beta^n \xi_{n+k}{}^{l-1}}{\sum_{n=0}^{\infty} \mathcal{A}_{n+k}\beta^n}, \qquad (k = 0, 1, 2, \cdots; l = 1, 2, \cdots), \quad (1)$$

where $\beta = \exp(-\pi/Q) \approx 1 - (\pi/Q)$ for large Q.

The conditions that must be satisfied by $\xi_k{}^{l-1}$ are the following:

$$|\xi_k{}^{l-1} - \xi_{k-1}{}^{l-1}| \ll \pi/Q \qquad \text{for all } k. \quad (2)$$

These conditions are unaltered when (1) is slightly modified to include the jitter introduced by each repeater [cf. (3)]. It is therefore very important to make sure that (2) holds in order for any of the results obtained on the basis of (1) to be valid. The quantity required to be small in (2) is the spacing jitter, whose behavior is important in that if it ever becomes too large there will occur errors and distortion in the decoded signal.[2,4] It is our intention here to investigate thoroughly the behavior of the spacing jitter. Equation (1) represents the way in which jitter propagates along a chain of repeaters. The jitter accumulation properties of a chain are, therefore, a consequence of this basic equation. Of course, at every repeater there is jitter injected in addition to the one propagated from previous repeaters. Since we are not discussing here effects of random noise, the sources of the injected jitter are signal-dependent, and they are identical if we assume identical repeaters. The jitter injected at every repeater by signal-dependent sources is thus identical. For such sources, this injection of jitter is simply additive either at the input end of the repeater[5] or at the output end.[2,3,6,7] For example, dispersion in the channel with a consequent imperfect detection of the pulse positions would be an additive source of jitter at the input of the repeater.[5] Certain nonlinear operations (such as limiting) on the response of the tuned circuit would inevitably alter the zero crossings, and this may be represented as an additive source at the output end of the repeater. If the mistuning of the resonant circuit is small, it can be shown that mistuning represents an additive source at the output end. Finite pulse widths also represent an additive source at the output.[2] In any case, such injection of jitter depends on the signal and repeater parameters. Since these are the same at every repeater, the injected jitter is the same at every re-

peater. Additivity of these sources is usually a consequence of the fact that the injected jitter is small. For convenience, we will refer all these sources to the output end and represent the injected jitter by $\{\xi_k^1\}$, the output of the first repeater, where the original pulse train is assumed to be jitter-free. It is somewhat inconvenient to refer all sources to the input, since it would involve inverting the functional relationship of (1). Our interest in this article is not to discuss the quantitative evaluation of $\{\xi_k^1\}$. For simple sources such as mistuning and finite pulse widths, it is relatively easy to evaluate $\{\xi_k^1\}$. We shall discuss elsewhere the question of determining $\{\xi_k^1\}$ when all sources are included. However, let us emphasize that we do not restrict the sources of jitter to short memory mechanisms.[5]

In this paper, our interest is to determine the behavior of $\{\xi_k^l\}$ for large l when the basic equation (1) is modified to include the injected jitter:

$$\xi_k^{\,l} = \frac{\sum_{n=0}^{\infty} \alpha_{n+k}\beta^n \xi_{n+k}^{\,l-1}}{\sum_{n=0}^{\infty} \alpha_{n+k}\beta^n} + \xi_k^{\,1}, \tag{3}$$

subject to condition (2), and where we assume that $\xi_k^0 = 0$. Condition (2) is satisfied by $\{\xi_k^0\}$ since they are all zero and by $\{\xi_k^1\}$ because for a practical repeater the jitter injected by a single repeater must be extremely small. We have made no assumptions on the nature of the signal. Thus, if it can be shown that condition (2) is valid for every l, then the results obtained by using (3) are valid. This is an important point which cannot be overemphasized. We will therefore pay particular attention to the behavior of the spacing jitter.

III. RECAPITULATION

In this section, we recapitulate the basic formulation of the problem developed in Parts I and II of this article. For details, the reader is referred to the original discussion. Define a vector

$$X_l = \{\xi_0^{\,l}, \xi_1^{\,l}, \xi_2^{\,l}, \cdots\}, \qquad (l = 1, 2, \cdots); \tag{4}$$

and $X_0 = 0$. Then, the basic equation (3) may be written as

$$X_l = T_0 X_{l-1} + X_1, \tag{5}$$

where

$$T_0 = \begin{bmatrix} \frac{\mathfrak{A}_0}{s_0} & \frac{\mathfrak{A}_1\beta}{s_0} & \frac{\mathfrak{A}_2\beta^2}{s_0} & \cdots \\ 0 & \frac{\mathfrak{A}_1}{s_1} & \frac{\mathfrak{A}_2\beta}{s_1} & \cdots \\ 0 & 0 & \frac{\mathfrak{A}_2}{s_2} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \tag{6}$$

and

$$s_i = \sum_{n=0}^{\infty} \mathfrak{A}_{n+i}\beta^n. \tag{7}$$

We have thus expressed our original problem in terms of an operator T_0 which maps the Banach space $\mathbf{1}_\infty$ (the space of bounded sequences) into itself. We are interested in the behavior of X_l as l approaches infinity. The operator T_0 is the most general one. However, there is some interest in the behavior of the jitter when the pulse trains are periodic and the operator T_0 under steady-state periodic condition (cf. Part I) becomes

$$A_0 = \begin{bmatrix} \frac{\mathfrak{A}_0}{s_0'} & \frac{\mathfrak{A}_1\beta}{s_0'} & \cdots & \frac{\mathfrak{A}_{m-1}\beta^{m-1}}{s_0'} \\ \frac{\mathfrak{A}_0\beta^{m-1}}{s_1'} & \frac{\mathfrak{A}_1}{s_1'} & \cdots & \frac{\mathfrak{A}_{m-1}\beta^{m-2}}{s_1'} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\mathfrak{A}_0\beta}{s_{m-1}'} & \cdots & \cdots & \frac{\mathfrak{A}_{m-1}}{s_{m-1}'} \end{bmatrix}, \tag{8}$$

where $s_n' = (1 - \beta^m)s_n$ for all n. Finally, it is more convenient to write the operator T_0 explicitly to indicate only those positions where the pulses are present. This leads us to the operator

$$T = \begin{bmatrix} \frac{1}{S_0} & \frac{\beta^{i_1}}{S_0} & \frac{\beta^{i_1+i_2}}{S_0} & \cdots \\ 0 & \frac{1}{S_1} & \frac{\beta^{i_2}}{S_1} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \tag{9}$$

where

$$S_0 = s_0 \tag{10}$$

and

$$S_{n-1} = 1 + \beta^{i_n} S_n \,. \tag{11}$$

The vectors X_l are also assumed to be suitably modified. We will use the operator T of (9) to represent our quantities of interest.* The operator T_0 will give identical results. The special restriction to the periodic case will be of interest when we discuss computational aspects. The operator A_0 is also assumed suitably modified to A. Several parameters of interest may now be expressed in terms of the operator T and the injected jitter element X_1.

3.1 *Absolute Jitter*

From (5), we have

$$X_l = \left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1 \,, \tag{12}$$

and in the limit

$$Y = \lim_{l\to\infty} \left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1 \,. \tag{13}$$

3.2 *Dispersion*

This is obtained by subtracting the average delay from the absolute jitter. A delay element is represented to within a constant by $\{1,1,1, \cdots\}$. This element happens to be an eigenvector of T corresponding to an eigenvalue at $\lambda = 1$. Under the condition that $\lambda = 1$ is a pole (cf. Part II) of T, we can represent the dispersion by

$$X_l^{\,D} = (I - E_1)X_l = \left[I - E_1\right]\left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1 \,, \tag{14}$$

where E_1 is the projection operator E_1 $(\lambda = 1;\ T)$ which takes on the value "one" in the neighborhood of $\lambda = 1$ and zero elsewhere.

---

* It should be observed that the condition (2) is also modified. The right-hand side of (2) now becomes π/Q times the number of basic periods τ between two adjacent pulses.

3.3 *Spacing Jitter*

This is obtained by subtracting from the absolute jitter at one pulse position the absolute jitter at the adjacent pulse position (not necessarily the adjacent time slot).

$$X_l^s = \left[I - S\right]\left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1, \tag{15}$$

where S is the shift operator given by

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{16}$$

3.4 *Alignment Jitter*

This quantity is given by the difference of X_{l+1} and X_l :

$$X_{l+1}^A = X_{l+1} - X_l = T^l X_1 . \tag{17}$$

Finally, we summarize here without proof the properties of the operator T which were determined in Part II of this article.

(a) The operator T is a bounded operator mapping $\mathbf{1}_p$ into itself for $1 \leqq p \leqq \infty$. In particular, the norm of T in $\mathbf{1}_\infty$ is equal to one (i.e., $| T | = 1$).

(b) The spectrum of T is a subset of the unit disk (i.e., $| \sigma(T) | \leqq 1$), and any pole λ of T with $| \lambda | = 1$ has order one.

(c) All points in the unit circle except $\lambda = 1$ are in $\rho(T)$, the resolvent of T. The point $\lambda = 1$ is an eigenvalue of T with the eigenvector $\{1, 1, \cdots\}$. The dimension of the eigenmanifold is one in this case.

(d) The point $\lambda = 1$ is the limit point of the point spectrum of T if the domain of T is unrestricted. It is a pole of T for very special cases, such as periodic pulse trains, truncated pulse trains, etc.

IV. ABSOLUTE JITTER AND DISPERSION

The results for absolute jitter follow immediately from (13)

$$Y = \lim_{l \to \infty} \left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1 .$$

Observe that $\lambda = 1$ is an eigenvalue of T, and since in general X_1 is any element of $\mathbf{1}_\infty$, the limit in the above equation approaches infinity as l

approaches infinity. In particular, the norm of the operator

$$K_l = \left[\sum_{\nu=0}^{l-1} T^\nu\right] \tag{18}$$

is l. We thus observe that, in general, the absolute jitter grows linearly with the number of repeaters.

If $\lambda = 1$ is a pole of T, the dispersion is given by

$$\begin{aligned} X_{l+1}{}^D &= \left[I - E_1\right]\left[\sum_{\nu=0}^{l} T^\nu\right] X_1 , \\ &= \left[\sum_{\nu=0}^{l} T^\nu\right]\left[I - E_1\right] X_1 \qquad \text{(since } E_1T = TE_1\text{)}, \\ &= \left[\sum_{\nu=0}^{l} D^\nu\right]\left[E_D X_1\right] \qquad \text{(cf. Part II)}; \end{aligned} \tag{19}$$

where $D = TE_D$ and $E_D = (I - E_1)$. Here $E_D X_1$ is the dispersion element due to the first repeater. The norm of the operator

$$K_{l+1}{}^D = \left[\sum_{\nu=0}^{l} D^\nu\right] \tag{20}$$

is bounded and converges to a finite value as l approaches infinity. This follows from the previous discussion (see Part II) where it was shown that

$$| D^m | \leqq M{\alpha_0}^m , \tag{21}$$

where M is a positive constant and $\alpha_0 < 1$. We therefore find that the dispersion is bounded provided that $\lambda = 1$ is a pole of T. This is true for certain highly constrained situations. In particular, this is true when the domain of T is restricted to a finite dimensional subspace of $\mathbf{1}_\infty$ which is invariant under T. Examples of such cases occur when the pulse trains are periodic, finite, etc.

On the other hand, when $\lambda = 1$ is not a pole of T the projection E_1 does not commute with T. In this case,

$$X_{l+1}{}^D = \left[I - E_1\right]\left[\sum_{\nu=0}^{l} T^\nu\right] X_1 \tag{22}$$

where

$$E_1T \neq TE_1 . \tag{23}$$

Furthermore, it can be shown that

$$| (I - E_1)T^m | = 1 \qquad \text{for all } m. \tag{24}$$

This is a consequence of the spectral properties of T summarized in the previous section, viz., the point $\lambda = 1$ is a limit point of the point spectrum of T. In fact, we show in Part II of this paper that there exist elements in $\mathbf{1}_\infty$ such that the norm of the operator

$$K_{l+1}{}^D = \left[I - E_1\right]\left[\sum_{\nu=0}^{l} T^\nu\right] \tag{25}$$

is $(l + 1)$. It follows, therefore, that the dispersion grows without bound in the case of purely unconstrained pulse trains. The result is not altered even if some form of coding is provided to eliminate indefinitely long strings of zeros in the pulse trains.

V. SPACING JITTER

At the output of the lth repeater, the spacing jitter is given by (15),

$$X_l{}^s = \left[I - S\right]\left[\sum_{\nu=0}^{l-1} T^\nu\right] X_1 = [K_l{}^s]X_1 . \tag{26}$$

In particular, we are interested in knowing whether the quantity $X_l{}^s$ remains bounded as l approaches infinity. Secondly, if it remains bounded we wish to determine the least upper bound for each l. Since there are no restrictions on X_1 (other than the requirement of boundedness), we are interested in determining the norm in the limit of the operator $K_l{}^s$; or,

$$\lim_{l\to\infty} | K_l{}^s | = \lim_{l\to\infty} | (I - S)K_l | \tag{27}$$

when we know that

$$\lim_{l\to\infty} | K_l | \to \infty. \tag{28}$$

For physical systems we are also interested in $| K_l{}^s |$ for all l.

All of our results in this section depend upon an important lemma concerning the operator $K_l{}^s$. We assert that $K_l{}^s$ has a representation simpler than the one given in (26) and prove this assertion by verification. Define an operator

$$B = \text{diag}\cdot\{(S_0 - 1)^{-1}, (S_1 - 1)^{-1}, (S_2 - 1)^{-1}, \cdots\}, \tag{29}$$

where $S_n \neq 1$ are defined in (11).* Then we assert the following *Lemma:*

$$K_l^s = \left[I - S\right]\left[\sum_{\nu=0}^{l-1} T^\nu\right] = B[I - T^l]T^{-1}. \tag{30}$$

Proof: Observe that in (30)

$$B(I - T^l)T^{-1} = B(I - T)T^{-1}\left[\sum_{\nu=0}^{l-1} T^\nu\right].$$

So, we need merely show that

$$[I - S] = B(T^{-1} - I).$$

Or,

$$B^{-1} - B^{-1}S = T^{-1} - I,$$

where

$$\begin{aligned} B^{-1} &= \text{diag}\cdot\{(S_0 - 1), (S_1 - 1), \cdots\} \\ &= \text{diag}\cdot\{S_0, S_1, S_2, \cdots\} - I. \end{aligned}$$

Thus we need to show that

$$\text{diag}\cdot\{S_0, S_1, S_2, \cdots\} - B^{-1}S = T^{-1}.$$

But from (11),

$$(S_{n-1} - 1) = \beta^{i_n} S_n,$$

so

$$B^{-1} = \text{diag}\cdot\{\beta^{i_1} S_1, \beta^{i_2} S_2, \cdots\}.$$

Therefore, the lemma is proven if

$$T^{-1} = \begin{bmatrix} S_0 & -\beta^{i_1}S_1 & 0 & 0 & 0 & \cdots \\ 0 & S_1 & -\beta^{i_2}S_2 & 0 & 0 & \cdots \\ \cdots & 0 & S_2 & -\beta^{i_3}S_3 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{31}$$

The truth of the above statement is verified directly by considering the products $T^{-1}T = TT^{-1} = I$. The validity of (30) is thus proven.

* It is quite possible that the S_n approach unity. If any $S_n = 1$, it implies a finite pulse train, and the question is analyzed very simply in a finite dimensional space as was done for the periodic case. Such cases, however, do not give us information for the infinite pulse trains which are required for long chains of repeaters.

We can now state the following

Theorem: The operator K_l^s is bounded for all l if $\inf_n S_n \geqq \alpha > 1$.

Proof: From (30), we have

$$\begin{aligned} | K_l^s | &= | B(I - T^l)T^{-1} | \\ &\leqq | B | \, | I - T^l | \, | T^{-1} |. \end{aligned}$$

We know that $| T^l | = 1$ for all l and $| T^{-1} |$ is finite from (31). Hence $| K_l^s |$ is bounded if B is bounded, which it is if $\inf_n S_n \geqq \alpha > 1$. The theorem is thus proven.

Corollary:

$$\lim_{l \to \infty} | K_l^s | < \infty.$$

This follows trivially since the boundedness of K_l^s was proven independently of l. Our assumption that $\inf_n S_n \geqq \alpha > 1$ is a simple assertion of the fact that indefinitely long strings of zeros are ruled out on any communication channel.

Next, we wish to determine what precisely is the norm of the element

$$X_l^s = K_l^s X_1$$

as it relates to the norm of X_1. Let us recall that X_1 is the jitter injected (at a single repeater) referred to the output of the repeater. The injected jitter referred to the input of the repeater is

$$X = T^{-1}X_1, \tag{32}$$

where T^{-1} is defined in (31). We find it more convenient to work with X in what follows. It is obvious, of course, that our entire discussion could have been carried out in terms of X from the very start. We chose not to do so in order to avoid a premature discussion of T^{-1}. The sequence represented by X naturally satisfies condition (2). In fact, the use of X allows a much simpler comparison of the spacing jitter at different repeaters along the chain. We are interested in the behavior of

$$X_l^s = [K_l^s T]X \tag{33}$$

for each l. This is obtained by a precise evaluation of the norm of

$$R_l^s = K_l^s T = B(I - T^l). \tag{34}$$

Let us first define $S_{\text{inf}} = \inf_n S_n$. Then, we state the following

Theorem: The norm of R_l^s is given by

$$| R_l^s | = \frac{2}{S_{\text{inf}}} \sum_{\nu=0}^{l-1} \frac{1}{S_{\text{inf}}^{\nu}}. \tag{35}$$

Proof:

$$R_l^s = B(I - T^l).$$

Let us consider the representation of $(I - T^l)$. The diagonal elements of $(I - T^l)$ are of the form $[1 - (1/S_n{}^l)]$, whereas the off-diagonal elements are all negative. Also, the sum of the elements in each row must be zero. If we multiply $(I - T^l)$ on the left by B, then the diagonal elements are of the form $[1/(S_n - 1)][1 - (1/S_n{}^l)]$. Again the off-diagonal elements are all negative and the sum of the elements in each row of $B(I - T^l)$ is zero. Hence, in $\mathbf{1}_\infty$

$$| B(I - T^l)| = 2\left(\frac{1}{S_{\text{inf}} - 1}\right)\left(1 - \frac{1}{S_{\text{inf}}^l}\right)$$

$$= \frac{2}{S_{\text{inf}}} \sum_{\nu=0}^{l-1} \frac{1}{S_{\text{inf}}^{\nu}}.$$

The theorem is thus proven. Observe that this is not just a bound but a precise norm. This value in the norm is taken by the spacing jitter X_l^s for some X whose norm is one. In other words, the value in (35) represents the maximum magnification of X that is possible to yield the value of the spacing jitter. It is interesting to note that this worst case occurs for each l for the same element X, viz., $\{\cdots, 1, -1, -1, \cdots\}$, where the $+1$ corresponds to the position of $1/S_{\text{inf}}$ in the matrix representing the operator T. Finally, to observe the maximum possible growth of the spacing jitter as it compares with the maximum possible spacing jitter at a single repeater, we compare the results for $l = 1$ and $l = \infty$.

$$R_1^s = \frac{2}{S_{\text{inf}}} \tag{36}$$

$$R_\infty^s = \frac{2}{S_{\text{inf}} - 1}. \tag{37}$$

The ratio of the quantities R_∞^s to R_1^s is less than two if S_{inf} is at least greater than two, which is to be expected in most physical situations. For example, if Q is of the order of 100 and there is at least one pulse

present in fifteen time slots, then $S_{\inf}$ is at least two or greater. Thus, we observe that not only is there a bound on the growth of the spacing jitter but that the growth is monotonic and levels off rather fast. Of course, this is a rather general result which, if desired, can be more specifically stated in terms of the a priori probability distributions of the binary signal. Furthermore, the importance of this result lies in specifying the conditions for the validity of our model. Under most realistic situations it should be clear that conditions (2) are met at every repeater. Of course, there are situations when these conditions are not met and the results obtained by the use of our model [cf. (3)] can no more be relied upon. However, we show that under most situations the results are reliable and very optimistic, as shown by (36) and (37).

The above results are the most crucial ones in this paper. The rest of the paper is devoted to a varied miscellany that has some bearing on different aspects of the timing problem.

VI. ALIGNMENT JITTER

The alignment jitter in the $(l + 1)$th repeater is given by (17),

$$X_{l+1}{}^{A} = T^{l} X_1 .$$

Obviously, for all l the alignment jitter is no greater in the norm than the absolute jitter X_1. This follows trivially since $\| T^{l} \|$ does not exceed one for any l.

For more detailed insight into the behavior of the alignment jitter, we need to discuss specific situations.

(a) If $\lambda = 1$ is a pole of T, then T^{l} converges to T^{∞} and $X_{l+1}{}^{A}$ settles down to a flat delay element for large l.

(b) If $\lambda = 1$ is not a pole of T, and $\inf_n S_n = 1$, T^{l} does not converge to T^{∞}. However, for large l the alignment jitter is slowly-varying.

(c) If $\lambda = 1$ is not a pole of T, and $\inf_n S_n = \alpha > 1$, the alignment jitter (for large l) varies even more slowly than it does in (b).

All the above results follow from the properties of T. If $\lambda = 1$ is a pole of T, the result is obvious (cf. Part I). If it is not a pole of T, the results follow from the fine structure of the spectrum of T. For example, in situation (b), all points in the point spectrum except $\lambda = 1$ are poles of T, whereas this is not true in situation (c). The corresponding eigenvectors have different structures for the two cases (cf. Part II).

It follows that the situation of (a) is to be preferred over that of (c), which in turn is preferable to that of (b).

VII. NONIDENTICAL REPEATERS

This is a rather difficult matter to discuss with any great generality. What we hope to do here is to briefly indicate perhaps the most convenient formulation and to give some reasons for believing that the orders of magnitude of the jitter parameters are not changed for small differences in the repeaters.

There are essentially three possible ways in which the repeaters may differ: (*i*) injected jitter, (*ii*) Q of the repeater, and (*iii*) mistuning. These differences appear mathematically in terms of different operators, multiplicative coefficients in the power series, and so on. We examine each of these separately.

If we assume that the injected jitter differs slightly at each repeater, then we may write (5) as

$$X_l = TX_{l-1} + X_{\text{av}} + \Delta_l , \tag{38}$$

where X_{av} is the average injected jitter and Δ_l represents the deviation from this average in the lth repeater. The norm of X_l differs at most from the previous case of identical repeaters by

$$| \Delta_l + T\Delta_{l-1} + T^2\Delta_{l-2} + \cdots + T^{l-1}\Delta_1 | \leqq \sum_{i=1}^{l} | \Delta_i |. \tag{39}$$

If the $| \Delta_i |$ are quite small, it is clear that the results will not be appreciably different from the previous ones.

If the Q's are different, then our basic operator is different for each repeater. It would be almost impossible to analyze such a case in general. However, we can make certain observations if we put

$$T_l = T + K_l , \tag{40}$$

where T_l is the operator representing the lth repeater, which is assumed to be an operator T perturbed by an operator K_l. It is reasonable to expect $| K_l | \ll 1$. Then (5) becomes

$$X_l = TX_{l-1} + X_1 + K_lX_{l-1} . \tag{41}$$

If $| K_l | \leqq \epsilon \ll 1$, then the norm of X_l does not differ from the previous results by more than

$$\left(\frac{\epsilon}{1-\epsilon}\right) | X_1 | \approx \epsilon | X_1 |. \tag{42}$$

Again, we see that the results are not appreciably different.

In the case of mistuning, (5) takes essentially the same form as (38),

$$X_l = TX_{l-1} + X_A + \epsilon_l X_B , \tag{43}$$

where X_A is the injected jitter due to sources other than mistuning, ϵ_l is the mistuning in the lth repeater, and $\epsilon_l X_B$ is the jitter due to mistuning (cf. Part I). Then

$$X_l = \left[\sum_{\nu=0}^{l-1} T^\nu\right] X_A + \left[\sum_{\nu=0}^{l-1} \epsilon_{l-\nu} T^\nu\right] X_B . \tag{44}$$

The contribution due to X_A is unaltered and the contribution due to X_B would depend on the specifications of ϵ_i. However, if we assume that the magnitudes of ϵ_i do not exceed one, then the contribution in the norm due to X_B cannot exceed l times the norm of X_B. Thus the worst case for the absolute jitter does indeed arise from the assumption of equal ϵ_i at their maximum values.

For spacing jitter we can make a slightly different statement for the contribution due to X_B. For one repeater, the worst case (in the norm sense) occurs for the maximum value of $\epsilon_1 = 1$, then the worst case for two repeaters is obtained by setting $\epsilon_2 = 1$. Setting the first two repeaters with $\epsilon_1 = \epsilon_2 = 1$, the worst case for a string of three repeaters is obtained by setting $\epsilon_3 = 1$ and so on. The statement is a simple consequence of the inequality

$$| (I - T^{l-1}) | < | (I - T^l) |. \tag{45}$$

It is believed that a similar statement can be made for the alignment jitter.

We thus observe that, when the differences in the repeaters are small, it is reasonable to expect that the results are not appreciably different from those obtained by assuming identical repeaters.

VIII. CONCLUSION

We have presented a general and rigorous theory of the jitter accumulation in a chain of regenerative repeaters. The sources of jitter are assumed to be the signal-dependent sources, as distinguished from purely random sources independent of the signal.

Our results show that while the absolute jitter and its dispersion grow without bound with the number of repeaters, the spacing and the alignment jitter remain bounded. In particular, the spacing jitter bounds are quite optimistic for most practical situations, viz., no greater than twice the absolute jitter injected at a single repeater. This result is of importance in that it ensures proper decoding of the binary signal. Its further importance lies in the fact that it does ensure, in most cases, the validity of the basic model and thus the validity of other results obtained by that model. One such result shows that the alignment jitter is slowly-varying

for repeaters further along the chain. Finally, a brief discussion shows that the assumption of identical repeaters leads to results which are of the same order of magnitude as would be obtained if the repeaters differed by not too great an amount. Some results which would be of use in computations are to be found in the Appendix, together with an example.

In our discussion so far, we have investigated the accumulation properties of jitter due to repeated regeneration. We have made no attempt to determine the jitter introduced by a single repeater. Analytically, this problem is complicated not only by the nonlinearities involved, but also by a lack of complete knowledge as to the actual mechanisms involved. We propose instead, in a later paper, an experimental approach which allows these measurements to be carried out under steady-state conditions. This experiment also has some bearing on the question of simulation of long chains of repeaters.

IX. ACKNOWLEDGMENTS

I wish to take this opportunity to express my appreciation for the help I received in this entire endeavor from several of my colleagues. In particular, I wish to thank B. Liu and M. R. Aaron for many helpful discussions and criticisms. For their advice and encouragement, my thanks go to J. A. Young and M. Karnaugh.

APPENDIX

Eigenvalues and Eigenvectors of T

Practical systems call for the evaluation of jitter when the number of repeaters is finite. In such cases, we need not concern ourselves with infinite pulse trains. So long as the pulse trains are much longer than the effective memory of the system, the results obtained by considering finite pulse trains will be reliable. The results will also be reliable if the pulse trains are considered periodic with the period being greater than the memory of the system. Actually, as we shall see, the periodic pulse trains are much more difficult to work with than the finite ones. However, we shall discuss the periodic case in detail since much of the experimental work is carried out using periodic pulse trains. Finally, another case of interest is that in which there is a certain quiescent pattern which changes to a different one. This would include a periodic pattern changing to either a nonperiodic or a different periodic pattern. In each case, our interest is in determining the set of eigenvectors. Computations can then be carried out by expressing the injected jitter element in terms of the

eigenvectors (cf. Part I). If the set of eigenvectors is not complete, we employ the standard procedure and use the so-called generalized eigenvectors. It should also be observed that the case of nonidentical repeaters is also handled using the same techniques. We present here certain simple algorithms for determining the eigenvalues and the eigenvectors for the several cases of interest.

A.1 *Truncated Pulse Trains*

This is the case where the pulse train is finite. The matrix T in (9) is now finite, upper triangular and the diagonal element in the last row is unity. Such a matrix also arises in the case of pattern transitions where originally the quiescent pattern is periodic with only one pulse present in each period. This is also the more realistic case because the tuned circuit is thus properly excited. We consider the truncated case, therefore, together with the pattern transition case.

A.2 *Pattern Transitions*

Here we have a steady-state periodic pattern which changes to a different pattern. The operator T can be represented as

$$T = \begin{bmatrix} P & C \\ 0 & A \end{bmatrix},$$

where P, representing the new pattern, is an upper triangular square matrix, A represents the quiescent periodic pattern, 0 is the null matrix, and C is the connecting matrix. The matrix A is either an arbitrary positive stochastic matrix for an arbitrary periodic pattern or it is a scalar (viz., unity) for the case of only one pulse present in each period. The latter case also occurs when the tuned circuit is excited by a reference pulse train such as $101010 \cdots$.

In either case, the eigenvalues of T are given by the eigenvalues of P and those of A. The eigenvalues of the matrix A are discussed in the section dealing with the periodic case. The eigenvalues of P are given by the diagonal elements S_n^{-1}. If A is a scalar, the only other eigenvalue is unity. All the eigenvalues are thus determined.

Next, we observe that the eigenvalues of P are distinct. If not, for some n (say, $n = 0$),

$$S_0 = 1 + \beta^{i_1} + \beta^{i_1+i_2} + \cdots + [\beta^{i_1+i_2+\cdots+i_m}]S_0 ,$$

which implies a steady-state periodic pulse train, contradicting the

transient nature of P. Thus the eigenvalues of T are distinct unless, of course, there is a periodic pattern involved. Let us reserve the periodic case for the next section. Then, the eigenvectors of T form a basis for the space of jitter vectors both for the truncated case and the pattern transition case when the quiescent pattern has only one pulse present in each period.

The eigenvectors are given by the algorithm

$$\xi_{n+1} = \frac{S_n - \frac{1}{\lambda}}{S_n - 1}\,\xi_n$$

(cf. Part II) for each eigenvalue λ.

The eigenvalues and the corresponding eigenvectors are thus very simply determined when either the reference pattern has only one pulse in each period or the pulse train is finite.

A.3 *Periodic Patterns*

Let us start by assuming that the period is m and there are n pulses in a period. Then

$$m = i_1 + i_2 + \cdots + i_n .$$

Let $\alpha_0 = (1 - \beta^m)$, and $D_i = S_i\alpha_0$. Then,

$$A = \begin{bmatrix} \frac{1}{D_0} & \frac{\beta^{i_1}}{D_0} & \frac{\beta^{i_1+i_2}}{D_0} & \cdots & \frac{\beta^{i_1+\cdots+i_{n-1}}}{D_0} \\ \frac{\beta^{i_2+\cdots+i_n}}{D_1} & \frac{1}{D_1} & \frac{\beta^{i_2}}{D_1} & \cdots & \frac{\beta^{i_2+\cdots+i_{n-1}}}{D_1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\beta^{i_n}}{D_{n-1}} & \frac{\beta^{i_n+i_1}}{D_{n-1}} & \cdots & \cdots & \frac{1}{D_{n-1}} \end{bmatrix},$$

$$\Delta = \det(\lambda I - A)$$

$$= \frac{1}{D_0 D_1 \cdots D_{n-1}} \begin{vmatrix} \lambda D_0 - 1 & -\beta^{i_1} & \cdots & -\beta^{i_1+\cdots+i_{n-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ -\beta^{i_n} & -\beta^{i_n+i_1} & \cdots & \lambda D_{n-1} - 1 \end{vmatrix}.$$

$$\Delta = \frac{1}{D_0 D_1 \cdots D_{n-1}} \cdot \begin{vmatrix} (\lambda D_0 - \alpha_0) & -\beta^{i_1} D_1 \lambda & 0 & 0 & \cdots & 0 \\ 0 & (\lambda D_1 - \alpha_0) & -\beta^{i_2} D_2 \lambda & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & & \cdots & \cdots \\ -\beta^{i_n} & -\beta^{i_n + i_1} & \cdots & & \cdots & (\lambda D_{n-1} - 1) \end{vmatrix}$$

$$= \left[-\frac{1}{D_0} \beta^m \lambda^{n-1} + \left(\lambda - \frac{\alpha_0}{D_0}\right)\left(-\frac{1}{D_1} \beta^m \lambda^{n-2}\right) + \cdots \right.$$

$$+ \left(\lambda - \frac{\alpha_0}{D_0}\right)\left(\lambda - \frac{\alpha_0}{D_1}\right) \cdots \left(\lambda - \frac{\alpha_0}{D_{n-3}}\right)\left(-\frac{1}{D_{n-2}} \beta^m \lambda\right)$$

$$\left. + \left(\lambda - \frac{\alpha_0}{D_0}\right) \cdots \left(\lambda - \frac{\alpha_0}{D_{n-2}}\right)\left(\lambda - \frac{1}{D_{n-1}}\right) \right]$$

$$= P(\lambda) + Q(\lambda),$$

where

$$P(\lambda) = \sum_{i=0}^{n-1} \left(\lambda - \frac{\alpha_0}{D_i}\right)$$

and

$$Q(\lambda) = \sum_{\nu=0}^{n=1} - \frac{1}{D_\nu} \beta^m \lambda^{n-\nu-1} \prod_{k=1}^{\nu} \left(\lambda - \frac{\alpha_0}{D_{\nu-k}}\right).$$

Finally, after some manipulation it can be shown that

$$\Delta = \left(\frac{\alpha_{0-1}}{\alpha_0}\right) \lambda^n + \frac{1}{\alpha_0} P(\lambda).$$

Thus the eigenvalues of A are given by the zeros of the polynomial

$$R(\lambda) = P(\lambda) - \beta^m \lambda^n,$$

when

$$P(\lambda) = \prod_{i=0}^{n-1} (\lambda - S_i^{-1}).$$

The zeros of $R(\lambda)$ can be obtained from those of $P(\lambda)$ by root-loci considerations or other numerical methods. Since β^m is usually small, this can be handled easily on a digital computer.

Next, let us discuss the eigenvectors corresponding to the eigenvalues given by the zeros of $R(\lambda)$. We show that for each distinct eigenvalue of A, the eigenvector is given by the algorithm

$$\xi_{k+1} = \frac{S_k - \frac{1}{\lambda}}{S_k - 1}\,\xi_k .$$

This would be true if $\xi_{k+n} = \xi_k$ for all k or, if

$$\frac{\left(S_{n-1} - \frac{1}{\lambda}\right)\left(S_{n-2} - \frac{1}{\lambda}\right)\cdots\left(S_0 - \frac{1}{\lambda}\right)}{(S_{n-1} - 1)(S_{n-2} - 2)\cdots(S_0 - 1)} = 1.$$

The above is true if

$$\begin{aligned} S_{n-1}S_{n-2}\cdots S_0\left(\frac{1}{\lambda^n}\right)P(\lambda) &= \prod_{k=1}^{n}(S_{k-1} - 1) \\ &= \prod_{k=1}^{n}(\beta^{i_k}S_k) \\ &= \beta^m S_1 S_2 \cdots S_n \\ &= \beta^m S_0 S_1 S_2 \cdots S_{n-1} \qquad (\text{since } S_n = S_0), \end{aligned}$$

or if

$$P(\lambda) = \beta^m \lambda^n,$$

which is indeed true for every eigenvalue λ. For a repeated eigenvalue λ, if it exists, one must find a generalized eigenvector in the usual way.

We have thus given simple algorithms for determining the eigenvalues and the corresponding eigenvectors of A.

It should be mentioned that the algorithm for eigenvectors is the same as the one given above when A is a submatrix of T as in Section A.2.

A.4 *Example*

Let us consider a simple example to illustrate some of the points. Consider a system with repeaters having $Q = 100$ and signals having at least one pulse present in 10 time slots. Consider the case of a quiescent pattern (101010 $\cdots$ 10) suddenly changing to a new periodic pattern with one pulse in 10 time slots. We are interested in determining the behavior of the first pulse after the transition.

The operator T has the form

$$T = \begin{bmatrix} \frac{1}{1+\beta^{10}(1+\beta^2+\beta^4+\cdots)} & \frac{\beta^{10}}{1+\beta^{10}(1+\beta^2+\beta^4+\cdots)} & \cdots & \cdots \\ 0 & \frac{1}{(1+\beta^2+\beta^4+\cdots)} & \frac{\beta^2}{(\ )} & \cdots \\ 0 & 0 & \frac{1}{(1+\beta^2+\beta^4+\cdots)} & \cdots \\ \cdots & \cdots & \cdots & \end{bmatrix}.$$

Let the jitter vector $X_1 = \{\xi_0, \xi_1, \xi_1, \xi_1, \cdots\}$, where ξ_1 is the reference phase jitter for the quiescent pattern and $(\xi_0 - \xi_1)$ is the change due to the transition.

In our simple example it is clear that there are only two eigenvalues of T, viz., $\lambda_0 = (1 - \beta^2)/(1 - \beta^2 + \beta^{10})$ and $\lambda_1 = 1$. The eigenvectors corresponding to these eigenvalues are $e_0 = \{(\xi_0 - \xi_1), 0, 0, \cdots\}$ and $e_1 = \{\xi_1, \xi_1, \xi_1, \cdots\}$.

The absolute jitter obviously increases without bound with the number of repeaters. The dispersion remains bounded, since $\lambda_1 = 1$ is a pole of T in our simple case. In the limit, the dispersion is given by

$$X^D = \left(\frac{1}{1 - \lambda_0}\right) e_0 = \left(\frac{1 - \beta^2 + \beta^{10}}{\beta^{10}}\right) e_0.$$

The spacing jitter in the limit is given by

$$\left(\frac{1 - \beta^2 + \beta^{10}}{\beta^{10}}\right) (\xi_0 - \xi_1)$$

in the zeroth position and zero elsewhere. Finally, the alignment jitter approaches e_1 in the limit.

The validity of the results is assured if

$$\left(\frac{1 - \beta^2 + \beta^{10}}{\beta^{10}}\right) (\xi_0 - \xi_1) \ll 10 \frac{\pi}{Q},$$

or if

$$(\xi_0 - \xi_1) \ll 10 \frac{\pi}{Q} \left(\frac{\beta^{10}}{1 - \beta^2 + \beta^{10}}\right) \approx \frac{\pi}{10}.$$

Thus, if the *jump* in the phase jitter, for a single repeater, due to the transition in pattern is much smaller than 18°, our results are valid. This requirement can be expected to be satisfied by most practical repeaters.[5]

REFERENCES

1. Sunde, E. D., Self-Timing Regenerative Repeaters, B.S.T.J., **36,** July, 1957, pp. 891–938.
2. Rowe, H. E., Timing in a Long Chain of Regenerative Binary Repeaters, B.S.T.J., **37,** Nov., 1958, pp. 1543–1598.
3. Kinariwala, B. K., Timing Errors in a Chain of Regenerative Repeaters, I and II, B.S.T.J., **41,** Nov., 1962, pp. 1769–1797.
4. Bennett, W. R., Statistics of Regenerative Digital Transmission, B.S.T.J., **37,** Nov., 1958, pp. 1501–1542.
5. Byrne, C. J., Karafin, B. J., and Robinson, D. B., Jr., Systematic Jitter in a Chain of Digital Regenerators, B.S.T.J., **42,** Nov., 1963, pp. 2679–2714.
6. Rice, S. O., unpublished work.
7. Aaron, M. R., and Gray, J. R., Probability Distribution for the Phase Jitter in Self-Timed Reconstructive Repeaters for PCM, B.S.T.J., **41,** Mar., 1962, pp. 503–558.
8. Delange, O. E., The Timing of High-Speed Regenerative Repeaters, B.S.T.J., **37,** Nov., 1958, pp. 1455–1486.
9. Delange, O. E., and Pustelnyk, M., Experiments on the Timing of Regenerative Repeaters, B.S.T.J., **37,** Nov., 1958, pp. 1487–1500.

The Maser Rate Equations and Spiking

By D. A. KLEINMAN

(Manuscript received March 6, 1964)

The rate equations of Statz and De Mars giving the time development of the inversion and photon number in a maser or laser are discussed analytically with the aid of a mechanical analogy in which a particle moves in a potential well under the influence of a viscous damping force. The coordinate of this particle is analogous to the logarithm of the light output of the laser, and the amplitude, period, and damping of the motion can be directly related to the parameters of the rate equations. Simple analytic approximations are developed for all of the quantities of experimental interest in the spiking pattern of a laser. Four relationships are given, which do not contain any of the rate equation parameters, whereby a spike pattern can be tested to determine if it is consistent with the usual rate equations. Systematic procedures are described for extracting all of the information contained in spike patterns.

I. INTRODUCTION

The most fruitful approach for the discussion of maser and laser[1] behavior has been through rate equations describing the time rates of change of the atomic populations and the photon numbers of the electromagnetic field. Bloembergen[2] introduced rate equations for the populations in a paramagnetic maser and based his discussion on the steady-state solutions without explicitly considering the photon field. On the other hand, Shimoda, Takahasi, and Townes[3] have considered the photon rate equations and on this basis have given a theory of maser amplification without explicitly considering the atomic populations. Statz and De Mars[4] have shown that the transient behavior of masers depends upon coupled rate equations for both the populations and the photons. A number of authors have rederived these equations and discussed their applications to various maser systems. Considerable attention has been given to the question of whether these equations have periodic (undamped) solutions. It has been shown by Makhov[5] and by Sinnett[6] that the small-signal solutions are always damped, and it has been pre-

sumed and often confirmed by numerical computations that the same is true in the large-signal domain. Statz, et al.[7] have suggested that the damping of experimentally observed "spikes" in the output of lasers is probably sensitive to coherence and noise conditions, and not necessarily related to the damping predicted by the usual rate equations. They also suggest that the complicated spiking patterns frequently observed[8] are due to oscillation in many modes of the laser cavity. Despite these doubts which have been cast on the adequacy of the Statz-De Mars rate equations for describing laser behavior, they still remain the logical starting point for any discussion of the power output of lasers, whether it be transient or steady-state, and of the dependence of power on the quality of the cavity, the intensity of the pumping light, or the linewidth, relaxation time, and concentration of the active atoms.

Several ingenious suggestions have been made for modifying the rate equations so as to obtain periodic solutions. Statz and De Mars[4] and Makhov[5] propose that periodic solutions are obtained if terms are added to the rate equations representing cross relaxation in the inhomogeneously broadened maser transition. On the other hand, Shimoda[9] suggests that periodic spiking can result if the losses due to absorption in the cavity can be partially saturated by the buildup of laser oscillations. Although these and other modifications may ultimately prove to be justified and necessary in laser theory, we shall confine our attention in this paper to the original Statz-De Mars equations which relate the photons in a single cavity mode to a single quantity, the *inversion*, describing the atomic populations. We shall make it our task to understand as fully as possible the damped oscillatory solutions of these equations and how they may be applied to the study of the spiking phenomenon seen[8] in solid-state lasers.

Rate equations in the simple form have been successfully applied by McClung and Hellwarth[10] and by Vuylsteke[11] to the giant-pulse laser, which produces a single very short and very intense burst of radiation. Wagner and Lengyel[12] have shown that an exact analytic solution can be obtained to a simplified rate equation which neglects spontaneous emission and pumping during the pulse. By also neglecting the loss of photons in the cavity Dunsmuir[13] has obtained a still simpler analytic solution which is applicable to the rising portion of a spike or a giant pulse. We shall not consider further these exact solutions or the giant-pulse laser in this paper, but confine ourselves to those solutions representing repetitive pulsations, or spikes, in the ordinary laser. It is in this field that the greatest need now exists for an analytical discussion of the solutions of the rate equations.

A number of authors have endeavored to put the rate equations on a firmer theoretical basis. Following Anderson[14] and Clogston,[15] who first described masers in terms of the density matrix, treatments using the density matrix to derive rate equations have been given by Fain, et al.,[16] Kaplan and Zier,[17] and Pao.[18] By considering directly, without explicit use of the density matrix, the correlation functions of the electromagnetic field which are measured in simple maser experiments, McCumber[19] has shown that the maser medium acts like a dielectric of negative conductivity; he concludes that the field and the dielectric satisfy rate equations of the usual form, providing that the populations change by a small fractional amount during a coherence time (reciprocal linewidth) of the atomic system. Another formal theory, based on a successive approximation approach to the quantum mechanical equations of motion, has been applied by Haken and Sauermann[20] to the frequency shifts and interactions of cavity modes in the laser. An extensive survey with bibliography of the early formal work has been given by Lamb.[21] In the present paper we shall not go into the theoretical basis for the rate equations, but confine ourselves entirely to the problem of solving the equations in the large-signal domain.

Despite the fact that the rate equations are generally accepted, and that the general nature of the solutions has been familiar from digital computer calculations for some time,[5,13,17,22] it is still quite inconvenient in any particular case to compare observed spiking patterns in solid-state lasers with predictions of the rate equations. It has been necessary either to use the small-signal solutions and hope that they are not too inaccurate in the large-signal domain, or to resort to machine calculations and try to fit three or more parameters to the data by trial and error. To be sure, much of the data on spiking is not amenable to analysis, consisting apparently of random spikes with widely varying amplitude, duration, and interval. Nevertheless, several laser systems are now known to give very regular pulsations of the type that might be consistent with the rate equations. Regular spiking patterns have been observed in $CaF_2:U^{+3}$ by Sorokin and Stevenson[23] and by O'Connor and Bostick,[24] and in $CaWO_4:N_d{}^{+3}$ by Johnson and Nassau.[25] Recently the effect has also been seen in a highly perfect ruby by Nelson and Remeika,[26] and in a confocal ruby by Johnson, et al.[27] More extensive studies of highly regular spiking have been reported by Gürs[28] and by Hercher.[29] Thus it is clear that good data on spiking patterns can be obtained, and it therefore becomes cogent to inquire into practical and convenient means for analyzing this data and obtaining information from it.

The information contained in spike patterns is of two distinct kinds,

which we may call qualitative and quantitative. Qualitatively, we can determine quickly by means of relationships given here whether a spike pattern is *consistent* with the rate equations. If we determine that certain patterns are not consistent, we are spared the waste of time that would result from attempting to fit these patterns numerically by trial and error. It is not obvious at this writing that any of the regular patterns that have been reported are consistent, because the application of the consistency relationships requires a measurement of the ratio of the peaks to the valleys in the light output, and only the peaks are seen in pictures published so far. Thus we suggest that by extending spike studies to include the valleys new and interesting information can be obtained. It is to be expected that patterns which are regular but yet not consistent will turn out to be physically the most interesting of all, since they will point the way to new understanding of laser behavior. The quantitative information can only be obtained from patterns which are consistent, at least in some average sense, and consists in obtaining values for the physical constants which appear in the rate equations:

N = the *number* of active laser atoms in the optical cavity

t_r = the *relaxation* time of the upper laser level, usually due to spontaneous emission

t_p = the *photon* lifetime in the laser mode of the cavity

t_m = the *mode* time, the time for spontaneous emission into the laser mode

t_g = the *ground* state time, the time spent by an atom in the ground state before being excited by the pump

s = the *source* strength, the rate of production of laser photons by spontaneous emission, the pumping light, or any other noise source in the cavity, or any signal applied to the cavity. Although s may vary with time, it is convenient here to consider it with the constants.

In principle all of these quantities except s could be directly measured or calculated from independent measurements on the laser material, the cavity, and the pump. The spiking data would then serve as confirmatory evidence. In many cases, however, spike patterns may prove to be the most convenient method of measurement. The last quantity, s, is in some respects the most interesting. The first assumption would be that s is due entirely to spontaneous emission into the laser mode; if so, s could be calculated and comparison with the measured value would reveal the verity of the assumption. Experiments could be carried out with an external weak signal from a monochromator to test the response of the laser as observed in its spiking patterns. Thus we hope that the analysis given here will help to stimulate new experiments by making rate equation analysis more convenient for the experimentalist.

II. FORMULATION OF THE RATE EQUATIONS

The laser is a system consisting essentially of an *optical cavity* with very high Q in a few modes, low Q in all other modes, the *laser medium* containing the active atoms, and a *pump*, usually an intense light source, to excite the atoms into a broad band of excited states. It is a property of the laser medium that the atoms decay from these states in an extremely short time by nonradiative processes to one or more very sharp excited states, called the upper laser levels. The upper laser levels can decay radiatively to a sharp lower level, called the lower laser level, which may be the ground state. If the lower laser level is not the ground state, we shall assume that very rapid nonradiative decay processes return the atom to the ground state. Thus we may always neglect the population of the pumping band and of the lower laser level if the latter is not the ground state. The laser transition takes place from the lowest of the upper laser levels, but it may sometimes be necessary to take into account the populations of nearby levels in thermal equilibrium with this level. The statistical weights of the laser levels must also be taken into account.[30]

The rate equations may always be written in the form (p = photon number, n = inversion)

$$\frac{dp}{dt} = -\frac{p}{t_p} + \frac{pn}{t_m} + s \tag{1}$$

$$\frac{dn}{dt} = \frac{n_0 - n}{t_0} - a\frac{pn}{t_m} \tag{2}$$

as long as we consider only a single mode of the cavity, the laser mode, and neglect all atomic populations except the upper laser level and the ground state. The time t_0 might be called the pumping relaxation time

$$\frac{1}{t_0} = \frac{1}{t_g} + \frac{1}{t_r}, \tag{3}$$

since it represents the characteristic time in the response of the population inversion n to the pump. The inversion n may always be written

$$n = N_u - (1/w)N_l, \tag{4}$$

where N_u, N_l are the populations of the upper and lower laser levels respectively and w is the statistical weight of the lower relative to that of the upper laser level. In view of our assumptions, w will come in only when the lower laser level is the ground state. Let us suppose that the upper laser level is in thermal equilibrium[26,30] with certain other states not directly involved in the laser transition such that there is a tempera-

ture-dependent probability P that an excited atom is in the upper laser level. Then we have for a *three-level system*, in which the lower laser level is the ground state,

$$\begin{aligned} a &= P + \frac{1}{w} \\ n_0 &= PN\, t_0 \left(\frac{1}{t_g} - \frac{1}{Pwt_r} \right); \end{aligned} \tag{5}$$

and for a *four-level system*, in which the lower laser level is not the ground state, we have

$$\begin{aligned} a &= P \\ n_0 &= PN\, t_0/t_g\,. \end{aligned} \tag{6}$$

If the relaxation of the upper laser level is predominantly by spontaneous emission to the lower laser level, there is a simple relation between t_r, t_m, and the linewidth $\Delta\nu$ (cps) of the laser transition

$$t_m = (8\pi\nu^2 n_{\text{ref}}{}^3/c^3)\, V\, t_r \Delta\nu, \tag{7}$$

where V is the volume of the cavity and n_{ref} is the refractive index. This relation should not be taken too literally, since it takes no account of the anisotropy of the laser medium or the polarization properties of the transition, and V would have to be replaced by a suitable effective "optical volume" if the laser material does not fill the cavity. Nevertheless, it points out the important fact that t_m varies with temperature in the same way as $\Delta\nu$ and therefore is subject to control in spiking studies. Another constant subject to convenient control is t_g, since $1/t_g$ is proportional to the pump intensity. Even when flash lamps are used for pumping it is still approximately valid to assume t_g is constant, since the time constant for the flash will usually be much longer than the interval between spikes. To assure that this is true, it would be advantageous to have the spike pattern commence when the flash is at its maximum intensity. It is not valid to assume without investigation that $1/t_g$ is proportional to the total energy dissipated in the flash. Spiking data should always include a record of the flash as a function of time and an indication of when the spike pattern occurred. Also subject to experimental control is t_p, the photon lifetime in the laser mode of the cavity. Presumably the losses in a good laser cavity can be estimated rather reliably, so that t_p can usually be directly calculated. The total number N of active laser atoms can ordinarily be determined in a given sample, but N does not appear to be a convenient parameter to vary in laser experi-

ments. The signal or source strength s is best determined from an analysis of the spiking data as described in this paper. Although $s = s(t)$ will in general be a function of time, it will be shown that the spiking pattern depends only on the value of s at the instant when the net losses in the laser mode vanish due to the buildup of inversion. We shall denote this time by t_1 and the critical inversion by n_1, where

$$n_1 = t_m/t_p \tag{8}$$

is the celebrated Schawlow-Townes[1] criterion for the buildup of laser oscillation.

We now introduce dimensionless variables and parameters; in terms of the variables

$$\begin{aligned} \tau &= t/t_p \\ \eta &= n(t_p/t_m) = n/n_1 \\ \rho &= p(a\ t_0/t_m), \end{aligned} \tag{9}$$

and the parameters

$$\begin{aligned} \omega &= t_p/t_0 \\ \xi &= n_0(t_p/t_m) = n_0/n_1 \\ \sigma &= s(a\ t_p t_0/t_m), \end{aligned} \tag{10}$$

the rate equations (1), (2) become

$$\dot{\rho} = \frac{d\rho}{d\tau} = \sigma - \rho + \rho\eta \tag{11}$$

$$\dot{\eta} = \frac{d\eta}{d\tau} = \omega(\xi - \eta - \rho\eta). \tag{12}$$

Here ρ represents the *photons*, η the *inversion*, and τ the *time*, while ω represents the *pumping rate*, ξ the *limiting inversion* toward which the pump is tending to drive the system, and σ the *source*. We see that there are really only three parameters in the rate equations; it follows that three relationships among the six relevant physical parameters with which we started can be obtained from spiking studies. Although we shall assume in our analysis that ω and ξ are constant, our results will provide a useful adiabatic approximation for the case of slowly varying ω, ξ. Typical values of the parameters for a ruby laser will be given in the discussion of a numerical example. For the present we need only mention that σ is relevant only in the initial growth of ρ prior to the onset of laser gain.

If we regard σ as a constant, the steady-state solution of (11) and (12) is

$$\begin{aligned}\rho_\infty &= \tfrac{1}{2}\{(\xi + \sigma - 1) + [(\xi + \sigma - 1)^2 + 4\sigma]^{\frac{1}{2}}\} \\ \eta_\infty &= \tfrac{1}{2}\{(\xi + \sigma + 1) - [(\xi + \sigma - 1)^2 + 4\sigma]^{\frac{1}{2}}\}.\end{aligned} \tag{13}$$

The sign of the radical is determined by the requirement that $\rho \geqq 0$. In the limit $\sigma \to 0$ we obtain two cases, depending on whether $\xi < 1$ or $\xi > 1$. For $\xi < 1$

$$\begin{aligned}&\rho_\infty \underset{\sigma \to 0}{\to} \sigma/(1 - \xi) \\ &\eta_\infty \to \xi(1 - \rho_\infty)\end{aligned} \qquad \xi < 1; \tag{14}$$

this is the case in which the limiting inversion n_0 is less than n_1 given by (8), and laser oscillation does not occur. For $\xi > 1$

$$\begin{aligned}&\rho_\infty \underset{\sigma \to 0}{\to} \xi - 1 \\ &\eta_\infty \to 1\end{aligned} \qquad \xi > 1; \tag{15}$$

this describes the steady state of laser oscillation, which is usually approached through a series of sharp pulses in $\rho(\tau)$ called *relaxation oscillations*, or *spikes*.

III. THE NATURE OF THE SPIKING SOLUTIONS TO THE RATE EQUATIONS

In this section we shall consider the nature of the solutions to (11) and (12) when $\xi > 1$ and the initial conditions on ρ and η correspond to very few photons and a small inversion $n \ll n_1$. This may be contrasted with the situation in the giant-pulse laser in which immediately after switching $n \gg n_1$. We shall also assume the *spiking condition*

$$\omega\xi \ll 1, \tag{16}$$

which will be satisfied whenever spiking can be observed. It will be apparent later that when (16) is not satisfied there will be no spikes, but ρ will smoothly approach ρ_∞. This is the case in gas lasers, where the pumping rate $1/t_g$ has to be very high to overcome the high relaxation rate $1/t_r \sim 10^8$ sec^{-1}.

In view of (16) and (12), $\eta(\tau)$ will increase slowly with time and $\dot{\rho}$ in (11) can be neglected; thus we have

$$\rho(\tau) \approx \bar{\rho}(\tau) = \sigma(\tau)/(1 - \eta(\tau)) \tag{17}$$

until τ approaches τ_1 and η approaches unity. The behavior of $\rho(\tau)$ is

shown in Fig. 1 for the initial condition $\rho(0) = 0$. Initially ρ rises with slope $\sigma(0)$ and asymptotically approaches the adiabatic solution $\bar{\rho}(\tau)$ which it follows for a relatively long time until $\tau \to \tau_1$. It follows that the initial condition on $\rho(\tau)$ is unimportant. As τ passes through τ_1, where

$$\eta(\tau_1) = 1, \tag{18}$$

$\rho(\tau)$ remains finite and is no longer given by (17); we then leave the *adiabatic phase* and enter the *spiking phase* of the time development of $\rho(\tau)$. According to (11) the slope of $\rho(\tau)$ at τ_1 is $\sigma(\tau_1) \equiv \sigma_1$. Thus we construct an approximate solution by smoothly joining (17) to a line of slope σ_1 as shown in Fig. 1. It follows that

$$\rho(\tau_1) = \rho_1 = 2\sigma_1/(\omega\beta)^{\frac{1}{2}}, \tag{19}$$

where

$$\beta = \xi - 1. \tag{20}$$

This is our first important result. It shows that the source occurs in spiking theory only as a parameter, the single value σ_1. Thus we may drop the subscript on σ and let $\sigma_1 = \sigma$, a constant.

Once we enter the spiking phase, $\eta(\tau)$ remains close to unity, fluctuat-

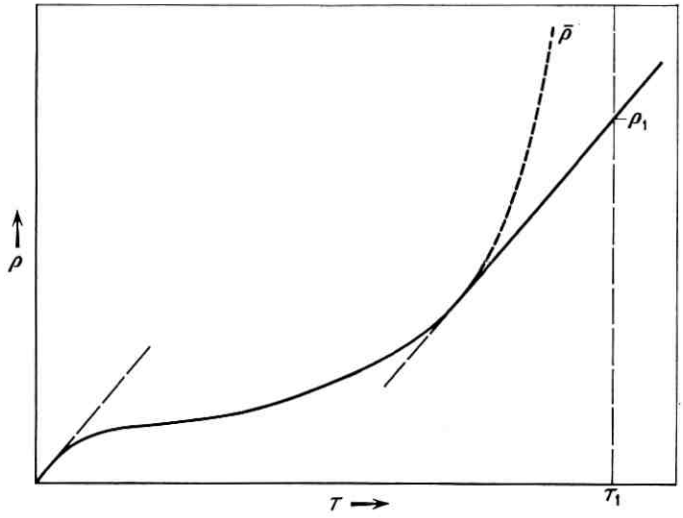


Fig. 1 — The adiabatic phase of the growth of $\rho(\tau)$. The curve is the adiabatic solution (17), and the line segments of slope σ are constructions to approximate $\rho(\tau)$ near $\tau = 0$ and $\tau = \tau_1$ where $\eta = \eta_1 = 1$.

ing up and down and eventually settling down to its steady-state value $\eta_\infty = 1$. Thus we make the approximations

$$\eta - 1 \approx \ln \eta, \tag{21}$$

and

$$\frac{\xi}{\eta} - 1 \approx \beta + \xi(1 - \eta). \tag{22}$$

The rate equations (11), (12) can now be written

$$\frac{d}{d\tau} \ln \rho = \ln \eta \tag{23}$$

$$\frac{d}{d\tau} \ln \eta = \omega(\beta - \xi \ln \eta - \rho). \tag{24}$$

Since the inversion is not directly observed we eliminate $\ln \eta$ from (23), (24); the result is most conveniently written

$$\ddot{\Psi} = \omega\beta(1 - e^{\Psi}) - \omega\xi\dot{\Psi} \tag{25}$$

in terms of the *logarithmic light output*

$$\Psi = \ln (\rho/\beta). \tag{26}$$

The discussion of (25) is greatly facilitated by a mechanical analogy which is shown in Fig. 2. We regard Ψ as the coordinate of a particle of unit mass moving in a one dimensional potential field.

$$\begin{aligned} V(\Psi) &= -\omega\beta \int_0^{\Psi} (1 - e^{\Psi})\, d\Psi \\ &= \omega\beta(e^{\Psi} - \Psi - 1). \end{aligned} \tag{27}$$

There is also a dissipative resistive force $\omega\xi\dot{\Psi}$ as if the particle were moving through a viscous medium. For the moment let us disregard the viscous force, in which case the total energy E of the particle is conserved

$$E = V(\Psi) + \tfrac{1}{2}\dot{\Psi}^2. \tag{28}$$

The particle executes a periodic motion between extreme points $\Psi_m < 0$ and $\Psi_M > 0$ such that

$$V(\Psi_m) = V(\Psi_M) = E. \tag{29}$$

In the spiking phase it is permissible to neglect $\exp(\Psi_m)$. From (27)

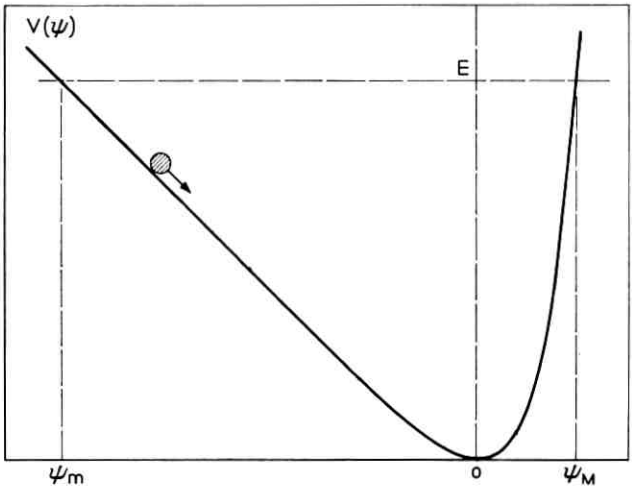


Fig. 2 — The spiking phase of the time development of $\rho(\tau)$ considered in terms of a mechanical analogy in which a particle with coordinate $\Psi = \ln(\rho/\beta)$ moves in a potential $V(\Psi)$ given by (27). The extremes of the motion Ψ_m and Ψ_M are determined by the total energy E of the particle.

and (29) we then obtain a simple relation between Ψ_m and Ψ_M

$$\Psi_M - \Psi_m = \exp(\Psi_M). \tag{30}$$

This is our second important result.

Let us define

$$\Psi_1 = \ln(\rho_1/\beta); \tag{31}$$

then it follows from (28) that

$$\Psi_m = \Psi_1 - (\sigma_1/\rho_1)^2/2\omega\beta, \tag{32}$$

since $\dot{\Psi}_1 = \sigma_1/\rho_1$. It will be clear from the numerical example in the next section that the second term can be neglected, and we can write

$$\Psi_m \approx \Psi_1. \tag{33}$$

This says that the small kinetic energy of the particle at the start of the spiking phase can be neglected.

It is convenient to denote the successive times when $\eta(\tau) = 1$ by $\tau_1, \tau_2, \tau_3, \cdots$. From (23) we see that these times correspond to extrema in the motion of Ψ; according to this convention the successive minima and maxima of Ψ are

$$\begin{aligned} &\text{minima: } (\Psi_1), \Psi_3, \Psi_5, \Psi_7, \cdots \\ &\text{maxima: } \Psi_2, \Psi_4, \Psi_6, \cdots. \end{aligned} \tag{34}$$

We have placed Ψ_1 in parentheses, since it is a minimum only in the mechanical analogy and not in the observed light output. In the absence of damping, of course, we would have $\Psi_1 = \Psi_3 = \Psi_5 = \cdots$, and $\Psi_2 = \Psi_4 = \Psi_6 = \cdots$. If (16) is satisfied, the damping will be small enough so that the motion is approximately periodic, or quasi-periodic. Thus to a good approximation we can compute the *first maximum* Ψ_2 from Ψ_1 by means of (30). Before considering the damping we shall consider other quantities of experimental interest which are characteristic of the periodic motion.

The *maximum velocity* $\dot{\Psi}_{\max} = \dot{\Psi}_0$ is capable of being measured experimentally from spiking patterns in which the spikes are well resolved in time. It follows from (28), (29) and (33) that

$$\begin{aligned} \dot{\Psi}_{\max} = \dot{\Psi}_0 &= [2\omega\beta(e^{\Psi_1} - \Psi_1 - 1)]^{\frac{1}{2}} \\ &\approx [2\omega\beta(-\Psi_1 - 1)]^{\frac{1}{2}}. \end{aligned} \tag{35}$$

In numerical applications this can be used to ascertain the validity of (21), since according to (23)

$$\dot{\Psi}_0 = \ln \eta_{\max} \approx \eta_{\max} - 1. \tag{36}$$

In general we can write near Ψ_1

$$\begin{aligned} \dot{\Psi} &= \{2[E - V(\Psi)]\}^{\frac{1}{2}} \\ &\approx [2\omega\beta(\Psi - \Psi_1)]^{\frac{1}{2}}; \end{aligned} \tag{37}$$

thus the time dependence near a minimum is given by

$$\Psi(\tau) \approx \Psi_1 + \tfrac{1}{2}\omega\beta(\tau - \tau_1)^2. \tag{38}$$

Let us denote by m the full width in time of the minimum measured between points e times the minimum in light output; this is the same as the full width of $\Psi(\tau)$ measured between points $\Psi_m + 1$. Thus the *duration of the minima* according to (38) is

$$m = (8/\omega\beta)^{\frac{1}{2}}. \tag{39}$$

This applies to all minima regardless of damping, which provides a very convenient means for determining almost by inspection whether or not a spike pattern is consistent with the assumption of constant $\omega\beta$. Even if the minima are not observed to have the same durations, it may be meaningful, in the sense that our theory provides an adiabatic approximation, to apply (39) to each minimum separately and deduce the variation of $\omega\beta$.

Near the maximum Ψ_2 we write instead of (37)

$$\begin{aligned}\dot{\Psi} &\approx [2\omega\beta(e^{\Psi_2} - 1)(\Psi_2 - \Psi)]^{\frac{1}{2}} \\ &\approx [2\omega\beta(\Psi_2 - \Psi_1 - 1)(\Psi_2 - \Psi)]^{\frac{1}{2}},\end{aligned} \tag{40}$$

where use has been made of (30); thus

$$\Psi(\tau) \approx \Psi_2 - \tfrac{1}{2}\omega\beta(\Psi_2 - \Psi_1 - 1)(\tau - \tau_2)^2. \tag{41}$$

The *duration M of a maximum* will be defined as the time interval measured between points at $1/e$ times the maximum in light output; this is the same as the interval between points at $\Psi_M - 1$. We see from the factor $(\Psi_2 - \Psi_1 - 1)$ in (41) that M will depend upon damping; it may be written in a general way

$$M = m(\Psi_M - \Psi_m - 1)^{-\frac{1}{2}}. \tag{42}$$

There is in this relation a certain ambiguity which is inherent in our method of regarding the motion as quasi-periodic. In applying the relation we may wonder whether the minimum is the one preceding or following the maximum. Within the accuracy of the quasi-periodic approximation it makes no difference: either may be used, or the average of the two.

The most readily observed quantity in spiking experiments is the *interval* between spikes, which can be identified with the period of the quasi-periodic motion

$$I = \oint d\Psi/\dot{\Psi}, \tag{43}$$

where the integral is over one cycle. To evaluate I we use the approximations (37) and (40), which are accurate near the turning points Ψ_1 and Ψ_2 respectively where $1/\dot{\Psi}$ is large. Upon comparing (35) and (37) we see that (37) is reasonably accurate even at $\Psi = 0$ providing $|\Psi_1| \gg 1$. However, (40) is only accurate near Ψ_2, say in the range

$$\Psi_2 - 1 < \Psi \leqq \Psi_2 .$$

Thus we shall use (37) in the range $\Psi_1 \leqq \Psi \leqq \Psi_c$ and (40) in the range $\Psi_c < \Psi \leqq \Psi_2$, where Ψ_c is a crossover point which will be determined presently. The integral (43) can now be carried out to obtain

$$I(\Psi_c) \approx m\left[(\Psi_c - \Psi_1)^{\frac{1}{2}} + \frac{(\Psi_2 - \Psi_c)^{\frac{1}{2}}}{(\Psi_2 - \Psi_1 - 1)^{\frac{1}{2}}}\right]. \tag{44}$$

Since both our approximations tend to underestimate $1/\dot{\Psi}$, we must

choose Ψ_c so as to maximize I, which gives the condition

$$\Psi_c = \Psi_2 - 1. \tag{45}$$

Thus the *interval I between spikes* is given by

$$\begin{aligned} I &= m[(\Psi_M - \Psi_m - 1)^{\frac{1}{2}} + (\Psi_M - \Psi_m - 1)^{-\frac{1}{2}}] \\ &= M(\Psi_M - \Psi_m). \end{aligned} \tag{46}$$

It is logical in this case to choose the minimum between the two maxima between which I is measured. There is still an ambiguity, however, in the choice of maxima. Since our approximation tends to underestimate I, it is good to use the larger of the two maxima.

We now return to the equation of motion (25) and consider the damping force $-\omega\xi\dot{\Psi}$. If (16) holds, the damping will be small, and can be computed from the work done per cycle against the damping force.

$$W = \omega\xi \oint \dot{\Psi} d\Psi. \tag{47}$$

With damping present, energy is no longer conserved, but decreases by W every cycle of the motion until the particle eventually settles down at its equilibrium position $\Psi = 0$, corresponding to the steady-state light output given by (15). Near Ψ_1 we can neglect e^{Ψ}, so that (27) gives

$$\omega\beta(\Psi_3 - \Psi_1) = W. \tag{48}$$

We evaluate (47) just as we did (43), using (37) and (40) with a cross-over point ψ_c, determined this time by the condition that W should be a minimum; the result is

$$W = (4\sqrt{2}/3)(\omega\xi)(\omega\beta)^{\frac{1}{2}}[(\Psi_M - \Psi_m - 1)^{\frac{3}{2}} + (\Psi_M - \Psi_m - 1)^{\frac{1}{2}}]. \tag{49}$$

Let us denote damping by the increment $\Delta\Psi_m$ between successive minima, or $\Delta\Psi_M$ between successive maxima. From (48) and (49) the general formula for damping of minima is

$$\Delta\Psi_m = (4\sqrt{2}/3)\omega\xi(\omega\beta)^{-\frac{1}{2}}[(\Psi_M - \Psi_m - 1)^{\frac{3}{2}} + (\Psi_M - \Psi_m - 1)^{\frac{1}{2}}]. \tag{50}$$

The choice of Ψ_m is ambiguous, but Ψ_M refers to the maximum between the two minima of $\Delta\Psi_m$. It is obvious from the shape of the potential $V(\Psi)$ shown in Fig. 2 that the maxima will be less damped than the minima. For small damping we have

$$\omega\beta(e^{\Psi_2} - 1)(\Psi_2 - \Psi_4) = W, \tag{51}$$

where W is now computed by integrating (47) from Ψ_2 around the cycle

and back to Ψ_2. The result is

$$\Delta\Psi_M = (4\sqrt{2}/3)\omega\xi(\omega\beta)^{-\frac{1}{2}}[(\Psi_M - \Psi_m - 1)^{\frac{1}{2}} + (\Psi_M - \Psi_m - 1)^{-\frac{1}{2}}] \quad (52)$$

with the familiar ambiguity in choice of Ψ_M. From (52), (46) and (39) we obtain the very convenient formula

$$\Delta\Psi_M = \tfrac{2}{3}(\omega\xi)I \quad (53)$$

relating the damping directly to the interval. We note that all ambiguity has disappeared from (53).

We now have a complete arsenal of formulas with which to attack experimental spiking patterns. Our formulas give all of the minima and maxima of Ψ as well as the durations and intervals and the maximum of $\dot{\Psi}$ in terms of the dimensionless rate equation parameters ω, ξ, σ and $\beta = \xi - 1$. These formulas are valid in the spiking phase where

$$\Psi_M - \Psi_m - 1 \gg 1. \quad (54)$$

As the spikes damp out the solution finally enters the *small-signal phase*

$$\Psi_M - \Psi_m \to 0. \quad (55)$$

The small-signal solution is well known,[5,6,13,22,29,31] so there is no need to discuss it here, but we give it for ready reference:

$$\begin{aligned} \eta(\tau) &= 1 + Ae^{-\frac{1}{2}\omega\xi\tau}[\Omega \cos(\Omega\tau + \varphi) - \tfrac{1}{2}\omega\xi \sin(\Omega\tau + \varphi)] \\ \rho(\tau) &= \beta[1 + Ae^{-\frac{1}{2}\omega\xi\tau} \sin(\Omega\tau + \varphi)] \\ \Omega^2 &= \omega\beta - \tfrac{1}{4}(\omega\xi)^2 \end{aligned} \quad (56)$$

where A is an arbitrary real small amplitude and φ is an arbitrary real phase. Exactly the same small-signal solution is obtained from (25), thereby justifying the approximation (22). It is now easy to see that when (16) breaks down the frequency Ω of the small-signal solution becomes imaginary and there are no oscillations. This may be the case in the gas laser, where we may have $\xi \sim 1$, $\beta \ll 1$, and $\omega > 4\beta$.

IV. A NUMERICAL EXAMPLE

Before attempting to apply our formulas to the analysis of spiking patterns we wish to discuss their accuracy with the aid of a machine calculation. For numerical integration the rate equations (11), (12) are best written in the form

$$\begin{aligned} \dot{x} &= \omega(\xi - x - xe^{y}) \\ \dot{y} &= \sigma e^{-y} + x - 1, \end{aligned} \quad (57)$$

where

$$\begin{aligned} x &= \eta \\ y &= \ln \rho. \end{aligned} \tag{58}$$

As already pointed out following (17), the initial condition on x,y is not critical; it is convenient to start the solution on the adiabatic solution (17), so we take

$$y(0) = \ln \sigma, \qquad x(0) = 0. \tag{59}$$

The constants will have the values

$$\begin{aligned} \omega &= 7.12 \times 10^{-6} \\ \xi &= 5 \\ \sigma &= 2 \times 10^{-9} \end{aligned} \tag{60}$$

which are typical for a ruby laser. The numerical integration of (57) has been performed on the IBM 7090 computer using Hamming's[32] predictor-corrector method. The program provides for automatic halving and doubling of the integration interval under the control of the fractional error of the increment and a preselected tolerance (5×10^{-5}). The stability and accuracy of the program for this problem were checked using initial conditions which lead to the small-signal solution (56).

The results are shown in Fig. 3. The solid curve and scale on the right give $y(\tau)$, while the dotted curve and scale on the left give $x(\tau)$ for τ in the range 3×10^4 to 4×10^4. The origin for τ is completely arbitrary. From (26), (58) and (60) we have

$$y = \Psi + \ln \beta = \Psi + 1.387. \tag{61}$$

We are concerned primarily with y, since the inversion is not ordinarily observed experimentally. The main features of the computed results are summarized in Table I, which lists the values of τ_n, y_n, and m_n or M_n for $n = 1, \cdots, 9$ for the first nine extrema in the notation of (34). These values were obtained from the computed points by fitting a parabola to the three points nearest the extremum. The spacing of the computed points was sufficiently small ($\Delta\tau = 0.005 \times 10^4$) that in all cases all three points lay well within the validity of the parabolic approximations (38) or (41).

We shall now attempt to calculate the information of Table I by the formulas of the preceding section. In every case we shall indicate the

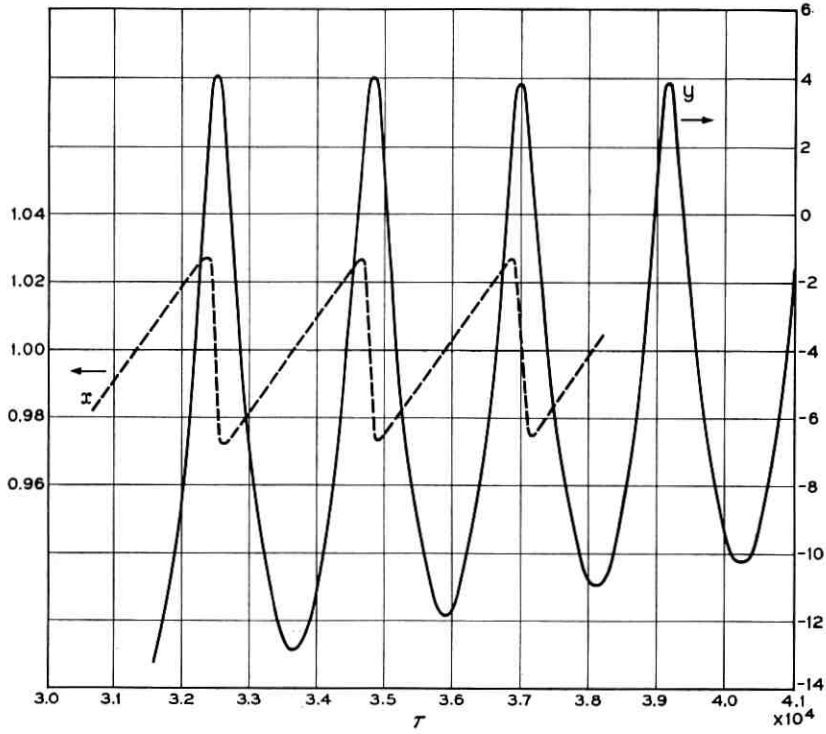


Fig. 3 — A machine calculation of $y(\tau)$, solid curve with scale on the right, and $x(\tau)$, dotted curve with scale on the left, satisfying the rate equations (57) for $\omega = 7.12 \times 10^{-6}$, $\xi = 5$, $\sigma = 2 \times 10^{-9}$. The important results are summarized in Table I. Here $x(\tau)$ represents the inversion and $y(\tau)$ the logarithmic light output.

Table I—Computed Results of Numerical Example

Summary of results of machine solution to the rate equations for $\omega = 7.12 \times 10^{-6}$, $\xi = 5$, $\sigma = 2 \times 10^{-9}$. Also obtained were $x_{max} - 1 = 0.02830$, $\dot{y}_{max} = 0.0282$.

n	τ_n	y_n	m_n	M_n
1	3.1344×10^4	−14.570	—	
2	3.2521	4.240		0.01416×10^4
3	3.3642	−12.815	0.0532	
4	3.4806	4.179		0.01415
5	3.5906	−11.795	0.0529	
6	3.7013	4.053		0.0151
7	3.8100	−10.986	0.0531	
8	3.9186	4.064		0.0152
9	4.0239	−10.303	0.0531	

correct value as computed by machine on the right in parentheses. According to (19)

$$\begin{aligned} \rho_1 &= 7.50 \times 10^5 \\ y_1 &= \ln \rho_1 = -14.09 \qquad (-14.57). \end{aligned} \tag{62}$$

Thus we *overestimate* y_1 by 0.48 or 3.3 per cent. From (61) the theoretical value for Ψ_1 is

$$\Psi_1 = -15.48. \tag{63}$$

Putting $\Psi_m = \Psi_1$ and $\Psi_M = \Psi_2$ in (30) gives

$$\Psi_2 = 2.91, \tag{64}$$

or

$$y_2 = 4.30 \qquad (4.240). \tag{65}$$

Thus we *overestimate* y_2 by 0.06 or 1.4 per cent. For the excursion of y we obtain

$$y_2 - y_1 = \Psi_2 - \Psi_1 = 18.39 \qquad (18.81), \tag{66}$$

which is an *underestimate* by 0.42 or 2.2 per cent. Henceforth we need not consider Ψ, but only y. From (50) using $\Psi_M - \Psi_m = y_2 - y_1$ we obtain

$$y_3 - y_1 = 0.965 \qquad (1.355) \tag{67}$$

$$y_3 = -13.12 \qquad (-12.815). \tag{68}$$

We are *underestimating* y_3 by 0.30 or 2.2 per cent. From (52) using $\Psi_M - \Psi_m = y_2 - y_3$ we obtain

$$y_2 - y_4 = 0.0541 \qquad (0.061) \tag{69}$$

$$y_4 = 4.25 \qquad (4.179). \tag{70}$$

By repeating steps (67) and (69) we could obtain values for any number of succeeding maxima and minima

$$y_5 - y_3 = 0.881 \qquad (1.020) \tag{71}$$

$$y_5 = -12.24 \qquad (-11.80) \tag{72}$$

$$y_4 - y_6 = 0.0523 \qquad (0.126) \tag{73}$$

$$y_6 = 4.19 \qquad (4.053). \tag{74}$$

We observe from Table I that $y_4 - y_6$ is unusually large (0.126),

while $y_6 - y_8$ is negative (-0.011); the average of these is $\frac{1}{2}(y_4 - y_8) = (0.058)$ in better agreement with the machine-computed $y_2 - y_4$ and with theory. Thus we suspect that the discrepancy in (73) is not significant. There is apparently a slight inaccuracy in the computed solution on the third and fourth spike. One might expect that noise of various kinds would have a similar effect on real lasers, making the damping from spike to spike unreliable. Averaging over several spikes, however, would still give a meaningful value for Δy_M, as it does in our computed spike pattern.

The durations of all the minima should be given by (39)

$$m = 0.0526 \times 10^4 \qquad (0.0531 \times 10^4). \tag{75}$$

We see in Table I that m is constant within 0.4 per cent and the average over the four minima is 0.0531×10^4. The durations of the maxima will be calculated from (42), using for $\Psi_M - \Psi_m$ the values of the excursions just calculated, with the minimum preceding the maximum

$$\begin{aligned} M_2 &= 0.0127 \times 10^4 \qquad (0.01416 \times 10^4) \\ M_4 &= 0.0130 \qquad\qquad (0.01415) \\ M_6 &= 0.0134 \qquad\qquad (0.0152). \end{aligned} \tag{76}$$

Here we have *underestimated* the spike durations by 11–12 per cent. The intervals are given by (46)

$$\begin{aligned} I_{2\text{-}4} &= 0.226 \times 10^4 \qquad (0.2285 \times 10^4) \\ I_{4\text{-}6} &= 0.220 \times 10^4 \qquad (0.2207 \times 10^4). \end{aligned} \tag{77}$$

The agreement here may be considered perfect. According to (53) we should have

$$\Delta y_M / I = 2.37 \times 10^{-5} \tag{78}$$

for all consecutive spikes; the computed results are as follows:

$$\begin{aligned} 2 \to 4 &\qquad (2.66 \times 10^{-5}) \\ 4 \to 6 &\qquad (5.70 \times 10^{-5}) \\ 6 \to 8 &\qquad (0.51 \times 10^{-5}) \\ \tfrac{1}{2}[(4 \to 6) + (6 \to 8)] &\qquad (2.63 \times 10^{-5}). \end{aligned} \tag{79}$$

The last line is considered more significant than the discrepancies in the third and fourth lines. Thus our theory underestimates $\Delta y_M / I$ by

11 per cent. Finally, from (35)

$$\begin{gathered} \dot{y}_{\max} = x_{\max} - 1 = 0.0288 \\ (0.0282) \qquad (0.0283), \end{gathered} \tag{80}$$

which shows that the approximation (21) is excellent.

We may conclude from this confrontation between our formulas and a machine computation that the accuracy is probably adequate for the analysis of experimental spike patterns.

V. APPLICATION OF THE FORMULAS

Experimental data on spiking are ordinarily obtained in the form of an oscilloscope trace proportional to the light output. The trace is proportional to $p(t)$, the number of photons in the laser mode of the cavity, but it is usually considered impractical to calibrate the equipment so as to obtain the absolute value of $p(t)$. And even if $p(t)$ could be measured absolutely it would not fix the absolute value of $\rho(\tau)$ defined in (9). We shall assume that only relative values of $\rho(\tau)$ can be measured. The most significant measurements of ρ are the *peak-to-valley ratios*

$$R = \rho_M/\rho_m , \tag{81}$$

the ratio of the maximum light output of a spike to the minimum output in a neighboring minimum. In analyzing a spike pattern with many spikes, some convention must be adopted on which minimum to choose. From (26) and (61)

$$\Psi_M - \Psi_m = y_M - y_m = \ln R. \tag{82}$$

Thus the excursions of the model particle in the potential $V(\Psi)$ are uniquely fixed by the data. It follows from (30) that both Ψ_M and Ψ_m are fixed by R

$$\begin{aligned} \Psi_M &= \ln (\ln R) \\ \Psi_m &= \ln (\ln R) - \ln R. \end{aligned} \tag{83}$$

Even Ψ_1, the hypothetical minimum not actually observed, can be determined by extrapolating the other minima Ψ_m. Thus all of the extrema may be regarded as immediately fixed by the data. From the extrema alone we obtain 2 relations satisfied by the three parameters ω,ξ,σ, namely (19) and (50) or (52).

It is now clear that a single spike pattern does not contain enough information to determine ω,ξ,σ. The reason for this is that we cannot

make use of the measured durations or intervals, since we do not know the ratio t_p between real time t and dimensionless time τ. We might think that we could obtain information from time ratios such as (M/m) or (m/I). It turns out that we can, but it is a different kind of information, the kind we have called "qualitative" in Section I. Consider the three quantities

$$A = (M/I) \ln R \tag{84}$$

$$B = (m/I)[(\ln R - 1)^{\frac{1}{2}} + (\ln R - 1)^{-\frac{1}{2}}] \tag{85}$$

$$C = (\Delta\Psi_M/\Delta\Psi_m)(\ln R - 1), \tag{86}$$

all of which can be determined immediately from the data since they depend only on ratios. It is to be expected that C might have to be averaged over several spikes to get a meaningful value. Except for this difficulty, values of A, B, and C may be calculated for every spike in the data after adopting some convention to handle the usual ambiguity in the definitions (84), (85), (86). One such convention is illustrated by the following example:

$$\begin{aligned} \ln R_2 &= \Psi_2 - \Psi_3 \\ A_2 &= (M_2/I_{2\text{-}4}) \ln R_2 \\ B_2 &= (m_3/I_{2\text{-}4})[(\ln R_2 - 1)^{\frac{1}{2}} + (\ln R_2 - 1)^{-\frac{1}{2}}] \\ C_2 &= [(\Psi_2 - \Psi_4)/(\Psi_3 - \Psi_1)](\ln R_2 - 1). \end{aligned} \tag{87}$$

From (42), (46), (50) and (52) we find the simple relations

$$A = B = C = 1 \tag{88}$$

which do not involve the parameters ω,ξ,σ. Therefore (88) should hold even if ω,ξ are slowly-varying functions of time within our adiabatic approximation. We call the relations (88) *consistency relations*, since they can be used to determine whether data are consistent with the rate equations. If (88) is satisfied reasonably well, at least in an average sense over the spikes, it is a foregone conclusion that a reasonable fit to the data can be obtained from the rate equations. The converse is also true: if (88) is not satisfied the data cannot be fitted from the rate equations.

The importance of observing the minima (valleys) as well as the maxima (peaks) in a spike pattern is abundantly clear. Without the valleys we cannot determine R, m, or $\Delta\Psi_m$, all of which appear in the consistency relations. Therefore a great deal of information is lost unless

the valleys can be seen above noise. This presents some difficulty, because ordinarily $R \gg 1$ is much too large to be measured from an oscilloscope trace that responds linearly to the light output and contains both the peaks and the valleys. In spike patterns the valleys usually just correspond to the noise level in the experiment. It is not our purpose to go into experimental details except to point out that with care the valleys should be observable. The basic experimental requirement is that the acceptance cone of the detector should correspond to the radiation cone of the laser so as to exclude extraneous light from the pump and the spontaneous emission of the laser medium. If it is not possible to measure the valleys, there is still one consistency relation that can be applied to the peaks alone. From (46), (82) and (83) we have

$$\begin{aligned} \ln R &= I/M \\ -\Psi_m &= (I/M) - \ln (I/M). \end{aligned} \tag{89}$$

The use of (46) is equivalent to assuming $A = 1$. Thus the minima can be calculated from I/M if we assume $A = 1$. We could obtain the durations m of the minima by putting $B = 1$. The damping of the minima would not be given very reliably by (89), but can be obtained from (86) by putting $C = 1$. Thus we can deduce R, m, and $\Delta\Psi_m$ from the peaks alone, but we lose all of our consistency relations (88). However, from (89), (42), (39) and (35) it follows that

$$D = 1, \tag{90}$$

where

$$D = (M\dot{\Psi}_{\max}/4)[(I/M) - 1]^{\frac{1}{2}}/[(I/M) - \ln (I/M) - 1]^{\frac{1}{2}} \tag{91}$$

can be determined from the peaks if the time resolution is good enough to give a good value for $(M\dot{\Psi}_{\max})$. If (90) is satisfied, it is probably good evidence that the laser obeys the rate equations, and a rate equation analysis is meaningful. However, if all that is desired is to apply the rate equations blindly to obtain quantitative information, it is not necessary to measure $M\dot{\Psi}_{\max}$. All of the quantitative information in a consistent spike pattern can be deduced from I, M, and $\Delta\Psi_M$.

We now consider practical ways of obtaining quantitative information from spike patterns. The *one-pattern method* is to measure t_p by an independent experiment. The measurement of the cavity losses has been discussed by several authors.[26,33] Suffice it to say here that t_p can be measured from the dependence of the threshold flash energy for producing laser action on the temperature and on losses deliberately introduced

into the laser mode. Once t_p is measured all of the times m, M and I become known in dimensionless time. From (39)

$$\omega\beta = \omega(\xi - 1) = 8/m^2, \tag{92}$$

and from (53)

$$\omega\xi = \tfrac{3}{2}\Delta\Psi_M/I. \tag{93}$$

These equations can be solved for ω,ξ

$$\begin{aligned} \xi &= \gamma m/(\gamma m - 1) \\ \omega &= 8(\gamma m - 1)/m^2, \end{aligned} \tag{94}$$

where

$$\gamma = \tfrac{3}{16}(m/I)\Delta\Psi_M \tag{95}$$

is independent of t_p. Now σ can be obtained from (19) and (89)

$$\sigma = \tfrac{1}{2}\beta(\omega\beta)^{\frac{1}{2}}(I/M)e^{-(I/M)} \tag{96}$$

with $M = M_2$ and $I = I_{2\text{-}4}$. This procedure makes use of the duration m of the valleys but not the peak-to-valley ratio R. If only the peaks are observed we write instead of (92)

$$\omega\beta = \omega(\xi - 1) = (8/MI)/[1 - (M/I)]. \tag{97}$$

Solving (93) and (97) for ω,ξ gives

$$\begin{aligned} \xi &= \delta M/(\delta M - 1) \\ \omega &= 8(\delta M - 1)/(I - M)M, \end{aligned} \tag{98}$$

where

$$\delta = \tfrac{3}{16}[1 - (M/I)]\Delta\Psi_M \tag{99}$$

is independent of t_p. This kind of analysis can be applied to every spike in a spike pattern. It may be found that ξ and especially ω vary slowly from spike to spike, which is permissible within our adiabatic approximation. If the data satisfy the consistency relations (88), the same values of ω,ξ will be obtained from (94) and (98). The greatest weakness of this method is that (94) fails completely if $\gamma m \leqq 1$; (98) fails if $\delta M \leqq 1$. Thus a great deal depends on the accuracy of the formulas as well as that of the measurements of t_p and γ or δ. It follows that the method will fail whenever $\xi \gg 1$.

As an example we shall apply this method to the machine calculation considered in Section IV. We consider the solid curve of Fig. 3 to be the

data, which implies that t_p is known and both the peaks and valleys have been studied. The relevant numbers have already been given in parentheses in (69), (75), (76) and (77); we repeat them here without parentheses

$$\begin{aligned} \Delta\Psi_M &= y_2 - y_4 = 0.061 \\ m &= 0.0526 \times 10^4 \\ M &= M_2 = 0.0142 \times 10^4 \\ I &= I_{2\text{-}4} = 0.2285 \times 10^4. \end{aligned} \tag{100}$$

From (94), (95) and (96) we obtain

$$\begin{aligned} \xi &= 3.4 && (5.0) \\ \omega &= 12 \times 10^{-6} && (7.12 \times 10^{-6}) \\ \sigma &= 10 \times 10^{-9} && (2 \times 10^{-9}). \end{aligned} \tag{101}$$

The lack of accuracy is primarily due to the fact that $\xi = 5$ is a little too large to give good results with this method. Using only the peaks, we obtain from (96), (98) and (99)

$$\begin{aligned} \xi &= 2.9 && (5.0) \\ \omega &= 14 \times 10^{-6} && (7.12 \times 10^{-6}) \\ \sigma &= 8 \times 10^{-9} && (2 \times 10^{-9}). \end{aligned} \tag{102}$$

We now describe a *two-pattern method* which does not require the measurement of t_p. In fact, it yields a value for t_p and may in some cases give better results for ω,ξ than the one just described. It is based upon observing the valleys in two or more spike patterns for which the *relative* values of ω and ξ are known. We must assume that the linewidth $\Delta\nu$ is known as a function of temperature. Let us suppose that we measure the valley durations m and m' in two spike patterns in which the ratios (ξ'/ξ) and (ω'/ω) are known. We know (ξ'/ξ) from the temperatures at which the patterns were obtained. We can obtain ω'/ω from the relative pump intensities at the times when spiking occurred. It immediately follows from (92) that

$$\xi = \frac{(\omega m^2/\omega' m'^2) - 1}{(\omega m^2/\omega' m'^2) - (\xi'/\xi)}. \tag{103}$$

This result is meaningful providing that the data give $(\omega m^2/\omega' m'^2)$ lying between (ξ'/ξ) and unity. Once ξ is determined ω can be calculated from

(92) and (93)

$$\omega = \frac{9}{32}\frac{(\xi - 1)}{\xi^2}\left(\frac{m}{I}\right)^2 (\Delta\Psi_M)^2, \tag{104}$$

and σ is again obtained from (96).

As an example of this method we shall apply it to two machine-calculated spike patterns, one of which is that of Fig. 3 with the parameters (60), and the other has the parameters

$$\begin{aligned} \omega' &= \omega \\ \xi' &= 4 \\ \sigma' &= \sigma. \end{aligned} \tag{105}$$

From the machine-calculated pattern we find

$$m' = m_3' = 0.0614 \times 10^4. \tag{106}$$

The value of $m = m_3$ is given in (75). We assume that the ratio

$$\xi'/\xi = 0.8 \tag{107}$$

is known from the temperatures at which the data were taken, and the ratio

$$(\omega m^2/\omega' m'^2) = 0.748 \tag{108}$$

is calculable from (ω/ω') and the two valleys. We now obtain from (103) the value

$$\xi = 4.85 \qquad (5.0). \tag{109}$$

From (104) we now obtain

$$\omega = 9.0 \times 10^{-6} \qquad (7.12 \times 10^{-6}) \tag{110}$$

using the values in parentheses from (75), (77) and (69). From (39), (109) and (110) we obtain the *absolute value* of m

$$m = 0.048 \times 10^4 \qquad (0.0531 \times 10^4). \tag{111}$$

It follows that t_p is given by

$$t_p = m_{\text{exp}}/m, \tag{112}$$

where m_{exp} is the measured duration in laboratory time. We conclude that the two-pattern method is to be preferred to the one-pattern method as a general approach to spike analysis. It should be mentioned that

there seems to be no two-pattern method involving only the peaks which is sufficiently accurate to give meaningful results.

The most convenient experiments to perform involve changing the pump intensity $1/t_0$ while all other parameters remain fixed. If $t_0 \ll t_r$, as is usually the case in flash-pumped lasers, ξ will be independent of t_0 and ω will vary as $1/t_0$. Varying ω does not give quantitative information such as we obtained from varying ξ in the two-pattern method. Nevertheless, it is of interest to consider what effects are to be expected. From (30), (26) and (19) we have roughly

$$\begin{aligned} \rho_M &\sim -\beta\psi_m \\ &\sim \beta \ln (\beta\sqrt{\omega\beta}/2\sigma). \end{aligned} \tag{114}$$

Thus the maxima in ρ depend only logarithmically on pumping power. It follows from (9) that the observed light output proportional to p should vary as

$$p \propto 1/t_0 . \tag{115}$$

The time intervals m, M, and I should vary as

$$m \propto M \propto I \propto t_0^{\frac{1}{2}}. \tag{116}$$

VI. SUMMARY

We have now outlined a comprehensive program for the study of lasers by means of their spiking patterns. The rate equations have been formulated in terms of the light output and the atomic inversion and five physical parameters in (1) and (2). We have written the equations in a general form valid for three- and four-level systems and taking into account statistical weights and thermalization of the upper laser level. These equations were then reduced to dimensionless form in terms of three parameters in (11) and (12). All of the properties of the spiking solutions of experimental interest were then deduced analytically by means of the mechanical analogy shown in Fig. 2, in which a particle moves in a potential well in a viscous medium.

The formulas obtained were illustrated and tested for accuracy against a machine-computed solution to the rate equations. The value of the formulas was confirmed by this comparison, and we went on to discuss their application to experimentally observed spiking patterns. Four relations were given whereby spike patterns can be tested for consistency with the rate equations. These consistency relations do not contain any parameters, only ratios of times and of light outputs.

Two methods were described for obtaining quantitative information. In the one-pattern method everything is deduced from a single spike pattern, but it is necessary to measure the photon lifetime t_p in the cavity by independent experiments. It is possible to apply this method to data in which only the peaks are observed. It is emphasized, however, that the observation of the valleys in light output should be perfectly feasible and very much worthwhile. In the two-pattern method all the parameters and t_p are deduced from two patterns obtained at different temperatures. In this method, the more accurate of the two, it is essential that the valleys be observed.

Our objective has been to provide the researcher with a set of tools for applying the rate equations to experimental spiking data. The spiking phenomenon in solid state lasers can be utilized in research now that it is beginning to come under good experimental control. We have tried here to point out what kind of spiking data is needed, and what additional information is needed, to get the maximum information from spiking. Our analysis applies to the sharp spike region of time in which the rate equations are highly nonlinear. It complements the well known small-signal analysis in which the rate equations become linear. The entire discussion is based upon the rate equations of Statz and De Mars, which we regard as reasonable, relevant and widely accepted. We have not gone into the derivation of these equations or the modifications that have been proposed or might be proposed. We prefer to leave that to the future, when we may reasonably expect that systematic spiking studies will have clearly revealed the adequacy or inadequacy of these equations, and indicated the direction which these modifications must take.

VII. ACKNOWLEDGMENTS

This work actually began several years ago in a joint effort with R. J. Collins, now of the University of Minnesota. Circumstances intervened to prevent the completion of our collaboration, but I am pleased to acknowledge the great benefits that accrued to me through it. For interest, encouragement, and informative discussions I am also indebted to D. F. Nelson, W. K. Kaiser and D. E. McCumber of Bell Telephone Laboratories.

REFERENCES

1. We shall use the term "laser" to mean an optical maser as described by Schawlow, A. L., and Townes, C. H., Phys. Rev., **29**, 1958, p. 1940.
2. Bloembergen, N., Phys. Rev., **104**, 1956, p. 324.
3. Shimoda, K., Takahasi, H., and Townes, C., J. Phys. Soc. Japan, **12**, 1957, p. 686.

4. Statz, H., and De Mars, G., *Quantum Electronics*, Columbia University Press New York, 1960, ed. C. H. Townes, p. 530.
5. Makhov, G., J. Appl. Phys., **33,** 1962, p. 202.
6. Sinnett, D. M., J. Appl. Phys., **33,** 1962, p. 1578.
7. Statz, H., Luck, C., Shafer, C., and Ciftan, M., *Advances in Quantum Electronics*, Columbia University Press, New York, 1961, ed. J. R. Singer, p. 342.
8. Collins, R. J., Nelson, D. F., Schawlow, A. L., Bond, W., Garrett, C. G. B., and Kaiser, W., Phys. Rev. Letters, **5,** 1960, p. 303.
9. Shimoda, K., in *Symposium on Optical Masers*, Interscience Publishers, New York, 1963, ed. Jerome Fox.
10. McClung, F. J., and Hellwarth, R. W., J. Appl. Phys., **33,** 1962, p. 828; and Hellwarth, R. W., *Advances in Quantum Electronics*, Columbia University Press, New York, 1961, ed. J. R. Singer, p. 334.
11. Vuylsteke, A., J. Appl. Phys., **34,** 1963, p. 1615.
12. Wagner, W. G., and Lengyel, B. A., J. Appl. Phys., **34,** 1963, p. 2040.
13. Dunsmuir, R., J. Electronics and Control, **10,** 1961, p. 453.
14. Anderson, P. W., J. Appl. Phys., **28,** 1957, p. 1049.
15. Clogston, A. M., J. Phys. Chem. Solids, **4,** 1958, p. 271.
16. Fain, V. M., and Khanin, Y. I., Soviet Physics JETP (translation), **14,** 1962, p. 1069.
17. Kaplan, J., and Zier, R., J. Appl. Phys., **33,** 1962, p. 2372.
18. Yoh-han Pao, J. Opt. Soc. Am., **52,** 1962, p. 871.
19. McCumber, D. E., Phys. Rev., **130,** 1963, p. 675.
20. Haken, H., and Sauermann, H., Z. Physik, **173,** 1963, p. 261; **176,** 1963, p. 47.
21. Lamb, W. E., Jr., in *Lectures in Theoretical Physics II*, University of Colorado Summer School, 1959, ed. W. E. Brittin and B. W. Downs, Interscience Publishers, Inc., New York, 1960, p. 435.
22. Burch, J. M., in *Proceedings of the Conference on Optical Instruments*, 1961, ed. K. J. Habell, John Wiley, New York, 1963, p. 463.
23. Sorokin, P., and Stevenson, M., Phys. Rev. Letters, **5,** 1960, p. 557; also *Advances in Quantum Electronics*, Columbia University Press, New York, 1961, ed. J. R. Singer, p. 65.
24. O'Connor, J., and Bostick, H., Proc. I.R.E., **50,** 1962, p. 219.
25. Johnson, L. F., and Nassau, K., Proc. I.R.E., **49,** 1961, p. 1704.
26. Nelson, D. F., and Remeika, J., J. Appl. Phys., **35,** 1964, p. 522.
27. Johnson, McMahon, Oharek, and Sheppard, Proc. I.R.E., **49,** 1961, p. 1942.
28. Gürs, K., Z. Naturforsch, **17a,** 1962, p. 990; **18a,** 1963, p. 510; **18a,** 1963, p. 418; Gürs, K., and Müller, R., Physics Letters, **5,** 1963, p. 179.
29. Hercher, M. M., unpublished thesis, 1963, University of Rochester.
30. Collins, R. J., and Nelson, D. F., Ref. 22, p. 441.
31. Kaiser, W., Garrett, C. G. B., and Wood, D. L., Phys. Rev., **123,** 1961, p. 766.
32. Hamming, R. W., J. Assoc. Computing Machinery, **6,** 1959, p. 37.
33. Masters, J. I., Nature, **199,** 1963, p. 442; D'Haenens, I. J., and Asawa, C. K., J. Appl. Phys., **33,** 1962, p. 3201.

Design of Wideband Sampled-Data Filters

By R. M. GOLDEN and J. F. KAISER

(Manuscript received March 6, 1964)

A design procedure is presented for readily obtaining sampled-data filter representations of continuous filters. The procedure utilizes the bilinear z transformation and preserves the essential amplitude characteristics of the continuous filter over the frequency range between zero and one-half the sampling frequency. It is shown that the procedure can yield meaningful sampled-data filter designs for many of those filters where the standard z transform cannot be used directly.

I. INTRODUCTION

Sampled-data filter representations for continuous filters can be obtained using several different design procedures.[1] A particular design method utilizing the bilinear transformation is developed herein. The method is especially useful in designing wideband* sampled-data filters which exhibit relatively flat frequency-magnitude characteristics in successive pass and stop bands. Filters of this type are widely used in network simulation and data processing problems.[2] The design method possesses two chief advantages over the standard z transform.[3] The first is that the transformation used is purely algebraic in form. This means it can be applied easily to a continuous filter having a rational transfer characteristic expressed in either polynomial or factored form. The second advantage is the elimination of aliasing[4] errors inherent in the standard z transform. Thus, the sampled-data filter obtained by this design method exhibits the same frequency response characteristics as the continuous filter except for a nonlinear warping of the frequency scale. Compensation for this warping can be made by a suitable frequency scale modification. Some of the more common filter networks to which the design method can be applied effectively are the Butterworth, Bessel, Chebyshev, and elliptic filter structures.

The essential properties of the bilinear transformation are presented

* A sampled-data filter design will be termed "wideband" if the frequency range of useful approximation approaches half the sampling frequency.

in the next section. For comparison purposes, properties of the standard z transform are also given. This is followed by a detailed description of a filter design procedure using frequency transformations. Examples illustrating the design procedure are then presented. A short discussion concerning computer simulation of the obtained sampled-data filters is also included.

II. THE STANDARD AND BILINEAR z TRANSFORMATIONS

In this section it is assumed that a satisfactory rational expression is known for the transfer function of a continuous filter for which a sampled-data approximation is sought. What is then necessary is a transformation which will convert this transfer characteristic into a sampled-data transfer function rational in z^{-1}, the unit delay operator.

Two transformations applicable to this problem are the standard z transform and the bilinear z transformation.

The standard z transform applied to $H(s)$, the transfer function of the filter, is[3]

$$H^*(s) = \sum_{m=-\infty}^{\infty} H(s + jm\omega_s) \tag{1}$$

or equivalently in terms of the impulse response, $h(t)$, of the filter

$$\mathfrak{IC}^*(z) = T \sum_{l=0}^{\infty} h(lT)z^{-l} \tag{2}$$

where

$$\begin{aligned} s &= \sigma + j\omega \\ H(s) &= \text{Laplace transform of } h(t) \\ \omega_s &= 2\pi/T = \text{radian sampling frequency} \\ H^*(s) &= \text{Laplace transform of the sampled filter impulse response} \\ z^{-1} &= \exp(-sT) = \text{the unit delay operator} \\ \mathfrak{IC}^*(z) &= H^*(s) \mid_{s=(\ln z)/T} = z \text{ transform of } h(t). \end{aligned}$$

The behavior of $H(s)$ for s greater than some critical frequency $j\omega_c$ is assumed to be of the form

$$H(s) \mid_{\text{all } s > j\omega_c} = K/s^n, \qquad n > 0 \tag{3}$$

where K is a determined constant.

Equation (1) or (2) is the transfer function of a sampled-data filter which *approximates* the continuous filter. In the time domain, the im-

pulse response of the sampled-data filter is the sampled impulse response of the continuous filter. This can be shown by taking the inverse transform of (2). Equation (1) shows that in the baseband

$$(-\omega_s/2 \leqq \omega \leqq \omega_s/2)$$

the frequency response characteristics of the sampled-data filter, $H^*(s)$, differ from those of the continuous filter, $H(s)$. The difference is the amount added or "aliased"[4] in through terms of the form

$$H(s + jm\omega_s),\ m \neq 0.$$

If $H(s)$ is bandlimited to the baseband, i.e., $|H(s)| = 0$ for $\omega > \omega_s/2$, then there is no aliasing error and the sampled-data filter frequency response is identical to that of the continuous filter. Unfortunately, when $H(s)$ is a rational function of s, it is not bandlimited and therefore $H(s) \neq H^*(s)$ in the baseband.

The magnitude of the errors resulting from aliasing is directly related to the high-frequency asymptotic behavior of $H(s)$ as defined in (3). If n is large and $\omega_c \ll \omega_s/2$, then the aliasing errors will be small and the standard z transform generally will yield a satisfactory sampled-data filter design. However, in wideband designs, ω_c is usually an appreciable fraction of $\omega_s/2$. Furthermore, many continuous filter designs result in transfer functions in which n is no greater than 1. These two conditions, namely $\omega_c \approx \omega_s/2$ and $n = 1$, can create large aliasing errors in the frequency response characteristics, thus yielding an unusable result.

Fortunately, even when $\omega_c \approx \omega_s/2$ and $n = 1$, a design method employing the bilinear z transformation* may provide satisfactory wideband designs. This z form is defined from the mapping transformation,

$$s = (2/T) \tanh (s_1 T/2) \tag{4}$$

where

$$s_1 = \sigma_1 + j\omega_1 .$$

The right-hand side of (4) is periodic in ω_1 with period $2\pi/T$. Considering only the principal values of ω_1, $-\pi/T < \omega_1 < \pi/T$, it is seen that the transformation given by (4) maps the *entire* complex s plane into the strip in the s_1 plane bounded by the lines $\omega_1 = -\pi/T$ and $\omega_1 = +\pi/T$. For this reason the bilinear transformation can be looked upon as a bandlimiting transformation. Therefore, when this transformation is applied to a transfer function $H(s)$, the entire s-plane frequency characteristics of $H(s)$ are uniquely carried over into the frequency characteristics of $H(s_1)$.

* This transformation will be referred to as the bilinear z form or z form.

With the substitution

$$z^{-1} = e^{-s_1 T},$$

(4) can be written immediately as

$$s = \frac{2}{T}\frac{(1 - z^{-1})}{(1 + z^{-1})}. \tag{5}$$

Thus,

$$\mathfrak{IC}(z) \equiv H(s)\Big|_{s = \frac{2}{T}\frac{(1-z^{-1})}{(1+z^{-1})}} \tag{6}$$

where $\mathfrak{IC}(z)$ denotes a sampled-data transfer function obtained by the use of the bilinear z form.*

The transfer function $\mathfrak{IC}(z)$ obtained by means of the bilinear z form and the function $\mathfrak{IC}^*(z)$ obtained by means of the standard z transform are both rational in z^{-1} and of the same denominator order as the continuous filter. These two functions, $\mathfrak{IC}(z)$ and $\mathfrak{IC}^*(z)$, become essentially equal to each other as the sampling frequency becomes large compared to the moduli of each of the poles of the continuous filter function, $H(s)$. When the sampling frequency is not large, representation of some filters by the standard z transform can be quite unsatisfactory because of aliasing errors. However, the bilinear z form, with its absence of aliasing errors, may give a satisfactory representation for these filters. A particular set of filters to which the bilinear z form always can be applied successfully are those which exhibit relatively flat frequency-magnitude characteristics in successive pass and stop bands. This follows directly from (6).

Thus, sampled-data filters designed by using the bilinear z form preserve the essential amplitude characteristics of the continuous filter. In the baseband ($-\omega_s/2 \leqq \omega \leqq \omega_s/2$), the frequency characteristics of the sampled-data filter are identical to those of the continuous filter *except* for a nonlinear warping of the frequency scale.

This warping is found from (4) upon substituting $j\omega$ for s and is

$$\omega = (2/T) \tan (\omega_1 T/2). \tag{7}$$

For small values of ω_1, the relation is essentially linear, producing

* It should be noted that the bilinear transform is used here in a distinctly different way than it is commonly used in sampled-data control system design. In the control system literature it is used to transform the sampled-data function $\mathfrak{IC}^*(z)$ from the discrete domain back to the continuous domain for conventional stability and frequency response analysis. See for example J. T. Tou, *Digital and Sampled-Data Control Systems*, McGraw-Hill, New York, 1959, pp. 244–247 and pp. 466–470.

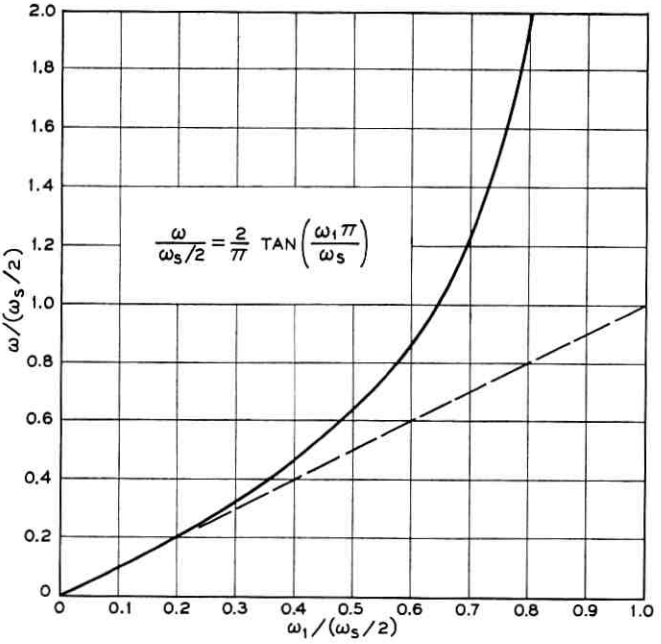


Fig. 1 — The frequency scale warping of the bilinear z transformation.

negligible warping at the lower end of the frequency scale. Fig. 1 shows the nature of this warping. Compensation for the effect of warping can be made by prewarping the band-edge frequencies of the continuous filter in such a way that application of the z-form transformation will shift the band-edge frequencies back to the desired values. The incorporation of this prewarping compensation into a sampled-data filter design procedure is discussed in the next section.

III. A SAMPLED-DATA FILTER DESIGN METHOD

A sampled-data filter design may be obtained by applying the z-form transformation of (6) to the rational transfer characteristic for a continuous filter. However, in order to compensate for the frequency warping imposed by the z form, the frequency characteristics of the continuous filter first must be altered or prewarped. Hence the transfer characteristic for the continuous filter must be redesigned such that the band-edge (cutoff) and maximum loss frequencies are computed according to,

$$\omega_c = (2/T) \tan (\omega_d T/2) \tag{8}$$

where:

ω_c = computed cutoff or loss frequency
ω_d = desired cutoff or loss frequency.

The redesign of the continuous transfer characteristic cannot be accomplished simply by applying (8) to each pole and zero of the original transfer characteristic. On the contrary a completely new transfer characteristic must be obtained for the continuous filter. It is then possible to obtain the desired sampled-data filter by applying the z-form of (6) directly to the redesigned transfer characteristic of the continuous filter. The sampled-data filter so obtained will then have the desired magnitude-frequency characteristics.

Compensation for frequency warping becomes especially simple to apply if the original continuous filter design was obtained by applying a frequency-band transformation to a suitable low-pass design such as Butterworth, Chebyshev, etc. Thus the extensive literature available on tabulated low-pass filter designs can be used to great advantage to simplify the filter design problem. The well-known frequency transformations which convert a normalized low-pass filter to a low-pass, a bandpass, a bandstop, or a high-pass design are

$$s_n = s/\omega_u \text{ low-pass to low-pass} \tag{9}$$

$$s_n = \frac{(s^2 + \omega_u\omega_l)}{s(\omega_u - \omega_l)} \text{ low-pass to bandpass} \tag{10}$$

$$s_n = \frac{s(\omega_u - \omega_l)}{(s^2 + \omega_u\omega_l)} \text{ low-pass to bandstop} \tag{11}$$

$$s_n = \omega_u/s \text{ low-pass to high-pass} \tag{12}$$

where:

s_n = the complex variable of the normalized low-pass filter transfer function
s = the complex variable of the desired filter transfer function
ω_u = the upper radian cutoff frequency
ω_l = the lower radian cutoff frequency.

When continuous filters are designed with the aid of these transformations, prewarping is accomplished by properly choosing the cutoff frequencies used in the frequency transformations. The choice of these cutoff frequencies is determined from the desired cutoff frequencies by means of (8). Using these values, the new prewarped transfer function is determined by applying (9), (10), (11) or (12) to the original low-pass

function. Transformation is made to a sampled-data filter by applying (5) to the prewarped continuous function. This sampled-data filter will now have the correct cutoff frequencies. The transfer function thus obtained can be used directly in a digital computer simulation.

IV. SIMULATION OF SAMPLED-DATA FILTERS

Application of either the standard z transform or the bilinear z form to a rational transfer function yields a transfer function rational in z^{-1} for the sampled-data filter. The programming or simulation of this sampled-data filter on a digital computer can be accomplished by either the direct, the cascade or the parallel form. These forms, as commonly defined, are shown in Fig. 2. In this figure $\mathcal{G}(z)$ and $\mathcal{F}(z)$ represent finite polynomials in z^{-1} for feed-forward and feedback transmissions re-

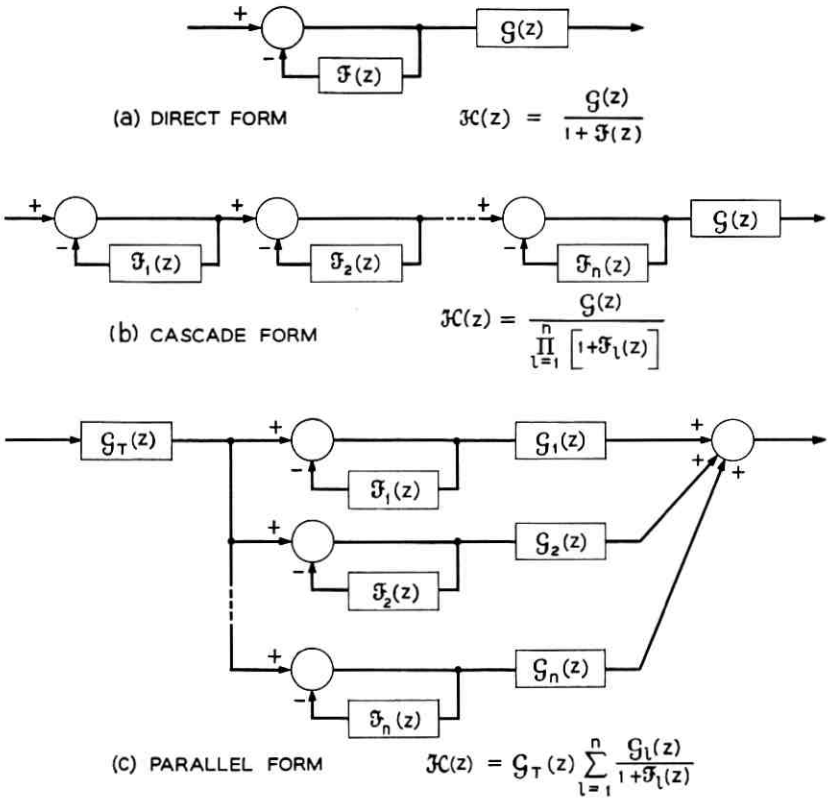


Fig. 2 — Some possible simulation forms for sampled-data filters.

spectively; where subscripted, the order of the polynomial is at most second.

The choice of which of the three forms to use for simulation of the sampled-data filter depends on the complexity of the filter function $H(s)$, on the form of $H(s)$, and on the particular z transformation used. Generally, simulation by the direct form requires considerably greater accuracy in the determination of the filter parameters than either of the other two forms. This is especially true when the order of $H(s)$ is large and when $H(s)$ has poles with real parts that are a very small fraction of the sampling frequency. For this reason either the cascade or parallel form may be preferred.

The choice between using the cascade or the parallel form depends largely on which z-transform method is used to obtain the sampled filter and how that particular method is applied. Realization in the cascade form requires calculation of the numerator polynomial, $\mathcal{G}(z)$, or its factors. This computation consists of a simple algebraic substitution when the bilinear z form is applied to a filter function $H(s)$ expressed in the form,

$$H(s) = \frac{G(s)}{\prod_{k=1}^{m} (s - \alpha_k)}. \tag{13}$$

Determination of $\mathcal{G}(z)$ in polynomial or product form respectively allows either of the following cascade realizations to be made:

$$\mathcal{H}(z) = \mathcal{G}(z) \prod_{k=1}^{n} \left(\frac{1}{1 + b_{1k}z^{-1} + b_{2k}z^{-2}} \right) \tag{14}$$

or

$$\mathcal{H}(z) = \prod_{k=1}^{n} \left(\frac{a_{0k} + a_{1k}z^{-1} + a_{2k}z^{-2}}{1 + b_{1k}z^{-1} + b_{2k}z^{-2}} \right). \tag{15}$$

If the numerator $G(s)$ is in polynomial form, considerable care must be taken in the calculation of the coefficients of the polynomial $\mathcal{G}(z)$, as this computation involves differencing nearly equal numbers.

For realization in the parallel form the partial fraction expansion of $H(s)$ must be known. Since in the standard z transform method obtaining the partial fraction expansion is a necessary step, simulation of filters designed by this method is most directly accomplished in the parallel form. Here the continuous filter transfer characteristic is represented by

$$H(s) = \sum_{k=1}^{N} \frac{P_{1k}s + P_{0k}}{Q_{2k}s^2 + Q_{1k}s + Q_{0k}}. \tag{16}$$

Transforming this expression by use of the standard z transformation yields

$$\mathcal{H}(z) = \sum_{k=1}^{N} \frac{A_{1k}z^{-1} + A_{0k}}{B_{2k}z^{-2} + B_{1k}z^{-1} + 1} \tag{17}$$

whereas transforming by use of the bilinear z form yields the similar expression

$$\mathcal{H}(z) = (1 + z^{-1}) \sum_{j=1}^{N} \frac{A_{1j}z^{-1} + A_{0j}}{B_{2j}z^{-2} + B_{1j}z^{-1} + 1}. \tag{18}$$

Each rational function in either of the above summations can be synthesized by the recursive structure shown in functional block diagram form in Fig. 3(a). This recursive structure uses only two delays, four multiplications, and five additions. The complete realization of (18) is shown in Fig. 3(b).

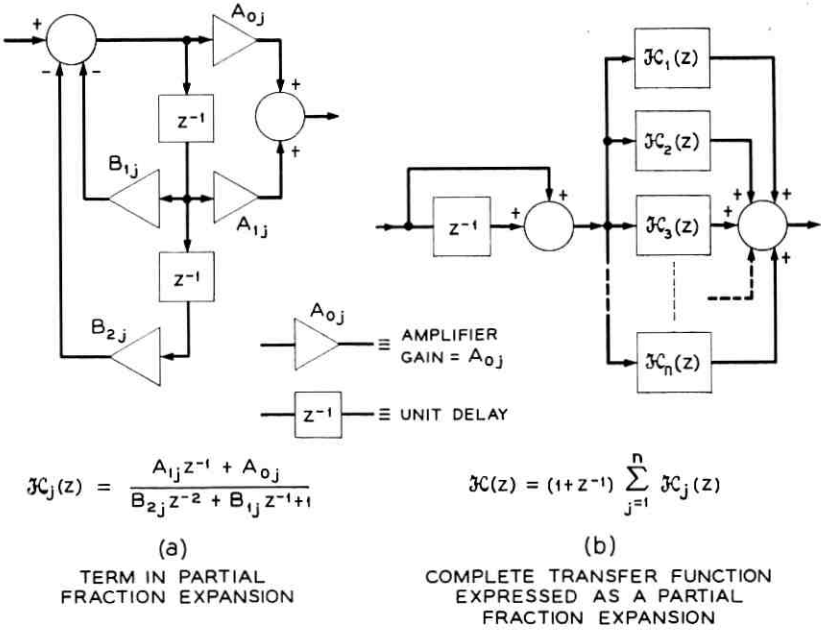


Fig. 3 — Simulation in parallel form of a sampled-data filter obtained by the bilinear z transformation.

The programming of these sampled-data filters for computer simulation can be greatly simplified if a compiler such as the Block Diagram (BLODI) compiler[5] developed at Bell Telephone Laboratories is used. The compiler permits specification of a sampled-data system in functional block diagram form.

In the following section an example is presented for a filter designed, synthesized, and simulated by the foregoing method.

V. DESIGN, SYNTHESIS, AND SIMULATION OF A WIDEBAND BANDSTOP SAMPLED-DATA FILTER

As an example of the application of the bilinear z-form to the simulation of a practical filter, consider the design of a particular bandstop filter. The filter is to exhibit at least 75 db loss in a rejection band which extends between 2596 cps and 2836 cps. Below 2588 cps and above 2844 cps, the loss is to be between 0 and +0.5 db. The sampling rate of the discrete filter is to be 10 kc. The complexity or order of the filter is to be held to a minimum. Fig. 4 shows a sketch of the amplitude response characteristics desired of the filter.

Minimum filter complexity and sharp transition between pass and stop bands suggest the use of an elliptic filter[6] (equiripple) as the basic low-pass type. However, before a suitable low-pass structure can be determined, the above specified critical frequencies must be prewarped by

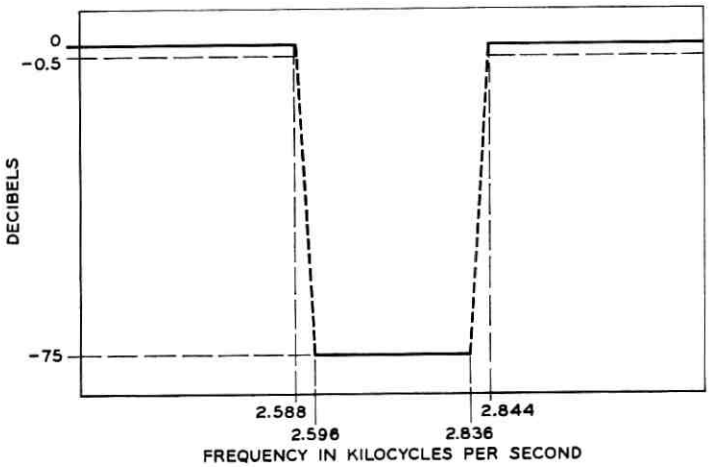


Fig. 4 — Desired amplitude response characteristic of a bandstop filter.

means of (8). The warped values are:

f_{lp} = lower cutoff frequency in passband (2588 cps) = 3364.15 cps
f_{lr} = lower cutoff frequency in rejection band (2596 cps)
= 3381.13 cps
f_{ur} = upper cutoff frequency in rejection band (2836 cps)
= 3937.54 cps
f_{up} = upper cutoff frequency in passband (2844 cps) = 3957.84 cps.

The warped values at the lower band edge require a low-pass filter with a transition ratio of 0.93792, while the values at the upper band edge require a transition ratio[6] of 0.93658. Therefore, to meet the original specifications, the larger of the two transition ratios must be chosen. Hence specifications required for the basic low-pass elliptic filter are:

in-band ripple = 0.5 db
out-of-band minimum attenuation = 75.0 db
transition ratio = 0.93792.

Application of elliptic filter design procedure with these specifications yields a basic low-pass structure of eleventh order. The poles and zeros for the transfer function of this low-pass filter are listed in Table I. The low-pass filter has been normalized to have a cutoff frequency of one radian per second and amplitude gain of unity at zero frequency.

TABLE I—POLES AND ZEROS OF NORMALIZED ELEVENTH-ORDER ELLIPTIC LOW-PASS FILTER

In-band ripple = 0.500 db
Minimum attenuation = 76.504 db
Transition ratio = 0.937917

Gain factor = 0.0011060

Poles
$-0.0069130 \pm j\ 1.0010752$
$-0.0257616 \pm j\ 0.9756431$
$-0.0615122 \pm j\ 0.9063786$
$-0.1269215 \pm j\ 0.7504391$
$-0.2142976 \pm j\ 0.4483675$
-0.2611853

Zeros
$\pm j\ 1.0695414$
$\pm j\ 1.1009005$
$\pm j\ 1.1946271$
$\pm j\ 1.4652816$
$\pm j\ 2.5031313$

The desired bandstop filter is obtained next by applying the low-pass-to-bandstop transformation, given in (11), to the normalized elliptic low-pass filter. The cutoff frequencies used are the warped values obtained above for f_{lp} (3364.15 cps) and f_{up} (3957.84 cps). The resulting bandstop filter is then transformed by the bilinear z form to yield the required sampled-data filter. Table II lists the coefficients of the resulting sampled-data filter needed for parallel realization of the form shown in Fig. 3. Fig. 5 shows the frequency response characteristics of this sampled-data filter. It is seen that the original filter specifications are met by the sampled-data filter. For comparison purposes, the standard z transform was applied directly to the twenty-second-order continuous filter. The frequency response characteristic of this filter is shown in Fig. 6. It is seen that the standard z transform has yielded an unusable result.

VI. SUMMARY

The need for sampled-data filters in wideband simulations of many processing systems has led to a synthesis method which overcomes the shortcomings of the standard z transform. The method presented consists of directly transforming a suitable continuous transfer function to a sampled-data filter by means of the bilinear z form. For wideband filters the method is particularly suited to those filters that exhibit relatively constant magnitude-frequency characteristics in successive pass and stop bands. Conventional design techniques of continuous filters are used

TABLE II—PARTIAL-FRACTION EXPANSION COEFFICIENTS FOR PARALLEL REALIZATION OF SAMPLED-DATA BANDSTOP FILTER

Term	Numerator Coefficients		Denominator Coefficients	
	A_1	A_0	B_2	B_1
1	0.0001628	0.0008827	0.9987854	0.1106416
2	−0.0009283	−0.0001764	0.9989898	0.4285348
3	−0.0024098	−0.0027894	0.9956089	0.1063723
4	0.0031774	0.0026966	0.9957459	0.4317548
5	0.0102446	0.0026026	0.9879911	0.0940731
6	−0.0037799	−0.0112135	0.9883051	0.4414974
7	−0.0277640	0.0127415	0.9651789	0.0616261
8	−0.0108027	0.0289421	0.9661438	0.4663508
9	0.0272223	−0.1163873	0.8694592	−0.0204564
10	0.1206914	−0.0054765	0.8742300	0.5186036
11	0.2973946	−0.2973227	0.5283651	0.2074591

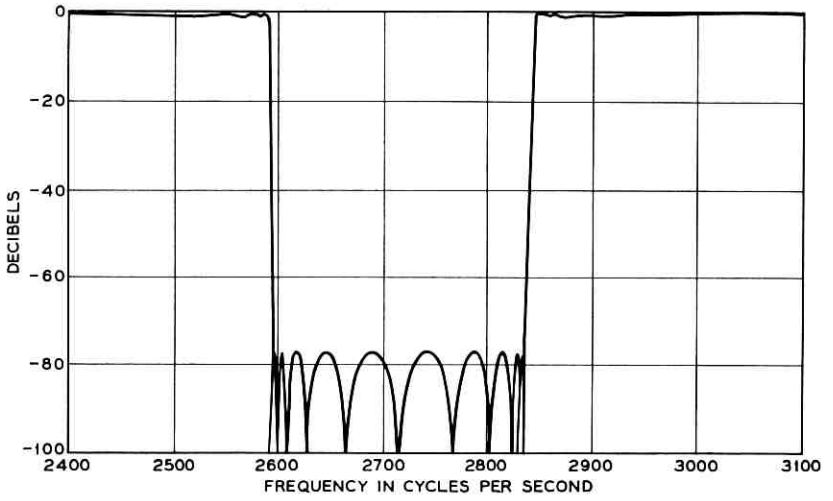


Fig. 5 — Frequency response characteristics of the sampled-data bandstop filter designed by the bilinear z transformation.

directly in the synthesis procedure of the sampled-data filters. (Thus the synthesized filters have frequency characteristics comparable to those of continuous filters.) An example has been presented of a filter function synthesized by this procedure and easily programmed for a simulation. Results obtained from this example demonstrate the usefulness and accuracy of the bilinear z-form method.

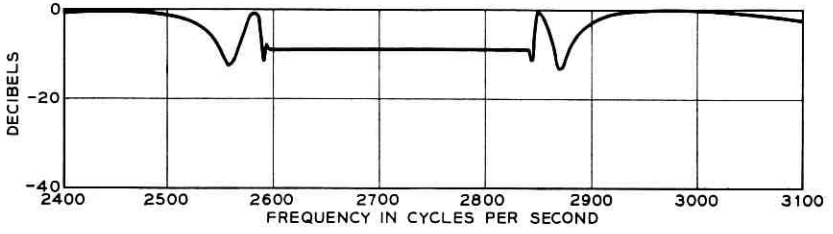


Fig. 6 — Frequency response characteristics of the sampled-data bandstop filter designed by the standard z transformation.

REFERENCES

1. Kaiser, J. F., Design Methods for Sampled-Data Filters, Proc. First Allerton Conference on Circuit and System Theory, Nov., 1963, Monticello, Illinois.

2. Golden, R. M., Digital Computer Simulation of a Sampled-Data Voice-Excited Vocoder, J. Acoust, Soc. Am., **35,** Sept., 1963, pp. 1358–1366.
3. Wilts, C. H., *Principles of Feedback Control,* Addison-Wesley, 1960, pp. 197–207.
4. Hamming, R. W., *Numerical Methods for Scientists and Engineers,* McGraw-Hill, New York, 1962, pp. 277–280.
5. Kelly, J. L., Jr., Lochbaum, C., and Vyssotsky, V. A., A Block Diagram Compiler, B. S. T. J., **40,** May, 1961, pp. 669–676.
6. Storer, J. E., *Passive Network Synthesis,* McGraw-Hill, New York, 1957, pp. 293–296.

The ALPAK System for Nonnumerical Algebra on a Digital Computer — III: Systems of Linear Equations and a Class of Side Relations

By J. P. HYDE

(Manuscript received March 6, 1964)

This is the third and last in a series of papers describing the ALPAK system for nonnumerical algebra on a digital computer. The first paper[1] *is concerned with polynomials in several variables and truncated power series with polynomial coefficients. The second paper*[2] *is concerned with rational functions in several variables and truncated power series with rational-function coefficients. The present paper discusses systems of linear equations with rational-function coefficients and a certain class of side relations.*

The ALPAK system has been programmed within the BE-SYS-4 monitor system on the IBM 7090 computer, but the language and concepts are machine independent. Several practical applications are described in Ref. 1.

This paper is divided into six sections. The first two assume that the reader has no knowledge of computers or computer programming and the last four assume that the reader is familiar with basic computer programming and Refs. 1 and 2. Section I is a general description of ALPAK and this paper; Section II discusses the different forms in which a linear system can occur, including canonical form; Section III describes the ALPAK linear system operations for converting these forms; Section IV discusses side relations; Section V describes list naming operations; and Section VI discusses possible future developments and improvements.

I. INTRODUCTION

This is the third and last in a series of papers describing the ALPAK system, a programming system for performing routine manipulations of algebraic expressions on a digital computer. The system can perform the operations of addition, subtraction, multiplication, division, sub-

stitution, and differentiation. The first paper[1] is concerned with polynomials in several variables and truncated power series with polynomial coefficients. The second paper[2] is concerned with rational functions in several variables and truncated power series with rational-function coefficients. The present paper describes the ALPAK facilities for manipulating and solving by Gaussian elimination systems of equations linear in certain variables with coefficients which are rational functions of other variables. The facilities for handling a certain class of side relations are also described.

The ALPAK system has been programmed within the BE-SYS-4 monitor system on the IBM 7090 computer, but the language and concepts are machine independent. Several practical applications are described in Ref. 1.

This paper is divided into six sections, of which the first two do not presuppose any knowledge of computers or computer programming and the last four assume that the reader is familiar with the basic concepts of computer programming and Refs. 1 and 2. Section I is a general description of ALPAK and deals with basic concepts. Section II describes the different forms in which a linear system can occur, including especially the canonical form of a linear system. Section III discusses the ALPAK linear system operations for converting these forms.

Section IV describes the way in which ALPAK has been programmed to simplify a rational function, using a certain class of side relations. The most important relations in the allowed class are of the form $X^2 = C$ (C a rational function independent of X). This includes in particular $i^2 = -1$ and $s^2 = 1 - c^2$ where s and c can stand for $\sin \alpha$ and $\cos \alpha$, respectively. The simplification is done by a special rearrangement of the ALPAK format statement and has certain limitations.

Section V discusses list naming operations, a convenient set of auxiliary operations for handling arguments of ALPAK subroutines which are lists (one-dimensional arrays). Finally, Section VI discusses possible future developments and improvements.

1.1 *An Example of the ALPAK Language*

The simplicity of handling linear systems by ALPAK is illustrated in the process of solving the following system of two linear equations, *EQ1* and *EQ2*, in two unknowns, *X1* and *X2*, with polynomials in a as coefficients.

$$\text{EQ1: } 3aX1 + 2aX2 - 1 = 0$$

$$\text{EQ2: } 2aX1 + 5a^2X2 - 3 = 0$$

We first extract a coefficient matrix, SYS, for the equations with -1 and -3 moved to the right side.

$$\text{SYS:} \left\| \begin{matrix} 3a & 2a & 1 \\ 2a & 5a^2 & 3 \end{matrix} \right\|.$$

The matrix is then put into canonical form using Gaussian elimination.[3,4]

$$\text{SYS:} \left\| \begin{matrix} 1 & 0 & \dfrac{-6 + 5a}{-4a + 15a^2} \\ 0 & 1 & \dfrac{7}{-4a + 15a^2} \end{matrix} \right\|.$$

The fact that the original coefficient matrix and the canonical form matrix both have the name *SYS* does not imply that they are equal but rather that the latter replaces the former *physically* in the computer. The expressions for the unknowns are then extracted from the coefficient matrix.

$$\text{X1:} \quad \frac{-6 + 5a}{-4a + 15a^2}$$

$$\text{X2:} \quad \frac{7}{-4a + 15a^2}.$$

The following program illustrates how these operations are performed by ALPAK.

SYS	SYSRES	2,2	Reserve space in the computer for the physical representation of the set of system coefficients which will be obtained from two linear equations in two unknowns and name the set *SYS*.
	SYSFRM	SYS,(EQ1,EQ2),=2	Extract the 2×3 coefficient matrix from the equations *EQ1* and *EQ2* and place it in *SYS*. The "=2" says that there are two unknowns and the third column of the matrix is used for the terms of the equations which are independent of the unknowns.
	SYSPRT	SYS	Print the system coefficients.

SYSCFM SYS

Put the system into canonical form (described in next section) using Gaussian elimination. If the system were triangular, the row selection strategy would cause this to be done in the obvious way. The canonical system retains the same name as the original system.

SYSSLV (X1,X2),SYS

Fill *X1* and *X2* with the solutions for the unknowns in *SYS*. The operation SYSSLV assumes that *SYS* is in canonical form.

RFNPRT X1
RFNPRT X2

Print *X1* and *X2*.

X1
X2

These are names of single cells in memory which will be filled in with "pointers"* to the physical representations of the solutions in the computer.

The usefulness of the linear system operations was demonstrated in a problem from queuing theory, proposed by L. Takacs,† in which a truncated power series of 813 terms was involved in forming a system of nine linear equations in nine unknowns. One of these unknowns was the third moment of a probability distribution. Its numerator had 200 terms in five variables with maximum degrees 1, 1, 3, 7, and 9 and its denominator had 39 terms in two variables with maximum degrees 7 and 10.

II. LINEAR SYSTEMS

In this section are discussed the different forms of linear systems as they are dealt with in the ALPAK context. It is important in writing ALPAK programs to remember what these forms are. The next section discusses the ALPAK subroutines for changing one form to another.

* See Ref. 2, p. 795.
† See Ref. 1, pp. 2090–2092.

2.1 *System of Equations*

A linear system of m equations in n unknowns, x_j, is a set of m rational functions, (1), of v variables ($v \geqq n$), each of which is linear in the x_j and is implicitly equal to zero.

$$\sum_{j=1}^{n} \lambda_{ij} x_j - c_i = 0 \qquad (1 \leqq i \leqq m). \tag{1}$$

Thus for each i in (1) the λ_{ij} are the coefficients of the x_j and together with c_i may be thought of as $n + 1$ rational functions with a common denominator.

2.2 *System Coefficients*

Consider (1) written in the form:

$$\sum_{j=1}^{n} \lambda_{ij} x_j = c_i \qquad (1 \leqq i \leqq m). \tag{2}$$

The λ_{ij} and the c_i of (2) shall be referred to as the *system coefficients* of the linear system (1). In ALPAK they form an array of $m(n + 1)$ rational functions stored row-wise and forwards.

2.3 *System Canonical Form*

Let $x_{a_1}, \cdots, x_{a_r}$ be a subset of the unknowns $x_1, \cdots, x_n$ which we shall call the *dependent set*, and let $x_{a_{r+1}}, \cdots, x_{a_n}$ be the remaining unknowns, which we shall call the *independent set*. The dependent set is said to be *valid* if r is the rank of the system and if the associated columns of system coefficients are linearly independent over the field to which they belong. The system

$$x_{a_i} + \sum_{j=r+1}^{n} \lambda_{ij} x_{a_j} = c_i \qquad (1 \leqq i \leqq r) \tag{3}$$

and its array of system coefficients are both said to be in *canonical form* with respect to such a dependent set.* It can be shown that for any linear system and a given valid dependent set, there exists a unique canonical form which is obtainable from the original system and which is satisfied by the same values of the x_{a_i}. One obtains this canonical form by Gaussian elimination; i.e., operating on the system by suitably chosen row operations and column interchanges.† When there is a choice of row interchange, the row with the most zero coefficients is selected to minimize the work involved. If, in the derived canonical form, $r < m$,

* The dependent set in (3) is clearly valid.
† See Refs. 3 and 4.

the last $m - r$ rows should be of the form $0 = 0$. If they are not, the system is said to be *inconsistent*. If $r < n$, the system is said to be *singular*.

2.4 *System Solution*

The solution of a linear system is a set of r rational functions, (4), of v variables ($v \geqq n - r$), each of which is linear in the x_{a_j} ($r + 1 \leqq j \leqq n$) and is implicitly equal to x_{a_i} .

$$x_{a_i} = c_i - \sum_{j=r+1}^{n} \lambda_{ij} x_{a_j} \qquad (1 \leqq i \leqq r). \tag{4}$$

The solution is easily produced once the system is in canonical form, and if the system is nonsingular the solution is of the form $x_{a_i} = c_i$ ($1 \leqq i \leqq n$).

III. LINEAR SYSTEM OPERATIONS

3.1 *General Remarks*

In this section are discussed the ALPAK subroutines for converting the different linear system forms discussed in Section II. The name of a set of system coefficients must be defined by operations SYSNAM or SYSRES if it is to be used in any other operations. This name is the BSS address of a three-word system heading in which are stored the five system parameters. These parameters are the BSS address of the system coefficients, the number of equations, the number of unknowns, an ALPAK format address, and the number of leading variables in this format of which the equations are independent. They are set at assembly time by operations SYSNAM and SYSRES or at run time by operations SYSSET and SYSMPR.

The $m(n + 1)$ system coefficient pointers are stored row-wise and forwards, and a block of $n + 1$ cells must immediately follow to be used by ALPAK as work space. In the ALPAK format statement of the system equations, the n unknowns must have consecutive variable numbers $k + 1$ through $k + n$ ($k \geqq 0$). If $k > 0$, the system equations must be independent of the first k variables, and thus the system coefficients are independent of the first $k + n$ variables. The system parameter *fmt* is normally this ALPAK format statement and is referenced in any system operations involving the names of the unknowns. If it is not supplied by SYSNAM, SYSRES, SYSFRM, or SYSSET, all such operations must refer to variables by number (VARTYP NUM or VARTYP NUM*).

Those arguments of operations SYSFRM, SYSCFM, and SYSSLV which are lists are specified according to the conventions established

in Section V. System parameters and system names are not indexable, but the addresses where they are stored or to be stored are. As in other ALPAK operations, index registers are preserved with the exception of index register four.

3.2 *Notational Conventions*

The following conventions of notation are used in descriptions of instructions. Upper-case letters are used for operation codes (including macro names) and for any parameters which must appear exactly as shown. Dummy parameters are indicated by lower-case letters. A dummy parameter usually stands for the symbolic address of a cell or block of cells in the program where the argument is stored. Those dummy parameters which are the arguments themselves are in boldface. Finally, optional parameters are enclosed in brackets, and parameters which usually have subarguments are enclosed in parentheses. All integer arguments are decimal. By this notation, then, the instructional description

```
sys     SYSRES     m,n,[fmt],[k]
```

specifies certain properties and restrictions about the arguments of the following call:

```
COEFF   SYSRES     9,9,,INDEP
```

Thus, only SYSRES must appear exactly as shown and all other parameters are dummies with the third one omitted, as it is optional. The number of equations is nine, but the number of leading variables of which the equations are independent is in the cell whose symbolic address is INDEP.

3.3 *Linear System Operations*

sys	SYSNAM	bss,**m**,**n**,[fmt],[k]	name	(a)
sys	SYSRES	**m**,**n**,[fmt],[k]	reserve	(b)
	SYSPRT	sys	print	(c)
	SYSFRM	sys,(listr),n,[k]	form	(d)
	SYSCFM	sys,[(listv)],[**inc**],[**ids**]	canonical form	(e)
	SYSSLV	(listr),sys,[(listv)]	solve	(f)
	SYSOBT	[(abss)],[(m)],[(n)],[(afmt)], [(k)],sys	obtain parameters	(g)
	SYSSET	sys,[abss],[m],[n],[afmt],[k]	set parameters	(h)
	SYSMPR	sys,[(**op** oper)], [(**op** oper)] [(**op** oper)], [(**op** oper)], [(**op** oper)]	modify parameters	(i)

sys = name of system (symbolic address of heading)
bss = BSS address of the array of system coefficients
abss = address where *bss* is or is to be stored
m = the number of equations in the system
n = the number of unknowns in the system
fmt = the address of the system's ALPAK format statement
afmt = address where *fmt* is or is to be stored
k = the number of leading variables in this format statement of which the system equations are independent
listr = list of rational functions (see Section V)
listv = list of variables (specified in the manner indicated by the last previous VARTYP declaration — see Section V)
(op oper) = a 7090-94 machine operation and an operand separated by a blank
inc = inconsistency return
ids = invalid dependent set return.

3.4 *Descriptions*

(a) sys SYSNAM bss,**m**,**n**,[fmt],[k]

Declare a block of length $(m + 1)$ $(n + 1)$ starting at *bss* to be a set of linear system coefficients and work space, and name it *sys* by reserving remotely a three-word system heading. This heading is filled in with *bss*, *m*, *n*, *fmt*, and *k*. If *fmt* and/or *k* is omitted, the corresponding fields in the system heading are filled in with zeros.

(b) sys SYSRES **m**,**n**,[fmt],[k]

Reserve remotely a block of length $(m + 1)$ $(n + 1)$ for a set of system coefficients and work space, and name the set *sys* by reserving remotely a three-word system heading as in SYSNAM. *sys* is to be filled in at run time (e.g., by SYSFRM).

(c) SYSPRT sys

Print the set *sys* of system coefficients.

(d) SYSFRM sys,(listr),n,[k]

Replace *sys* by the set of system coefficients formed from the set *listr* of system equations and remove the common factors between the coefficients of any given equation (*listr* is destroyed). The contents of *n* and *k* and the number of rational functions in *listr* together with their format are copied into the heading of *sys*. If *k* is not supplied, it

is assumed to be zero. If SYSFRM is not used to fill in *sys*, the system parameters must be filled in with operations (a), (b), or (h).

(e) SYSCFM sys,[(listv)],[inc],[ids]

Replace the set *sys* of system coefficients by its associated canonical set, using Gaussian elimination. *listv* is a list of unknowns (specified in the manner indicated by the last previous VARTYP declaration) to be included in a valid dependent set. If *listv* is not supplied, the list is assumed to be empty. If *sys* is found to be inconsistent, control will be transferred to *inc* (or to the REMARK subroutine if *inc* is not supplied) and *sys* will have a canonical form with an inconsistency. If the set of unknowns in *listv* cannot be included in a valid dependent set, control will be transferred to *ids* (or to the REMARK subroutine if *ids* is not supplied) and *sys* will have a canonical form with some subset of *listv* in the dependent set. At *inc* or *ids* it is possible to call SYSSLV, SYSPRT, or to go to some other part of the program.

(f) SYSSLV (listr),sys,[(listv)]

Replace *listr* (whose length must not be less than that of *listv*) by the solutions for the list of unknowns *listv* (specified in the manner indicated by the last previous VARTYP declaration). *sys* is assumed to be in canonical form. If *listv* is not supplied, all the unknowns in the dependent set are solved for in the order in which they were at the start of SYSCFM.

(g) SYSOBT [(abss)],[(m)],[(n)],[(afmt)],[(k)],sys

Obtain the system coefficient parameters of the system whose name is *sys*. Each optional argument is a memory location in whose address field the parameter is to be stored. Thus the parameter *bss* is stored in the location *abss* specified by SYSOBT, etc. Each optional argument may actually be several arguments, and if an argument is an integer equal to seven or less, it refers to an index register.

(h) SYSSET sys,[abss],[m],[n],[afmt],[k]

Set the system coefficient parameters of the system whose name is *sys* from the locations specified by the bracketed arguments. Thus the parameter *bss* is set to the contents of the location *abss* specified by SYSSET, etc.

(i) SYSMPR sys,[(**op** oper)],[(**op** oper)]

[(**op** oper)], [(**op** oper)], [(**op** oper)]

Modify the system coefficient parameters of the system whose name is *sys* using the 7090-94 machine operations *op* with operands *oper*. Thus, the parameter *bss* is modified by the first operation and operand, *m* is modified by the second, *n* by the third, *fmt* by the fourth, and *k* by the fifth. Each operation and operand may be different. Typically, the operation is ADD or SUB and the operand is the address of some increment or decrement.

IV. SIDE RELATIONS

4.1 *General Remarks*

The ALPAK programmer may find that expressions involving radicals occur in his problem. A radical can be handled by assigning it a variable name and writing the rational functions using this name. Thus in the polynomial $a + 2a\sqrt{3}$, we let $X = \sqrt{3}$ and the expression becomes $a + 2aX$. The problem is that in the outputs of arithmetic operations involving such rational functions, X can have an exponent greater than one and the fact that $X^2 = 3$, $X^3 = 3X$, $X^4 = 9$, $\cdots$ will be ignored. The implicit equation $X^2 = 3$ is called a *side relation* of degree two on X. A subroutine is provided for simplifying rational functions using side relations of the general form

$$X^{2^j} = C \qquad \begin{matrix} (j \text{ an integer} \geqq 1) \\ (C \text{ a rational function independent of } X). \end{matrix} \tag{5}$$

This category includes especially $i^2 = -1$ and $s^2 = 1 - c^2$ where s and c can stand for $\sin \alpha$ and $\cos \alpha$, respectively.

4.2 *Limitations*

Many limitations exist in the present handling of side relations. In the relation $X^n = C$, n must be a power of two and X a single variable. Dependencies between relations are not observed; i.e., $R^2 = 2$ and $S^2 = 3$ and $T^2 = 6$ will not result in the implicit relation of $T = \pm RS$. Moreover, relations are not handled automatically by the lowest-level subroutines, thus causing exponents to grow unnecessarily until simplification is done at main program level. A more sophisticated version of ALPAK would prevent this by including the relations as part of the format statement. To repair these limitations would require a great deal of extra programming, and it turns out in practice that these limitations do not usually matter. The general problem of dependencies is especially difficult, as it involves algebraic extensions of the field of rational functions of several variables.

4.3 *Implementation*

The simplification of a rational function, *RF*, by a side relation $X^{2^j} = C$ is accomplished with the aid of a specially constructed temporary ALPAK format statement. The temporary format is the same as the original one except that the exponent field of X is split into two parts. The right j bits are assigned the name X and the remaining left-hand bits form a temporary exponent field which is assigned any name and stands effectively for X^{2^j}. The rational function C is then substituted for the temporary variable by the call SIDREL. If several side relations are defined on several variables, then a single temporary format statement can be used to split up these variables. There will then be a list of rational functions to be substituted for the temporary variables by a single call to SIDREL (see Section 3.2).

SIDREL rf,(listv),(listr),tfmt

rf = rational function to be simplified
listv = list of temporary variables (specified in the manner indicated by the last previous VARTYP declaration — see Section V)
listr = list of rational functions to be substituted for these variables and specified in the same order (see Section V)
tfmt = address of temporary format.

The format of *rf* after simplification is the format of the items in *listr*, or if none of these items has a format, the format of *rf* is unaltered.

4.4 *Example*

Suppose it is desired to simplify the function *RF* using the side relation $I^2 = -1$. This is done by the following program.

```
FMT     POLCVF    (X,5,Y,5,I,5,Z,21)
                                  Permanent format.
TFMT    POLCVF    (X,5,Y,5,ISQ,4,I,1,Z,21)
                                  Temporary format with I split up
                                  into ISQ with four bits and I with
                                  one bit.
        VARTYP    NAM
        POLSTC    MON,=-1
        SIDREL    RF,ISQ,MON,TFMT
          ⋮
RF
MON
```

Testing equality of rational functions R and S which have been simplified by a side relation should always be done by subtracting and testing for zero as follows:

```
RFNSUB   TEMP,R,S
SIDREL   TEMP,ISQ,MON,TFMT
   ⋮                The side relation applied to R and
                    S is applied to TEMP.
RFNZET   TEMP
                    Test if TEMP is zero.
```

This procedure will recognize that the expressions $(1 + i)/(1 - i)$ and i are equal.

V. LIST NAMING OPERATIONS

5.1 *General Remarks*

Whenever an ALPAK subroutine argument is a list, the list may be specified either by actually listing the contents; e.g.,

```
SYSSLV      (P,Q,R),SYS,(X,Y,Z)
```

or by name; e.g.,

```
SYSSLV      (LISTP,*),SYS,(LISTV,*)
```

Here the asterisk indicates that the list has been specified by name. Both methods may be used within the same command; e.g.,

```
SYSSLV      (P,Q,R)SYS(LISTV,*)
```

This section describes a set of operations LSTNAM, LSTMAK, and LSTRES for assigning names to lists and blocks of storage, thus enabling one subsequently to call them by these names in the appropriate subroutines. The operations LSTOBT, LSTSET, and LSTMPR serve as auxiliary operations. The facilities are especially useful whenever the items of the list are to be filled in at run time or whenever the items do not form a contiguous block in core. A list has two parameters, which are its BES address and its length. These can be set at assembly time by LSTNAM, LSTMAK, and LSTRES or changed at run time by LSTSET and LSTMPR (see Section 3.2). List parameters and list names are not indexable, but the addresses where they are stored or to be stored are. Each item in the specified contents of a list may be tagged. Index registers are preserved with the exception of index register four.

5.2 *List Naming Operations*

ttl	LSTNAM	bes,**lng**,[VAR]	name	(a)
ttl	LSTMAK	(items),[VAR]	make	(b)
ttl	LSTRES	**lng**,[VAR]	reserve	(c)
	LSTOBT	[(abes)],[(lng)],ttl	obtain	(d)
	LSTSET	ttl,[abes],[lng]	set	(e)
	LSTMPR	ttl,[(**op** oper)],[(**op** oper)]	modify	(f)

ttl = name for list
bes = BES address of list
abes = address where "bes" is or is to be stored
lng = length of list
items = contents of list
(**op** oper) = 7090–94 machine operation and an operand separated by a blank.

5.3 *Descriptions*

(a) ttl LSTNAM bes,**lng**,[VAR]

Declare a set of items in a contiguous block of length *lng* to be a list whose BES address is *bes*, name it *ttl*, and set the list parameters to *bes* and *lng*. If *VAR* is present, the list is assumed to consist of variables (specified in the manner indicated by the last previous VARTYP declaration). If *VAR* is not supplied, the list is assumed to consist of rational functions, polynomials, etc. (i.e., of symbolic addresses of pointers.)

(b) ttl LSTMAK (items),[VAR]

Declare the set whose elements are the subarguments in *items* to be a list, name it *ttl*, and set the list parameters accordingly. *VAR* is as described in LSTNAM. The items need not be in a contiguous block as in LSTNAM.

(c) ttl LSTRES **lng**,[VAR]

Reserve remotely a block of length *lng* for a list, name it *ttl*, and set the list parameters to the BES address of the block and *lng*. *VAR* is as described in LSTNAM. The list is to be filled in at run time (e.g., by SYSSLV).

(d) LSTOBT [(abes)],[(lng)],ttl

Store the BES address of the list whose name is *ttl* in location *abes* and store its length in *lng*. The bracketed arguments may actually

consist of several subarguments, and if an argument is an integer equal to seven or less, it refers to an index register.

(e) LSTSET ttl,[abes],[lng]

Set the BES address of the list whose name is *ttl* to the contents of *abes* and set its length to the contents of *lng*.

(f) LSTMPR ttl,[(**op** oper)],[(**op** oper)]

Modify the BES address of the list whose name is *ttl* using the 7090-94 machine operation *op* with operand *oper* specified by the first bracketed argument. Modify the length of the list in a similar manner as indicated by the second bracketed argument. Typically, the operation is ADD or SUB and the operand is the address of some increment or decrement.

5.4 *Example*

The following example shows how list naming can be used to good advantage. We are given polynomials (A_1, $\cdots$, A_y; $y \leqq 15$), a set of variable names (m_1, $\cdots$, m_y; $y \leqq 15$), and polynomials (F_1, $\cdots$, F_y; $y \leqq 15$). It is desired to form a set of polynomials (G_1, $\cdots$, G_y; $y \leqq 15$) in the following way, where m_i: A_i means A_i is substituted for m_i

$$\begin{aligned} G_1 &= F_1(m_1{:}A_1) \\ G_2 &= F_2(m_1{:}A_1, m_2{:}A_2) \\ &\vdots \quad\quad \vdots \\ G_y &= F_y(m_1{:}A_1, m_2{:}A_2, \cdots, m_y{:}A_y). \end{aligned}$$

Assume that the F_i's, A_i's, M_i's, and G_i's are stored forwards in blocks whose BES addresses are F, A, M and G respectively and that the parameter y is in location Y. The following program will perform the substitution.

```
POLS    LSTNAM   A,15
VARS    LSTNAM   M,15
                          Define the lists thus setting the list
                          parameters to (A,15) and (M,15)
        LSTMPR   POLS(SUB  Y) (SUB  =15)
        LSTMPR   VARS(SUB  Y) (SUB  =15)
                          Initialize the list parameters to (A -
                          y,0) and (M - y,0).
        LXA      Y,1
```

```
LOOP    LSTMPR   POLS(ADD  =1) (ADD  =1)
        LSTMPR   VARS(ADD  =1) (ADD  =1)
        POLSST   (G,1) (F,1) (POLS,*) (VARS,*)
        TIX      LOOP,1,1
```

Increment the list parameters by one at each repetition of the above loop.

```
        ⋮
F       BES      15
A       BES      15
M       BES      15
G       BES      15
Y
```

VI. OUTLOOK

Our experience has shown us that the present handling of linear systems has its limitations. Large linear systems are always difficult to put into canonical form, and even a relatively simple set of system coefficients can grow quite rapidly throughout the course of SYSCFM and cause some form of overflow. This growth becomes coupled with the growth produced by the greatest common divisor algorithm,* thus making the inadequacies of the latter most apparent. The success or failure of SYSCFM depends less on the dimensions of the system and more on the internal structure and size of the individual coefficients. Moreover, it is very difficult to tell by looking at the input array whether the structure and size at a later stage of the reduction will cause trouble. This difficulty is illustrated in that SYSCFM succeeded in reducing a 9×9 array with large, apparently complex, entries,† but failed in a related queuing theory problem to reduce a 10×10 array whose entries averaged only two or three terms. It can at least be said that there will be no GCD problems if the original array consists of all rational numbers.

The subroutine SYSCFM is perhaps too comprehensive. A series of orders which would enable one to perform the Gaussian elimination method a step at a time, leaving the choice of row and column permutations completely up to the user, might be useful. SYSPRT could then be called at any time during the reduction. A routine for evaluating determinants, if available, would enable the solution of nonsingular systems by Cramer's method as an alternative.

* See Ref. 2, pp. 791–794.
† See Ref. 1, pp. 2090–2092.

The growth problem could be reduced by allowing multiple precision polynomial coefficients and by allowing a polynomial to be represented as a product of polynomials (not necessarily irreducible). Thus one could compute the GCD as a product of simpler GCD's. To do this would require the ability to have a data structure hierarchy in the data buffer more complicated than that of a rational function.* This ability would also enable a linear system itself to be such a data structure rather than an array in the main program.

A new version of ALPAK (to be called ALPAKB) is now being developed. Its foundation is a programming system[5] called OEDIPUS (Operating Environment with Dynamic storage allocation, Input-output, Public push down list, Unhurried diagnostics, and Symbolic snaps) which provides for the dynamic storage allocation of such data structures, among other things. ALPAKB will also include multiple precision integer arithmetic which will handle polynomial coefficient overflow.

VII. ACKNOWLEDGMENT

I would like to thank W. S. Brown for many valuable suggestions and discussions concerning every aspect of this paper.

REFERENCES

1. Brown, W. S., The ALPAK System for Nonnumerical Algebra on a Digital Computer — I: Polynomials in Several Variables and Truncated Power Series with Polynomial Coefficients, B.S.T.J., **42**, Sept., 1963, p. 2081.
2. Brown, W. S., Hyde, J. P., and Tague, B. A., The ALPAK System for Nonnumerical Algebra on a Digital Computer — II: Rational Functions of Several Variables and Truncated Power Series with Rational-Function Coefficients, B.S.T.J., **43**, March, 1963, p. 785.
3. Hyde, J. P., unpublished work.
4. Stoll, Robert R., *Linear Algebra and Matrix Theory*, McGraw-Hill, New York, 1952. (See especially Chap. 1.)
5. Brown, W. S., and Leagus, D. C., OEDIPUS: Operating Environment with Dynamic storage allocation, Input-output, Public push down list, Unhurried diagnostics, and Symbolic snaps, to be published.

* See Ref. 2, p. 794.

A Technique for Measuring Small Optical Loss Using an Oscillating Spherical Mirror Interferometer

By A. J. RACK and M. R. BIAZZO

(Manuscript received March 19, 1964)

The measurement of very small optical losses (the order of a few per cent) by conventional methods becomes very difficult because of the extreme accuracy required. This article shows that both high mirror reflectances and low transmission losses can be readily measured using an oscillating mirror interferometer as a frequency spectrum analyzer. The theory developed shows that when this type of interferometer is excited by a continuous gaseous laser, the total optical loss is proportional to the frequency resolution or the finesse. The theory also shows that the first-order velocity effect produced by having the mirror move at a velocity of one foot per hour can be large if the total optical loss is about 0.25 per cent. For the velocities and optical losses we have encountered so far in our measurement system, the first-order mirror velocity effect can be neglected. The range of reflectance of mirrors we have measured is from 94 to 99.5 per cent, and the measurements for the optical transmission loss range from 0.2 to 3 per cent. The accuracy to which a 1 per cent loss can be repeated is 1.0 ± 0.1 per cent. It was found that the transmission loss through an optical grade of fused quartz (Homosil) at 6328 Å is about 1 db per meter, and that for Plexiglas II is about 2 db per meter.

I. INTRODUCTION

In developing a long-distance optical communication system employing a large number of components, it is essential to be able to measure accurately the optical transmission loss of each component. These components may include mirrors with reflectances in the order of 99 per cent, Brewster angle output windows, various lenses, and other passive elements with optical transmission losses of 1 per cent or less. For such small losses, the conventional measuring techniques become increasingly difficult because of the extreme accuracy required.

In conventional measuring systems, the loss (or reflectance) is calculated by comparing the electrical output of a photodetector for two different optical conditions, first with the unknown in the system, and then with the unknown removed. As the magnitude of the optical loss decreases, this comparison becomes more and more inaccurate, as it requires the measurement of a small difference between two relatively large photodetector outputs. In these measuring systems, a number of methods have been developed to minimize these errors, which are principally optical.[1] However, it is believed that measuring techniques to be described below will more readily measure very small optical losses.

The proposed method uses a frequency spectrum analyzer at optical frequencies. Such an analyzer can be obtained by using a Fabry-Perot type of interferometer as a transmission element between an optical source and a photodetector.[2,3] When the mirror separation of the interferometer is varied periodically by moving one of the mirrors linearly, a large photocell output will be obtained whenever the optical cavity is in resonance at any frequency that may be present in the optical source. If the photodetector output is observed on an oscilloscope whose sweep is synchronized with the mirror drive, the scope will display the energy distribution of an optical source as a function of frequency. If this type of Fabry-Perot interferometer is illuminated by a continuous laser with its extremely narrow line output (one or two cycles wide), the linewidth of the pattern displayed on the scope is determined by the optical losses in the cavity itself, and by the velocity of the moving mirror. If the excursion of the moving mirror is several optical wavelengths, the scope pattern will repeat several times during a single sweep trace. That is, as the mirror separation increases, the mth harmonic of the cavity becomes resonant with the source; a short time later, the $(m + 1)$th harmonic is resonant with the same optical frequency, then the $(m + 2)$th, and so on. At each resonant point, there will be several output scope pulses, since the laser usually has an output at more than one frequency. The ability of any interferometer to separate or resolve two adjacent optical frequencies is determined by the finesse of the system. In the moving-mirror interferometer, the finesse is equal to the ratio of the fundamental pulse group spacing to the half-power-height width of any pulse. It will be shown below that the total power loss of the Fabry-Perot cavity, expressed as a ratio, is equal to π divided by the finesse.

This interferometer method of measuring small optical losses has been suggested on several occasions,[4] but it is believed that this is the first time such a system has been so fully developed. Since this system of

measurements requires an optical source with an output linewidth very narrow compared to that of the optical cavity, loss measurements cannot be made over a continuous range of wavelengths, but only at the discrete wavelengths at which cw lasers have been developed. Since some light must be transmitted through the cavity, the reflectance can be measured only for those mirrors not coated with an opaque reflecting surface, such as the multiple-layer dielectric coated mirror.

The interferometer loss measuring technique is very sensitive to building and floor vibrations, and to air currents. The reason for this is as follows: if the separation between the two mirrors in the cavity is changed by one half wavelength, the resonant frequency of the interferometer is changed by the natural frequency of the cavity, which in our system is about 1 kmc, whereas the bandwidth of the cavity for 1 per cent optical loss is only about 3.0 mc. Thus, even very small random variations in the mirror spacing in both the laser and in the cavity, produced by either vibrations or air currents, will cause the output pulse pattern displayed on the scope to have large random time position variation.

II. THEORY

The relation for which the mirror reflectance and other optical losses can be calculated from measurable quantities is derived in the Appendix, and will be given briefly here. The effects of the velocity of the moving mirror were included in the Appendix.

Let

R_1, R_2 = power reflectance of the two mirrors expressed as a ratio
g = power loss per single pass through any material within the cavity, also expressed as a ratio
T_c = fundamental pulse group spacing
T_p = half-power output pulse width
v = velocity of moving mirror
d_0 = mirror spacing
t = time
λ = optical wavelength in free space
c = velocity of light in free space

$$\beta_0 = \frac{4\pi d_0}{\lambda}\frac{v}{c}$$

$$x = g(R_1R_2)^{\frac{1}{2}}$$

$$\alpha \equiv -\frac{4\pi vt}{\lambda}$$

$$g_0 \equiv [g(1 - R_1)(1 - R_2)]^{\frac{1}{2}}.$$

An exceedingly good approximation for the time response of the interferometer appropriate for computer calculation is given by (5) of the Appendix:

$$I_T \approx \left[\sum_{n=0}^{\infty} g_0 x^n \cos(\alpha n + \beta_0 n^2)\right]^2 + \left[\sum_{n=0}^{\infty} g_0 x^n \sin(\alpha n + \beta_0 n^2)\right]^2. \quad (5)$$

Since x is very nearly unity, a large number of terms must be taken in the series. For a maximum error of ϵ in stopping the series after the first $N - 1$ terms, we have

$$N \approx \frac{\log(2/\epsilon)}{1 - x}. \quad (6)$$

For example

$$\epsilon = 1/1000 \quad \text{and} \quad x = 0.9975, \qquad N = 3040.$$

The values of I_T shown in Fig. 1 were calculated for the following three

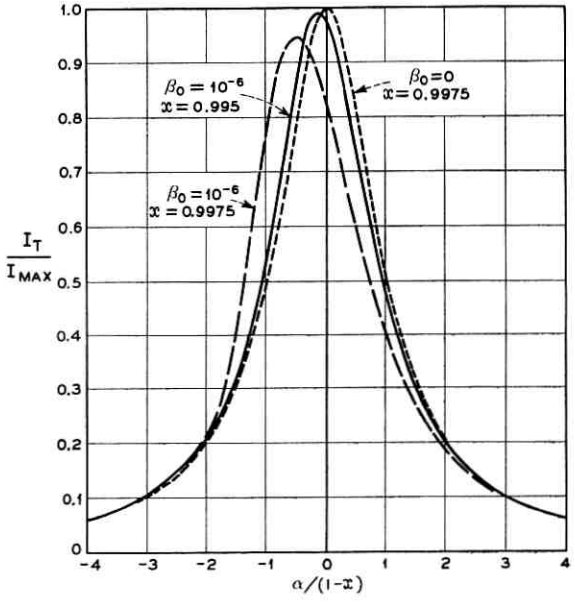


FIG. 1 — Interferometer time response as affected by mirror velocity.

conditions: (a) $\beta_0 = 10^{-6}$, and $x = 0.9975$, (b) $\beta_0 = 0$, $x = 0.9975$, and (c) $\beta_0 = 10^{-6}$ and $x = 0.995$. The curves given in Fig. 1 show that the first-order mirror velocity effect is to decrease the maximum response, increase the half-power-height pulse width, and make the response unsymmetrical about the maximum response.

From a number of computed values for (5), of which only a few are given in Fig. 1, it can be shown that the first-order effect of the moving mirror's velocity is to increase the half-power pulse width according to the relation (see Appendix)

$$\frac{Tm}{Tp} \approx \left[1 + 13.8 \frac{\beta_0^2}{(1 - x)^4}\right]^{\frac{1}{2}}. \tag{7}$$

This relation holds only when T_m is within a few per cent of T_p. For larger values, the computed results are less than those given in (7). Thus, in order to have the width increase by less than 1 per cent,

$$\frac{\beta_0}{(1 - x)^2} \leqq \frac{1}{25}.$$

This relation can readily be satisfied unless the total loss becomes extremely small. In our measurements, the mirror spacing is varied about one micron at a 20-cycle frequency. The mirror spacing is about 15 cm and the wavelength is 6328 Å. Then, if $x = 0.995$ (the largest we have measured),

$$v = \tfrac{1}{250} \text{ cm/sec} \approx 0.4 \text{ ft/hour}$$

and

$$\frac{\beta_0}{(1 - x)^2} = \frac{0.4 \times 10^{-6}}{25 \times 10^{-6}} = \frac{1}{62.5}.$$

Hence, for optical losses of 0.5 per cent or greater, we can neglect the first-order mirror velocity effects.

If mirror velocity effects can be neglected, then the system response given by (10) and (11) of the Appendix can be used. The relation between the optical loss and the finesse of the system is

$$g(R_1R_2)^{\frac{1}{2}} = 1 - (\pi T_p/T_c) + \tfrac{1}{2}(\pi T_p/T_c)^2 + \cdots \tag{11}$$

where T_c/T_p is defined as the finesse. This equation points out the potential accuracy with which small losses can be measured. The interferometer method actually measures how much the combined loss, $g(R_1R_2)^{\frac{1}{2}}$, differs from unity. Obviously, the smaller this difference, the greater can be the experimental errors to obtain a fixed accuracy in

$g(R_1R_2)^{\frac{1}{2}}$. For a value of $g(R_1R_2)^{\frac{1}{2}}$ near 0.99, a 10 per cent error made in the measurement of T_p/T_c would give an error of only one part in a thousand in the value of $g(R_1R_2)^{\frac{1}{2}}$.

To obtain a good signal-to-noise ratio in the electrical output of the photodetector, it is important to obtain a maximum amount of light transmission through the interferometer. From (10), the fraction of the incident light transmitted through the cavity is given by (see Appendix)

$$I_m = \frac{g(1 - R_1)(1 - R_2)}{[1 - g(R_1R_2)^{\frac{1}{2}}]^2}. \tag{10}$$

Equation (10) shows that for $g = 1$ and $R_1 = R_2$, I_m is unity for any value of reflectance, but if $R_1 \neq R_2$, I_m will be less than unity. For example, if $R_1 = 0.995$, $R_2 = 0.97$ and $g = 1$, then $I_m = 0.49$. The remaining 51 per cent of the light is reflected back towards the source. Therefore the two mirror reflectances should be identical for maximum transmission. For measurements of transmission losses in the cavity, there is some advantage to be obtained by not having the mirror reflectances too great. As an example, for $g = 0.97$ and $R_1 = R_2 = 0.99$, $I_m = 0.062$. For the same loss and $R_1 = R_2 = 0.97$, $I_m = 0.25$. In the second case, however, there is a greater chance of making an error in measuring g since it is a smaller fraction of the total loss.

The shape of the output pulse expressed as a function of time is given by (12) of the Appendix

$$I(t) \approx \frac{I_m}{1 + 4t^2/T_p^2}. \tag{12}$$

The frequency spectrum of this time pulse is

$$F(\omega) = \frac{\pi}{2} T_p I_m \exp\left(-\frac{T_p}{2}|\omega|\right).$$

Now, the value of T_p is a function of the velocity of the moving mirror and hence, subject to the limitations on velocity discussed above, may be made as large or small as is desired. In our laboratory, the motion of the moving mirror was so selected that for a 1 per cent total cavity loss, the pulse width is about 30 μsec, and the frequency spectrum is down to 1 per cent of its low-frequency value at about 50 kc. The required bandwidth of the photodetector and its associated electrical circuits, including the viewing oscilloscope, is increased by a decrease in the total optical loss, as this decreases the pulse width, T_p. In order that the system be capable of measuring optical losses as small as 0.25 per cent, the over-all bandwidth of the electrical components should be at least 200 kc.

The above theory was developed for the assumptions that the incident

beam of monochromatic light was collimated and that the plane surface mirrors were infinite in size. For finite-size mirrors with plane or spherical surfaces, and for a finite-size input light beam diameter, the electromagnetic energy inside the interferometer can be described by the familiar TEM modes.[5,6,7] For a given input light beam condition and for a given set of mirrors, the energy within the cavity can be characterized by selecting the appropriate amplitudes of the TEM modes. If the input beam spot size is too large or too small, a large portion of the input energy will be found in higher transverse modes of quite large order. If the laser is adjusted to operate in only the TEM_{00q} mode, then it has been shown[8] that the light energy will be principally in the fundamental TEM_{00q} mode in the cavity, if the spot size and the surfaces of constant phases of the input beam are both equal to those for the TEM_{00q} resonant cavity mode. As is usual in matching problems, these conditions are not too critical.

When the measuring interferometer has spherical mirrors at nonconfocal spacing, an incident light ray at a small angle off the system axis will produce repeated reflections, which in general will trace an ellipse on the mirrors.[9] Under special conditions, the points of reflection lie on a circle and are displaced by some angle after every round trip. When this angle is a multiple of 2π, the rays will exactly retrace their paths, and the trace of reflections on a mirror will break up in a number of separate and equally spaced dots. Under these reentrant conditions, the cavity will become resonant not only at a multiple of the fundamental cavity frequency, but also at multiples of a much lower frequency which is not quite a subharmonic of the fundamental cavity frequency. For these conditions, the oscilloscope pattern of the photocell output will show, in addition to the main response, a number of smaller equally spaced pulses. Since the response of one of the "off-axis" modes can coincide (or nearly coincide) with the main response, the measurement of the loss under these conditions can be considerably in error unless these off-axis mode responses are made very small.

In general, the system should be designed to be nondegenerate. That is, the length of the interferometer must be so selected as to avoid any possible overlapping of the cavity resonant response of the different orders of the TEM modes to any of the several optical frequencies present in the laser source. Usually, this is not difficult to accomplish.

III. MEASUREMENT SYSTEM

A block diagram of the components in the interferometer measuring system is given in Fig. 2. The He-Ne laser has external spherical mirrors

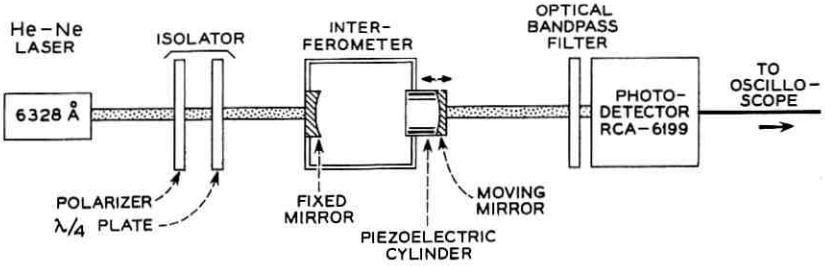


FIG. 2 — Schematic diagram of the interferometer measuring system.

and Brewster angle output windows, and operates at 6328 Å. To reduce the effects of vibrations, the laser and the cavity structures were made very rigid by using a construction similar to that used by Bennett in his magnetostrictively tuned laser.[10] The interferometer mirrors are mounted with suitable tilt controls in 6-inch square steel end blocks, 1 inch thick. The two blocks are tied rigidly together at each of the four corners by 1-inch diameter Invar rods 16 cm long. The interferometer cavity, shown in Fig. 3, uses a piezoelectric transducer to vary the mirror spacing in the cavity. It is a ceramic cylinder $1\frac{1}{4}$ inches ID, $1\frac{1}{2}$ inches OD, and $1\frac{1}{2}$ inches long.[11] The mirror holder is epoxied to one end of the cylinder. The other end is epoxied to a mounting plate which is fastened through suitable tilt controls to the steel end block. The laser, interferometer, and other optical components were fastened rigidly to a heavy steel optical table. This 4 by 8-foot table is supported on six small airplane inner tubes encased in heavy canvas covers. For this type of support, the natural frequency of the table is about four cycles per

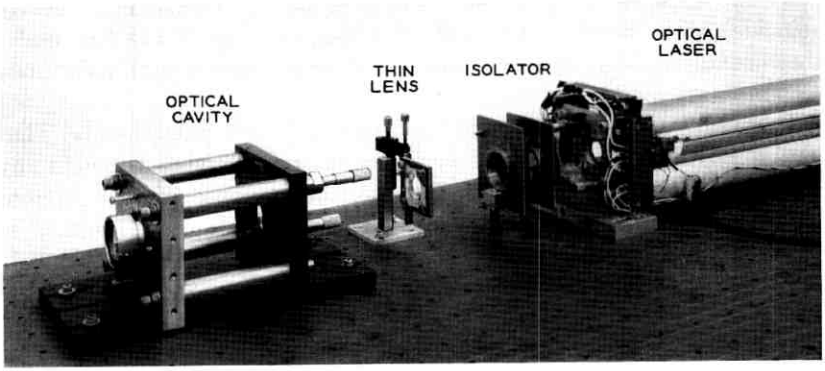


FIG. 3 — Photograph of the interferometer apparatus

second. To avoid air currents, both the laser and the optical cavity are enclosed separately in Plexiglas boxes.

The isolator between the laser and the cavity was found to be essential to prevent interaction between the two optical cavities. It is a circular polarizer consisting of a polaroid analyzer and a quarter-wavelength plate. This circular polarizer will absorb any light reflected back from the interferometer, as the sense of rotation of the circularly polarized light is inverted upon reflection and hence will not be transmitted back through the polaroid analyzer. For this type of isolator, all loss measurement must be made with circularly polarized light.

The photodetector is a standard electron multiplier phototube with an S-20 cathode. A bandpass optical filter, 200 Å wide, centered at 6300 Å, is placed between the source and the detector to eliminate most of the background light.

A 20-cycle triangular wave shape generator delivers 500 volts pp to the piezoelectric mirror drive. For this voltage, the motion of the mirror is about one micron, which is about three half-wavelengths of the 6328 Å laser source. The motion of the mirror is parallel and was checked by using an alignment telescope with a flat mirror placed in the movable mirror holder. The fringe pattern of the alignment telescope reflected back from the moving mirror showed no discernible change when 1000 volts dc or ac was applied to the driver. Therefore any mirror tilt variation must be less than 5 seconds of arc.

IV. EXPERIMENTAL PROCEDURE

When spherical mirrors are used in the interferometer, the ray of incident light must be on a line passing through the centers of curvature of both mirrors. With the aid of a thin optical lens to vary the incident angle, the cavity can readily be aligned to minimize the "off-axis" modes.[2,9]

The visible red gaseous laser, which was one meter long, was so adjusted that it oscillated only in the fundamental transverse mode and at several longitudinal modes.

In spite of all the precautions taken to eliminate building vibrations and air currents, the output pulse pattern on the output scope shows a considerable amount of time position jitter for any one pulse whenever the scope sweep speed is increased to be able to measure the half-power pulse widths. When the sweep speeds are made 10 μsec per cm to view a 30-μsec output pulse, the excursion of the time jitter is about ± 50 μsec, and the jitter frequency is about two cycles or less.

To overcome the effects of the time jitter, a photographic method was

developed to determine the pulse width. For a single period of the 20-cycle drive on the piezoelectric driver, the time jitter is small. The half height of the output pulse was determined by using the following electronic circuits. In the amplifier following the photocell, a fast transistor switch, operating at a one-megacycle rate, reduces the voltage gain of the network periodically by a factor of two (6 db). With this arrangement, a single photograph shows simultaneously both the full-height and the half-height pulses, as shown in Fig. 4. The time position jitter in the pulse pattern is small enough that the fundamental pulse group spacing can be determined directly from the scope. The measurement of the finesse of the cavity does depend upon the accuracy of the various sweep rate calibrations. According to the manufacturer of the oscilloscope, these sweeps, once calibrated, should remain accurate to several per cent for several months.

In order to obtain repeatable loss measurements, it is important to have the laser operate with a stable mode pattern output. At times, this was found to be difficult because adjacent longitudinal modes would

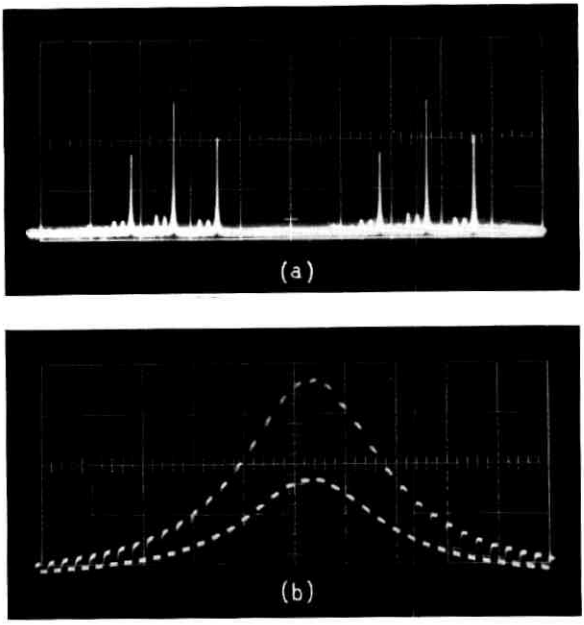


FIG. 4 — (a) Fundamental cavity spacing. Laser source has three output frequencies. Sweep rate is 2 msec per cm. (b) Half width of one of the cavity responses. Sweep rate is 10 μsec per cm. The indicated half width is 34 μsec.

compete with each other, thereby producing an erratic pulse pattern output from the interferometer. This condition was improved by reducing the length of the laser to 70 cm to separate the adjacent modes farther in frequency.

V. EXPERIMENTAL RESULTS

The reflectance was measured for a number of the multiple dielectric coated mirrors whose radius of curvature varied from 2 meters to infinity. The highest value of reflectance was 0.995 and the lowest was 0.940. The average mirror reflectance was about 0.990. The accuracy to which the reflectance of 0.990 can be repeated was about ± 0.001. Since the reflectance for an individual mirror was calculated from the three loss measurements for three mirrors taken two at a time, the error in the individual mirror reflectance was probably twice that of the single measurement, or ± 0.002 maximum. This leaves something to be desired. The largest source of error is in determining the half-power-height pulse width. The photographic method permits a determination of this width to about ± 5 per cent under ideal conditions.

The transmission loss through a fused quartz (Homosil) Brewster angle window with fairly high-quality surfaces was found to be about 0.25 per cent. These and all other transmission loss measurements were taken with two specimens at opposite Brewster angles in the optical cavity, so that the deflection of the light ray passing through the samples canceled out. The transmission loss through the quartz was measured by comparing the loss of a sandwich of three quartz blanks to that of just two windows, where an index of refraction matching liquid was used to overcome the surface irregularities at the interfaces in both cases. This loss was about 0.2 per cent $\pm$ 0.05 per cent for a $\frac{3}{4}$ cm optical path length in the Homosil. This would give a transmission loss of about 1.0 db per meter for high-quality optical Homosil. Using the same technique, the transmission loss through cast Plexiglas II, properly annealed, is about 2 db per meter.

VI. CONCLUSIONS

The interferometer method has proven to be capable of measuring very small optical losses. It requires a number of special precautions, such as a stable mode pattern output from the laser source, a careful and correct alignment of the interferometer, all possible reduction of the effects of building vibrations, and a large degree of optical isolation between the cavity and the laser. This method measures the optical loss at

only one small spot in the cavity, and this is measured using circularly polarized light. It is believed that the system accuracy can be further increased by developing an improved method of measuring the finesse of the system.

VII. ACKNOWLEDGMENTS

All computer tabulations were programmed by J. F. Ossanna, Jr., and Mrs. L. M. Calamia. The authors would like to thank D. R. Herriott and B. P. Bogert for many interesting discussions and for their invaluable help in developing this small optical loss measurement technique.

APPENDIX

The theory for the loss measuring optical cavity using plane mirrors is that of the Fabry-Perot interferometer, given many times before,[12] modified to include loss within the cavity, different reflectivities for the two mirrors, and the first-order effects of the velocity of the moving mirror. The exact theory for the effect of the moving mirror's velocity upon the interferometer response was developed. However, it was found that after neglecting some of the higher-order velocity terms, the same expression for the first-order velocity effects could be more readily obtained by assuming the mirror spacing to be fixed and by linearly varying the input frequency. Only the simpler theory will be given here. These first-order velocity effects for the interferometer, which has a Lorentzian frequency response, will be shown to be appreciably different from those previously calculated for the Gaussian filter.[13]

We assume that the incident light beam is collimated, monochromatic, and perpendicular to the mirrors. It is also assumed that the index of refraction of the space between the two mirrors is equal to that of free space, and that the mirrors have no loss. The list of all symbols and definition of all terms used in the following theory are given below:

$a = 2\pi(\Delta f/\Delta t)$ = angular sweep rate
$\alpha = (-4\pi vt/\lambda)$
B = half-power bandwidth
$\beta_0 \equiv 2m\pi v/c$
c = velocity of light in free space
d_0 = fixed mirror spacing
E_T = total combined electric field of all the output light rays assuming a unit input
g = power loss per single pass through any material within the cavity expressed as a ratio

$g_0 \equiv [g(1 - R_1)(1 - R_2)]^{\frac{1}{2}}$
I_T = total output light intensity assuming unit light input
I_m = maximum output from the interferometer assuming a unit light input
λ = free-space wavelength of light source
$m \approx (2d_0/\lambda)$ = large integer
R_1, R_2 = power reflectance of the two mirrors expressed as a ratio
t = real time
T_p = pulse width in time of the output pulse at half peak power level
T_c = fundamental pulse group spacing in time
$v = (\Delta d/\Delta t)$ = velocity of moving mirror
$x = g(R_1R_2)^{\frac{1}{2}}$.

Let the instantaneous angular frequency input be

$$\omega = \omega_0 - at. \tag{1}$$

Then the instantaneous phase is $\varphi = \omega_0 t - (at^2/2)$.

If the fixed mirror spacing is d_0 and the input light ray is normal to the mirrors as indicated in Fig. 5, then the total time delay for the nth output ray is

$$\tau_n + \tau_0 = 2n\frac{d_0}{c} + \frac{d_0}{c}$$

where the delay for the initial ray is

$$\tau_0 = \frac{d_0}{c}.$$

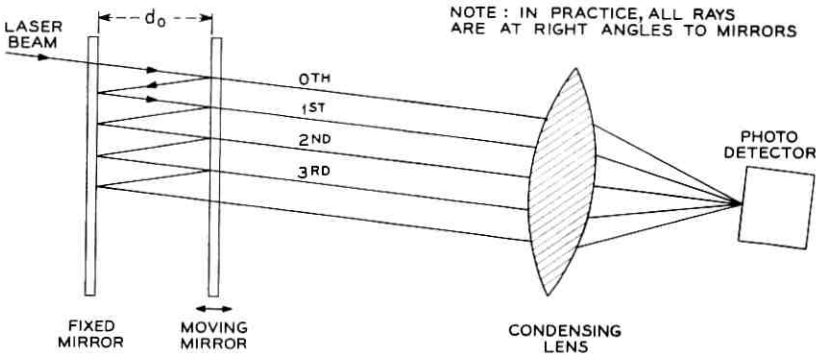


FIG. 5 — Light path through interferometer mirrors to output.

Thus the phase of the nth output ray is

$$\varphi_n = \omega_0\left[t - \frac{2nd_0}{c} - \frac{d_0}{c}\right] - \frac{a}{2}\left[t - \frac{2nd_0}{c} - \frac{d_0}{c}\right]^2.$$

The voltage for the nth output ray is

$$E_n = g_0 x^n \exp\,(-j\varphi_n).$$

The term independent of n can be neglected since its magnitude is unity. Thus the total output voltage is given by

$$E_T = \sum_{n=0}^{\infty} g_0 x^n \exp\left\{-j\,\frac{2nd_0}{c}\left[\omega_0 - a\left(t - \frac{d_0}{c}\right)\right] - j2an^2\,\frac{d_0^2}{c^2}\right\}. \quad (2)$$

The relation between the rate of change of the angular frequency, a, and the velocity of the mirror may be found as follows

$$a = 2\pi\,\frac{\Delta f}{\Delta t} = \frac{2\pi}{\Delta t}\,\frac{2\Delta d}{\lambda}\,\frac{c}{2d_0} = \frac{\omega_0 v}{d_0}. \quad (3)$$

Now, the velocity of the mirror is so small that in any reasonable length of time the variation in the mirror spacing is very small compared to the initial spacing. Hence, the exponent of (2) can be written as

$$\frac{2nd_0}{c}\left[\omega_0 - a\left(t - \frac{d_0}{c}\right)\right] + 2an^2\,\frac{d_0^2}{c^2} = n[2m\pi + \alpha] + n^2\beta_0 \quad (4)$$

where α is small,

$$m \approx 2d_0/\lambda = \text{large integer},$$

and

$$\beta_0 = 2\pi mv/c.$$

Thus from (2) and (4), the total output voltage is given by

$$E_T = \sum_{n=0}^{\infty} g_0[x \exp\,(-j\alpha)]n \exp\,(-j\beta_0 n^2).$$

Now

$$I_T = |\,E_T\,|^2.$$

Thus

$$I_T \approx \left[\sum_{n=0}^{\infty} g_0 x^n \cos\,(\alpha n + \beta_0 n^2)\right]^2 + \left[\sum_{n=0}^{\infty} g_0 x^n \sin\,(\alpha n + \beta_0 n^2)\right]^2. \quad (5)$$

The above expression is a good approximation to that obtained by the

exact theory for $v/c < 10^{-8}$. At the present time, the series in (5) can only be summed numerically, since any other approximation results in a slowly convergent infinite series. The number of terms required by the series may be calculated as follows:

$$\frac{I_T}{g_0^2} \leqq \left[\sum_{n=0}^{\infty} x^n\right]^2 = \left[\sum_{n=0}^{N-1} x^n + \sum_{n=N}^{\infty} x^n\right]^2$$
$$\leqq \frac{(1 - x^N)^2 + 2x^N(1 - x^N) + x^{2N}}{(1 + x)^2}.$$

Thus the ratio of the total error in stopping the series after $N - 1$ terms to the actual value is

$$\epsilon = 2x^N - x^{2N}.$$

The number of terms required is then given by

$$N = \frac{\log (2/\epsilon)}{\log (1/x)} \approx \frac{\log (2/\epsilon)}{1 - x}. \tag{6}$$

Thus if

$$\epsilon = 1/1000 \quad \text{and} \quad x = 0.9975,$$
$$N = 3040.$$

Hence for $x \approx 1$ a very large number of terms must be used in the series.

The series in (5) was computed for a number of values of x, and β_0, some of which are given in Fig. 1. From these values, it was found that the reduction in the maximum of the output response, I_p/I_m, and the increase in the half-height pulse width, T_m/T_p, produced by the mirror velocity can be expressed as

$$(I_m/I_p)^2 = T_m/T_p \approx \{1 + [13.8\ \beta_0^2/(1 - x)^4]\}^{\frac{1}{4}}. \tag{7}$$

The above expression matches the computed values only when the first-order velocity effects are the order of a few per cent. For higher values, the computed results are less than those given by (7).

The relation given in (7) may be compared to that developed for a Gaussian filter which might be used in a spectrum analyzer. It has been shown[13] that the loss in sensitivity, S/S_0, and the increase in apparent bandwidth, B_m/B, produced by sweeping the frequency in the spectrum analyzer is given by

$$\left(\frac{S_0}{S}\right) = \frac{B_m}{B} = \left[1 + 0.195\left(\frac{1}{B^2}\frac{\Delta f}{\Delta t}\right)^2\right]^{\frac{1}{4}} \tag{8}$$

where the above relation was calculated on a power basis.

In the interferometer, the mirror spacing is varied linearly with time. Hence the ratio of any two time intervals may be replaced by the ratio of their appropriate frequency differences. Thus from (3) and (11), it can be readily shown that

$$\frac{1}{B^2}\frac{\Delta f}{\Delta t} = \frac{\pi\beta_0}{(1-x)^2}. \tag{8a}$$

From (8a), (7) becomes

$$\left(\frac{I_m}{I_p}\right)^2 = \frac{T_m}{T_p} = \left[1 + 1.34\left(\frac{1}{B^2}\frac{\Delta f}{\Delta t}\right)^2\right]^{\frac{1}{4}}$$

which is an order of magnitude different from the relation given in (8).

When the mirror velocity is small enough to be neglected, the output from the interferometer is given by the first terms of (5). Thus for $\beta_0 = 0$,

$$I_T = \frac{g_0^2}{1 - 2x\cos\alpha + x^2} = \frac{g_0^2}{(1-x)^2 + 4x\sin^2(\alpha/2)}. \tag{9}$$

Now the maximum value of I_T occurs at $\alpha = 2\pi k$, $k = 0, 1, 2, 3, \cdots$, hence

$$I_m = (I_T)_{\max} = \frac{g_0^2}{(1-x)^2} \equiv \frac{g(1-R_1)(1-R_2)}{[1 - g(R_1R_2)^{\frac{1}{2}}]^2}. \tag{10}$$

From these relations, it can readily be shown that for x nearly unity the finesse of the system is given by[12]

$$F \equiv \frac{T_c}{T_p} = \frac{\pi\sqrt{x}}{1-x}.$$

The solution for x is then

$$x \equiv g(R_1R_2) \approx 1 - \frac{\pi T_p}{T_c} + \frac{1}{2}\left(\frac{\pi T_p}{T_c}\right)^2 + \cdots. \tag{11}$$

This equation relates the optical loss to measurable quantities for the pulse response of the interferometer. The time function for the output pulse can be obtained from (9),

$$I_T \approx \frac{I_m}{1 + (4t^2/T_p^2)} \tag{12}$$

for x close to unity.

REFERENCES

1. Bennett, H. E., and Koehler, W. F., J. Opt. Soc. Am., **50,** 1960, No. 1, pp. 1–6.
2. Herriott, D. R., Appl. Opt., **2,** 1963, No. 8, pp. 865–866.
3. Tolansky, S., and Bradley, D. J., Interferometry, Nat. Phys. Labs. Symposium No. 11, Her Majesty's Stationery Office, London, 1960, p. 375.
4. Herriott, D. R., and Gordon, E. I., unpublished work.
5. Fox, A. G., and Li, T., B.S.T.J., **40,** 1961, pp. 453–488.
6. Boyd, G. D., and Gordon, J. P., B.S.T.J., **40,** 1961, pp. 489–508.
7. Boyd, G. D., and Kogelnik, H., B.S.T.J., **41,** 1962, pp. 1347–1369.
8. Fork, R. L., Herriott, D. R., and Kogelnik, H., to be published.
9. Herriott, D. R., Kogelnik, H., and Kompfner, R., Appl. Opt., **3,** 1964, pp. 523–526.
10. Bennett, W. R., and Kindlmann, P. J., Rev. Sci. Instr., **33,** 1962, pp. 601–605.
11. PTZ-4 ceramic manufactured by the Clevite Corporation.
12. Born, M., and Wolf, E., *Principles of Optics*, Pergamon Press, New York, 1959, pp. 322–327.
13. Chang, S. L., Proc. I.R.E., **42,** 1954, pp. 1278–1282.

On the $\mathcal{L}_2$-Boundedness of Solutions of Nonlinear Functional Equations

By I. W. SANDBERG

(Manuscript received April 8, 1964)

Let $\mathcal{E}_N$ denote the set of N-vector-valued functions of t defined on $[0, \infty)$ such that for any real positive number y, the square of the modulus of each component of any element is integrable on $[0, y]$, and let $\mathcal{L}_{2N}(0, \infty)$ denote the subset of $\mathcal{E}_N$ with the property that the square of the modulus of each component of any element is integrable on $[0, \infty)$.

In the study of nonlinear physical systems, attention is frequently focused on the properties of one of the following two types of functional equations

$$g = f + \mathbf{K}\mathbf{Q}f$$

$$g = \mathbf{K}f + \mathbf{Q}f$$

in which $\mathbf{K}$ and $\mathbf{Q}$ are causal operators, with $\mathbf{K}$ linear and $\mathbf{Q}$ nonlinear, $g \in \mathcal{E}_N$, and f is a solution belonging to $\mathcal{E}_N$. Typically, f represents the system response and g takes into account both the independent energy sources and the initial conditions at $t = 0$.

It is often important to determine conditions under which a physical system governed by one of the above equations is stable in the sense that the response to an arbitrary set of initial conditions approaches zero (i.e., the zero vector) as $t \to \infty$. In a great many cases of this type, g belongs to $\mathcal{L}_{2N}(0, \infty)$ and approaches zero as $t \to \infty$ for all initial conditions, and, in addition, it is possible to show that if $f \in \mathcal{L}_{2N}(0, \infty)$, then $f(t) \to 0$ as $t \to \infty$.

In this paper we attack the stability problem by deriving conditions under which $g \in \mathcal{L}_{2N}(0, \infty)$ and $f \in \mathcal{E}_N$ imply that $f \in \mathcal{L}_{2N}(0, \infty)$. From an engineering viewpoint, the assumption that $f \in \mathcal{E}_N$ is almost invariably a trivial restriction.

As a specific application of the results, we consider a nonlinear integral equation that governs the behavior of a general control system containing linear time-invariant elements and an arbitrary finite number of time-varying nonlinear elements. Conditions are presented under which every solution of this equation belonging to $\mathcal{E}_N$ in fact belongs to $\mathcal{L}_{2N}(0, \infty)$ and approaches zero as $t \to \infty$.

I. NOTATION AND DEFINITIONS

Let M denote an arbitrary matrix. We shall denote by M', M^*, and M^{-1}, respectively, the transpose, the complex-conjugate transpose, and the inverse of M. The positive square-root of the largest eigenvalue of M^*M is denoted by $\Lambda\{M\}$.

The set of complex measurable N-vector-valued functions of the real variable t defined on $[0,\infty)[(-\infty,\infty)]$ is denoted by $\mathfrak{IC}_N(0,\infty)[\mathfrak{IC}_N(-\infty,\infty)]$, and

$$\mathfrak{L}_{2N}(0,\infty) = \left\{f \mid f \varepsilon \mathfrak{IC}_N(0,\infty), \int_0^\infty f^*f\,dt < \infty\right\}.$$

In order to be consistent with standard notation, we let $\mathfrak{L}_2(0,\infty) = \mathfrak{L}_{2N}(0,\infty)$ when $N = 1$. We shall not distinguish between elements of $\mathfrak{IC}_N(0,\infty)[\mathfrak{IC}_N(-\infty,\infty)]$ that agree almost everywhere on $[0,\infty)$ $[(-\infty,\infty)]$. The range of any operator considered in this article is assumed to be contained in either $\mathfrak{IC}_N(0,\infty)$ or $\mathfrak{IC}_N(-\infty,\infty)$.

The inner product of two elements of $\mathfrak{L}_{2N}(0,\infty)$, $f = (f_1, f_2, \cdots, f_N)'$ and $g = (g_1, g_2, \cdots, g_N)'$, is denoted by $\langle f, g\rangle$ and is defined by

$$\langle f, g\rangle = \int_0^\infty f^*g\,dt.$$

The norm of $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$ is denoted by $\|f\|$ and is defined by

$$\|f\| = \langle f,f\rangle^{\frac{1}{2}}.$$

The norm of a linear operator $\mathbf{T}$ defined on $\mathfrak{L}_{2N}(0,\infty)$ is denoted by $\|\mathbf{T}\|$.

Let $y \varepsilon (0,\infty)$, and define f_y by

$$f_y(t) = f(t) \qquad \text{for } t \varepsilon [0,y]$$
$$= 0 \qquad \text{for } t > y$$

for any $f \varepsilon \mathfrak{IC}_N(0,\infty)$, and let

$$\mathcal{E}_N = \{f \mid f \varepsilon \mathfrak{IC}_N(0,\infty), \quad f_y \varepsilon \mathfrak{L}_{2N}(0,\infty) \text{ for } 0 < y < \infty\}.$$

The set of real vector-valued functions is denoted by $\mathfrak{R}$, and $\mathbf{I}$ and 1_N, respectively, denote the identity operator on $\mathfrak{L}_{2N}(0,\infty)$ and the identity matrix of order N.

With A an arbitrary measurable $N \times N$ matrix-valued function of t with elements $\{a_{nm}\}$ defined on $[0,\infty)$, let $\mathfrak{K}_{pN}$ $(p = 1,2)$ denote

$$\left\{A \mid \int_0^\infty |a_{nm}(t)|^p\,dt < \infty \qquad (n,m = 1,2,\cdots,N)\right\}.$$

Let $\psi[f(t),t]$ denote

$$(\psi_1[f_1(t),t],\ \psi_2[f_2(t),t],\ \cdots,\ \psi_N[f_N(t),t])', \qquad f \in \mathfrak{R} \cap \mathfrak{IC}_N(0,\infty)$$

where $\psi_1(w,t)$, $\psi_2(w,t)$, $\cdots$, $\psi_N(w,t)$ are real-valued functions of the real variables w and t for $-\infty < w < \infty$ and $0 \leqq t < \infty$ such that

(*i*) $\psi_n(0,t) = 0$ for $t \in [0,\infty)$ and $n = 1, 2, \cdots, N$

(*ii*) there exist real numbers α and β with the property that

$$\alpha \leqq \frac{\psi_n(w,t)}{w} \leqq \beta \qquad (n = 1, 2, \cdots, N)$$

for $t \in [0,\infty)$ and all real $w \neq 0$.

(*iii*) $\psi_n[w(t),t]$ $(n = 1, 2, \cdots, N)$ is a measurable function of t whenever $w(t)$ is measurable.

The symbol s denotes a scalar complex variable with $\sigma = \mathrm{Re}[s]$ and $\omega = \mathrm{Im}[s]$.

We shall say that a (not necessarily linear) operator $\mathbf{T}$ with domain $\mathfrak{D}(\mathbf{T}) \subset \mathfrak{IC}_N(0,\infty)$ is *causal* if an only if for an arbitrary $\delta > 0$,

$$(\mathbf{T}f)(t) = (\mathbf{T}g)(t) \quad \text{a.e. on } (0,\delta)$$

whenever $f,g \in \mathfrak{D}(\mathbf{T})$ and $f(t) = g(t)$ a.e. on $(0,\delta)$.

II. INTRODUCTION

In the study of nonlinear physical systems, attention is frequently focused on the properties of one of the following two types of functional equations

$$g = f + \mathbf{K}\mathbf{Q}f \tag{1}$$

$$g = \mathbf{K}f + \mathbf{Q}f \tag{2}$$

in which $\mathbf{K}$ and $\mathbf{Q}$ are causal operators, with $\mathbf{K}$ linear and $\mathbf{Q}$ nonlinear, $g \in \mathcal{E}_N$, and f is a solution belonging to $\mathcal{E}_N$. Typically, f represents the system response and g takes into account both the independent energy sources and the initial conditions at $t = 0$.

It is often important to determine conditions under which a physical system governed by one of the above equations is stable in the sense that the response to an arbitrary set of initial conditions approaches zero (i.e., the zero vector) as $t \to \infty$. In a great many cases of this type, g belongs to $\mathcal{L}_{2N}(0,\infty)$ and approaches zero as $t \to \infty$ for all initial conditions, and, in addition, it is possible to show that if $f \in \mathcal{L}_{2N}(0,\infty)$, then $f(t) \to 0$ as $t \to \infty$.

In this paper we attack the stability problem by deriving conditions under which $g \varepsilon \mathfrak{L}_{2N}(0,\infty)$ and $f \varepsilon \mathfrak{E}_N$ imply that $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$. From an engineering viewpoint, the assumption that $f \varepsilon \mathfrak{E}_N$ is almost invariably a trivial restriction.

As a specific application of the abstract results of Section III, we consider, in Section IV, the following integral equation which governs the behavior of a general control system containing linear time-invariant elements and an arbitrary finite number of time-varying nonlinear elements:

$$g(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau),\tau]\,d\tau, \qquad t \geqq 0 \tag{3}$$

in which $k \varepsilon \mathfrak{K}_{1N} \cap \mathfrak{K}_{2N}$, $\psi[\cdot,\cdot]$ is as defined in Section I, and $g \varepsilon \mathfrak{L}_{2N}(0,\infty)$. We prove that every solution f of (3) belonging to $\mathfrak{R} \cap \mathfrak{E}_N$ in fact belongs to $\mathfrak{L}_{2N}(0,\infty)$ and approaches zero as $t \to \infty$ if, with

$$K(s) = \int_0^\infty k(t)e^{-st}\,dt \qquad \text{for} \qquad \sigma \geqq 0,$$

(*i*) $\det [1_N + \tfrac{1}{2}(\alpha+\beta)K(s)] \neq 0$ for $\sigma \geqq 0$

(*ii*) $\tfrac{1}{2}(\beta-\alpha) \sup_\omega \Lambda\{[1_N + \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega)\} < 1.$

An analogous result is proved for the integral equation

$$g(t) = \psi[f(t),t] + \int_0^t k(t-\tau)f(\tau)\,d\tau, \qquad t \geqq 0.$$

For $N = 1$, the key condition (*ii*) possesses a simple geometric interpretation: it is satisfied if and only if the locus of $[K(i\omega)]^{-1}$ for $-\infty < \omega < \infty$ lies outside the circle of radius $\tfrac{1}{2}(\beta-\alpha)$ centered in the complex plane at $[-\tfrac{1}{2}(\alpha+\beta),0]$.†

In Section V we consider two direct applications to nonlinear differential equations. One of our results asserts that if f is any real-valued function of t defined and twice-differentiable on $[0,\infty)$ such that

$$\frac{d^2f}{dt^2} + a\frac{df}{dt} + \psi[f,t] = g$$

for almost all $t \varepsilon [0,\infty)$, where $g \varepsilon \mathfrak{R} \cap \mathfrak{L}_2(0,\infty)$, $\psi[\cdot,\cdot]$ is as defined in Section I with $N = 1$ and $\alpha > 0$, and a is a real constant such that $a > \sqrt{\beta} - \sqrt{\alpha}$, then $f \varepsilon \mathfrak{L}_2(0,\infty)$ and $f(t) \to 0$ as $t \to \infty$.

† For some earlier results concerned with frequency-domain conditions for the stability of nonlinear or time-varying systems, see Refs. 1–4.

III. KEY RESULTS

Assumption 1: It is assumed throughout that

(*i*) $\mathbf{K}$ is a linear causal operator with domain $\mathfrak{D}(\mathbf{K})$ such that $\mathfrak{L}_{2N}(0,\infty) \subset \mathfrak{D}(\mathbf{K}) \subset \mathfrak{C}_N(0,\infty)$

(*ii*) $\mathbf{K}$ maps $\mathfrak{L}_{2N}(0,\infty)$ into itself, and is bounded on $\mathfrak{L}_{2N}(0,\infty)$

(*iii*) $\mathbf{Q}$ is a (not necessarily linear) causal operator with domain $\mathfrak{D}(\mathbf{Q}) \subset \mathfrak{C}_N(0,\infty)$.

The following two theorems are the key results of the paper.

Theorem 1: Let $f \in \mathfrak{D}(\mathbf{Q}) \cap \mathcal{E}_N$ *such that* $\mathbf{Q}f \in \mathfrak{D}(\mathbf{K}) \cap \mathcal{E}_N$, $\mathbf{KQ}f \in \mathcal{E}_N$, *and* $g = f + \mathbf{KQ}f$, *where* $g \in \mathfrak{L}_{2N}(0,\infty)$. *Let* f *not be the zero-element of* $\mathcal{E}_N$, *and let* $y_0 = \inf\{y \mid y > 0, \|f_y\| \neq 0\}$.

Suppose that $\{f_y, 0 < y < \infty\} \subset \mathfrak{D}(\mathbf{Q})$ *and that there exists a real or complex number* x *such that*

(*i*) *on* $\mathfrak{L}_{2N}(0,\infty)$, $(\mathbf{I} + x\mathbf{K})^{-1}$ *exists and is causal*

(*ii*) $$\|(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K}\| \sup_{y>y_0} \frac{\|(\mathbf{Q}f_y)_y - xf_y\|}{\|f_y\|} < 1.$$

Then $f \in \mathfrak{L}_{2N}(0,\infty)$ *and*

$$\|f\| \leqq (1 - r)^{-1} \|(\mathbf{I} + x\mathbf{K})^{-1}g\|,$$

in which

$$r = \|(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K}\| \sup_{y>y_0} \frac{\|(\mathbf{Q}f_y)_y - xf_y\|}{\|f_y\|}.$$

Theorem 2: Let $f \in \mathfrak{D}(\mathbf{Q}) \cap \mathfrak{D}(\mathbf{K}) \cap \mathcal{E}_N$ *such that* $\mathbf{K}f \in \mathcal{E}_N$, $\mathbf{Q}f \in \mathcal{E}_N$, *and*

$$g = \mathbf{K}f + \mathbf{Q}f$$

where $g \in \mathfrak{L}_{2N}(0,\infty)$. *Let* f *not be the zero-element of* $\mathcal{E}_N$, *and let* $y_0 = \inf\{y \mid y > 0, \|f_y\| \neq 0\}$.

Suppose that $\{f_y, 0 < y < \infty\} \subset \mathfrak{D}(\mathbf{Q})$ *and that there exists a real or complex number* x *such that*

(*i*) *on* $\mathfrak{L}_{2N}(0,\infty)$, $(x\mathbf{I} + \mathbf{K})^{-1}$ *exists and is causal*

(*ii*) $$\|(x\mathbf{I} + \mathbf{K})^{-1}\| \sup_{y>y_0} \frac{\|(\mathbf{Q}f_y)_y - xf_y\|}{\|f_y\|} < 1.$$

Then $f \in \mathfrak{L}_{2N}(0,\infty)$ *and*

$$\|f\| \leqq (1 - q)^{-1} \|(x\mathbf{I} + \mathbf{K})^{-1}g\|,$$

in which

$$q = \|(x\mathbf{I} + \mathbf{K})^{-1}\| \sup_{y>y_0} \frac{\|(\mathbf{Q}f_y)_y - xf_y\|}{\|f_y\|}.$$

3.1 *Proof of Theorem 1*

It is convenient to introduce the operator $\mathbf{P}$ defined on $\mathfrak{C}_N(0,\infty)$ by $\mathbf{P}f = f_y$, where y is an arbitrary real positive number.

From $g = f + \mathbf{KQ}f$, we clearly have

$$g_y = f_y + \mathbf{PKQ}f.$$

Since $\mathbf{K}$ is causal,

$$g_y = f_y + \mathbf{PKPQ}f.$$

Similarly, since $\mathbf{Q}$ is causal,

$$\begin{aligned} g_y &= f_y + \mathbf{PKPQP}f \\ &= f_y + \mathbf{PKPQ}f_y . \end{aligned}$$

Thus,

$$g_y = \mathbf{P}(\mathbf{I} + x\mathbf{K})f_y + \mathbf{PKP}(\mathbf{Q} - x\mathbf{I})f_y .$$

Since on $\mathfrak{L}_{2N}(0,\infty)$, $(\mathbf{I} + x\mathbf{K})^{-1}$ exists and is causal,

$$\mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{P}(\mathbf{I} + x\mathbf{K})f_y = f_y ,$$

and hence,

$$f_y = -\mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{PKP}(\mathbf{Q} - x\mathbf{I})f_y + \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}g_y .$$

It follows that

$$\| f_y \| \leqq \| \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{PK} \| \cdot \| \mathbf{PQ}f_y - xf_y \| + \| \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}g_y \|.$$

Moreover, in view of the causality of $(\mathbf{I} + x\mathbf{K})^{-1}$,

$$\mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\,\mathbf{PK} = \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K},$$

and hence, using the fact that $\mathbf{P}$ is a projection on $\mathfrak{L}_{2N}(0,\infty)$,

$$\| \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}\mathbf{PK} \| \leqq \| (\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K} \|.$$

Similarly,

$$\| \mathbf{P}(\mathbf{I} + x\mathbf{K})^{-1}g_y \| \leqq \| (\mathbf{I} + x\mathbf{K})^{-1}g \|.$$

Thus, with r as defined in the statement of the theorem,

$$\| f_y \| \leqq r \| f_y \| + \| (\mathbf{I} + x\mathbf{K})^{-1}g \|$$

or

$$\| f_y \| \leqq (1 - r)^{-1} \| (\mathbf{I} + x\mathbf{K})^{-1}g \|.$$

Since this inequality is valid for arbitrary positive y, it follows that $f \varepsilon \mathcal{L}_{2N}(0, \infty)$ and

$$\| f \| \leqq (1 - r)^{-1} \| (\mathbf{I} + x\mathbf{K})^{-1} g \|.$$

3.2 *Proof of Theorem 2*

The argument is essentially the same as the one used in the proof of Theorem 1.

We have, with $\mathbf{P}$ as defined in the proof of Theorem 1,

$$\begin{aligned} g_y &= \mathbf{PK}f + \mathbf{PQ}f \\ &= \mathbf{PKP}f + \mathbf{PQP}f = \mathbf{PK}f_y + \mathbf{PQ}f_y \\ &= \mathbf{P}(x\mathbf{I} + \mathbf{K})f_y + \mathbf{P}(\mathbf{Q} - x\mathbf{I})f_y . \end{aligned}$$

Using the fact that on $\mathcal{L}_{2N}(0, \infty)$, $(x\mathbf{I} + \mathbf{K})^{-1}$ exists and is causal

$$f_y = -\mathbf{P}(x\mathbf{I} + \mathbf{K})^{-1}\mathbf{P}(\mathbf{Q} - x\mathbf{I})f_y + \mathbf{P}(x\mathbf{I} + \mathbf{K})^{-1} g_y .$$

Thus, with q as defined in the statement of the theorem,

$$\| f_y \| \leqq q \| f_y \| + \| (x\mathbf{I} + \mathbf{K})^{-1} g \|,$$

or

$$\| f_y \| \leqq (1 - q)^{-1} \| (x\mathbf{I} + \mathbf{K})^{-1} g \|.$$

This inequality is valid for arbitrary positive y. Hence $f \varepsilon \mathcal{L}_{2N}(0, \infty)$ and

$$\| f \| \leqq (1 - q)^{-1} \| (x\mathbf{I} + \mathbf{K})^{-1} g \|.$$

Remark: A moment's reflection concerning the proofs of Theorems 1 and 2 will show that by simply reinterpreting the symbols, analogous results can be obtained for other function spaces.

3.3 *Conditions under Which the Hypotheses of Theorems 1 or 2 Concerning x Are Satisfied*

The following theorem asserts that the hypotheses of Theorems 1 or 2 concerning x are satisfied in certain special but very important cases. The implications of the theorem are of direct interest in the theory of passive nonlinear electrical networks.

Theorem 3: Let f be as defined in Theorem 1 [Theorem 2]. Suppose that (i) there exist a nonnegative constant k_1 and a positive constant k_2 such that

$$Re \langle \mathbf{Q}f_y, f_y \rangle \geqq k_1 \| f_y \|^2, \qquad \| \mathbf{Q}f_y \|^2 \leqq k_2 \| f_y \|^2 \qquad \text{for } 0 < y < \infty$$

(*ii*) $\mathbf{K}$ *maps* $\mathfrak{L}_{2N}(0,\infty)$ *into itself such that there exists a nonnegative constant* c *with the property that*

$$Re\,\langle \mathbf{K}h,h\rangle \geqq c \parallel h \parallel^2$$

for all $h \in \mathfrak{L}_{2N}(0,\infty)$.

Then the hypotheses concerning x *of Theorem 1* [*Theorem 2*] *are satisfied if either:*

$$k_1 > 0 \quad \textit{and} \quad c \geqq 0,$$

or

$$k_1 = 0 \quad \textit{and} \quad c > 0.$$

3.4 *Proof of Theorem 3*

Lemmas 1, 2, and 3 (below) imply that for real positive x the operators $(\mathbf{I} + x\mathbf{K})$ and $(x\mathbf{I} + \mathbf{K})$ possess causal inverses on $\mathfrak{L}_{2N}(0,\infty)$ and

$$\parallel (\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K} \parallel^2 \leqq \frac{2(\parallel \mathbf{K} \parallel - c) + x \parallel \mathbf{K} \parallel^2}{x(1 + x \parallel \mathbf{K} \parallel)^2}$$

$$\parallel (x\mathbf{I} + \mathbf{K})^{-1} \parallel^2 \leqq \frac{1}{x(x + 2c)}.$$

With x real and positive,

$$\begin{aligned} \parallel (\mathbf{Q}f_u)_u - xf_u \parallel^2 &\leqq \parallel \mathbf{Q}f_u - xf_u \parallel^2 \\ &\leqq \parallel \mathbf{Q}f_u \parallel^2 - 2x\,\mathrm{Re}\,\langle \mathbf{Q}f_u, f_u\rangle + x^2 \parallel f_u \parallel^2 \\ &\leqq (k_2 - 2xk_1 + x^2) \parallel f_u \parallel^2. \end{aligned}$$

It is a simple matter to verify that if $(k_1 + c) > 0$, there exist positive values of x such that

$$\frac{2(\parallel \mathbf{K} \parallel - c) + x \parallel \mathbf{K} \parallel^2}{x(1 + x \parallel \mathbf{K} \parallel)^2} (k_2 - 2xk_1 + x^2) < 1,$$

and there exist positive values of x such that

$$\frac{k_2 - 2xk_1 + x^2}{x(x + 2c)} < 1.$$

Hence, it remains to prove Lemmas 1, 2, and 3.

Lemma 1: Let $\mathbf{T}$ *be a bounded linear mapping of* $\mathfrak{L}_{2N}(0,\infty)$ *into itself*

such that there exists a constant $c_1 > -1$ *with the property that*

$$Re\ \langle \mathbf{T}f,f\rangle \geqq c_1 \| f \|^2$$

for all $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$. *Then* $(\mathbf{I}+\mathbf{T})$ *possesses an inverse on* $\mathfrak{L}_{2N}(0,\infty)$.
Proof:

Since $c_1 > -1$, it is evident that there exists a positive constant k_1 such that

$$\operatorname{Re}\langle(\mathbf{I}+\mathbf{T})f,f\rangle \geqq k_1 \| f \|^2$$

for all $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$. This, together with the boundedness of $\mathbf{T}$, implies that $(\mathbf{I}+\mathbf{T})^{-1}$ exists (see Ref. 5, for example).

Lemma 2: Let $\mathbf{T}$ *be an invertible bounded linear mapping of* $\mathfrak{L}_{2N}(0,\infty)$ *into itself such that* $\mathbf{T}$ *is causal and* $Re\ \langle \mathbf{T}h,h\rangle \geqq 0$ *for all* $h \varepsilon \mathfrak{L}_{2N}(0,\infty)$. *Then* $\mathbf{T}^{-1}$ *is causal.*
Proof:

A bounded linear mapping $\mathbf{A}$ of $\mathfrak{L}_{2N}(0,\infty)$ into itself is causal if and only if[6]

$$\operatorname{Re}\int_0^y (\mathbf{A}f)^* f\,dt \geqq -\|\mathbf{A}\| \int_0^y f^* f\,dt$$

for all real $y \geqq 0$ and all $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$. Thus, to prove the lemma it suffices to point out that the causality of $\mathbf{T}$ implies that

$$\operatorname{Re}\int_0^y (\mathbf{T}g)^* g\,dt = \operatorname{Re}\langle \mathbf{T}g_y, g_y\rangle \geqq 0,$$

for all real $y \geqq 0$ and all $g \varepsilon \mathfrak{L}_{2N}(0,\infty)$, and hence that

$$\operatorname{Re}\int_0^y h^* \mathbf{T}^{-1} h\,dt \geqq 0$$

for all real $y \geqq 0$ and all $h \varepsilon \mathfrak{L}_{2N}(0,\infty)$.

Lemma 3:† *Let* $\mathbf{T}$ *be a bounded linear mapping of* $\mathfrak{L}_{2N}(0,\infty)$ *into itself such that* $(\mathbf{I}+\mathbf{T})$ *is invertible and there exists a real constant* c_2 *with the property that*

$$Re\ \langle \mathbf{T}f,f\rangle \geqq c_2 \| f \|^2$$

for all $f \varepsilon \mathfrak{L}_{2N}(0,\infty)$. *Then, for* $c_2 \geqq -\frac{1}{2}$,

† Lemmas 1 and 3, and their proofs, remain valid if $\mathfrak{L}_{2N}(0,\infty)$ is replaced with an arbitrary Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

$$\| (\mathbf{I} + \mathbf{T})^{-1}\mathbf{T} \| \leqq [1 - (2c_2 + 1)(1 + \| \mathbf{T} \|)^{-2}]^{\frac{1}{2}}$$

and, for $c_2 > -\frac{1}{2}$,

$$\| (\mathbf{I} + \mathbf{T})^{-1} \| \leqq (1 + 2c_2)^{-\frac{1}{2}}.$$

Proof:

In order to establish the first inequality, let $g = (\mathbf{I} + \mathbf{T})^{-1}\mathbf{T}f$ and, using the fact that $g = f - (\mathbf{I} + \mathbf{T})^{-1}f$, observe that

$$\langle g,g \rangle = \langle f,f \rangle - 2 \operatorname{Re} \langle \mathbf{T}z,z \rangle - \langle z,z \rangle,$$

where $z = (\mathbf{I} + \mathbf{T})^{-1}f$. Since

$$2 \operatorname{Re} \langle \mathbf{T}z,z \rangle + \langle z,z \rangle \geqq (2c_2 + 1) \| z \|^2$$

and

$$\| z \| \geqq \| (\mathbf{I} + \mathbf{T}) \|^{-1} \| f \| \geqq (1 + \| T \|)^{-1} \| f \|,$$

it follows that

$$\langle g,g \rangle \leqq [1 - (2c_2 + 1)(1 + \| \mathbf{T} \|)^{-2}] \langle f,f \rangle$$

for all $f, g \in \mathfrak{L}_{2N}(0, \infty)$ such that $g = (\mathbf{I} + \mathbf{T})^{-1}\mathbf{T}f$. Thus

$$\| (\mathbf{I} + \mathbf{T})^{-1}\mathbf{T} \| \leqq [1 - (2c_2 + 1)(1 + \| \mathbf{T} \|)^{-2}]^{\frac{1}{2}}.$$

The second inequality follows directly from the fact that if $g = (\mathbf{I} + \mathbf{T})^{-1}f$,

$$\| f \|^2 = \| g \|^2 + 2 \operatorname{Re} \langle \mathbf{T}g,g \rangle + \| \mathbf{T}g \|^2 \geqq (1 + 2c_2) \| g \|^2.$$

IV. APPLICATIONS TO NONLINEAR INTEGRAL EQUATIONS

In this section our primary objective is to prove the following two theorems.

Theorem 4: Let $k \in \mathfrak{K}_{1N}$ *and let*

$$g(t) = f(t) + \int_0^t k(t - \tau)\psi[f(\tau), \tau] \, d\tau, \qquad t \geqq 0$$

where $g \in \mathfrak{L}_{2N}(0, \infty)$ *and* $f \in \mathfrak{R} \cap \mathfrak{E}_N$. *Let*

$$K(s) = \int_0^\infty k(t)e^{-st} \, dt, \qquad \sigma \geqq 0.$$

Suppose that

(*i*) $\det [1_N + \frac{1}{2}(\alpha + \beta)K(s)] \neq 0$ *for* $\sigma \geqq 0$

(*ii*) $\frac{1}{2}(\beta - \alpha) \sup_\omega \Lambda\{[1_N + \frac{1}{2}(\alpha + \beta)K(i\omega)]^{-1}K(i\omega)\} < 1.$

Then $f \in \mathfrak{L}_{2N}(0,\infty)$.

Theorem 5: Let $k \in \mathfrak{K}_{1N}$ *and let*

$$g(t) = \int_0^t k(t-\tau)f(\tau)\,d\tau + \psi[f(t),t], \qquad t \geqq 0$$

where $g \in \mathfrak{L}_{2N}(0,\infty)$ *and* $f \in \mathfrak{R} \cap \mathcal{E}_N$. *Let*

$$K(s) = \int_0^\infty k(t)e^{-st}\,dt, \qquad \sigma \geqq 0.$$

Suppose that

(*i*) $\det[\frac{1}{2}(\alpha+\beta)1_N + K(s)] \neq 0$ *for* $\sigma \geqq 0$

(*ii*) $\frac{1}{2}(\beta-\alpha)\sup_\omega \Lambda\{[\frac{1}{2}(\alpha+\beta)1_N + K(i\omega)]^{-1}\} < 1.$

Then $f \in \mathfrak{L}_{2N}(0,\infty)$.

4.1 *Proof of Theorems 4 and 5*

In Theorems 1 and 2 let $\mathbf{Q}$ denote the operator defined by

$$(\mathbf{Q}g)(t) = \psi[g(t),t], \qquad 0 \leqq t < \infty$$

where g is an arbitrary element of $\mathfrak{R} \cap \mathfrak{C}_N(0,\infty)$. This operator maps $\mathfrak{R} \cap \mathfrak{L}_{2N}(0,\infty)$ into itself and possesses the property that for any real x

$$\| \mathbf{Q}h - xh \| \leqq \eta(x) \| h \|, \qquad h \in \mathfrak{R} \cap \mathfrak{L}_{2N}(0,\infty)$$

where

$$\eta(x) = \max[(x-\alpha), (\beta - x)].$$

Thus, with $\mathbf{K}$ defined on $\mathfrak{L}_{2N}(0,\infty)$ by[7]

$$\mathbf{K}h = \int_0^t k(t-\tau)h(\tau)\,d\tau, \qquad h \in \mathfrak{L}_{2N}(0,\infty),$$

condition (*ii*) of Theorem 1 and the corresponding condition of Theorem 2, respectively, are satisfied if there exists a real x such that

$$\| (\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K} \| \eta(x) < 1,$$

and

$$\| (x\mathbf{I} + \mathbf{K})^{-1} \| \eta(x) < 1.$$

Lemmas 4 and 5 (below) imply at once that if the assumptions of Theorem 4 (Theorem 5) are met, then hypotheses (*i*) and (*ii*) of Theorem 1 (Theorem 2) are satisfied with $x = \frac{1}{2}(\alpha+\beta)$. It can be shown[8] (with the

aid of Lemma 4) that this choice of x is optimal in the sense that if there exists a real x such that $(\mathbf{I} + x\mathbf{K})$ possesses a causal inverse on $\mathcal{L}_{2N}(0, \infty)$ and

$$\| (\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K}\| \, \eta(x) < 1,$$

then $[\mathbf{I} + \frac{1}{2}(\alpha + \beta)\mathbf{K}]$ possesses a causal inverse on $\mathcal{L}_{2N}(0, \infty)$ and

$$\| (\mathbf{I} + x\mathbf{K})^{-1}\mathbf{K} \| \, \eta(x) \geqq \| [\mathbf{I} + \tfrac{1}{2}(\alpha + \beta)\mathbf{K}]^{-1}\mathbf{K} \| \, \eta[\tfrac{1}{2}(\alpha + \beta)].$$

This choice of x is similarly optimal with regard to the statement of the conditions in Theorem 5.

Before proceeding to the statement and proofs of the lemmas, it is convenient to introduce a few definitions.

4.2 *Definitions*

With τ an arbitrary positive constant, let $\mathbf{S}_\tau$ denote the mapping of $\mathcal{L}_{2N}(0, \infty)$ into itself defined by

$$\begin{aligned} (\mathbf{S}_\tau f)(t) &= 0, && t \in [0,\tau) \\ &= f(t - \tau), && t \in [\tau, \infty) \end{aligned}$$

for any $f \in \mathcal{L}_{2N}(0, \infty)$.

Let

$$\mathcal{L}_{2N}(-\infty, \infty) = \left\{ f \mid f \in \mathcal{H}_N(-\infty, \infty), \int_{-\infty}^{\infty} f^* f \, dt < \infty \right\}.$$

We take as the definition of the Fourier transform of $f \in \mathcal{L}_{2N}(-\infty, \infty)$:

$$\hat{f} = \text{l.i.m.} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt,$$

and consequently,

$$2\pi \int_{-\infty}^{\infty} f^* f \, dt = \int_{-\infty}^{\infty} \hat{f}^* \hat{f} \, d\omega.$$

By the Fourier transform of an $f \in \mathcal{L}_{2N}(0, \infty)$ we mean simply

$$\text{l.i.m.} \int_{0}^{\infty} f(t) e^{-i\omega t} \, dt.$$

4.3 *Lemmas 4 and 5*

Lemma 4: Let $\mathbf{A}$ *be an invertible linear mapping of* $\mathcal{L}_{2N}(0, \infty)$ *into itself*

such that for an arbitrary $f \in \mathfrak{L}_{2N}(0, \infty)$

$$\mathbf{A}\mathbf{S}_\tau f = \mathbf{S}_\tau \mathbf{A} f, \qquad \tau > 0.$$

Then $\mathbf{A}^{-1}$ *is causal.*

Proof:

Suppose that, on the contrary, $\mathbf{A}$ possesses an inverse on $\mathfrak{L}_{2N}(0, \infty)$, but that it is not causal. Then there exist elements $z_1, z_2 \in \mathfrak{L}_{2N}(0, \infty)$, and a $\delta > 0$ such that $z_1(t) \neq 0$ on some positive-measure subset of $(0,\delta)$, and $\mathbf{A}z_1 = \mathbf{S}_\delta z_2$. Since $\mathbf{A}$ is assumed to possess an inverse, there exists a unique $z_3 \in \mathfrak{L}_{2N}(0, \infty)$ such that $z_2 = \mathbf{A}z_3$. Thus,

$$\mathbf{A}\mathbf{S}_\delta z_3 = \mathbf{S}_\delta \mathbf{A} z_3 = \mathbf{S}_\delta z_2 .$$

Clearly, $\mathbf{S}_\delta z_3 \neq z_1$, which contradicts the assumption that $\mathbf{A}$ possesses an inverse. This proves the lemma.

Lemma 5: Let $u \in \mathfrak{K}_{1N}$ *and let* $\mathbf{U}$ *be the mapping of* $\mathfrak{L}_{2N}(0, \infty)$ *into itself defined by*

$$\mathbf{U}f = \int_0^t u(t - \tau) f(\tau)\, d\tau, \qquad f \in \mathfrak{L}_{2N}(0, \infty).$$

Let

$$U(s) = \int_0^\infty u(t) e^{-st}\, dt, \qquad \sigma \geqq 0.$$

Suppose that $\det [1_N + U(s)] \not\equiv 0$ *for* $\sigma \geqq 0$. *Then*

(*i*) $(\mathbf{I} + \mathbf{U})$ *possesses an inverse on* $\mathfrak{L}_{2N}(0, \infty)$

(*ii*) $\| (\mathbf{I} + \mathbf{U})^{-1}\mathbf{U} \| \leqq \sup_\omega \Lambda\{[1_N + U(i\omega)]^{-1} U(i\omega)\}$

$\| (\mathbf{I} + \mathbf{U})^{-1} \| \leqq \sup_\omega \Lambda\{[1_N + U(i\omega)]^{-1}\}.$

Proof:

Consider first the invertibility of the operator $(\bar{\mathbf{I}} + \bar{\mathbf{U}})$ defined on $\mathfrak{L}_{2N}(-\infty, \infty)$ by

$$(\bar{\mathbf{I}} + \bar{\mathbf{U}})f = f + \int_{-\infty}^t u(t - \tau) f(\tau)\, d\tau, \qquad f \in \mathfrak{L}_{2N}(-\infty, \infty).$$

The assumption that $u \in \mathfrak{K}_{1N}$ implies that the elements of $U(i\omega)$ approach zero as $|\omega| \to \infty$, and that they are uniformly bounded and uniformly continuous for $\omega \in (-\infty, \infty)$. Thus, $\det [1_N + U(i\omega)]$ approaches unity as $|\omega| \to \infty$, and is uniformly continuous for

$\omega \varepsilon (-\infty, \infty)$. It follows that $\det [1_N + U(i\omega)] \neq 0$ for all ω implies that

$$\inf_{\omega} | \det [1_N + U(i\omega)] | > 0,$$

and hence that

$$\sup_{\omega} \Lambda\{[1_N + U(i\omega)]^{-1}\} < \infty.$$

Let $\hat{g}$ denote the Fourier transform of an arbitrary $g \varepsilon \mathcal{L}_{2N}(-\infty, \infty)$. Then,

$$\int_{-\infty}^{\infty} \hat{g}^*[1_N + U(i\omega)]^{-1^*}[1_N + U(i\omega)]^{-1}\hat{g}\, d\omega$$

$$\leqq \int_{-\infty}^{\infty} \Lambda^2\{[1_N + U(i\omega)]^{-1}\}\hat{g}^*\hat{g}\, d\omega$$

$$\leqq \sup_{\omega} \Lambda^2\{[1_N + U(i\omega)]^{-1}\} \int_{-\infty}^{\infty} \hat{g}^*\hat{g}\, d\omega < \infty,$$

and hence, by the Riesz-Fischer theorem, there exists an $f \varepsilon \mathcal{L}_{2N}(-\infty, \infty)$ with Fourier transform

$$\hat{f} = [1_N + U(i\omega)]^{-1}\hat{g}.$$

This establishes the existence of $(\bar{\mathbf{I}} + \bar{\mathbf{U}})^{-1}$.

Since $\det [1_N + U(s)] \neq 0$ for $\sigma \geqq 0$, and $U(s) \to 0$ as $| s | \to \infty$ uniformly in the closed right-half plane, every element of $[1_N + U(s)]^{-1}$ is analytic and uniformly bounded for $\sigma > 0$. Thus, $(\bar{\mathbf{I}} + \bar{\mathbf{U}})^{-1}$ maps

$$\{f \mid f \varepsilon \mathcal{L}_{2N}(-\infty, \infty), f(t) = 0 \quad \text{for} \quad t < 0\}$$

into itself,[9,10] and hence the operator $(\mathbf{I} + \mathbf{U})$ defined on $\mathcal{L}_{2N}(0, \infty)$ possesses an inverse.

To establish the first of the inequalities stated in the lemma, let $f \varepsilon \mathcal{L}_{2N}(0, \infty)$ and let

$$g = (\mathbf{I} + \mathbf{U})^{-1}\mathbf{U}f.$$

Then, with $\hat{g}$ and $\hat{f}$, respectively, the Fourier transforms of g and f,

$$\hat{g} = [1_N + U(i\omega)]^{-1}U(i\omega)\hat{f}.$$

Thus,

$$\int_{-\infty}^{\infty} \hat{g}^*\hat{g}\, d\omega = \int_{-\infty}^{\infty} \hat{f}^*U(i\omega)^*[1_N + U(i\omega)]^{-1^*}[1_N + U(i\omega)]^{-1}U(i\omega)\hat{f}\, d\omega$$

$$\leqq \int_{-\infty}^{\infty} \Lambda^2\{[1_N + U(i\omega)]^{-1}U(i\omega)\}\hat{f}^*\hat{f}\, d\omega$$

$$\leqq \sup_{\omega} \Lambda^2\{[1_N + U(i\omega)]^{-1}U(i\omega)\} \int_{-\infty}^{\infty} \hat{f}^*\hat{f}\, d\omega,$$

from which, using Plancherel's identity,

$$\| g \| \leqq \sup_{\omega} \Lambda\{[1_N + U(i\omega)]^{-1}U(i\omega)\} \| f \|.$$

Thus,

$$\| (\mathbf{I} + \mathbf{U})^{-1}\mathbf{U} \| \leqq \sup_{\omega} \Lambda\{[1_N + U(i\omega)]^{-1}U(i\omega)\}.$$

By simply repeating this argument with $(\mathbf{I} + \mathbf{U})^{-1}\mathbf{U}$ and $[1_N + U(i\omega)]^{-1} \cdot U(i\omega)$, respectively, replaced with $(\mathbf{I} + \mathbf{U})^{-1}$ and $[1_N + U(i\omega)]^{-1}$, we find that

$$\| (\mathbf{I} + \mathbf{U})^{-1} \| \leqq \sup_{\omega} \Lambda\{[1_N + U(i\omega)]^{-1}\}.$$

This proves the lemma.

4.4 *Remarks*

It can easily be shown that conditions (i) and (ii) of Theorems 4 and 5 are satisfied if $\alpha > 0$ and

$$K(i\omega) + K(i\omega)^*$$

is nonnegative definite for all ω.

A moment's reflection concerning the proof of Theorems 4 and 5 will show that those theorems remain valid if $\psi[f(t),t]$ and f, respectively, are replaced with $(\mathbf{Q}f)(t)$ and $f \in \mathfrak{D}(\mathbf{Q})$, with the understanding that

(a) $\mathbf{Q}$ is a causal mapping of a subset $\mathfrak{D}(\mathbf{Q})$ of $\mathcal{E}_N$ into $\mathcal{E}_N$ such that $\mathbf{Q}h \in \mathcal{L}_{2N}(0,\infty)$ whenever $h \in \mathfrak{D}(\mathbf{Q}) \cap \mathcal{L}_{2N}(0,\infty)$, and there exist real constants α and β $(\beta > \alpha)$ with the property that

$$\| \mathbf{Q}h - \tfrac{1}{2}(\alpha + \beta)h \| \leqq \tfrac{1}{2}(\beta - \alpha) \| h \|$$

for all $h \in \mathfrak{D}(\mathbf{Q}) \cap \mathcal{L}_{2N}(0,\infty)$.

(b) if $h \in \mathfrak{D}(\mathbf{Q})$, $\{h_y, 0 < y < \infty\} \subset \mathfrak{D}(\mathbf{Q})$.

4.5 *Conditions under Which $f(t) \to 0$ as $t \to \infty$*

Theorem 6: Suppose that the hypotheses of Theorem 4 are satisfied, that

$g(t) \to 0$ as $t \to \infty$, and that $k \in \mathcal{K}_{2N}$. Then $f(t) \to 0$ as $t \to \infty$.

Proof:

Observe first that the N-vector-valued function with values

$$\psi[f(t),t], \qquad 0 \leqq t < \infty,$$

is an element of $\mathcal{R} \cap \mathcal{L}_{2N}(0,\infty)$. Thus it suffices to show that if $h \in \mathcal{L}_{2N}(0,\infty)$,

$$\int_0^t k(t-\tau)h(\tau)\,d\tau \to 0 \quad \text{as} \quad t \to \infty.$$

In terms of $K(i\omega)$ and $\hat{h}(i\omega)$, respectively, the Fourier transforms of k and h,

$$\int_0^t k(t-\tau)h(\tau)\,d\tau = \frac{1}{2\pi}\int_{-\infty}^{\infty} K(i\omega)\hat{h}(i\omega)e^{i\omega t}\,d\omega.$$

Since $k \in \mathcal{K}_{2N}$, it follows that the modulus of each element of the N-vector $K(i\omega)\hat{h}(i\omega)$ is integrable on the ω-set $(-\infty,\infty)$. Thus, by the Riemann-Lebesgue lemma

$$\int_0^t k(t-\tau)h(\tau)\,d\tau \to 0 \quad \text{as} \quad t \to \infty.$$

This proves Theorem 6.

It is obvious that essentially the same argument suffices to prove the following corresponding result relating to Theorem 5.

Theorem 7: Suppose that the hypotheses of Theorem 5 are satisfied, that $g(t) \to 0$ as $t \to \infty$, and that $k \in \mathcal{K}_{2N}$. Then $\psi[f(t),t] \to 0$ as $t \to \infty$.

V. APPLICATIONS TO NONLINEAR DIFFERENTIAL EQUATIONS

Theorem 8: Let A be an $N \times N$ matrix of real constants, let $\psi[\cdot,\cdot]$ be as defined in Section I with α and β, respectively, replaced with $\hat{\alpha}$ and $\hat{\beta}$, and let f denote a real N-vector-valued function of t defined and differentiable on $[0,\infty)$ such that

$$\frac{df}{dt} + Af + \psi[f,t] = g$$

for almost all $t \in [0,\infty)$, where $g \in \mathcal{R} \cap \mathcal{L}_{2N}(0,\infty)$.

Suppose that

(i) *det* $[s1_N + \frac{1}{2}(\hat{\alpha}+\hat{\beta})1_N + A] \neq 0$ for $\sigma \geqq 0$

(ii) $\frac{1}{2}(\hat{\beta}-\hat{\alpha})\sup_{\omega} \Lambda\{[(i\omega)1_N + \frac{1}{2}(\hat{\alpha}+\hat{\beta})1_N + A]^{-1}\} < 1.$

Then $f \in \mathfrak{L}_{2N}(0,\infty)$ *and* $f(t) \to 0$ *as* $t \to \infty$.

Proof:

Clearly, $f \in \mathfrak{R} \cap \mathfrak{E}_N$. Using the well-known expression for the solution of an inhomogeneous system of linear first-order differential equations in terms of the solution of the corresponding matrix homgeneous differential equation, and regarding

$$g + \tfrac{1}{2}(\hat{\alpha} + \hat{\beta})f - \psi[f,t]$$

as the "forcing function," we find that f satisfies

$$e^{-Bt}c + \int_0^t e^{-B(t-\tau)}g(\tau)d\tau$$

$$= f(t) + \int_0^t e^{-B(t-\tau)}\left\{\psi[f(\tau),\tau] - \frac{1}{2}(\hat{\alpha} + \hat{\beta})f(\tau)\right\}d\tau$$

for $t \in [0,\infty)$, in which $B = \tfrac{1}{2}(\hat{\alpha} + \hat{\beta})1_N + A$, and c is a real constant N-vector.

In view of (i), the matrix e^{-Bt} is an element of $\mathfrak{K}_{1N} \cap \mathfrak{K}_{2N}$. By the argument used in the proof of Theorem 6,

$$\int_0^t e^{-B(t-\tau)}g(\tau)\,d\tau \to 0 \quad \text{as} \quad t \to \infty,$$

a property which is obviously shared by $e^{-Bt}c$. Thus, using the fact that

$$-\frac{1}{2}(\hat{\beta} - \hat{\alpha}) \leqq \frac{\psi_n(w,t) - \tfrac{1}{2}(\hat{\alpha} + \hat{\beta})w}{w} \leqq \frac{1}{2}(\hat{\beta} - \hat{\alpha})$$

$$(n = 1, 2, \cdots, N)$$

for all $t \in [0,\infty)$ and all real $w \neq 0$, the theorem follows from a direct application of Theorems 4 and 6 [$-\tfrac{1}{2}(\hat{\beta} - \hat{\alpha})$ and $\tfrac{1}{2}(\hat{\beta} - \hat{\alpha})$, respectively, play the roles of α and β in Theorem 4].

5.1 *Generalization of an Earlier Theorem Concerning a Linear Differential Equation with Periodic Coefficients*[11]

Theorem 9: Let $\psi[\cdot,\cdot]$ *be as defined in Section I with* $N = 1$ *and* α *and* β, *respectively, replaced with* $\hat{\alpha}$ *and* $\hat{\beta}$. *Let* f *denote a real-valued function of* t *defined and twice-differentiable on* $[0,\infty)$ *such that*

$$\frac{d^2f}{dt^2} + a\frac{df}{dt} + \psi[f,t] = g$$

for almost all $t \in [0,\infty)$, *where* $g \in \mathfrak{R} \cap \mathfrak{L}_2(0,\infty)$ *and* a *is a real constant. Then if* $\hat{\alpha} > 0$ *and* $a > \sqrt{\hat{\beta}} - \sqrt{\hat{\alpha}}$, $f \in \mathfrak{L}_2(0,\infty)$ *and* $f(t) \to 0$ *as* $t \to \infty$.

Proof:

Proceeding as in the proof of Theorem 8, we find that f satisfies

$$h(t) + \int_0^t k(t-\tau)g(\tau)\,d\tau = f(t) + \int_0^t k(t-\tau)\{\psi[f(\tau),\tau] - \frac{1}{2}(\hat{\alpha}+\hat{\beta})f(\tau)\}\,d\tau$$

for $t \varepsilon [0,\infty)$, in which h is a solution of

$$\frac{d^2h}{dt^2} + a\frac{dh}{dt} + \frac{1}{2}(\hat{\alpha}+\hat{\beta})h = 0,$$

and $k \varepsilon \mathcal{K}_{11} \cap \mathcal{K}_{21}$ with

$$K(s) = \int_0^\infty k(t)e^{-st}\,dt = \left[s^2 + as + \frac{1}{2}(\hat{\alpha}+\hat{\beta})\right]^{-1}, \qquad \sigma \geqq 0.$$

With $\alpha = -\frac{1}{2}(\hat{\beta}-\hat{\alpha})$ and $\beta = \frac{1}{2}(\hat{\beta}-\hat{\alpha})$, condition (i) of Theorem 4 is obviously satisfied, while condition (ii) reduces to

$$\tfrac{1}{2}(\hat{\beta}-\hat{\alpha}) < \inf_{\omega} |\tfrac{1}{2}(\hat{\alpha}+\hat{\beta}) - \omega^2 + ia\omega|.$$

It is a simple matter to show that this inequality is satisfied if $\hat{\alpha} > 0$ and $a > \sqrt{\hat{\beta}} - \sqrt{\hat{\alpha}}$. Hence $f \varepsilon \mathcal{L}_2(0,\infty)$.

Since $h(t) \to 0$ as $t \to \infty$ and, by the argument used to prove Theorem 6,

$$\int_0^t k(t-\tau)g(\tau)\,d\tau \to 0 \quad \text{as} \quad t \to \infty,$$

Theorem 6 implies that $f(t) \to 0$ as $t \to \infty$. This completes the proof of Theorem 9.

VI. FINAL REMARK

Some of the results and techniques of this paper are useful in establishing sufficient conditions for the existence and uniqueness of solutions of functional equations of the type that we have considered. The reader familiar with the contraction-mapping fixed-point theorem has probably recognized this fact.

REFERENCES

1. Popov, N. M., Absolute Stability of Nonlinear Systems of Automatic Control, Avtomatika i Telemekhanika, **22**, Aug., 1961, pp. 961–978.

2. Kalman, R. E., Lyapunov Functions for the Problem of Lur'e in Automatic Control, Proc. Natl. Acad. Sci., **49,** Feb., 1963, pp. 201–205.
3. Bongiorno, Jr., J. J., An Extension of the Nyquist-Barkhausen Stability Criterion to Linear Lumped Parameter Systems, IEEE-PTGAC, **AC-8,** No. 2, April, 1963, pp. 166–170.
4. Youla, D. C., Some Results in the Theory of Active Networks, Polytechnic Institute of Brooklyn Research Report No. 1063-62, Aug., 1962.
5. Sandberg, I. W., On the Properties of Some Systems That Distort Signals—I, B.S.T.J., **42,** September, 1963, p. 2033.
6. Sandberg, I. W., Conditions for the Causality of Nonlinear Operators Defined on a Function Space, to be published.
7. Bochner, S., and Chandrasekharan, *Fourier Transforms,* Princeton University Press, Princeton, N. J., 1949, p. 99.
8. Sandberg, I. W., A Note on the Application of the Contraction-Mapping Fixed-Point Theorem to a Class of Nonlinear Functional Equations, to be published.
9. Paley, R. E., and Wiener, N., *Fourier Transforms in the Complex Domain,* published by the American Mathematical Society, Providence, Rhode Island, p. 8.
10. Titchmarsh, E. C., *Introduction to the Theory of Fourier Integrals,* Clarendon Press, Oxford, 2nd ed., 1948, pp. 125 and 128.
11. Sandberg, I. W., On the Stability of Solutions of Linear Differential Equations with Periodic Coefficients, to be published in the SIAM Journal.

ERRATA

On the Theory of Linear Multi-Loop Feedback Systems, I. W. Sandberg, B.S.T.J., **42,** March, 1963, pp. 355–382.

On page 361, the expression (y_1+y_2), which appears twice, should be replaced in both positions with $(y_1 + \bar{y}_2)$, in which $\bar{y}_2$ denotes the value of y_2 when $y_1 = 0$.

On page 377, the left side of the first equation of Section 9.4 should be det $\mathbf{F}_{\bar{y}_1}$, not det $\mathbf{F}_1$.

ERRATA

On the Theory of Linear Multi-Loop Feedback Systems, I. W. Sandberg, B.S.T.J., **42,** March, 1963, pp. 355–382.

On page 361, the expression (y_1+y_2), which appears twice, should be replaced in both positions with $(y_1 + \bar{y}_2)$, in which $\bar{y}_2$ denotes the value of y_2 when $y_1 = 0$.

On page 377, the left side of the first equation of Section 9.4 should be det $\mathbf{F}_{\bar{y}_1}$, not det $\mathbf{F}_1$.

A Frequency-Domain Condition for the Stability of Feedback Systems Containing a Single Time-Varying Nonlinear Element

By I. W. SANDBERG

(Manuscript received May 6, 1964)

It is proved that a condition similar to the Nyquist criterion guarantees the stability (in an important sense) of a large class of feedback systems containing a single time-varying nonlinear element. In the case of principal interest, the condition is satisfied if the locus of a certain complex-valued function (a) is bounded away from a particular disk located in the complex plane, and (b) does not encircle the disk.

I. INTRODUCTION

The now well-known techniques introduced by Lyapunov have led to many very interesting results concerning the stability of time-varying nonlinear feedback systems governed by systems of differential equations. However, these methods have by no means led to a definitive theory of stability for even the simplest nontrivial time-varying nonlinear feedback systems. The general problem is, of course, one of considerable difficulty.

The unparalleled utility of the Nyquist stability criterion for single-loop, linear, time-invariant feedback systems is directly attributable to the fact that it is an explicit frequency-domain condition. The Nyquist locus not only indicates the stability or instability of a system, it presents the information in such a way as to aid the designer in arriving at a suitable design. The criterion is useful even in cases in which the system is so complicated that a sufficiently accurate analysis is not feasible, since experimental measurements can be used to construct the loop-gain locus.

The primary purpose of this article is to point out that some recently

obtained mathematical results,[1] not involving the theory of Lyapunov, imply that a condition similar to, and possessing the advantages of, the Nyquist criterion guarantees the stability (in an important sense) of feedback systems containing a single time-varying nonlinear element.*,†

II. THE PHYSICAL SYSTEM AND DEFINITION OF $\mathfrak{L}_2$-STABILITY

Consider the feedback system of Fig. 1. We shall restrict our discussion throughout to cases in which g_1, f, u, and v denote real-valued measurable functions of t defined for $t \geqq 0$.

The block labeled ψ is assumed to represent a memoryless time-varying (not necessarily linear) element that introduces the constraint $u(t) = \psi[f(t),t]$, in which $\psi(x,t)$ is a function of x and t with the

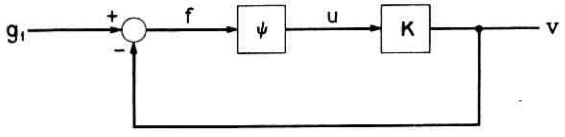


Fig. 1 — Nonlinear feedback system.

properties that $\psi(0,t) = 0$ for $t \geqq 0$ and there exist a positive constant β and a real constant α such that

$$\alpha \leqq \frac{\psi(x,t)}{x} \leqq \beta, \qquad t \geqq 0$$

for all real $x \neq 0$. In particular, we permit the extreme cases in which $\psi(x,t)$ is either independent of t or linear in x [i.e., $\psi(x,t) = \psi(1,t)x$].

The block labeled **K** represents the linear time-invariant portion of the forward path. It is assumed to introduce the constraint

$$v(t) = \int_0^t k(t - \tau)u(\tau)d\tau - g_2(t), \qquad t \geqq 0$$

in which k and g_2 are real-valued functions such that

$$\int_0^\infty |k(t)|\,dt < \infty, \qquad \int_0^\infty |g_2(t)|^2\,dt < \infty. \tag{1}$$

* The results of Ref. 1 relate to feedback systems containing an arbitrary finite number of time-varying nonlinear elements, but, with the exception of the case discussed here, they do not admit of a simple geometric interpretation.

† For results concerned with frequency-domain conditions for the global asymptotic stability (a sense of stability that is different from the one considered here) of nonlinear systems, see, for example, Refs. 2–4.

The function g_2 takes into account the initial conditions at $t = 0$. Our assumptions regarding **K** are satisfied, for example, if, as is often the case, u and v are related by a differential equation of the form

$$\sum_{n=0}^{N} a_n \frac{d^n v}{dt^n} = \sum_{n=0}^{N-1} b_n \frac{d^n u}{dt^n}, \qquad t \geqq 0$$

in which the a_n and the b_n are constants with $a_N \neq 0$, and

$$\sum_{n=0}^{N} a_n s^n \neq 0 \quad \text{for} \quad \mathrm{Re}[s] \geqq 0.$$

However, we *do not* require that u and v be related by a differential equation (or by a system of differential equations).

Assumption: We shall assume throughout that the response v is well defined and satisfies the inequality

$$\int_0^t | v(\tau) |^2 \, d\tau < \infty \tag{2}$$

for all *finite* $t > 0$, for each initial-condition function g_2 that meets the conditions stated above and each input g_1 such that

$$\int_0^\infty | g_1(t) |^2 \, dt < \infty.$$

Although this assumption plays an important role in the proof of the theorem to be presented, from an engineering viewpoint it is a trivial restriction (see Ref. 5).

Definition: We shall say that the feedback system of Fig. 1 is "$\mathfrak{L}_2$-stable" if and only if there exists a positive constant ρ with the property that the response v satisfies

$$\left(\int_0^\infty | v(t) |^2 \, dt \right)^{\frac{1}{2}} \leqq \rho \left(\int_0^\infty | g_1(t) + g_2(t) |^2 \, dt \right)^{\frac{1}{2}} + \left(\int_0^\infty | g_2(t) |^2 \, dt \right)^{\frac{1}{2}}$$

for every initial-condition function g_2 that meets the conditions stated above, and every input g_1 such that

$$\int_0^\infty | g_1(t) |^2 \, dt < \infty.$$

In particular, if the system is $\mathfrak{L}_2$-stable, then the response is square-integrable whenever the input is square-integrable.

It can be shown* that the response $v(t)$ approaches zero as $t \to \infty$ for any square-integrable input g_1, provided that the system is $\mathfrak{L}_2$-stable,

* See the proof of Theorem 6 of Ref. 1.

$g_2(t) \to 0$ as $t \to \infty$, and

$$\int_0^\infty | k(t) |^2 \, dt < \infty. \tag{3}$$

In addition, it follows at once from the Schwarz inequality that the response $v(t)$ is uniformly bounded on $[0, \infty)$ for any square-integrable input g_1, provided that the system is $\mathfrak{L}_2$-stable, $g_2(t)$ is uniformly bounded on $[0, \infty)$, and (3) is satisfied.

III. SUFFICIENT CONDITIONS FOR $\mathfrak{L}_2$-STABILITY

Theorem: Let

$$K(i\omega) = \int_0^\infty k(t) e^{-i\omega t} \, dt, \qquad -\infty < \omega < \infty.$$

The feedback system of Fig. 1 is $\mathfrak{L}_2$-stable if one of the following three conditions is satisfied:

(i) $\alpha > 0$; *and the locus of* $K(i\omega)$ *for* $-\infty < \omega < \infty$ *(a) lies outside the circle* C_1 *of radius* $\frac{1}{2}(\alpha^{-1} - \beta^{-1})$ *centered on the real axis of the complex plane at* $[-\frac{1}{2}(\alpha^{-1} + \beta^{-1}), 0]$, *and (b) does not encircle* C_1 *(see Fig. 2)*

(ii) $\alpha = 0$, *and* $Re[K(i\omega)] > -\beta^{-1}$ *for all real* ω

(iii) $\alpha < 0$, *and the locus of* $K(i\omega)$ *for* $-\infty < \omega < \infty$ *is contained within the circle* C_2 *of radius* $\frac{1}{2}(\beta^{-1} - \alpha^{-1})$ *centered on the real axis of the complex plane at* $[-\frac{1}{2}(\alpha^{-1} + \beta^{-1}), 0]$ *(see Fig. 3).*

Proof: Note first that

$$\int_0^\infty | u(t) |^2 \, dt \leqq \max(\beta^2, | \alpha |^2) \int_0^\infty | f(t) |^2 \, dt,$$

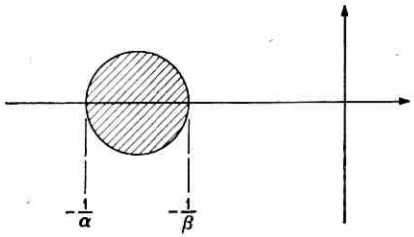


Fig. 2 — Location of the "critical circle" C_1 in the complex plane ($\alpha > 0$). The feedback system is $\mathfrak{L}_2$-stable if the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ lies outside C_1 and does not encircle C_1.

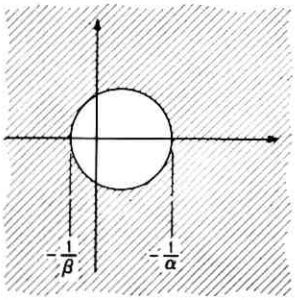


Fig. 3 — Location of the "critical circle" C_2 in the complex plane ($\alpha < 0$). The feedback system is $\mathfrak{L}_2$-stable if the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ is contained within C_2.

and hence, by a well-known result,

$$\int_0^\infty \left| \int_0^t k(t-\tau)u(\tau)d\tau \right|^2 dt \leqq \left(\int_0^\infty |k(t)|\,dt \right)^2 \int_0^\infty |u(t)|^2\,dt$$

$$\leqq \max\left(\beta^2, |\alpha|^2\right)\left(\int_0^\infty |k(t)|\,dt \right)^2 \int_0^\infty |f(t)|^2\,dt.$$

Using Minkowski's inequality,

$$\left(\int_0^\infty |v(t)|^2\,dt \right)^{\frac{1}{2}} \leqq \left(\int_0^\infty \left| \int_0^t k(t-\tau)u(\tau)d\tau \right|^2 dt \right)^{\frac{1}{2}}$$

$$+ \left(\int_0^\infty |g_2(t)|^2\,dt \right)^{\frac{1}{2}} \leqq \max(\beta, |\alpha|) \int_0^\infty |k(t)|\,dt$$

$$\cdot \left(\int_0^\infty |f(t)|^2\,dt \right)^{\frac{1}{2}} + \left(\int_0^\infty |g_2(t)|^2\,dt \right)^{\frac{1}{2}}.$$

Consider now the relation between $(g_1 + g_2)$ and f:

$$g_1(t) + g_2(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau),\tau]d\tau, \qquad t \geqq 0$$

and suppose that

$$\int_0^\infty |g_1(t) + g_2(t)|^2 dt < \infty.$$

According to the results of Ref. 1, our assumptions* imply that there

* In Ref. 1 it is assumed that

$$\int_0^t |f(\tau)|^2\,d\tau < \infty$$

exists a positive constant ρ_1 (which does not depend upon g_1 or g_2) such that

$$\int_0^\infty |f(t)|^2\,dt < \rho_1 \int_0^\infty |g_1(t) + g_2(t)|^2\,dt$$

provided that, with

$$K(s) = \int_0^\infty k(t)e^{-st}\,dt$$

and $\omega = \mathrm{Im}[s]$,

(i) $1 + \frac{1}{2}(\alpha + \beta)K(s) \neq 0$ for $\mathrm{Re}[s] \geqq 0$, and

(ii) $\frac{1}{2}(\beta - \alpha) \max_{-\infty<\omega<\infty} |K(i\omega)[1 + \frac{1}{2}(\alpha + \beta)K(i\omega)]^{-1}| < 1.$

Thus the feedback system of Fig. 1 is $\mathfrak{L}_2$-stable if conditions (i) and (ii) are satisfied.

According to the well-known theorem of complex-function theory that leads to the Nyquist criterion, condition (i) is satisfied if (and only if) the polar plot of $K(i\omega)$ for $-\infty < \omega < \infty$ does not encircle or pass through the point $[-2(\alpha + \beta)^{-1},0]$. It can easily be verified that condition (ii) is met if one of the following three conditions is satisfied.

(a) $\alpha > 0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ lies outside the circle C_1 of radius $\frac{1}{2}(\alpha^{-1} - \beta^{-1})$ centered in the complex plane at $[-\frac{1}{2}(\alpha^{-1} + \beta^{-1}),0]$.

(b) $\alpha = 0$, and $\mathrm{Re}[K(i\omega)] > -\beta^{-1}$ for all real ω.

(c) $\alpha < 0$, and the locus of $K(i\omega)$ for $-\infty < \omega < \infty$ is contained within the circle C_2 of radius $\frac{1}{2}(\beta^{-1} - \alpha^{-1})$ centered in the complex plane at $[-\frac{1}{2}(\alpha^{-1} + \beta^{-1}),0]$.

If $\alpha > 0$, the point $[-2(\alpha + \beta)^{-1},0]$ lies on the real-axis diameter of C_1, while if condition (b) or (c) is met, it is impossible for the polar plot of $K(i\omega)$ to encircle the point $[-2(\alpha + \beta)^{-1},0]$. Therefore, the conditions of the theorem guarantee that the feedback system is $\mathfrak{L}_2$-stable.

Remarks

With regard to the necessity of our sufficient conditions for $\mathfrak{L}_2$-stability, consider, for example, the case in which $\alpha > 0$ and suppose, for simplicity, that v and u are related by a differential equation of the type mentioned in Section II. Then, a moment's reflection shows that there exists a $\psi(x,t)$, in fact a $\psi(x,t)$ which is independent of t and linear in x,

for all finite $t > 0$. Our assumption that (2) is satisfied for all finite $t > 0$ implies that this condition is met.

that satisfies our assumptions and for which the feedback system is *not* $\mathfrak{L}_2$-stable, provided that for some value of ω, $K(i\omega)$ is a point on the real-axis diameter of C_1. This clearly shows that the condition is in the correct "ball park." Similar remarks can be made concerning our conditions for the cases in which $\alpha < 0$ and $\alpha = 0$.

IV. FURTHER PROPERTIES OF THE FEEDBACK SYSTEM OF FIG. 1

It is possible to say much more about the properties of the feedback system on the basis of frequency-domain information if our assumptions regarding $\psi(x,t)$ are strengthened.

For example, suppose that

$$\alpha \leqq \frac{\psi(x_1,t) - \psi(x_2,t)}{x_1 - x_2} \leqq \beta, \qquad \psi(0,t) = 0 \tag{4}$$

for $t \geqq 0$ and all real $x_1 \neq x_2$, and that one of the three conditions of our theorem is met. Let g_1 and $\hat{g}_1$ denote two arbitrary input functions such that

$$\int_0^t | g_1(\tau) |^2 d\tau < \infty \quad \text{and} \quad \int_0^t | \hat{g}_1(\tau) |^2 d\tau < \infty$$

for all finite $t > 0$, and

$$\int_0^\infty | g_1(\tau) - \hat{g}_1(\tau) |^2 d\tau < \infty.$$

Let v and $\hat{v}$, respectively, denote the (assumed well defined) responses due to g_1 and $\hat{g}_1$. Then if

$$\int_0^t | v(\tau) |^2 d\tau < \infty \quad \text{and} \quad \int_0^t | \hat{v}(\tau) |^2 d\tau < \infty$$

for all finite $t > 0$, and the assumptions of Section II are met, it follows* that

$$\int_0^\infty | v(\tau) - \hat{v}(\tau) |^2 d\tau < \infty$$

and that there exists a positive constant λ (which does not depend upon g_1 or $\hat{g}_1$) such that

$$\int_0^\infty | v(\tau) - \hat{v}(\tau) |^2 d\tau \leqq \lambda \int_0^\infty | g_1(\tau) - \hat{g}_1(\tau) |^2 d\tau.$$

* Consider Theorem 1 of Ref. 6 with $h_1(t) = f_1(t) = 0$ for $t < 0$.

Suppose now that $\psi(x,t)$ satisfies (4) and is either independent of t or periodic in t with period T for each x, and that one of the three conditions of our theorem is met. Assume that the initial-condition function $g_2(t)$ approaches zero as $t \to \infty$, and that the input $g_1(t)$ applied at $t = 0$ is a bounded periodic function with period T. Then it can be shown* that there exists a bounded periodic function p, with period T, which is independent of g_2 and such that the (assumed well defined) response $v(t)$ approaches $p(t)$ as $t \to \infty$, provided that the conditions of Section II are met, (2) is satisfied for all finite $t > 0$, and

$$\int_0^\infty \left| \int_t^\infty | k(\tau) | \, d\tau \right|^2 dt < \infty, \qquad \int_0^\infty | (1 + t)k(t) |^2 dt < \infty. \tag{5}$$

Observe that the conditions of (5) are satisfied if u and v are related by a differential equation of the form described in Section II.

REFERENCES

1. Sandberg, I. W., On the $\mathfrak{L}_2$-Boundedness of Solutions of Nonlinear Functional Equations, B.S.T.J., this issue, p. 1581.
2. Popov, V. M., Absolute Stability of Nonlinear Systems of Automatic Control, Avtomatika i Telemekhanika, **22,** Aug., 1961, pp. 961–978.
3. Kalman, R. E., Lyapunov Functions For the Problem of Lur'e in Automatic Control, Proc. Natl. Acad. Sci., **49,** Feb., 1963, pp. 201–205.
4. Rekasius, Z. V., A Stability Criterion for Feedback Systems with One Nonlinear Element, Trans. IEEE-PTGAC, **AC9,** Jan., 1964, pp. 46–50.
5. Tricomi, F. G., *Integral Equations,* Interscience Publishing, Inc., New York, 1957, p. 46.
6. Sandberg, I. W., and Beneš, V. E., On the Properties of Nonlinear Integral Equations That Arise in the Theory of Dynamical Systems, to be published.

* See Theorem 3 of Ref. 6.

The Resistance of an Infinite Slab with a Disk Electrode

By G. F. FOXHALL and J. A. LEWIS

(Manuscript received April 17, 1964)

We consider the resistance of an infinite slab measured between an electrode covering one face and a circular electrode attached to the other face by a uniform resistive film representing a contact resistance. Upper and lower bounds are found on the difference between the total resistance and the resistance of the film alone. The bounds are obtained by a combination of analysis and experiment, using an electrolytic tank. The results may be applied to determine contact resistance from a measured value of total resistance and a knowledge of the bulk resistivity of the slab material.

1. INTRODUCTION

We consider the resistance of an infinite conducting slab as measured between an electrode entirely covering one face and a circular electrode affixed to the other face by a resistive film. This resistance can be imagined to be made up of two resistances in series: namely, the film contact resistance and a resistance which is due to the body to which the electrode is attached, but which depends on the film resistance.

Lewis[1] has derived general upper and lower bounds on this body resistance. In the present case the upper bound may be calculated analytically. The lower bound is the resistance which would exist between the electrodes in the absence of the film. Calculation of the last-mentioned resistance involves a classical potential problem treated by Weber in 1873, but still not completely solved today.

We treat this problem by a combination of analysis and experiment, the latter in effect being an analog computation using an electrolytic tank. An asymptotic solution is found which converges rapidly for slab thicknesses as small as one disk radius, while for smaller thicknesses experimentally determined values of resistance are used.

The upper and lower bounds for the body resistance, which differ only by 8 per cent for a thick slab and tend to the same value for a thin slab,

provide a convenient estimate of slab resistance in the presence of a contact film. More important from the practical point of view, they provide a useful estimate of contact resistance for measured total resistance.

II. THE PROBLEM

We wish to determine the electrical resistance of the slab electrode configuration shown in Fig. 1. The entire base of the slab is in perfect contact with a highly conducting, grounded electrode, while the upper electrode, a circular disk of radius a, is separated from the slab by a thin

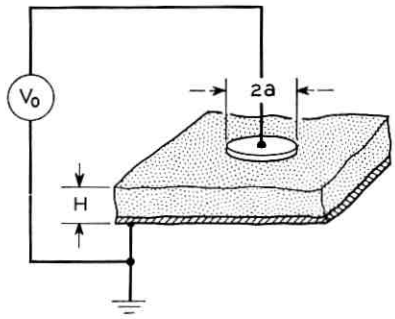


Fig. 1 — Infinite slab with disk electrode.

film of surface conductance c. The potential $V(R,Z)$ in the slab then satisfies Laplace's equation

$$\partial^2 V/\partial R^2 + \partial V/R\partial R + \partial^2 V/\partial Z^2 = 0, \tag{1}$$

for all R and $0 < Z < H$, and the boundary conditions

$$V(R,0) = 0, \tag{2}$$

$$\sigma\partial V(R,H)/\partial Z = \begin{cases} c[V_0 - V(R,H)], & \text{for } R < a \\ 0, & \text{for } R > a, \end{cases} \tag{3}$$

where σ is the conductivity of the slab.

The total input current I is given by

$$I = \int_0^a \sigma \frac{\partial V}{\partial Z}\,(R,H)2\pi R\,dR, \tag{4}$$

so that the resistance measured between the electrodes is

$$R_m = \frac{V_0}{I} = V_0 \Big/ \int_0^a \sigma \frac{\partial V}{\partial Z}\,(R,H)2\pi R\,dR. \tag{5}$$

If we set

$$r = R/a, \qquad z = Z/a, \qquad h = H/a, \qquad w(r,z) = V(R,Z)/V_0,$$

this becomes

$$\sigma a R_m = 1 \Big/ \int_0^1 \frac{\partial w}{\partial z}(r,h) 2\pi r \, dr, \tag{6}$$

where

$$\nabla^2 w = \partial^2 w/\partial r^2 + \partial w/r\partial r + \partial^2 w/\partial z^2 = 0, \tag{7}$$

in $0 < z < h$,

$$w(r,0) = 0, \tag{8}$$

for all r, and

$$\partial w(r,h)/\partial z = \begin{cases} (ca/\sigma)[1 - w(r,h)], & \text{for } r < 1 \\ 0, & \text{for } r > 1. \end{cases} \tag{9}$$

It is easy to show that R_m may be written in the form

$$R_m = R_c + R(w), \tag{10}$$

where R_c is the film resistance ($1/\pi a^2 c$, in the present case) and $R(w)$ is the ratio of average potential difference between electrodes to total current, i.e.,

$$\sigma a R(w) = 2 \int_0^1 w(r,h) r \, dr \Big/ \int_0^1 \frac{\partial w}{\partial z}(r,h) 2\pi r \, dr. \tag{11}$$

Thus, if $R(w)$ has been calculated and R_m measured, the film resistance R_c may be determined.

However, the calculation of $R(w)$ in general involves the solution of a difficult mixed boundary value problem. Furthermore, since w depends on the film conductance c, which is essentially the quantity to be determined by combined calculation and measurement, $R(w)$ must be calculated for a large number of values of c to make certain that the experimental range is covered. These difficulties can be circumvented, with only a moderate loss in accuracy in the present case, by the use of certain upper and lower bounds on $R(w)$.

III. UPPER AND LOWER BOUNDS

The bounds on $R(w)$ have the form[1]

$$R(u) \leqq R(w) \leqq R(v), \tag{12}$$

where

$$\sigma a R(u) = 1 \Big/ \int_0^1 \frac{\partial u}{\partial z}\,(r,h) 2\pi r\, dr, \tag{13}$$

$$\sigma a R(v) = (2/\pi) \int_0^1 v(r,h) r\, dr, \tag{14}$$

u and v satisfy (7) and (8), and

$$u(r,h) = \partial v(r,h)/\partial z = 1, \qquad \text{for } r < 1 \tag{15}$$

$$\partial u(r,h)/\partial z = \partial v(r,h)/\partial z = 0, \qquad \text{for } r > 1. \tag{16}$$

The lower bound $R(u)$ is the resistance as usually defined, i.e., the reciprocal of the total current for a unit potential difference applied uniformly between the electrodes. On the other hand, $R(v)$ is the *average* potential difference required to give unit total current, distributed uniformly over the input electrode. The former case may be realized by letting the film conductance c become very large; the latter by letting c tend to zero and V_0 tend to infinity in such a way that cV_0 tends to a finite value.

The calculation of $R(v)$ is straightforward, involving only the solution of an unmixed boundary value problem ($\partial v/\partial z$ specified all over $z = h$), but $R(u)$ involves the solution of a *mixed* boundary value problem and ultimately the solution of dual integral equations. An asymptotic solution of these integral equations, valid for large thickness ($h \gg 1$), has been obtained by Tranter.[2] The corresponding expression for $R(u)$ is rapidly convergent, so that it appears to be usable down to values of h of order unity. The range $h < 1$ is covered by measurements on an electrolytic tank analog.

IV. THE UPPER BOUND

If we set

$$v(r,z) = \int_0^\infty f(p) \frac{\sinh pz}{\cosh ph} J_0(pr) dp, \tag{17}$$

the conditions

$$\nabla^2 v = v(r,0) = 0$$

are satisfied and the remaining conditions become

$$\partial v(rh)/\partial z = \int_0^\infty pf(p) J_0(pr) dp = \begin{cases} 1, & r < 1 \\ 0, & r > 1. \end{cases} \tag{18}$$

Now

$$\int_0^\infty J_0(pr)J_1(p)dp = \begin{cases} 1, & r < 1 \\ 0, & r > 1, \end{cases} \tag{19}$$

so that, if we set

$$f(p) = J_1(p)/p,$$

we obtain the complete solution. In particular,

$$v(r,h) = \int_0^\infty \frac{\tanh ph}{p} J_1(p)J_0(pr)dp \tag{20}$$

and

$$\sigma aR(v) = (2/\pi) \int_0^\infty [J_1(p)/p]^2 \tanh ph\, dp. \tag{21}$$

For small h, $\sigma aR(v) \approx h/\pi$ or $R(v) \approx H/\pi a^2\sigma$, the resistance of a circular cylinder, while for large h, $\sigma aR(v) \approx 8/3\pi^2$, a result derived in Carslaw and Jaeger.[3]

Miss M. C. Gray has obtained an expansion in powers of h^{-1} whose first two terms, i.e.,

$$\sigma aR(v) \approx \frac{8}{3\pi^2} - \frac{\log 2}{2\pi h} \tag{22}$$

give reasonable accuracy down to $h = 1$. She has also evaluated the integral for $R(v)$ numerically for $0.1 < h < 10$. This is the curve labeled $R(v)$ in Fig. 2.

V. THE ASYMPTOTIC VALUE OF $R(u)$

If we assume $u(r,z)$ to have the same form as $v(r,z)$ in the previous section, the function $f(p)$ must now satisfy the dual integral equations

$$u(r,h) = \int_0^\infty f(p) \tanh phJ_0(pr)dp = 1, \qquad r < 1, \tag{23}$$

$$\partial u(r,h)/\partial z = \int_0^\infty pf(p)J_0(pr)dp = 0, \qquad r > 1. \tag{24}$$

We can no longer solve these equations by inspection, but for large h an approximate solution, due to Tranter,[2] is available. Tranter gives

$$f(p) \approx \left(\frac{2}{\pi}\right) A(h) \frac{\sin p}{p}, \tag{25}$$

where

$$A(h) = 1 + \frac{2 \log 2}{\pi h} + \left(\frac{2 \log 2}{\pi h}\right)^2 + O(h^{-3}). \tag{26}$$

For $A = 1$ ($h = \infty$) we obtain the classical result

$$[\sigma a R(u)]_\infty = \tfrac{1}{4},$$

and for large h

$$\sigma a R(u) \approx 1/4A(h). \tag{27}$$

This function is shown in Fig. 2 for $h \geqq 1$.

VI. THE EXPERIMENT

In order to verify and supplement the computed values of resistance, an experiment using an electrolytic solution as the conducting slab was devised. The apparatus, shown in Fig. 3, consisted of a 10 × 14 inch plastic tank into which a gold-plated brass plate was fitted. A 0.01 normal KCl solution was used to simulate the conducting slab, while the end of a 0.564-inch diameter gold-plated brass rod served as the disk electrode.

The experiment consisted of two parts. First resistance measurements were made for the slab configuration at various solution levels. Then a glass sleeve, closely fitted to the upper electrode, was used to constrain the current in the solution to a simple cylindrical geometry. This provided a measurement of solution conductivity and also of contact resist-

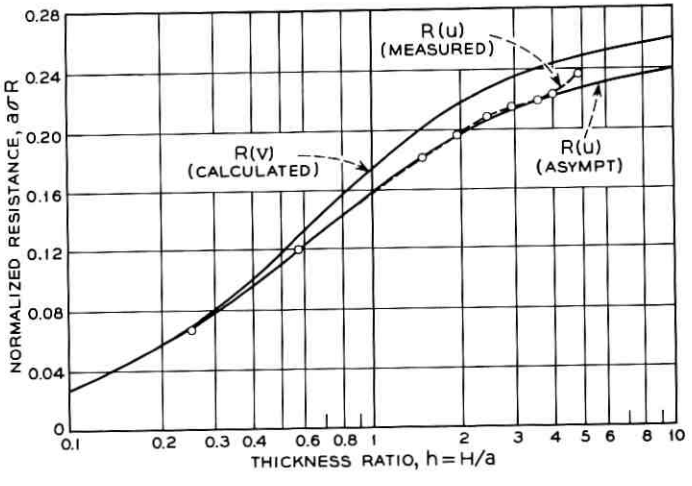


Fig. 2 — The resistance of an infinite slab.

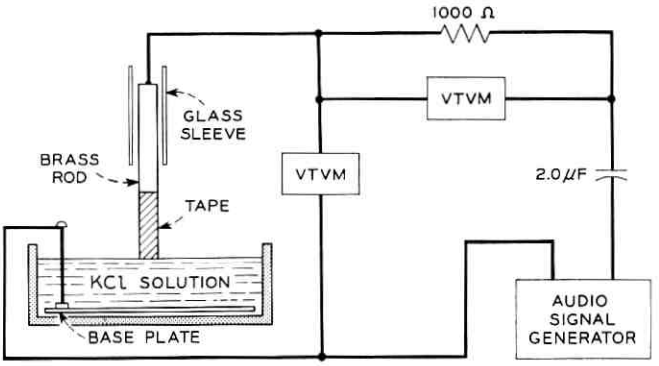


Fig. 3 — Electrolytic tank analog.

ance. The resulting measured values are shown in Fig. 4 as a function of solution depth. The slope gives a conductivity of 1.33×10^{-3} (ohm-cm)$^{-1}$, in close agreement with the tabulated value[4] for the solution temperature of 22°C. Extrapolation to zero solution depth indicates a negligible contact resistance.

As seen in Fig. 2, the measured values of resistance for the slab fall

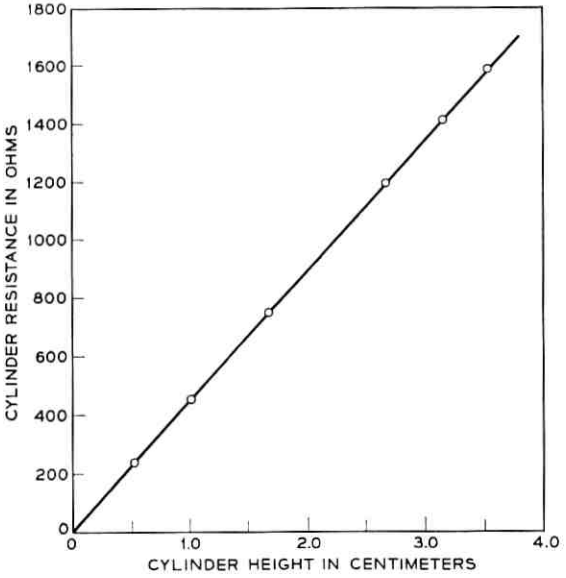


Fig. 4 — The resistance of a cylinder of electrolyte.

very close to the computed values in the range $2 \leqq h \leqq 4$, where the asymptotic form for $R(u)$ can be expected to be accurate. At smaller values of h, the accuracy of the asymptotic form decreases and the measured values may be taken as a good approximation to $R(u)$. At large values of h, on the other hand, the asymptotic form becomes very accurate, while the experimental model becomes less so. This may be seen in Fig. 2 for $h \approx 5$.

This divergence, which at first was attributed to the finite diameter of the tank, is now believed to be due to polarization effects produced by the nonuniform field near the upper electrode. The abrupt upturn of resistance near $h = 5$ cannot be due to finite tank diameter, which would yield a resistance-depth curve with a slope decreasing from a value for a cylinder having the same diameter as the upper electrode to one for a long cylinder having the same diameter as the tank. On the other hand, the current density distribution over the upper electrode is nonuniform for any depth, being infinite at the electrode edge in the mathematical idealization. This nonuniformity increases with depth, for more current is drawn from the electrode center at small depth than at large depth. Thus for fixed total current (the experimental condition) the current density at the edge increases with increasing depth until a depth is reached ($h \approx 5$ in the experiments) at which local polarization effects become appreciable. This dependence on depth also accounts for the different behavior of measured values for the cylinder (Fig. 4) and the slab (Fig. 2).

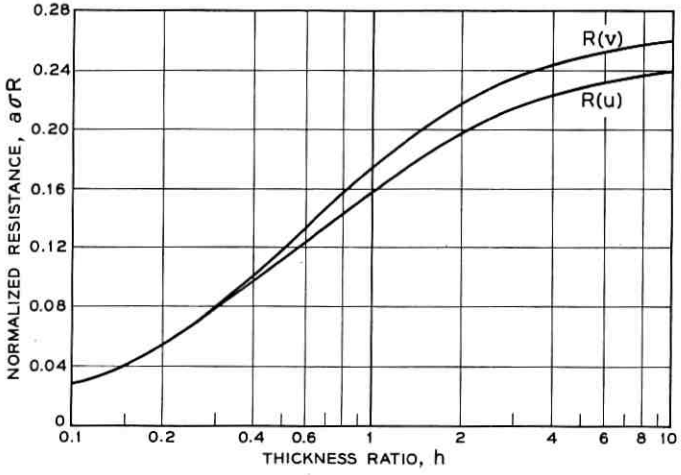


Fig. 5 — Upper and lower bounds on the resistance of an infinite slab.

VII. SUMMARY OF RESULTS

Fig. 5 summarizes the results of the analysis and the experiment. The lower curve, labeled $R(u)$, gives the resistance between the electrodes for zero contact resistance. It also gives a lower bound on the difference between the total resistance R_m with a resistive film between the upper electrode and the slab and the resistance $R_c = 1/\pi a^2 c$ of the film itself, while the upper curve, labeled $R(v)$, gives an upper bound on the same quantity. Thus, for all slab thicknesses and (constant) film conductances c,

$$R(u) \leqq R_m - R_c \leqq R(v), \tag{28}$$

or, assuming R_m to be determined by measurement,

$$R_m - R(v) \leqq R_c \leqq R_m - R(u) \tag{29}$$

giving an estimate of the contact resistance itself.

To supplement Fig. 3, the asymptotic forms

$$\sigma a R(u) \sim (0.250)/[1 + (0.441/h) + (0.441/h)^2], \tag{30}$$

$$\sigma a R(v) \sim (0.270) + (0.110/h), \tag{31}$$

may be used for thickness ratio $h \geqq 1$, while, for $h \leqq 0.1$, both $\sigma a R(u)$ and $\sigma a R(v)$ are closely approximated by the cylinder resistance h/π.

VIII. ACKNOWLEDGMENTS

The authors wish to acknowledge the help of Miss M. C. Gray with the evaluation of the integral for $R(v)$ and the advice of D. L. Klein on the experiment.

REFERENCES

1. Lewis, J. A., Generalized Thermal Resistance, Quart. Appl. Math., **19**, April, 1961, pp. 76–80.
2. Tranter, C. J., *Integral Transforms in Mathematical Physics*, 2nd Ed., Methuen, New York, 1956, pp. 117–120.
3. Carslaw, H. S., and Jaeger, J. C., *Conduction of Heat in Solids*, 2nd. Ed., Oxford, 1959, pp. 214–217.
4. *Handbook of Chemistry and Physics*, 32nd Ed., Chemical Rubber Publishing Company, Cleveland, 1950, p. 2147.

Permutation Groups, Complexes, and Rearrangeable Connecting Networks

By V. E. BENEŠ

(Manuscript received March 12, 1964)

In the interest of providing good telephone service with efficient connecting networks, it is desirable to have at hand a knowledge of some of the combinatorial properties of such networks. One of these properties is rearrangeability: a connecting network is rearrangeable if its permitted states realize every assignment of inlets to outlets, or alternatively, if given any state x of the network, any inlet idle in x, and any outlet idle in x, there is a way of assigning new routes (if necessary) to the calls in progress in x so that the idle inlet can be connected to the idle outlet.

A natural algebraic and combinatorial approach to the study of rearrangeable networks is described, with attention centered principally on two-sided networks built of stages of square crossbar switches, each stage having N inlets and N outlets. The approach is based in part on the elementary theory of permutation groups. The principal problem posed (and partly answered) is this: What connecting networks built of stages are rearrangeable? Sufficient conditions, including all previously known results, are formulated and exemplified.

I. INTRODUCTION

A connecting network is an arrangement of switches and transmission links through which certain terminals can be connected together in many combinations. Typical examples of connecting networks can be found in telephone central offices, where they are used to complete calls among the customers themselves, and between customers and outgoing trunks leading to other offices.

In the interest of providing good service with efficient connecting networks, it is desirable to have a thorough understanding of some of the combinatorial properties of such networks. In a previous paper,[1] we singled out three such combinatory properties as useful in assessing the performance of connecting networks. The weakest of these properties

was that of *rearrangeability*. A connecting network is rearrangeable if its permitted states realize every assignment of inlets to outlets, or alternatively, if given any state x of the network, any inlet idle in x, and any outlet idle in x, there is a way of assigning new routes (if necessary) to the calls in progress in x so as to lead to a new state of the network in which the idle inlet can be connected to the idle outlet.

Figs. 1 and 2 show the structure of two connecting networks built out of square crossbar switches, with each switch capable of connecting any subset of its inlets to an equinumerous subset of its outlets in any desired one-one combination. The network of Fig. 1 is often found in telephone central offices; we may call it the No. 5 crossbar network. It is *not* rearrangeable. The network of Fig. 2 *is* rearrangeable, but so far it has not found extensive practical use.

We shall describe a natural algebraic and combinatorial approach to the study of rearrangeability. For the most part we restrict attention to two-sided connecting networks that are built of stages of crossbar switches, and have the same number N of inlets as outlets. The approach is based in part on the elementary theory of permutation groups. The way the connection with group theory arises can be summarized as follows: a maximal state of the network is one in which no additional

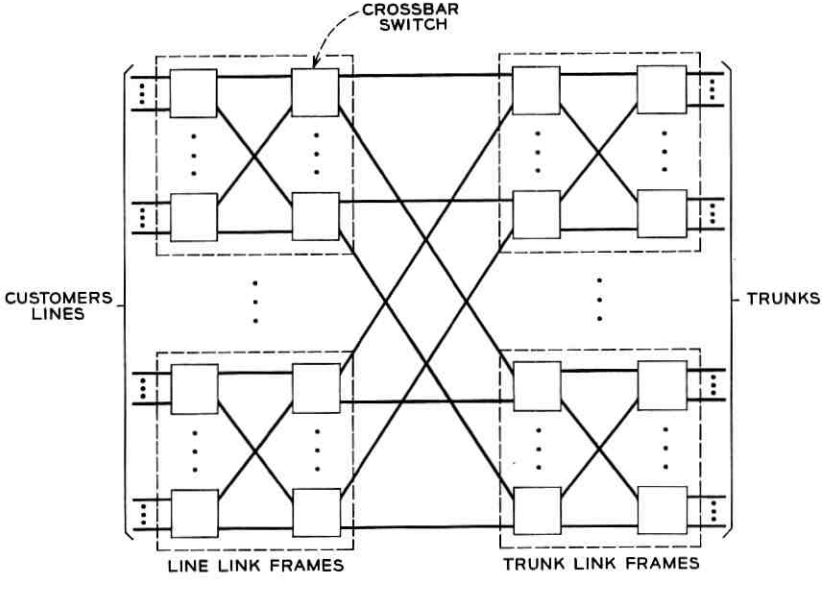


Fig. 1 — Structure of No. 5 crossbar network.

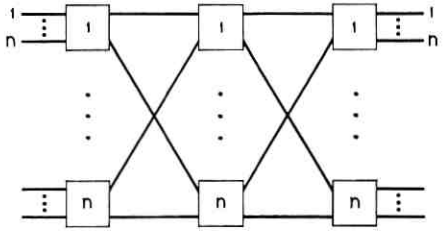


Fig. 2 — Rearrangeable network.

calls can be completed in the network; suppose that both the inlets and the outlets are numbered in an arbitrary way from 1 to N; each maximal state realizes some submap of a permutation on $\{1, \cdots, N\}$; the network is rearrangeable if and only if the whole group of all permutations of $\{1, \cdots, N\}$ is generated in this way by the maximal states of the network. Details are worked out in the main body of the paper.

It is not possible to explore in one paper all the possible uses of group theory in the study of connecting networks. Indeed, we shall restrict ourselves to formulating the fundamental problem of rearrangeable networks in terms of complexes of permutations, and to giving a partial answer. One of the difficulties with the approach is that it always seems to be easier to obtain results about groups by the few available methods known for rearrangeable networks, than *vice versa*.

A sequel[2] to the present paper is concerned with the problem of synthesizing a rearrangeable network (for N inlets and outlets), subject to certain structural conditions and to the condition that it have a minimum number of crosspoints.

II. SUMMARY

In Section III we define a precise general notion of a "stage" of switching in a connecting network, and, after describing how the networks which will be of interest are built out of stages by joining them together by patterns of links, we pose two problems: first, to discover what networks built in this way are rearrangeable; and second, to synthesize optimal rearrangeable networks of given size, optimal in the sense of having fewest crosspoints (among those in some class of networks having practical interest). (See Ref. 2.)

Section IV is devoted to giving a formulation of the first problem (discovering rearrangeable networks) in terms of partitions and permutation groups, using the notion of stage. In Section V we discuss how stages

generate complexes (in the group theory sense, i.e., sets of group elements). It is shown that a stage can generate a subgroup only if it contains a substage made of square switches, a result that indicates to some extent the "best possible" nature of stages made of square switches.

A known example, discussed in Section VI, indicates how a particular symmetric group S is generated by a rearrangeable network in the form

$$S = \left\{\prod_{i=1}^{3} \varphi_i\right\}$$

with φ_1 , φ_3 in a subgroup H and φ_2 in a certain subgroup $\varphi^{-1}H\varphi$ conjugate to H.

The remainder of the paper is devoted to proving two "rearrangeability" theorems for connecting networks built of stages of square switches. The first theorem gives sufficient conditions under which a set of stages of square switches connected by link patterns will give rise to a rearrangeable network. The second theorem indicates a simple way of describing link patterns and stages that satisfy the hypotheses of the first theorem, and so yield many specific rearrangeable networks, generalizations of those given by Paull.[3]

III. STAGES AND LINK PATTERNS

The switches in Figs. 1 and 2 are arranged in columns which we shall call *stages*, the switches in these stages being identical. Two adjacent stages are connected by a pattern of *links* or *junctors*. Along with the switches, the link patterns are responsible for the distributive characteristics of the network. They afford an inlet ways of reaching many outlets. Obviously, each outlet on a switch in a given stage is some inlet of the next stage, if there is one. Suppose that the N inlets are numbered in an arbitrary way, and that the N outlets are also numbered in an arbitrary way, both from 1 to N. Then it is clear that each link pattern, and each permitted way of closing the largest possible number of crosspoints in a stage, viz. N, can be viewed abstractly as a permutation on $\{1, \cdots, N\}$. Both the networks in Figs. 1 and 2 have the property that all maximal* states have the same number N of calls in progress, and any such maximal state realizes a permutation which is a *product* of certain of the permutations represented by the link patterns and the stages.

It will be convenient to generalize the usual notion of a "stage" of switching in a connecting network. By a stage (of switching) we shall

* I.e., states in which no additional calls can be completed.

mean a connecting network constructed as follows: with I the set of inlets, and Ω that of outlets, we choose an arbitrary subset $\mathcal{S}$ of $I \times \Omega$, and we place a crosspoint between all and only those inlets $u \in I$ and outlets $v \in \Omega$ such that $(u,v) \in \mathcal{S}$. We shall also speak of $\mathcal{S}$ itself as the "stage." Thus we make

Definition 1: A stage is a subset of $I \times \Omega$.

This terminology is easily seen to be an extension of the usual one, according to which, e.g., a column of switches in Fig. 1 forms a stage, the network having four stages separated (or joined) by three link patterns. Actually, a link pattern may be associated with one or the other (but usually not both) of the stages it connects, to define a new stage; we do not usually do this.

Definition 2: A stage $\mathcal{S}$ is made of square switches if and only if there is a partition Π of $\{1, \cdots, N\}$ such that

$$\mathcal{S} = \bigcup_{A \in \Pi} (A \times A).$$

Definition 3: A substage $\mathcal{S}'$ of $\mathcal{S}$ is a subset of $\mathcal{S}$.

Except in the trivial case in which $\mathcal{S}$ is actually a square N-by-N switch (i.e., $\mathcal{S} = I \times \Omega$), a stage $\mathcal{S}$ will not by itself give rise to a rearrangeable network. Still, it is known that several stages joined end to end by suitable link patterns can together give rise to such a network, e.g., that of Fig. 2. We can thus formulate two fundamental questions about connecting networks built out of stages:

(1) What stages and link patterns can be used to construct a rearrangeable network?

(2) What stages, and how many of them, should be used to construct a rearrangeable network that has a minimum number of crosspoints (switches) for a given number of terminals on a side?

Question (1) is studied in the present work, while question (2) is treated in another paper.[2]

IV. GROUP THEORY FORMULATION

We shall adopt some notational conventions from group theory to simplify our presentation. Let G be a group. It is customary to speak of a subset $K \subseteq G$ as a *complex*. If $x \in G$, then xK denotes the set of products xy with $y \in K$, Kx denotes the set of products yx with $y \in K$. Similarly, if K_1 and K_2 are complexes, K_1K_2 denotes the set of products yz with $y \in K_1$ and $z \in K_2$.

A group G of permutations is called *imprimitive*[4] if the objects acted

on by the permutations of G can be partitioned into mutually disjoint sets, called the sets of *imprimitivity*, such that every $\varphi \in G$ either permutes the elements of a set among themselves, or carries that set onto another. That is, there is a nontrivial partition Π of the set X of objects acted on such that $\varphi \in G$ and $A \in \Pi$ imply $\varphi(A) \in \Pi$. We shall extend this terminology as follows:

Definition 4: G is called *strictly imprimitive* if it is imprimitive, and each set A of imprimitivity is carried into itself by elements of G, i.e., there is a nontrivial partition Π of X such that $A \in \Pi$ implies $\varphi(A) = A$ for all $\varphi \in G$, so that $\varphi \in G$ is "nonmixing" on Π.

Consider a stage of switching that has N inlets and N outlets. It is evident that such a stage provides ways of connecting some of the inlets to some of the outlets. If the stage contains enough crosspoints it can be used to connect every inlet to some outlet in a one-to-one fashion, i.e., with no inlet connected to more than one outlet and vice versa. With the inlets and outlets both numbered $1, 2, \cdots, N$, such a setting of the switches corresponds to a permutation on $\{1, \cdots, N\}$. Indeed, there may be many ways of doing this, differing in what inlets are connected to what outlets, that is, corresponding to different permutations.

Definition 5: A stage $\mathcal{S}$ generates the permutation φ if there is a setting of N switches of $\mathcal{S}$ which connects each inlet to one and only one outlet in such a way that i is connected to $\varphi(i) i = 1, \cdots, N$, that is, if

$$(i,\varphi(i)) \in \mathcal{S}.$$

Definition 6: The set of permutations generated by a stage $\mathcal{S}$ is denoted by $P(\mathcal{S})$.

Definition 7: A network (with N inlets and N outlets) generates a permutation φ if there is a setting of the switches in the network which connects, by mutually disjoint paths, each inlet to one and only one outlet in such a way that i is connected to $\varphi(i)$, $i = 1, \cdots, N$.

If two stages $\mathcal{S}_1$, $\mathcal{S}_2$ are connected by a link pattern corresponding to a permutation φ_2, then the permutations that they generate together are those of the form

$$\varphi_1\varphi_2\varphi_3, \qquad \varphi_i \in P(\mathcal{S}_i), \qquad i = 1 \text{ or } 3.$$

If a network consists of two stages $\mathcal{S}_1$, $\mathcal{S}_2$ joined by a link pattern corresponding to a permutation φ, then it can be seen that it generates exactly the permutations in the set

$$P(\mathcal{S}_1)\varphi P(\mathcal{S}_2).$$

A network of s stages $\mathcal{S}_i$, $i = 1, \cdots, s$, with a link pattern corresponding to φ_i, $i = 1, \cdots, s-1$, between the ith and the $(i+1)$th stages, generates the complex

$$P(\mathcal{S}_1)\varphi_1 P(\mathcal{S}_2) \cdots \varphi_{s-1} P(\mathcal{S}_s).$$

We shall occasionally use the suggestive notation

$$\mathcal{S}_1 \varphi_1 \mathcal{S}_2 \cdots \varphi_{s-1} \mathcal{S}_s$$

to refer to or indicate such a network.

It is now possible to formulate a group-theoretic approach to the analysis and synthesis of rearrangeable connecting networks made of stages of switching joined by link patterns. Consider such a network, generating the complex

$$P(\mathcal{S}_1)\varphi_1 \cdots \varphi_{s-1} P(\mathcal{S}_s).$$

The factors $\varphi_i P(\mathcal{S}_{i+1})$, $i = 1, \cdots, s-1$, occurring herein are themselves again just complexes. Thus, given any product of complexes

$$\prod_{i=1}^{s} K_i$$

we seek to know whether the product is the whole symmetric group, and whether the factor complexes K_i can be written in the form

$$\varphi P(\mathcal{S})$$

where φ is a permutation and $\mathcal{S}$ is a stage. In this general form the problem is largely unsolved; however, special cases are worked out in the sequel.

V. THE GENERATION OF COMPLEXES BY STAGES

We start with this elementary result:

Remark 1: Let M be a complex (i.e., a set) of permutations. Define a stage $\mathcal{S}$ by

$$\mathcal{S} = \{(x,y): \varphi(x) = y \text{ for some } \varphi \in M\}.$$

Then $P(\mathcal{S}) \supseteq M$ and no smaller stage has this property.

In cases of practical importance, such as shown in Figs. 1 and 2, the stages are made of square switches, and it is clear that a stage $\mathcal{S}$ (with N inlets and N outlets) is capable of effecting certain special permutations on $X = \{1, \cdots, N\}$, and of course, all submaps thereof. (Indeed, for each switch there are numbers m and n with $m < n$ such that the switch

is capable of performing all the $(n - m + 1)!$ permutations of the numbers k in the range $m \leqq k \leqq n$ among themselves.) Since no inlet [outlet] is on more than one switch, these permutations form a *subgroup* of the symmetric group $S(X)$ of *all* permutations on $\{1, \cdots, N\}$. This subgroup has a property which might be described intuitively by saying that there exist sets on which the subgroup elements can mix "strongly," but which they keep separate. It is apparent, indeed, that the subgroup generated by a stage made of square switches is strictly imprimitive, the sets of imprimitivity being just the elements of the partition Π of $\{1, \cdots, N\}$ according to what switch an inlet [outlet] is on. This situation might also be described by saying that a permutation φ from the subgroup is *nonmixing* on Π.

Our second observation is

Remark 2: Let H be a strictly imprimitive group of permutations on $X = \{1, \cdots, N\}$, with sets of imprimitivity forming the partition Π. Let $\mathcal{S}$ be the smallest stage with $P(\mathcal{S}) \supseteq H$. Then

$$\mathcal{S} = \bigcup_{A \in \Pi} A \times A,$$

i.e., $\mathcal{S}$ is made of square switches.

The main result of this section states that a stage can generate a subgroup only if it contains a substage made of square switches. This suggests that stages made of square switches necessarily arise in the generation of the symmetric group by products of complexes some of which are subgroups.

Theorem 1: Let $\mathcal{S}$ be a stage, and let $P(\mathcal{S})$ contain a subgroup H of $S(X)$. Then there is a substage $\mathcal{R}$ of $\mathcal{S}$ which is made of square switches.

Proof: Define a relation $\mathcal{R}$ on $\{1, \cdots, N\}$ by the condition that $i\mathcal{R}j$ if and only if $j = \varphi(i)$ for some $\varphi \in H$. Since H is a subgroup, it must contain the identity permutation, i.e., $i\mathcal{R}i$ for all $i = 1, \cdots, N$. Let i, j, k be numbers in $\{1, \cdots, N\}$ such that $j = \varphi(i)$ and $k = \psi(j)$ for some permutations $\varphi, \psi \in H$. Then $\psi\varphi \in H$ and $k = \psi\varphi(i)$, that is, $i\mathcal{R}k$; hence $\mathcal{R}$ is transitive. Finally, if $j = \varphi(i)$ with $\varphi \in H$, we have $i = \varphi^{-1}(j)$ with $\varphi^{-1} \in H$, since H is a group. Hence $\mathcal{R}$ is an equivalence relation, and there is a partition Π such that

$$\mathcal{R} = \bigcup_{A \in \Pi} (A \times A).$$

Since $i\mathcal{R}j$ obviously implies $(i,j) \in \mathcal{S}$, we have

$$\mathcal{R} \subseteq \mathcal{S}.$$

$\mathcal{R}$ is clearly a substage of $\mathcal{S}$ made of square switches.

VI. AN EXAMPLE

As is well-known, elementary group theory contains many results that allow one to write a group as a product of complexes. These results often involve a *subgroup* of the group in question. We shall quote an elementary result of this kind, and interpret it in terms of a network that is known to be rearrangeable.

Let G be a group, and let H_1 and H_2 be subgroups of G, not necessarily distinct. A *double coset* is a complex of the form

$$H_1\varphi H_2, \qquad \varphi \in G$$

It is a known result[5] that two double cosets are either identical or disjoint. Thus there is at least one complex M with the properties

$$\bigcup_{\varphi \in M} H_1\varphi H_2 = G$$

$$H_1\varphi H_2 \cap H_1\psi H_2 = \theta \qquad \text{if } \varphi \neq \psi, \quad \text{with } \varphi, \psi \in M.$$

In particular

$$G = H_1 M H_2,$$

and we have factored G into a product of three complexes, two of which are subgroups. Now suppose that G is actually $S(X)$, the symmetric group of all permutations of N objects, and that m and n are positive integers such that $mn = N$. Let Π be a partition of $X = \{1, \cdots, N\}$ into m sets of n elements each, and let H be the largest strictly imprimitive subgroup of $S(X)$ whose sets of imprimitivity form Π. Also let φ be a self-inverse permutation, and Π_φ a partition, such that $A \in \Pi$, $B \in \Pi_\varphi$ imply*

$$|\varphi(A) \cap B| = 1.$$

Let K be the largest strictly imprimitive subgroup of $S(X)$ whose sets of imprimitivity form Π_φ.

By Remark 2, Section IV, H and K can each be generated by stages of square switches.

Returning to the earlier discussion leading to the factorization $G = H_1 M H_2$, we let $H_1 = H_2 = H$. Now it can be seen that the complex

$$H\varphi K\varphi H$$

is generated by a network of the form shown in Fig. 2. By the Slepian-Duguid theorem (Beneš, Ref. 1, p. 1484) this network is rearrangeable, so that

$$H\varphi K\varphi H = S(X).$$

* $|A|$ denotes the number of elements of a set A.

Since $\varphi = \varphi^{-1}$, the complex $\varphi K \varphi$ is itself actually the subgroup $\varphi^{-1} K \varphi$ conjugate to K. Thus for $G = S(X)$ and $H_1 = H_2 = K$, the factor M in

$$S(X) = HMH$$

can be chosen to be $\varphi^{-1} K \varphi$.

VII. SOME DEFINITIONS

The number of elements of a set A is denoted $| A |$. Let X, Y be arbitrary finite sets with $| X | = | Y |$, let B be a subset of Y, let Π_1, Π_2 be partitions of X, Y respectively, and let φ be a one-to-one map of Y onto X. Let θ be the null set.

Definition 8: $\varphi(B) = \{x \in X: \varphi^{-1}(x) \in B\}$.

Definition 9: φ *hits* Π_1 from B if and only if $A \in \Pi_1$ implies $\varphi(B) \cap A \neq \theta$.

Definition 10: φ *covers* Π_1 from Π_2 if and only if $B \in \Pi_2$ implies φ hits Π_1 from B.

Definition 11: $\varphi(\Pi_2)$ is the partition of X induced by φ acting on elements of Π_2, i.e.,

$$\varphi(\Pi_2) = \{\varphi(B): B \in \Pi_2\}.$$

Definition 12: ${}_B\varphi$ is the restriction of φ to B.

Let A be a subset of X.

Definition 13: ${}_A\Pi_1$ is the partition of A induced by Π_1, i.e.,

$${}_A\Pi_1 = \{C \cap A: C \in \Pi_1\}.$$

Definition 14: φ B-covers Π_1 from Π_2 if and only if ${}_B\varphi$ covers ${}_{\varphi(B)}\Pi_1$ from ${}_B\Pi_2$.

Definition 15: Let Π_0, Π_1 be partitions of X. Then $\Pi_1 > \Pi_0$ (read "pi-one refines pi-zero") if and only if every set in Π_0 is a union of sets in Π_1, and $\Pi_1 \neq \Pi_0$.

VIII. PRELIMINARY RESULTS

Lemma 1: Let X and Y be any sets with $| X | = | Y | < \infty$, let $\Pi_1 = \{A_1, \cdots, A_n\}$ and $\Pi_2 = \{B_1, \cdots, B_m\}$ be partitions of X and Y respectively, and suppose that for $k = 1, \cdots, n$ the union of any k elements of Π_1 has more elements than the union of any $k - 1$ elements of Π_2. Then

(*i*) $m \geqq n$.

(ii) For each one-to-one map f of X onto Y there exists a set of n distinct integers $k(1), \cdots, k(n)$ with $1 \leqq k(i) \leqq m$, $i = 1, \cdots, n$, and

$$f(A_i) \cap B_{k(i)} \neq \theta \qquad i = 1, \cdots, n.$$

Proof: Since Π_1 and Π_2 are partitions, and $|X| = |Y|$,

$$\sum_{j=1}^{n} |A_i| = \sum_{j=1}^{m} |B_j|.$$

If m were less than n, then the union of m B's has as many elements as the union of n A's, for $m < n$; this contradicts the hypothesis. Let

$$K_i = \{j: f(l) \in B_j \text{ for some } l \in A_i\}, \qquad i = 1, \cdots, n.$$

Also let $A_{i(1)}, \cdots, A_{i(k)}$ be any k elements of Π_1, $1 \leqq k \leqq n$, and set

$$T = \bigcup_{j=1}^{k} K_{i(j)}.$$

All of the

$$\sum_{j=1}^{k} |A_{i(j)}|$$

elements of

$$\bigcup_{j=1}^{k} A_{i(j)}$$

are mapped by f into $|T|$ sets of Π_2. Since no union of $k-1$ sets of Π_2 has

$$\sum_{j=1}^{k} |A_{i(j)}|$$

elements, it follows that $|T| \geqq k$. Thus the union of any k of the sets $\{K_i, i = 1, \cdots, n\}$ has at least k members. Hence by P. Hall's theorem[6] there is a set of n distinct representatives $k(1), \cdots, k(n)$ with

$$k(i) \in K_i \qquad i = 1, \cdots, n$$

$$k(i) \neq k(j) \qquad \text{for } i \neq j.$$

But clearly $k(i) \in K_i$ if and only if

$$f(l) \in B_{k(i)} \qquad \text{for some } l \in A_i,$$

that is, if and only if

$$f(A_i) \cap B_{k(i)} \neq \theta.$$

Lemma 2: Let Π_1, Π_2 *be partitions of sets* X, Y *respectively with the properties* $|X| = |Y|$ *and*

$$C_1, C_2 \in \Pi_1 \cup \Pi_2 \text{ implies } |C_1| = |C_2|.$$

Then for every one-to-one map φ *of* X *onto* Y *there is a set* $D \subseteq X$ *such that* φ *hits* Π_2 *from* D *and* φ^{-1} *hits* Π_1 *from* $\varphi(D)$.

Proof: We observe that $|\Pi_1| = |\Pi_2|$, and that the conditions of Lemma 1 are satisfied, with $m = n$. For each onto map φ there is a set $D \subseteq X$ with $|D| = |\Pi_1|$, such that

$$A \in \Pi_1 \text{ implies } D \cap A \neq \theta, \qquad (\varphi^{-1} \text{ hits } \Pi_1 \text{ from } \varphi(D))$$

$$B \in \Pi_2 \text{ implies } \varphi(D) \cap B \neq \theta, \qquad (\varphi \text{ hits } \Pi_2 \text{ from } D).$$

Lemma 3: Let X *be any set and let* Π_1 *and* Π_2 *be partitions of* X *such that* $A, B \in \Pi_1 \cup \Pi_2$ *implies* $|A| = |B|$. *Then for every permutation* φ *on* X *there is a partition* Π_φ *of* X *such that* (*i*) φ *covers* Π_2 *from* Π_φ, (*ii*) φ^{-1} *covers* Π_1 *from* $\varphi(\Pi_\varphi)$,

$$(iii) \quad |\Pi_\varphi| = |A|, \qquad A \in \Pi_1 \cup \Pi_2$$

$$|B| = |\Pi_1| = |\Pi_2|, \qquad B \in \Pi_\varphi.$$

Proof: Let φ be a permutation on X. The hypothesis implies that the union of any k elements of Π_1 has more elements than the union of any k-1 elements of Π_2, for $k = 1, \cdots, |\Pi_1|$. Hence by Lemma 2, with $X = Y$ and $m = n$, there is a set $D_1 \subseteq X$ such that

$$\varphi \text{ hits } \Pi_2 \text{ from } D_1$$

$$\varphi^{-1} \text{ hits } \Pi_1 \text{ from } \varphi(D_1).$$

Now we consider the sets $X_1 = X - D_1$ and $Y_1 = Y - \varphi(D_1)$ partitioned by

$${}_{X_1}\Pi_1 \quad \text{and} \quad {}_{Y_1}\Pi_2 \text{ respectively}$$

and we apply Lemma 2 to ${}_{X_1}\varphi$, i.e., to the restriction of φ to $X - D_1$. This gives a new set $D_2 \subseteq X$ such that again

$$\varphi \text{ hits } \Pi_2 \text{ from } D_2$$

$$\varphi^{-1} \text{ hits } \Pi_1 \text{ from } \varphi(D_2).$$

We proceed in this manner till X and Y are exhausted, and set $\Pi_\varphi = \{D_1, \cdots, D_n\}$, where $n = |A|$ for all $A \in \Pi_1 \cup \Pi_2$. It is clear that Π_φ has the stated properties.

Let φ be a permutation on X, and let Π_1, Π_2 be partitions of X.

Lemma 4: If φ covers Π_1 from Π_2, and $B \in \Pi_2$ implies $|B| = |\Pi_1|$, then $A \in \Pi_1$ implies $|A| = |\Pi_2|$.

Proof: Since φ covers Π_1 from Π_2, then $A \in \Pi_1$ and $B \in \Pi_2$ imply that there is an $x \in B$ with $\varphi^{-1}(x) \in A$. Thus $\varphi^{-1}(x) \in A$ for at least $|\Pi_2|$ distinct values of x, and so $|A| \geqq |\Pi_2|$. Since $B \in \Pi_2$ implies $|B| = |\Pi_1|$, it follows that $|\Pi_1|$ divides $|X|$ and

$$|\Pi_2| = \frac{|X|}{|\Pi_1|}.$$

Clearly

$$\sum_{A \in \Pi_1} |A| = |X|.$$

Since there are $|\Pi_1|$ sets in Π_1 each with at least $|\Pi_2|$ elements,

$$\sum_{A \in \Pi_1} |A| \geqq |\Pi_1| \cdot |\Pi_2| = |X|.$$

If any $A \in \Pi_1$ had more than $|\Pi_2|$ elements the sum would exceed $|X|$, which cannot be. Thus $A \in \Pi_1$ implies $|A| = |\Pi_2|$.

IX. GENERATING THE PERMUTATION GROUP

In this section we exhibit a sufficient condition on permutations $\varphi_1, \cdots, \varphi_{s-1}$ and stages $\mathcal{S}_1, \cdots, \mathcal{S}_s$ under which the complex

$$P(\mathcal{S}_1)\varphi_1 \cdots \varphi_{s-1}P(\mathcal{S}_s)$$

is actually the whole symmetric group, and the corresponding network (obtained by linking $\mathcal{S}_i$ and $\mathcal{S}_{i+1}$ by φ_i, $i = 1, \cdots, s-1$) is rearrangeable.

In order to focus on the mathematical character of the results, on their purely formal aspects divorced from physical considerations having to do with switches, etc., the conditions on the φ's and the $\mathcal{S}$'s purposely are phrased in a quite abstract way. Consequently, the practical implications and applications of the result may be unclear and require discussion. This discussion is given after the theorem has been stated, and is followed by its proof.

Theorem 2: Let $s > 3$ be an odd integer, let $\varphi_1, \cdots, \varphi_{s-1}$ be permutations on $X = \{1, \cdots, N\}$, let Π_k, $k = 1, \cdots, s$, and Π^k, $k = 1, \cdots, \frac{1}{2}(s-1)$, be partitions of X, and let

$$\Psi^k = \begin{cases} \{X\} & k = 0 \\ \varphi_k^{-1} \cdots \varphi_{\frac{1}{2}(s-1)}^{-1}(\Pi^k) & k = 1, \cdots, \frac{1}{2}(s-1) \\ \varphi_k \cdots \varphi_{\frac{1}{2}(s+1)}(\Pi^{s-k}) & k = \frac{1}{2}(s+1), \cdots, s-1, \\ \{X\} & k = s. \end{cases}$$

Suppose that

(*i*) *If* $s \geqq 3$, *then* $\Pi^k < \Pi^{k+1}$, $k = 0, \cdots, \frac{1}{2}(s-3)$.

(*ii*) $\Pi_{\frac{1}{2}(s+1)} = \Pi^{\frac{1}{2}(s-1)}$

(*iii*) *For* $k = 1, \cdots, \frac{1}{2}(s-1)$, *and every* $B \in \Psi^{k-1}$, φ_k^{-1} *B-covers* Π_k *from* φ_k *from* $\varphi_k(\Psi^k)$.

(*iv*) *For* $k = \frac{1}{2}(s+1), \cdots, s-1$, *and every* $B \in \Psi^{k+1}$, φ_k *B-covers* Π_{k+1} *from* $\varphi_k^{-1}(\Psi^k)$.

(*v*) *If* $A \in \Pi^k$ *and* $B \in \Psi^{k-1} \cup \Psi^{k+\frac{1}{2}(s+1)}$ *then* $|\, {}_B\Pi_k \,| = |\, {}_B\Pi_{s-k+1} \,| = |\, A \,|$, $k = 1, \cdots, \frac{1}{2}(s-1)$.

Let H_k, $k = 1, \cdots, s$ *be the largest strictly imprimitive subgroup of* $S(X)$ *whose sets of imprimitivity are exactly the elements of* Π_k (*i.e.*, $A \in \Pi_k$ *implies* $\{{}_A\varphi\colon \varphi \in H_k\} = S(A)$). *Then the complex* K *defined by*

$$K = H_1\varphi_1 H_2 \cdots H_{s-1}\varphi_{s-1}H_s$$

has the property $K = S(X)$, *and any network generating this complex is rearrangeable.*

The theorem given above does not provide any new designs of rearrangeable networks that are not already implicit in the work of M. C. Paull[3] and D. Slepian;[1] thus no new principle is involved. Rather, in formulating the result, we have sought insight by stating a generalized, purely combinatory form of these previous results. The theorem exhibits this generalization, first as providing a way of generating the symmetric group in a fixed number of multiplications of certain restricted group elements, and second as based on some purely abstract properties of some partitions and permutations.

As in A. M. Duguid's proof of the Slepian-Duguid theorem,[1] the basic combinatory theorem of P. Hall on distinct representatives of subsets is used repeatedly. This means (roughly) that the proof proceeds by showing that an arbitrary permutation (to be realized in the network) can be decomposed into submaps each of which can be realized in a disjoint part of the network, thereby not interfering with the realization of the other submaps. A significant departure from Ref. 3 is that we try to obtain rearrangeability directly from conditions that are stated for the network as a whole, as well as by building it up from rearrangeable subnetworks.

The following intuitive guides should be useful in understanding

Theorem 2. The permutations $\varphi_1, \cdots, \varphi_{s-1}$ are of course intended to be those corresponding to the link patterns between the stages of a network. The partition Π_k corresponds to the assignment of the terminals entering the kth stage to various square switches, all $u \in A$ for $A \in \Pi_k$ being on the same switch. The partitions Π^k are used in defining the submaps mentioned above.

The "covering" properties (*iii*) and (*iv*) of the φ_k in Theorem 2 ensure (roughly) that the φ_k are sufficiently mixing or distributive to be able to generate *all* permutations in the restricted ways permitted in the definition of K. They are generalizations of the property, exhibited in Fig. 2, that every middle switch is connected to every side switch by a link. The property (*v*), finally, implies that various sets of switches all have the same cardinality; this ensures (again, roughly) that if a crosspoint is not being used for a connection between one inlet-outlet pair, then it can be used for a connection between some other pair.

Proof of Theorem 2: We use induction on odd $s \geqq 3$. If $s = 3$, there is only one Π^k, viz. Π^1. Let φ be a permutation; we show that

$$\varphi \in H_1\varphi_1 H_2\varphi_2 H_3 .$$

The argument to be given is constructive, in that we do not use proof by contradiction, but actually give a kind of recipe for finding three permutations $\eta_i \in H_i$, $i = 1, 2, 3$, with

$$\varphi = \eta_1\varphi_1\eta_2\varphi_2\eta_3 .$$

To prove the theorem for $s = 3$, it is enough for $i = 1, 2, 3$ to exhibit a partition $\Pi(i)$ and to define η_i on $A \in \Pi(i)$, i.e., to give

$$_A\eta_i , \qquad A \in \Pi(i), \qquad i = 1, 2, 3.$$

Condition (*v*) for $k = 1$ ($= \frac{1}{2}(s - 1)$ here) tells us that for $A \in \Pi^1$

$$| \Pi_1 | = | \Pi_3 | = | A |.$$

However, the "middle-stage" condition (*ii*) states that $\Pi_2 = \Pi^1$. Hence $| \Pi_1 | \cdot | \Pi_2 | = N$. Since [condition (*iii*) now] φ_1^{-1} covers Π_1 from $\varphi_1(\Psi^1) = \varphi_1\varphi_1^{-1}(\Pi^1) = \Pi_2$, it follows that $B \in \Pi_1$ implies $| B | \geqq | \Pi_2 |$. If for some $B \in \Pi_1$ it was true that $| B | < | \Pi_2 |$, then

$$\sum_{B \in \Pi_1} | B | < | \Pi_1 | \cdot | \Pi_2 | = N$$

which is impossible. Thus $| B | = | \Pi_2 |$ for $B \in \Pi_1$. In exactly the same way, using condition (*iv*), we find that $C \in \Pi_3$ implies $| C | = | \Pi_2 |$. Therefore $B, C \in \Pi_1 \cup \Pi_3$ implies $| B | = | C |$.

Returning now to the chosen permutation φ, we apply Lemma 3 to conclude that there exists a partition Π_φ of X such that φ covers Π_3 from Π_φ, φ^{-1} covers Π_1 from $\varphi(\Pi_\varphi)$, $|\Pi_\varphi| = |A|$ for $A \in \Pi_1 \cup \Pi_3$, and $|B| = |\Pi_1| = |\Pi_3|$ for $B \in \Pi_\varphi$. Hence also $|\Pi_\varphi| = |\Pi_2|$.

Let $\mu: \Pi_\varphi \leftrightarrow \Pi_2$ be *any* map of Π_φ onto Π_2. The desired partitions $\Pi(i)$, $i = 1, 2, 3$ will be taken to be

$$\Pi(1) = \varphi_1(\Pi_2)$$

$$\Pi(2) = \{\mu(D): D \in \Pi_\varphi\} = \Pi_2$$

$$\Pi(3) = \Pi_\varphi$$

and the desired permutations η_i, $i = 1, 2, 3$, are defined so as to have these properties: for $D \in \Pi_\varphi$

$$\eta_3 : \varphi(D) \leftrightarrow \varphi_2^{-1}(\mu(D))$$

$$\eta_1 : \varphi_1(\mu(D)) \leftrightarrow D$$

$$\eta_2 : \varphi_2\eta_3(D) \leftrightarrow \varphi_1^{-1}\eta_1^{-1}\varphi(D).$$

That this can be done (uniquely, indeed) can be seen as follows: Let $D \in \Pi_\varphi$, and $\mu(D) = B \in \Pi_2$. Since φ^{-1} covers Π_1 from $\varphi(\Pi_\varphi)$, and φ covers Π_3 from Π_φ, it must be true that

(1) φ^{-1} hits Π_1 from $\varphi(D)$

(2) φ hits Π_3 from D.

But at the same time, by conditions (*iii*) and (*iv*), and the fact that $\Pi^1 = \Pi_2$, φ_1^{-1} covers Π_1 from Π_2 and φ_2 covers Π_3 from Π_2, and so

(3) φ_1^{-1} hits Π_1 from B

(4) φ_2 hits Π_3 from B.

Thus if $u \in D \cap A$ and $A \in \Pi_3$, there is a unique $v \in A$ such that $\varphi_2(v) \in B$, and we take $\eta_3(u) = v$. Similarly, if $z = \varphi(u) \in C$ and $C \in \Pi_1$, there is a unique $w \in C$ such that $\varphi_1^{-1}(w) \in B$, and we take $\eta_1(w) = z$. Finally, define η_2 so that $\eta_2(\varphi_2(v)) = \varphi_1^{-1}(w)$. Since $\mu(\cdot)$ is onto, each $D \in \Pi_\varphi$ deals with a unique $B = \mu(D) \in \Pi_2$, and the definition of η_i, $i = 1, 2, 3$ can be made for each D and its associated B, independently of the others. It is apparent that

$$\eta_1\varphi_1\eta_2\varphi_2\eta_3(D) = \varphi(D), \qquad D \in \Pi_\varphi,$$

or

$$\varphi = \eta_1\varphi_1\eta_2\varphi_2\eta_3.$$

Since η_i for $i = 1, 2, 3$ is onto, and is a subset of

$$\bigcup_{E \in \Pi_i} E \times E,$$

it follows that $\eta_i \in H_i$, $i = 1, 2, 3$, and thus that

$$K = S(X).$$

We now assume, as an hypothesis of induction, that the theorem is true for a given odd $s - 2 \geqq 3$, and that we are given permutations φ_k, and partitions $\{\Pi_k\}$ and $\{\Pi^k\}$, satisfying the conditions of the theorem.

No loss of generality is sustained if it is assumed that each $A \in \Pi^2$ is invariant under $\varphi_2, \cdots, \varphi_{s-2}$. This invariance can always be achieved by redefining the φ_i, without loss of properties (*iii*) to (*v*). It can now be seen that for $k = 2, \cdots, s - 2$ the restrictions

$$_A\Pi_k, \qquad _A\Pi^k, \qquad \text{with } A \in \Pi^2,$$

satisfy all the conditions on Π_k, Π^k (respectively) used in Theorem 2. Hence by the hypothesis of induction, for each $A \in \Pi^2$, the restriction of the complex

$$H_2\varphi_2 \cdots \varphi_{s-2}H_{s-1}$$

to A generates $S(A)$. The argument used for the case $s = 3$ can now be used to complete the induction, $\Psi^2 (= \Psi^{s-1}$ here) playing the role of Π_2.

X. CONSTRUCTION OF A CLASS OF REARRANGEABLE NETWORKS

We consider a network ν built of an odd number $s \geqq 3$ of stages,

$$\nu = \mathcal{S}_1\varphi_1 \cdots \varphi_{s-1}\mathcal{S}_{s-1},$$

satisfying the symmetry conditions

$$\left.\begin{aligned} \varphi_k &= \varphi_{s-k}^{-1} \\ \mathcal{S}_k &= \mathcal{S}_{s-k+1} \end{aligned}\right\} k = 1, \cdots, \tfrac{1}{2}(s - 1),$$

with each stage $\mathcal{S}_k$ made of identical square switches. The φ_k will be chosen in the following way: order the switches of each stage; to define φ_k for a given $1 \leqq k \leqq \frac{1}{2}(s - 1)$ take the first switch of $\mathcal{S}_k$, say with n outlets and n a divisor of N, and connect these outlets one to each of the first n switches of $\mathcal{S}_{k+1}$; go on to the second switch of $\mathcal{S}_k$ and connect its n outlets one to each of the *next* n switches of $\mathcal{S}_{k+1}$; when all the switches of $\mathcal{S}_{k+1}$ have one link on the inlet side, start again with the first switch; proceed cyclically in this way till all the outlets of $\mathcal{S}_k$ are assigned. (See Fig. 3.)

We shall show that a network ν constructed in this way is always rearrangeable.

Theorem 3: Let $s \geqq 3$ be an odd integer. Let $n_k, k = 1, \cdots, (s + 1)/2$, be any positive integers such that

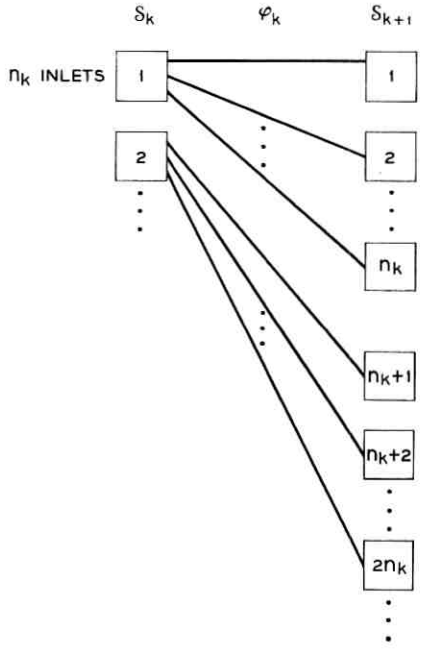


Fig. 3 — Assignment of inlets and outlets of rearrangeable network.

$$\prod_{k=1}^{\frac{1}{2}(s+1)} n_i = N \quad \text{and} \quad n_i \geqq 2.$$

For each $k = 1, \cdots, (s+1)/2$, *let* Π_k *be the partition of* $X = \{1, \cdots, N\}$ *into the* N/n_k *sets of the form*

$$A_{ki} = \{t: (i-1)n_k < t \leqq in_k\} \qquad i = 1, \cdots, N/n_k .$$

Let $\varphi_k, k = 1, \cdots, (s-1)/2$, *be permutations with the property that* $n \equiv t \pmod{N/n_{k+1}}$ *if and only if*

$$\varphi_k(n) \in A_{k+1,t} \qquad t = 0, \cdots, (N/n_{k+1}) - 1.$$

Define

$$\varphi_k = \varphi_{s-k}^{-1} \quad \text{for} \quad (s-1)/2 < k \leqq s-1.$$

Let S_k, $k = 1, \cdots, s$, *be the stages made of square switches defined by*

$$S_k = S_{s-k+1} \qquad k = 1, \cdots, (s-1)/2$$

$$S_k = \bigcup_{A \in \Pi_k} A \times A$$

and ν be the network constructed by putting a link pattern corresponding to φ_k, $k = 1, \cdots, s-1$ between stages $\mathcal{S}_k$ and $\mathcal{S}_{k+1}$. Then ν is rearrangeable.

Proof: It is readily seen that for $k = 1, \cdots, s$

$P(\mathcal{S}_k)$ = the largest strictly imprimitive subgroup with sets of imprimitivity Π_k,
= H_k,

in the notation of Theorem 2. Thus to prove the theorem by appeal to Theorem 2 it is enough to exhibit suitable partitions Π^k, $k = 1, \cdots, \frac{1}{2}(s-1)$, and to show that these, together with $\varphi_1, \cdots, \varphi_{s-1}$, satisfy the conditions of Theorem 2. For $k = 1, \cdots, \frac{1}{2}(s-1)$, set

$$\Pi^k = \text{class of all } \{j: (i-1)N/n_1 \cdots n_k < j \leqq iN/n_1 \cdots n_k, \quad i = 1, \cdots, n_1 n_2 \cdots n_k\}.$$

It is evident that the Π^k are successively finer partitions, i.e., that

$$\Pi^k < \Pi^{k+1} \qquad k = 1, \cdots, \tfrac{1}{2}(s-3).$$

Also, since $\Pi^{\frac{1}{2}(s-1)}$ consists of the $n_1 n_2 \cdots n_{\frac{1}{2}(s-1)}$ sets

$$\{j: (k-1)n_{\frac{1}{2}(s+1)} < j \leqq kn_{\frac{1}{2}(s+1)}, \ k = 1, \cdots, n_1 n_2 \cdots n_{\frac{1}{2}(s-1)}\}$$

it can be seen that

$$\Pi^{\frac{1}{2}(s-1)} = \{A_{\frac{1}{2}(s+1),i} : 1 \leqq i < N/n_{\frac{1}{2}(s+1)}\},$$

and hence that the middle stage condition (ii) in Theorem 2,

$$\Pi_{\frac{1}{2}(s+1)} = \Pi^{\frac{1}{2}(s-1)},$$

is satisfied.

The remainder of the proof, in which the requisite covering properties of the φ_k are demonstrated, is based on some auxiliary results.

Lemma 5: For $k = 2, \cdots, \frac{1}{2}(s-1)$, and $1 \leqq i \leqq N/n_n n_{n+1}$ the following identity holds

$$\bigcup_{\substack{t \equiv i-1 \\ \left(\mathrm{mod}\ \frac{N}{n_k n_{k+1}}\right)}} A_{k,t} = \varphi_k^{-1}\Big(\bigcup_{i-1 < \frac{t}{n_k} \leqq i} A_{k+1,t} \Big)$$

and the sets on the right are disjoint for different i.

Proof: Since

$$\varphi_k^{-1}(A_{k+1,t}) = \{n: n \equiv t \ (\mathrm{mod}\ N/n_{k+1}), \quad 1 \leqq n \leqq N\}$$

the union on the right in the lemma is the set of all n that are $\equiv t$ (mod

N/n_{k+1}) for some t with $(i-1)n_k < t \leqq in_k$. Consider such an n, with say

$$n = \frac{lN}{n_{k+1}} + t, \qquad \begin{matrix} i - 1 < \dfrac{t}{n_k} \leqq i \\ 0 \leqq l < n_{k+1}. \end{matrix}$$

Then

$$n = n_k \frac{lN}{n_k n_{k+1}} + (i-1)n_k + u,$$

with $0 < u \leqq n_k$, and so

$$n \in A_{k,(lN/(n_k n_{k+1}))+i-1)}$$

or

$$n \in A_{k,t} \quad \text{with} \quad t \equiv (i-1) \pmod{N/n_k n_{k+1}}.$$

Since the representation of n in terms in l and u is unique, the lemma follows.

The practical or physical import of the lemma is this: In any stage $k+1$, $1 \leqq k \leqq \frac{1}{2}(s-1)$, the ith block of n_k switches is connected by the link pattern φ_k to exactly those n_{k+1} switches (in the kth stage) whose number $t \equiv (i-1) \pmod{N/n_n n_{n+1}}$.

Definition: For $1 \leqq i \leqq n$, and $2 \leqq k \leqq m = \frac{1}{2}(s+1)$

$$B_{ik} = \bigcup_t \left\{ A_{k,t} : t \equiv r \pmod{n_1 n_2 \cdots n_{k-1}} \text{ for some } r \text{ with } (i-1) < \frac{r}{n_2 n_3 \cdots n_{k-1}} \leqq i \right\}$$

where $n_2 n_3 \cdots n_{k-1}$ is taken $= 1$ if $k = 2$.

Lemma 6: For $1 \leqq i \leqq n_1$ *and* $2 \leqq k < m$

$$B_{ik} = \varphi_k^{-1}(B_{i,k+1}).$$

Proof: We show that the right-hand side contains the left. Equality then follows from Lemma 5, since $B_{i,k+1}$ will always be a union of sets of the form

$$\bigcup_{j-1 < \frac{\tau}{n_k} \leqq j} A_{k+1,\tau}.$$

This is because $t \equiv r \pmod{u}$ if and only if $t + 1 \equiv (r + 1) \pmod{u}$, if $0 \leqq r < u$. Consider then an $n \in B_{ik}$. There is a $t \geqq 1$ congruent to an $r \pmod{n_1 n_2 \cdots n_{k-1}}$, with

$$i - 1 < \frac{r}{n_2 n_3 \cdots n_{k-1}} \leqq i,$$

such that $n \in A_{k,t}$. The latter fact implies that

$$(t - 1)n_k < n \leqq t n_k .$$

Now $\varphi_k(n) \in A_{k+1,\tau}$, where n is congruent to $\tau \pmod{N/n_{k+1}}$, so we can represent n in the form

$$n = \frac{aN}{n_{k+1}} + \tau.$$

Hence

$$(t - 1)n_k < \frac{aN}{n_{k+1}} + \tau \leqq t n_k .$$

Writing $t = l n_1 n_2 \cdots n_{k-1} + r$, with

$$(i - 1)n_2 n_3 \cdots n_{k-1} < r \leqq i n_2 n_3 \cdots n_{k-1},$$

we see that

$$q n_1 \cdots n_k + (r - 1)n_k < \tau \leqq q n_1 \cdots n_k + r n_k$$

where

$$q = l - \frac{aN}{n_1 n_2 \cdots n_{k+1}} .$$

It follows that τ is congruent $\pmod{n_1 n_2 \cdots n_k}$ to some integer p in the region

$$(i - 1)n_2 \cdots n_k < p \leqq i n_2 \cdots n_k ,$$

and thus $\varphi_k(n) \in B_{i,k+1}$, completing the proof of Lemma 6.

Now if $k = m$, the defining condition that $t \equiv r \pmod{n_1 \cdots n_{m-1}}$ for some r with

$$(i - 1)n_2 \cdots n_{m-1} < r \leqq i n_2 \cdots n_{m-1},$$

used in the definition of B_{ik}, can be put into a slightly different form. In this case we must have

$$1 \leqq t \leqq N/n_m = n_1 n_2 \cdots n_{k-1}$$

and so t can only be congruent to r by being equal to r, that is

$$(i - 1)n_2 \cdots n_{m-1} < t \leqq in_2 \cdots n_{m-1}.$$

Hence it can be seen that

$$\{B_{im}, i = 1, \cdots, n_1\} = \Pi^1.$$

Applying Lemma 6 $(m - 2)$ times we find that

$$\begin{aligned}\{B_{i2}, i = 1, \cdots, n_1\} &= \varphi_2^{-1} \cdots \varphi_m^{-1}(\Pi^1) \\ &= \varphi_1(\Psi^1).\end{aligned}$$

Now $n \equiv t \pmod{N/n_2}$ if and only if $\varphi_1(n) \in A_{2t}$, $t = 1, \cdots, N/n_2$. Also, by definition of B_{i2},

$$B_{i2} = \bigcup_{t \equiv i \pmod{n_1}} A_{2t}.$$

Let $\varphi_1(n) \in B_{i2}$, $\varphi_1(n) \in A_{2t}$. Then n has the form

$$n = \frac{aN}{n_1 n_2} n_1 + t = \left(\frac{aN}{n_1 n_2} + b\right) n_1 + i,$$

so $n \equiv i \pmod{n_1}$. Since $|B_{i2}| = |B_{im}| = n_2 \cdots n_m = N/n_1$, it follows that B_{12} is the φ_1 image of N/n_1 integers each of which is congruent to $i \pmod{n_1}$. Since each such integer must be in a different A_{1t}, it follows that φ_1^{-1} covers Π_1 from $\varphi_1(\Psi^1)$.

The remaining conditions in Theorem 3 can be demonstrated in essentially the same way; one has merely to identify the sets in question, and use Lemma 5 and an analog of Lemma 6. The details will be omitted.

REFERENCES

1. Beneš, V. E., On Rearrangeable Three-Stage Connecting Networks, B.S.T.J., **41**, 1962, pp. 1481–1492.
2. Beneš, V. E., Optimal Multistage Rearrangeable Connecting Networks, B.S.T.J., this issue, p. 1641.
3. Paull, M. C., Reswitching of Connection Networks, B.S.T.J., **41**, May, 1962, pp. 833–855.
4. Hall, M., Jr., *The Theory of Groups*, Macmillan, New York, 1959, p. 64.
5. *Ibid.*, p. 10.
6. Hall, P., On Representatives of Subsets, J. London Math. Soc., **10** (1935), pp. 26–30.

Optimal Rearrangeable Multistage Connecting Networks

By V. E. BENEŠ

(Manuscript received March 12, 1964)

A rearrangeable connecting network is one whose permitted states realize every assignment of inlets to outlets—that is, one in which it is possible to rearrange existing calls so as to put in any new call. In the effort to provide adequate telephone service with efficient networks it is of interest to be able to select rearrangeable networks (from suitable classes) having a minimum number of crosspoints. This problem is fully resolved for the class of connecting networks built of stages of identical square switches arranged symmetrically around a center stage: roughly, the optimal network should have as many stages as possible, with switches that are as small as possible, the largest switches being in the center stage; the cost (in crosspoints per inlet) of an optimal network of N inlets and N outlets is nearly twice the sum of the prime divisors of N, while the number of its stages is $2x - 1$, where x is the number of prime divisors of N, in each case counted according to their multiplicity. By using a large number of stages, these designs achieve a far greater combinatorial efficiency than has been attained heretofore.

I. INTRODUCTION

A study of rearrangeable connecting networks, begun in a previous paper,[1] is here continued; the object of the present work is to solve the synthesis problem of choosing, from a class of networks that are built of stages of square switches and satisfy some reasonable conditions on uniformity of switch size, a rearrangeable connecting network having a *minimum* number of crosspoints. Some of the terminology, notation, and results of Ref. 1 are used, and familiarity with it will be assumed from Section IV on.

Naturally, we do not pretend that minimizing the number of crosspoints (used to achieve a given end) is the only consideration relevant to the design of a connecting network. Other factors, like the number

of memory elements, the amount and placing of terminal equipment, the ease with which a network is controlled (e.g., the possibility of reliable end-marking), etc., may be of overriding significance, depending on the technology used. Still, it is important to know the limits of the region of possible designs, and these are obtained by optimizing on one variable without attention to others.

The problem of designing a good rearrangeable network was (probably first) considered in a paper of C. E. Shannon[2] investigating memory requirements in a telephone exchange. On the networks that he considered he imposed the realistic "separate memory condition" to the effect that in operation a separate part of the memory can be assigned to each call in progress. This means that completion of a new call or termination of an old call will not disturb the state of memory elements associated with any call in progress. Shannon showed that under this assumption a two-sided rearrangeable network, with N inlets and N outlets, and N a power of 2, requires at least

$$2N \log_2 N$$

memory elements (e.g., relays). He gave a design which actually realized this lower bound using

$$4(2^N - 1)\log_2 N$$

crosspoints (e.g., relay contacts). His design had the disadvantage of having very large numbers of contacts on certain relays. It is to be noted that Shannon was concerned with minimizing the number of memory elements, without regard to the number of crosspoints.

Shannon's separate memory condition is actually met by modern connecting networks that are of current practical interest, viz., by the networks made of stages of crossbar switches, considered here. For indeed, an inlet relay on an $n \times n$ crossbar switch is used to close any and each of n crosspoints: the exact one that closes depends on what outlet relay is simultaneously activated.

In this paper we consider the problem of minimizing the number of crosspoints in a network built of square switches, without attention to the number of relays. The following result (a consequence of Theorem 8) then complements Shannon's: For N a power of 2 it is possible to design a rearrangeable network with N inlets and N outlets using $4N \log_2 N - 6N$ relays and $4N(\log_2 N - 2)$ crosspoints. The figure for relays is roughly twice Shannon's while that for crosspoints is much smaller than his, for N large. In our design, no relay controls more than 4 contacts.

II. SUMMARY AND DISCUSSION

In Section III we discuss the notion of the combinatory power or efficiency of a connecting network, and propose to define it as the fraction r of permutations it can realize. According to this definition the four-stage No. 5 crossbar type of network with 10×10 switches has efficiency r close to zero, although it turns out that for the same number ($\approx$1000) of terminals there are networks that achieve $r = 1$ with a smaller number of crosspoints.[1] This greater efficiency is obtained by using many more stages than four.

Preliminaries are treated in Section IV. Particular attention is drawn to the class C_N of all two-sided networks having N inlets and N outlets, and built of stages of identical square switches symmetrically arranged around a center stage. The cost $c(\nu)$ of such a network ν is defined as the total number of crosspoints, divided by N. It is proposed to select rearrangeable networks ν from C_N that have minimal cost $c(\nu)$. This problem is attacked in Section V by defining (*i*) a map T from C_N to a special set A such that $c(\nu)$ is a function of $T(\nu) \in A$, and (*ii*) a partial ordering of A. It is then shown (Section VI) that (roughly) a network ν is optimal if and only if $T(\nu)$ is at the bottom of the partial ordering of A. This result allows one to identify (Section VII) the optimal networks in C_N. Their general characteristics are these: Except in some easily enumerated cases, the optimal network should have as many stages as possible, and switches that are as small as possible, the largest switches being in the middle stage; the cost $c(\nu)$ of an optimal network ν is very nearly twice the sum of the prime divisors of N, while the number of its stages is $2x - 1$, where x is the number of prime divisors of N.

Our chief conclusion is that by using many stages of small switches it is possible to design networks that are rearrangeable and cost less (in crosspoints per terminal) than networks in current use, which are far from being rearrangeable. The price paid for this great increase in combinatory power is the current difficulty of controlling networks of many stages. This difficulty is technological, though, and will decrease as improved circuits are developed.

III. THE COMBINATORY POWER OF A NETWORK

A principal reason why *rearrangeable* networks are of practical interest is (of course) that they can be operated as nonblocking networks. If the control unit of the connecting system using the rearrangeable network is made complex enough, it is in principle possible to rearrange calls in progress, repeatedly, in such a way that no call is ever blocked.

At present this possibility is being exploited in only a few special-purpose systems, because of the large amount of searching and data-processing it requires.

However, there is another reason why rearrangeable networks should evoke current interest. Even if we do not care to exploit it, the property of rearrangeability in a connecting network is an indication of its combinatory power or reach, and so can be used as a qualitative "figure of merit" for comparing networks. Other things being equal, a rearrangeable network is better than one which cannot realize all assignments of inlets to outlets. Rearrangeability expresses to some extent the efficiency with which crosspoints have been utilized in designing a connecting network for *distribution*, that is, for reaching many outlets from inlets.

If a numerical measure is called for, one can use the fraction of realizable maximal assignments. For a network ν with the same number N of inlets as outlets, and with inlets disjoint from outlets, this is just

$$r = \frac{\text{number of permutations realizable by } \nu}{N!}$$

$$= \text{combinatorial power of } \nu.$$

It is apparent that $0 \leqq r \leqq 1$, and that for a rearrangeable network $r = 1$. Also, r may be viewed as the chance that a permutation chosen at random will be realizable.

We shall calculate a bound on the combinatorial power r of the kind of connecting network most commonly found in modern telephone central offices. This is the network illustrated in Fig. 1. We choose the switch size $n = 10$ as a representative value; the network then has $N = 1000$ inlets, as many outlets, and 4×10^4 crosspoints. Clearly, the network can realize at most all the permutations that take exactly n terminals from each frame on the inlet side into each frame on the outlet side. Now a frame has n^2 inlets (outlets), and there are

$$\frac{n^2!}{(n!)^n}$$

ways of partitioning n^2 things into n groups of n each. Since there are $2n$ frames, there are

$$\left(\frac{n^2!}{(n!)^n}\right)^{2n}$$

ways of choosing n groups of n each on each frame, and assigning inlet groups to outlet groups (one-to-one and onto) in such a way that for

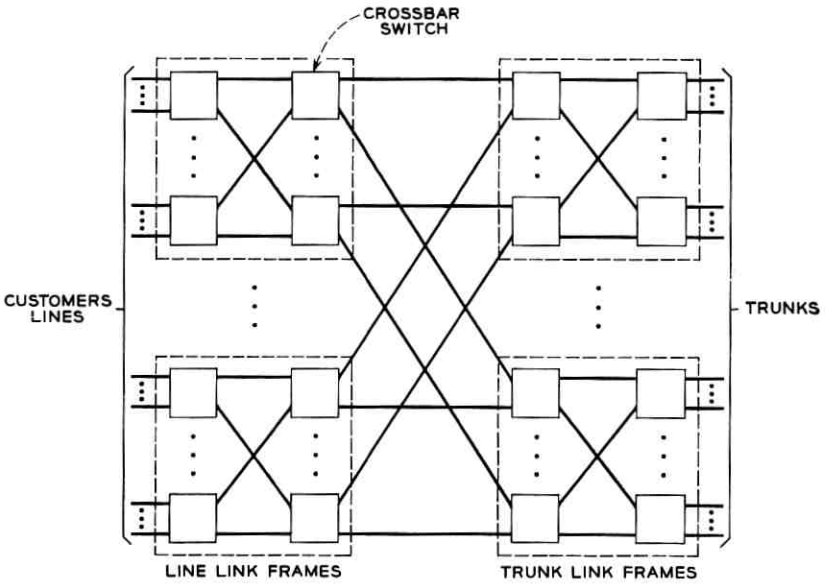


Fig. 1 — Structure of No. 5 crossbar network.

every inlet frame and every outlet frame exactly one group on the inlet frame is assigned to a group on the outlet frame. There are n^2 groups on a side (inlet or outlet), and within each group (at most) $n!$ permutations can be made, i.e., each inlet group can be mapped, terminal by terminal, in at most $n!$ ways onto its assigned outlet group. Hence at most

$$\left(\frac{(n^2!)}{(n!)^n}\right)^{2n} (n!)^{n^2}$$

permutations can be realized. There are $N = n^3$ terminals on a side, and a total of $n^3!$ possible permutations in all. Thus

$$r \leqq \frac{(n^2!)^{2n}}{n!^{n^2} n^3!} .$$

For $n = 10$, with

$$20 \log (100!) = 3159.4000$$

$$100 \log (10!) = 655.976$$

$$\tfrac{1}{2} \log 2\pi = 0.39959$$

$$\log (x!) \sim \tfrac{1}{2} \log 2\pi + (x + \tfrac{1}{2}) \log x - x \log_{10} e$$

we find roughly

$$r \leqq 10^{-64}.$$

Thus only a vanishingly small fraction of all possible permutations can actually be achieved by the No. 5 crossbar network (illustrated in Fig. 1) for $n = 10$, a reasonable switch size.

In the example calculated, the network has a "cost" of 40 crosspoints per terminal on a side. Much of the force of the example would be lost if it were in fact impossible to achieve high values of r (i.e., near 1) without incurring a great increase in the cost in crosspoints per terminal. This, however, is not the case. It follows from our Theorem 8 that a *rearrangeable* network ($r = 1$) can be designed for $N = 1024$ terminals on a side using only

$$4(\log_2 N - 2) = 32$$

crosspoints per terminal. Thus it is actually possible to *achieve* $r = 1$ for *more* than 1000 lines with *fewer* than 40 crosspoints per line. The network that does this turns out to have 17 stages instead of 4, an illustration of the way that allowing many stages can lead to vastly more combinatorially efficient network designs. The middle stage of this network consists of a column of 256 4×4 switches, and each of the other 16 stages, arranged symmetrically, consists of a column of 512 2×2 switches. For $k = 1, \cdots, 8$, the kth stage is connected to the $(k + 1)$th as follows: the first outlet of the first switch of stage k goes to the first switch of stage $(k + 1)$, the second outlet of the first switch of stage k goes to the second switch of stage $(k + 1)$, the first outlet of the second switch of stage k goes to the third switch of stage $(k + 1)$, etc., as in Fig. 2 with $1 \leqq k \leqq 7$; when each switch of stage $(k + 1)$ has 1 link on it the process starts over again with the first switch, and continues cyclically until all the links from stage k are assigned. The connections between stages k and $k + 1$ for $k = 9, \cdots, 17$ are the inverses of those for $k = 1, \cdots, 8$, so that the network is symmetric about the middle stage.[1]

IV. PRELIMINARIES

The symbol C_N, $N \geqq 2$, is used to denote the class of all connecting networks ν with the following properties:*

(1) ν is two-sided, with N terminals on each side

(2) ν is built of an odd number s of stages S_k, $k = 1, \cdots, s$, of

* Familiarity with Ref. 1 is assumed henceforth.

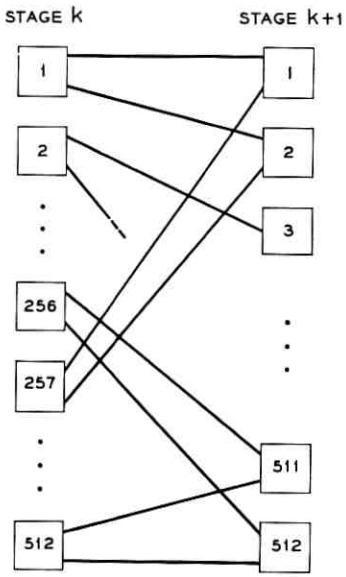


Fig. 2 — Link assignment.

square switches, i.e., there are permutations $\varphi_1, \cdots, \varphi_{s-1}$ such that

$$\nu = \mathcal{S}_1\varphi_1\mathcal{S}_2, \cdots, \varphi_{s-1}\mathcal{S}_s$$

(3) ν is *symmetric* in the sense that

$$\mathcal{S}_k = \mathcal{S}_{s-k+1} \qquad \text{for} \qquad k = 1, \cdots, \tfrac{1}{2}(s-1)$$

(4) With the notation

$$s = s(\nu) = \text{number of stages of } \nu$$

$$n_k = n_k(\nu) = \text{switch size in the } k\text{th stage of } \nu,$$

ν has N/n_k identical switches in stage k, i.e., each stage $\mathcal{S}_k$ is of the form

$$\bigcup_{A \in \Pi} A \times A$$

for some partition Π with $|A| = |B|$ for all $A, B \in \Pi$.

The defining conditions of C_N imply that

$$n_k = n_{s-k+1} \qquad \text{for} \qquad k = 1, \cdots, (s-1)/2$$

and that

$$\prod_{k=1}^{\frac{1}{2}(s+1)} n_k = N.$$

It is assumed throughout that $n_k(\nu) \geqq 2$ for all ν and all $k = 1, \cdots, s(\nu)$.

The *cost per terminal* (on a side) $c(\nu)$ of a network $\nu \in C_N$ is defined to be the total number of crosspoints of ν divided by the number N of terminals on a side. Since there are N/n_k $n_k \times n_k$ switches in stage k, the total number of crosspoints is (using the symmetry condition)

$$\begin{aligned}\sum_{k=1}^{s} (N/n_k) \cdot n_k^2 &= N \sum_{k=1}^{s} n_k \\ &= N\left(n_{\frac{1}{2}(s+1)} + 2 \sum_{k=1}^{\frac{1}{2}(s-1)} n_k\right)\end{aligned}$$

and so

$$c(\nu) = n_{\frac{1}{2}(s+1)} + 2 \sum_{k=1}^{\frac{1}{2}(s-1)} n_k .$$

A network ν is called *optimal* if

$$c(\nu) = \min \{c(\mu) : \mu \in C_N\}.$$

It is clear that the cost per terminal of a network $\nu \in C_N$ depends only on the switch sizes, and not at all on the permutations that define the link patterns between stages.

Also, it is apparent from Theorem 3 of Ref. 1 that given any network $\nu_1 \in C_N$ there is another network $\nu_2 \in C_N$ that is rearrangeable and differs from ν_1 only in the fixed permutations that are used to connect the stages; in particular, ν_1 and ν_2 have the same number of crosspoints. Thus the problem of selecting an optimal rearrangeable network from C_N is equivalent to that of choosing an optimal network from C_N, rearrangeable or not. A network in C_N can be made rearrangeable by changing its link patterns at no increase in cost.

We make

Definition 1: $m = m(\nu) = [s(\nu) + 1]/2$ = numerical index of the middle stage

$n = n(\nu) = n_{m(\nu)}$ = size of middle stage switches

Definition 2: $O(\nu) = \{n_1, \cdots, n_{m-1}\}$ = the set of switch sizes (with repetitions) in *outer* (i.e., nonmiddle) stages

Definition 3: $\omega(N) = \{O(\nu) : \nu \in C_N\}$.

Remark 1: $c(\nu) = n(\nu) + 2 \sum_{x \epsilon O(\nu)} x.$

Theorem 1: Let (A,n) *be a point (element) of*

$$\omega(N) \times X$$

with

$$n \prod_{y \epsilon A} y = N.$$

Then there exists a nonempty set $Y \subseteq C_N$ *such that*

$$T(\nu) = (A,n), \qquad \nu \epsilon Y.$$

The ν's in Y differ only in the permutations between the stages and in the placing of the outer stages, and at least one of them is rearrangeable. This result follows from the definition of C_N and from Theorem 2 of Ref. 1.

V. CONSTRUCTION OF THE BASIC PARTIAL ORDERING

The solution to the problem of synthesizing an optimal rearrangeable network from C_N will be accomplished as follows: we shall define a mapping T of C_N into $\omega(N) \times X$, with $X = \{1, \cdots, N\}$, and a *partial ordering* $\leqq$ of $T(C_N)$; the map T will have the property that $c(\nu)$ is a function of $T(\nu)$; then we shall prove that (roughly speaking) a network ν is optimal if and only if $T(\nu)$ is at the "bottom" of the partial ordering, i.e., that $c(\nu)$ is almost an isotone function of $T(\nu)$.

To define a partial ordering of a finite set, it is enough to specify consistently which elements *cover* which others. Let $Z, Z_0, Z_1, \cdots$ be sets of positive integers $\leqq N$ possibly containing repetitions.

Definition 4: Z_1 covers Z_2 if and only if there are positive integers j and k such that k occurs in Z_1, j divides k, and Z_2 is obtained from Z_1 by replacing an occurrence of k with one occurrence each of j and k/j.

Definition 5: $Z_0 \leqq Z$ if and only if there is an integer n and sets $Z_1, Z_2, \cdots, Z_n$ such that Z_{i+1} covers Z_i, $i = 0, 1, \cdots, n-1$ and $Z_n = Z$.

Definition 6: T: $\nu \rightarrow O(\nu), n(\nu)$.
A partial ordering $\leqq$ of $T(C_N)$ is defined by the following definition of covering:

Definition 7: Let μ, ν be elements of $C(N)$.
$T(\mu)$ covers $T(\nu)$ if and only if either

(*i*) $n(\nu) < n(\mu)$, $n(\nu)$ divides $n(\mu)$, and $O(\nu)$ results from $O(\mu)$ by adding an occurrence of $n(\mu)/n(\nu)$, or

(*ii*) $n(\nu) = n(\mu)$ and $O(\mu)$ covers $O(\nu)$.

VI. COST IS NEARLY ISOTONE ON $T(C_N)$

Theorem 2: If $T(\nu) \leqq T(\mu)$, *and* $n(\mu) > 6$, *then*

$$c(\nu) \leqq c(\mu).$$

Proof: It is enough to prove the result for μ and ν such that $T(\mu)$ covers $T(\nu)$.

Case (*i*): $n(\nu) < n(\mu)$, $n(\nu)$ divides $n(\mu)$, $n(\nu) \geqq 2$, and $O(\nu)$ results from $O(\mu)$ by adding an occurrence of $n(\mu)/n(\nu)$. Then

$$\begin{aligned} c(\nu) &= n(\nu) + 2 \sum_{x \epsilon O(\nu)} x \\ &= n(\nu) + \frac{2n(\mu)}{n(\nu)} + 2 \sum_{x \epsilon O(\mu)} x \\ &= c(\mu) + n(\nu) - n(\mu) + \frac{2n(\mu)}{n(\nu)} \\ &= c(\mu) + n(\nu)\left[1 - \frac{n(\mu)}{n(\nu)}\right] + \frac{2n(\mu)}{n(\nu)}. \end{aligned}$$

Thus $c(\nu) \leqq c(\mu)$ if and only if

$$n(\nu)\left(1 - \frac{n(\mu)}{n(\nu)}\right) + \frac{2n(\mu)}{n(\nu)} \leqq 0$$

that is, if

$$\frac{2y}{y-1} \leqq x$$

where $x = n(\nu)$ and $y = n(\mu)/n(\nu)$. Now $n(\mu) > 6$ implies that either

(*i*) $n(\nu) = 2$ and $\dfrac{n(\mu)}{n(\nu)} \geqq 4$

or

(*ii*) $n(\nu) = 3$ and $\dfrac{n(\mu)}{n(\nu)} \geqq 3$

or

(*iii*) $n(\nu) \geqq 3$.

The condition $2y/(y-1) \leqq x$ is fulfilled in all three cases, and so $c(\nu) \leqq c(\mu)$.

Case (*ii*): $n(\mu) = n(\nu)$ and $O(\mu)$ covers $O(\nu)$. There exist integers j,k such that j divides k, $j \geqq 2$ in $O(\mu)$, and $O(\nu)$ results from $O(\mu)$ by replacing one occurrence of k with one each of j and k/j. Then

$$\begin{aligned} c(\nu) &= n(\nu) + 2 \sum_{x \epsilon O(\nu)} x \\ &= n(\mu) - 2k + 2j + (2k/j) + 2 \sum_{x \epsilon O(\mu)} x \\ &= c(\mu) - 2k + 2j + (2k/j). \end{aligned}$$

Since j divides k and $j \geqq 2$, $k \geqq 2j$ and $k \geqq 2k/j$, so

$$k \geqq 2 \max\left(j, \frac{k}{j}\right) > j + \frac{k}{j}$$

and $c(\nu) < c(\mu)$.

Theorem 3: If $\nu \epsilon C_N$ and $O(\nu)$ does not consist entirely of prime numbers (possibly repeated), then there exists a network μ in C_N of $s(\nu) + 2$ stages with $c(\mu) < c(\nu)$, and ν cannot be optimal in C_N.

Proof: There is a value of k in the range $1 \leqq k \leqq n(\nu) - 1$ for which n_k is not a prime, say $n_k = ab$. Define stages $\mathcal{S}_j(\mu)$, $j = 1, \cdots, s(\nu) + 2$ as follows:

$$\mathcal{S}_{j+1}(\mu) = \mathcal{S}_j(\nu), \qquad j = k+1, \cdots, n(\nu);$$

let Π_a, Π_b be partitions of $X = \{1, \cdots, N\}$ with

$$|\Pi_a| = N/a \quad \text{and} \quad A \epsilon \Pi_a \Rightarrow |A| = a$$
$$|\Pi_b| = N/b \quad \text{and} \quad B \epsilon \Pi_b \Rightarrow |B| = b.$$

Set

$$\begin{aligned} \mathcal{S}_{k+1}(\mu) &= \bigcup_{A \epsilon \Pi_a} A^2 \\ \mathcal{S}_k(\mu) &= \bigcup_{B \epsilon \Pi_b} B^2 \\ \mathcal{S}_j(\mu) &= \mathcal{S}_j(\nu) && j = 1, \cdots, k-1 \\ \mathcal{S}_j(\mu) &= \mathcal{S}_{s(\nu)-j+1}(\mu) && \text{all } j = 1, \cdots, s(\nu) + 2 \end{aligned}$$

By Theorem 2 of Ref. 1 permutations $\varphi_1, \cdots, \varphi_{s(\nu)-1}$ can be found so that the network

$$\mu = \mathcal{S}_1\varphi_1, \cdots, \varphi_{s(\mu)-1}\mathcal{S}_{s(\mu)}$$

is in C_N and is rearrangeable. It is apparent that $n_{m(\mu)} = n_{m(\nu)}$ and that $O(\nu)$ covers $O(\mu)$. Hence the argument for case (ii) of Theorem 2 shows that μ has strictly lower cost than ν.

Corollary 1: If $N > 6$ and is not prime, then a network ν consisting of one square switch is not optimal.

VII. PRINCIPAL RESULTS

Definition 8: An element $T(\nu)$ of $T(C_N)$ is *ultimate* if there are no $\mu \in C_N$ such that $T(\nu)$ covers $T(\mu)$.

Remark 2: $T(\nu)$ is ultimate if and only if $n(\nu)$ is prime and $O(\nu)$ consists entirely of prime numbers.

Definition 9: An element $T(\nu)$ of $T(C_N)$ is *penultimate* if it covers an ultimate element.

Definition 10: p_n, $n = 1, 2, \cdots$, is the nth prime.

Definition 11: $\pi(n)$ is the prime decomposition of n, that is, the set of numbers (with repetitions) such that

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_l^{\alpha_l}$$

if and only if $\pi(n)$ contains exactly α_i occurrences of p_i, $i = 1, \cdots, l$, and nothing else.

Definition 12: p is the largest prime factor of N.

Lemma 1: If $p = 3$ and $N > 6$ is even, then the following conditions are equivalent:

(i) *ν is optimal*
(ii) *$T(\nu)$ is penultimate and $n(\nu) = 6$ or 4*
(iii) *$T(\nu) = (\pi(N/)6,)6$ or $(\pi(N/4),4)$.*

Proof: By Theorems 2, 3 only ν with $n(\nu) \leqq 6$ and $O(\nu)$ consisting entirely of primes can be optimal. Writing $N = 2^x 3^y$ with $x \geqq 1$ and $y \geqq 1$, it is seen that such ν must have a cost $c(\nu)$ having one of the forms

$$2 + 2[2(x - 1) + 3y] = 4x + 6y - 2,$$

$$3 + 2[2x + 3(y - 1)] = 4x + 6y - 3,$$

$$4 + 2[2(x - 2) + 3y] = 4x + 6y - 4 \quad \text{(only occurs if } x > 1),$$

$$6 + 2[2(x - 1) + 3(y - 1)] = 4x + 6y - 4.$$

The least of these is either of the last two, which correspond to $n(\nu) = 6$ if $x = 1$ or to $n(\nu) = 6$ or 4 if $x > 1$. It is apparent that (*ii*) is equivalent to (*iii*).

Lemma 2: If $p = 2$, and $N > 4$, then the following conditions are equivalent:

(*i*) *ν is optimal*
(*ii*) *$T(\nu)$ is penultimate and $n(\nu) = 4$*
(*iii*) $T(\nu) = (\pi(N/4), 4)$.

Proof: With $N = 2^r$ it can be seen as in Lemma 1 that only those ν can be optimal whose cost $c(\nu)$ has one of the forms

$$2 + 2[2(x - 1)],$$

$$4 + 2[2(x - 2)].$$

The second of these is the better, and corresponds to $n(\nu) = 4$.

Theorem 4: Let μ be a network such that a prime number $r > n(\mu)$ occurs in $O(\mu)$. Let M result from $O(\mu)$ by replacing one occurrence of r by $n(\mu)$. Then for any network ν with

$$T(\nu) = (M,r)$$

it is true that

$$c(\nu) < c(\mu)$$

i.e., ν is strictly better than μ. Among such ν, that is best for which r is largest.

Proof: Existence of a rearrangeable ν satisfying $T(\nu) = (M,r)$ is guaranteed by Theorem 1. For the rest of the proof, we observe that $r > n(\mu)$ and

$$\begin{aligned} c(\mu) &= n(\mu) + 2 \sum_{x \epsilon O(\mu)} x \\ &= n(\mu) + 2r - 2n(\mu) + 2 \sum_{x \epsilon M} x \\ &= r - n(\mu) + c(\nu). \end{aligned}$$

Theorem 5: If $n(\mu) \leqq 6$, $n(\mu) = 2^x 3^y 5^z$, some prime number $r > 3$ occurs in $0(\mu)$, and if M results from $0(\mu)$ by replacing one occurrence of r by x occurrences of 2, y occurrences of 3, and z occurrences of 5 then for any network $\nu \epsilon C_N$ with

$$T(\nu) = (M,r)$$

it is true that

$$c(\nu) \leqq c(\mu)$$

i.e., ν is at least as good as μ. Among such ν, that is best for which r is largest.

Proof: Existence of a rearrangeable $\nu \in C_N$ satisfying $T(\nu) = (M,r)$ is given by Theorem 1. For the rest of the proof, we observe that $r \geqq 5$ and

$$\begin{aligned} c(\mu) &= n(\mu) + 2 \sum_{u \in O(\mu)} u \\ &= n(\mu) + 2r - 4x - 6y - 10z + 2 \sum_{u \in M} u \\ &= r + n(\mu) - 4x - 6y - 10z + c(\nu). \end{aligned}$$

Since x, y, and z can only assume the values 0 and 1, with $z = 1$ if and only if $x = y = 0$, we have $c(\mu) \geqq c(\nu)$, the best ν corresponding to the largest r.

Definition 13: $Q = \{(A,r) : r \text{ a prime and } A = \pi(N/r)\}$.

Definition 14: $L = T^{-1}(Q)$.

Remark 2: Q consists of all the absolute minima in the partial ordering $\leqq$ of $T(C_N)$, i.e., $\nu \in L$ implies that there are no $\mu \in C_N$ for which

$$T(\mu) < T(\nu).$$

Theorem 6: If $p > 3$, then all optimal networks belong to L.

Proof: Let $\mu \in C_N - L$ be given. We show that there exists a $\nu \in L$ that is at least as good.

Case 1: There is a sequence $\mu = \mu_1, \mu_2, \cdots, \mu_n, \nu$ with $\mu_n \neq \nu$, $\nu \in L$, $n(\mu_n) > 6$,

$$T(\mu_1) \geqq T(\mu_2) \geqq \cdots \geqq T(\mu_n)$$

and such that $T(\mu_n)$ covers $T(\nu)$. Then the numbers $n(\mu_j)$, $j = 1, \cdots, n$ are all > 6, and the result follows from Theorem 2.

Case 2: All sequences $\mu = \mu_1, \mu_2, \cdots, \mu_n, \nu$ with $\mu_n \neq \nu$, $\nu \in L$, $T(\mu_1) \geqq T(\mu_2) \geqq \cdots \geqq T(\mu_n)$, and such that $T(\mu_n)$ covers $T(\nu)$, are such that $n(\mu_n) \leqq 6$. Consider such a sequence. Let i be the smallest index j for which $n(\mu_j) \leqq 6, j = 1, \cdots, n$. Then Theorem 2 gives $c(\mu) \geqq c(\mu_i)$. Since $n(\mu_i) \leqq 6$ and $T(\mu_n)$ covers $T(\nu)$, it follows that $O(\mu_i)$ contains

an occurrence of $p > 3$. Hence by Theorem 5 there exists a network $\eta \epsilon C_N$ with $n(\eta) = p$ and

$$c(\eta) \leqq c(\mu_i) \leqq c(\mu).$$

Let $\xi \epsilon L$ be such that $n(\xi) = p$ and $T(\eta)$ covers $T(\xi)$. Then $c(\xi) \leqq c(\eta)$ by case (ii) of Theorem 2. Hence

$$c(\xi) \leqq c(\mu)$$

$$\xi \epsilon L.$$

Theorem 7: If $N \leqq 6$ and ν is optimal, then ν is a square switch and $c(\nu) = N$.

Proof: For prime N with $2 \leqq N < 6$ the result is obvious. If $N = 6$ and $\nu \epsilon C_6$ then exactly one of the following alternatives obtains:

$$T(\nu) = (\theta, 6) \quad \text{and} \quad c(\nu) = 6$$

$$T(\nu) = (\{3\}, 2) \quad \text{and} \quad c(\nu) = 8$$

$$T(\nu) = (\{2\}, 3) \quad \text{and} \quad c(\nu) = 7.$$

The first alternative is optimal, and there is exactly one $\nu \epsilon C_6$ such that $T(\nu) = (\theta, 6)$, viz., the 6×6 square switch. Similarly, if $N = 4$ and $\nu \epsilon C_4$, then $T(\nu) = (\theta, 4)$ or $(\{2\}, 2)$; the former has cost 4, the latter 6.

Definition 15: For $n \geqq 2$, $D(n)$ is the sum of the prime divisors of n counted according to their multiplicity; thus if

$$n = 2^{\alpha_1} 3^{\alpha_2} \cdots p_k^{\alpha_k}$$

then

$$D(n) = \sum_{j=1}^{k} p_j \alpha_j = \sum_{x \epsilon \pi(n)} x.$$

Definition 16: $c(N) = \min \{c(\nu) : \nu \epsilon C_N\}$.

Theorem 8:

$$c(N) = \begin{cases} N \text{ if } N \leqq 6 & \text{or } N \text{ is prime} \\ p + 2D(N/p) & \text{if } N > 6 \text{ and either } p > 3 \text{ or } N \text{ is odd} \\ 2D(N/2) & \text{if } N > 6 \text{ in all other cases (i.e., } p = 2, \text{ or } \\ & \quad p = 3 \text{ and } N \text{ is even}). \end{cases}$$

Proof: Putting together Lemmas 1, 2 and Theorems 1, 2, 3, 4, 6, and 7

we obtain the following values for the minimal cost in crosspoints per terminal on a side for networks in C_N:

$$c(N) = \begin{cases} N \text{ if } N \leqq 6 \text{ or } N \text{ is prime} \\ p + 2 \sum_{x \epsilon \pi(N/p)} & \text{if } p > 3, N > 6 \\ 6 + 2 \sum_{x \epsilon \pi(N/6)} x = 2 \sum_{x \epsilon \pi(N/2)} x & \text{if } p = 3, N > 6, N \text{ even} \\ 3 + 2 \sum_{x \epsilon \pi(N/3)} x = 3 + 6(\log_3 N - 1) & \\ & \text{if } p = 3, N > 6, N \text{ odd} \\ 4 + 2 \sum_{x \epsilon \pi(N/4)} x = 4(\log_2 N - 2) = 2 \sum_{x \epsilon \pi(/2)} x & \\ & \text{if } p = 2, N > 6; \end{cases}$$

simplification gives Theorem 8.

REFERENCES

1. Beneš, V. E., Permutation Groups, Complexes, and Rearrangeable Connecting Networks, B.S.T.J., this issue, p. 1619.
2. Shannon, C. E., Memory Requirements in a Telephone Exchange, B.S.T.J., **29,** July, 1950, pp. 343–349.

Attitude Determination and Prediction of Spin-Stabilized Satellites

By L. C. THOMAS and J. O. CAPPELLARI

(Manuscript received February 17, 1964)

Techniques for both attitude determination and prediction for spin-stabilized satellites are developed. Their use is demonstrated using Telstar I and II satellite data. It is shown that an inclined dipole model of the earth's magnetic field and the method of averaging the gravitational and magnetic torques over each anomalistic period of the satellite permits attitude predictions to within a few tenths of a degree of determined values in most instances. In those few cases where departures are above one degree, explanations are presented to show the reason for such discrepancies.

The usefulness of combining optical flash and solar sensor data for attitude determination and their inherent accuracy are demonstrated. Optical flash data can provide loci with a resolution of 0.1°. Solar sensor loci are resolved to within 1°.

All techniques described have been consolidated into working computer programs which follow closely the mathematical analysis presented. A number of important supporting calculations such as the solar position, sidereal time, orbit updating, etc. are also developed. Because of the complexities of the mean torque and gyroscopic equations, the precessional techniques presented are most useful in computer embodiments.

TABLE OF CONTENTS

I. INTRODUCTION

To maintain a defined attitude in space, the Telstar I and II satellites were spin stabilized. By this method of passive attitude control, a satellite is rotated about an axis of symmetry and consequently exhibits the characteristics of a gyroscope. In the absence of disturbing torques, the satellite's spin axis maintains its spatial orientation fixed with respect to an inertial reference frame throughout its orbit. For the Telstar I and II satellites, this is desirable because of certain required attitude constraints. First of all, the satellite communication antenna is not omnidirectional. More energy is radiated along the equator of the satellite than along its spin poles, as shown by the antenna pattern of Fig. 1. This fact dictates an attitude for which the line of sight from

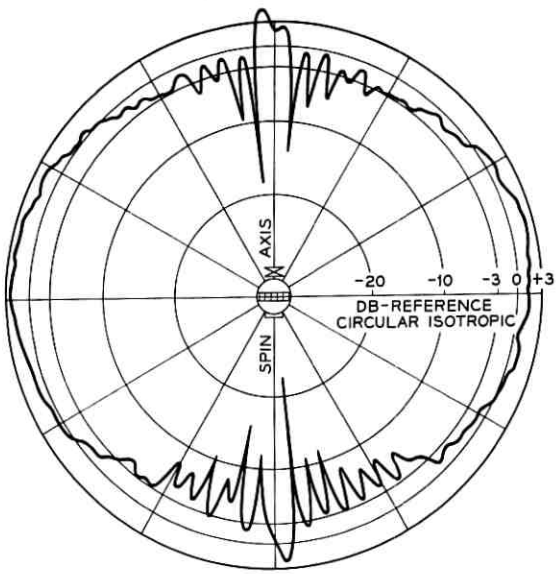


Fig. 1 — Antenna pattern at 4170 mc.

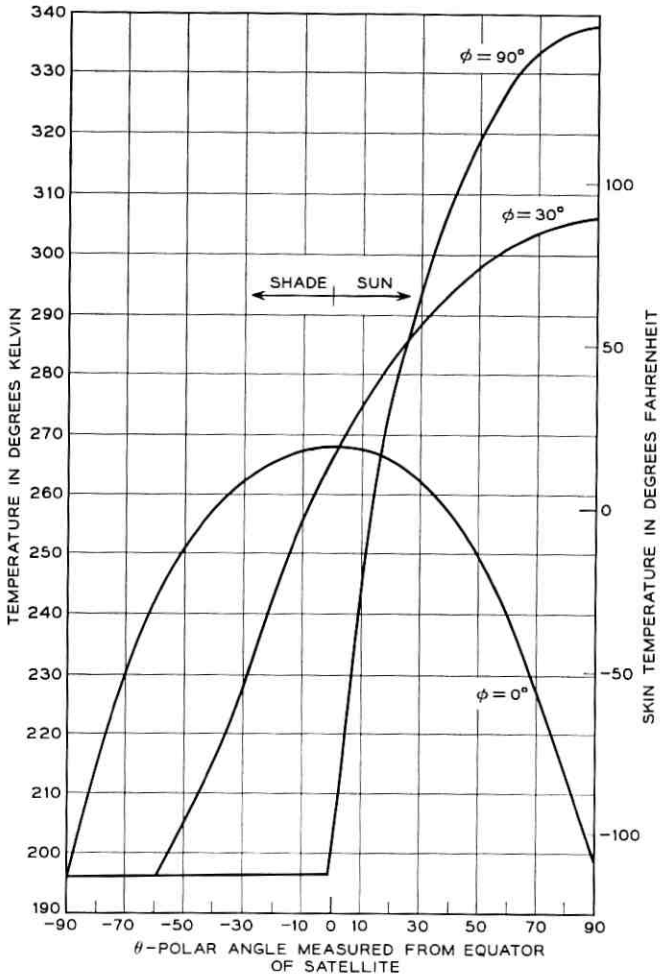


Fig. 2 — Skin temperature distribution vs polar angle.

a ground station to the satellite avoids its poles. The second attitude constraint involves temperature considerations. A good degree of temperature control is obtained by orienting the spin axis so that it is nearly perpendicular to the satellite-sun line. In this manner, temperature balance is maintained by the satellite's spin as indicated in Fig. 2. Here, φ is the solar offset angle, defined as the angular departure of the spin axis from perpendicularity with the satellite-sun line. Solar offsets of about 15° result in temperature deviations of about 150° (see Fig. 2) and are tolerable.

An ideal orientation from a communications standpoint would be to have the spin axis nearly parallel to the earth's surface as it passes over any ground station. A spin axis perpendicular to the orbital plane would accomplish this, but would produce a maximum axis tilt toward the sun equal to the sum of the orbital inclination and the earth's $23\frac{1}{2}°$ tilt with respect to the plane of the ecliptic (see Fig. 3). Under these conditions the spin axis of the first Telstar satellite would have a maximum solar tilt of 68°. This exceeds the 15° tilt limit dictated by temperature

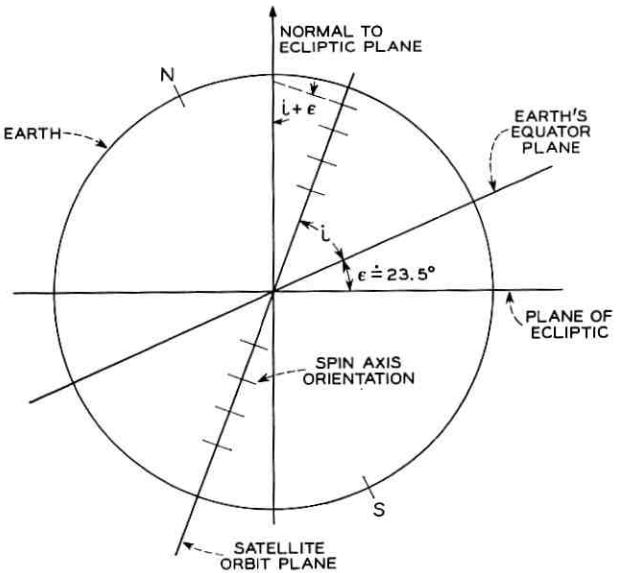


Fig. 3 — Geometry of spin axis perpendicular to orbit plane.

balance. On the other hand, if the spin axis is made perpendicular to the ecliptic plane (which automatically insures its perpendicularity to the sun), the axis tilt with respect to the line of sight from any ground station over which it may be passing will range from 90° to 90° minus the sum of the orbit inclination and ecliptic obliquity or 22° (see Fig. 4).* From a communications standpoint this is tolerable, since nulls in the antenna pattern are major only within about 15° of the spin

* With this orientation, stations south of the satellite's instantaneous earth latitude would experience angles less than 22°, such as station B of Fig. 4. The major ground stations for Telstar I and II, however, are all at latitudes above 43°. Since the inclinations of the Telstar I and II satellites are 45° and 42.7° respectively, these stations are almost always north of the satellite.

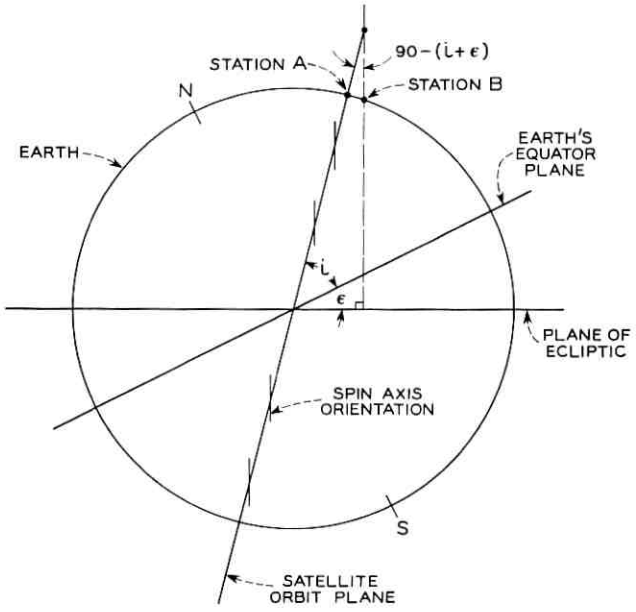


Fig. 4 — Geometry of spin axis perpendicular to ecliptic.

poles (see Fig. 1). Rising and setting satellites may experience a greater tilt to the ground station line of sight, but this is unavoidable.

Because of these considerations, the Telstar I and II satellites were launched with their spin axes as nearly perpendicular to the ecliptic as the powered flight trajectory of the Thor-Delta launch vehicle would permit and still meet certain orbital requirements, such as inclination, apogee height, and perigee height.[1] The predicted attitudes of these satellites at orbit injection are calculated from telemetry data from the first and second rocket stages. Table I lists these attitudes as the right ascension and declination direction of the north pole of the satellite.

TABLE I — ATTITUDE OF THE TELSTAR I AND II SATELLITES AT INJECTION, DETERMINED FROM POWERED FLIGHT DATA

Satellite	Initial Attitude	
	Right Ascension	Declination
Telstar I	83.73°	−66.80°
Telstar II	82.23°	−57.31°

The north pole is defined here as the direction of advance of a right-handed screw turning in the direction of the satellite's spin. It also happens to be the spin pole which carries a helical telemetry antenna.

The initial attitudes of the Telstar I and II satellites are changed by gyroscopic precession. This motion of spin-stabilized satellites is chiefly produced by both magnetic and gravitational torques. The former is a result of interaction between residual and eddy-current-produced magnetic fields of the satellite and the earth's magnetic field. The latter is produced by differential gravitational forces acting across the body of the satellite. In the present cases, the magnetic torques are several orders of magnitude in excess of the gravitational torques.

It is necessary to predict precession to allow proper scheduling and planning for satellite use, to resolve certain attitude determination ambiguities, and to sensibly plan alteration of satellite attitude in a prescribed manner when needed. (See Section V for a description of this technique.) It is the purpose of this paper to outline the methods of attitude determination and correction, to develop the precessional theory, and to show the application of these to the Telstar I and II satellites.

II. ATTITUDE DETERMINATION — GENERAL REMARKS

The attitude of the Telstar I satellite has been determined through the analysis of two sets of data: the time of optical flashes of sunlight from three mirrors attached to the surface of Telstar and the current produced by six on-board solar sensors located on the ends of three orthogonal axes.

The first of these sets of data in combination with the spatial position of the satellite, sun, and observer's position determines a locus of possible spin axis positions which would result in the observed flash. This locus describes a cone in space whose axis is the mirror normal. The six solar sensors are designed to determine the direction of the satellite-sun line with respect to a satellite frame of reference and thus to the spin axis. Knowing this angle, again there is defined a conical array of possible spatial spin axis directions. The intersection of the optical cone with this solar sensor cone should, therefore, determine two possible attitudes. A priori knowledge of the approximate attitude as provided by launch data and/or a succession of measurements and predictions over an interval of several days permits the determination of a unique spin axis direction in space.

In practice, the solar sensors present a few difficulties. First of all, the deduction of the sun's position from solar cell current entails inferring

a solar angle between each cell's normal and the sun's direction. Such a direction from each of three cells uniquely establishes this solar aspect.* Thus calibration curves relating electrical output to light intensity must be employed for each cell along with temperature corrections. While D. W. Hill[2] has prepared a computer program to take the drudgery out of this work, it remains difficult to calculate the solar angle for a cell if it is illuminated in a direction far from its normal. Moreover, the solar sensor data, which are reported every minute by telemetry, were found to be not always mutually consistent. Often, the solar direction calculated over 30 minutes from a succession of telemetry frames had a spread ranging from 1° to over 8°. A correlation between these deviations and the spatial position of the satellite exists which suggests a biasing of solar cell data, on occasion, by the earth's reflectivity.

To show the geometry of this situation, consider the satellite position shown in Fig. 5. Here one half of the satellite to the right of line AB is

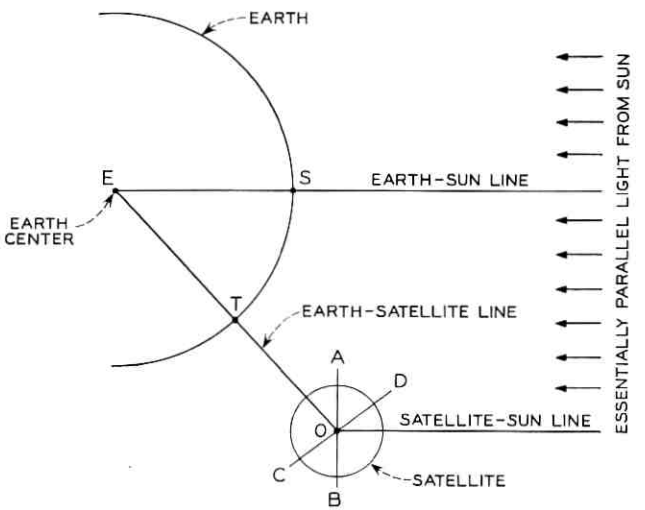


Fig. 5 — Satellite shadow geometry.

illuminated by the sun. At least one half, to the upper left of line CD, is illuminated by the earth in reflecting about 38 per cent† of the sun's total incident light. Thus a solar cell in region BOD would record only direct sunlight, one in region AOC would record only earthshine, and

* The solar aspect is the angle between the satellite's spin axis and the satellite-sun line.

† This is not the albedo but the earth's mean reflectivity or ratio of mean earth brightness measured at a spot along the earth-sun line relative to the brightness of a perfectly diffusing disk of the same size and at the same distance that the

one in region AOD would record both. There would be no light from either the sun or the earth in region COB. By the difference in solar illumination and earthshine, data from cells in regions BOD and AOC are easily separated and the latter disregarded. However, those cells in region AOD measure both sunlight and earthshine in an inseparable manner and therefore report erroneously the sun's position. If AOD is a large angle, the spin of the satellite will carry solar sensors into that region frequently and result in sizable variations in successive solar aspect determinations. Since AOD equals SET, as the subsatellite position (T) departs from the subsolar point (S), the solar cell data must be carefully interpreted to avoid false conclusions. Under these conditions, it would be wrong, for example, to simply calculate the mean of all solar aspect determinations over a succession of telemetry frames. This procedure, in general, would not yield a good estimate of the true aspect. A better technique would be to calculate the mean for only those solar aspect determinations in which all three cells entering into the determination exhibited current readings above those which could possibly be produced by either the earth's reflectivity or low angles of solar illumination. Operating in this fashion, the true solar aspect can be determined to within about 1°.

Since the maximum attitude resolution obtainable using optical flashes from the Telstar satellites is about 0.1°, it was decided to rely on these for attitude determinations insofar as possible. Two groups of flashes close together in time are needed for an attitude fix, however, and in cases where only a single one existed, the attitude was determined by a combination of mirror flashes and the solar sensor data previously described.

The optical reflections are characterized by a series of intermittent flashes provided by the spin of the Telstar satellites. The time midpoint of these flash series is determined by photoelectric equipment[3] at Bell Telephone Laboratories in Holmdel, N. J. The time of each flash series determines a conical locus of possible spin axis positions about the flashing mirror's normal. The tip of the spin axis vector, therefore, lies on a circle on the celestial sphere. One of the intersections of two such circles defines the attitude, provided the two corresponding flash series occur close enough together in time so that no appreciable precession occurs during the separation interval. Since the mirrors employed have their normals far removed from the spin axis (68° and 95°), it is

earth is from the given spot. Quite naturally a phase law applies which reduces the light reflected to points off the earth-sun line. Off-line brightness varies crudely as the ratio of observable illuminated area of the earth as seen from the point in question to the total observable illuminated area seen from a point on the earth-sun line.

always possible to select the valid locus intersection from past performance and attitude predictions.

III. ATTITUDE DETERMINATION — COMPUTATIONAL TECHNIQUE

A computational technique has been developed for the calculation of spin loci. It consists of four basic parts, which are:

(*i*) the determination of the right ascension and declination of the normal of the flashing mirror;

(*ii*) the determination of the right ascension and declination of the sun in the case of solar cell data;

(*iii*) the construction of a circular locus on the celestial sphere centered on the right ascension and declination of the mirror normal and having a radius of 68° or 95°, depending upon the mirror involved (in the case of solar aspect data from the sensors, the locus is centered on the right ascension and declination of the sun and has a radius equal to the measured solar aspect angle);

(*iv*) the plotting, by computer microfilm techniques, of these circles and others determined from additional flash series and solar aspect data to ascertain intersections and corresponding satellite attitudes.

3.1 *The Right Ascension and Declination of the Mirror Normal*

Consider the generalized mirror orientation shown in Fig. 6. Here, a ray from the sun strikes the mirror and is reflected toward a particular

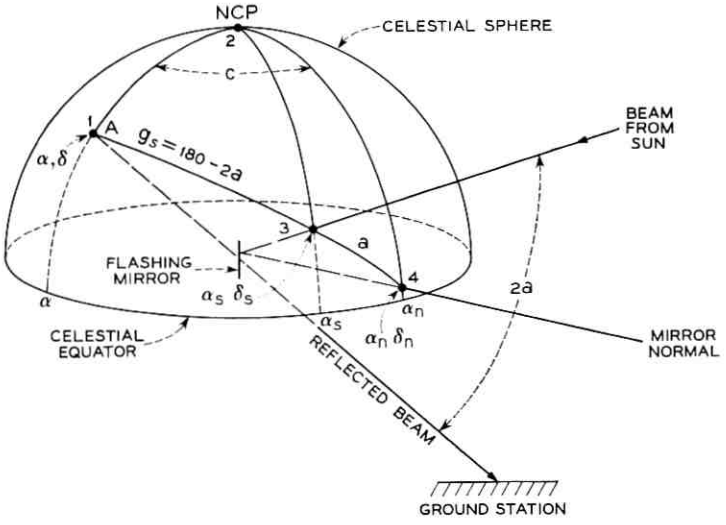


Fig. 6 — Geometry of mirror normal.

ground station. For this to be possible, the mirror normal must bisect the angle $2a$ and lie in the plane determined by the mirror-sun line and the mirror-ground station line.

Sketched around the mirror in Fig. 6 is the celestial sphere. On this sphere the solar direction, as seen from the mirror, is indicated by its right ascension, α_s, and its declination, δ_s. In like fashion, the direction of the satellite mirror as seen from the ground station is specified as α,δ and the direction of the mirror normal is $\alpha_n\delta_n$.

To determine α_n and δ_n, we begin by solving for arc g_s and A in the spherical triangle 1, 2, 3:

$$\cos g_s = \sin \delta \sin \delta_s + \cos \delta \cos \delta_s \cos (\alpha_s - \alpha) \tag{1}$$

$$\cos A = \frac{\sin \delta_s - \sin \delta \cos g_s}{\cos \delta \sin g_s} \tag{2}$$

where angles g_s and A may have values from 0° to 180°.

By triangle 1, 2, 4, one obtains the declination of the normal as

$$\delta_n = \sin^{-1} [\sin \delta \cos g_s + \cos \delta \sin g_s \cos A] \tag{3}$$

where δ_n may have values between ±90°.

Using the same triangle we may write

$$C = \sin^{-1} \left[\frac{\sin (g_s + a) \sin \delta_n}{\sin A} \right] \tag{4}$$

where

$$A = \frac{180 - g_s}{2}. \tag{5}$$

C may have values between 0° and 180°. Since α_n may either exceed α by C or be less by the same amount owing to the two possible orientations of triangle 1, 2, 4, we have

$$\alpha_n = \alpha + \text{SIGNF } (C,\alpha_s - \alpha) \tag{6}$$

where SIGNF, a common computer symbol, indicates that the algebraic sign to be affixed to C shall be determined by the quantity $\alpha_s - \alpha$.

Right ascension is measured eastward from the vernal equinox as a 0 reference through 360°, and if α_s and α should lie on opposite sides in this reference (C remaining less than 180°), (6) becomes

$$\alpha_n = \alpha + \text{SIGNF } (C,\alpha - \alpha_s). \tag{7}$$

Thus (6) is employed if $|\alpha_s - \alpha| - 180°$ is negative and (7) if otherwise.

3.2 *The Right Ascension and Declination of the Sun*

The apparent right ascension and declination of the sun at any specified time is computed from the mean orbital elements of the sun.[4] These may be expressed beginning with the true solar mean anomaly in degrees as

$$M_s = 358.47583 + 0.9856002670d' - 0.000150T^2 - 0.000003T^3$$

where

T = the time in Julian centuries of 36525 ephemeris days from January 0.5, 1900, ephemeris time[5]
d' = ephemeris days from same epoch.

We may express the above equation in a more useful form for the present calculations by changing the epoch to January 1.0, 1960 and writing an equivalent expression as

$$M_s \doteq 357.41283 + 0.985600267d$$

where

d = ephemeris days since the 1960 epoch
= mean solar days since 1960 epoch + 1 second for the years 1963, 1964, 1965.

One obtains the apparent mean solar anomaly, used in the present calculations to determine the apparent position of the sun, by antedating for the solar light transit time. For this reason d is increased by 0.005375 day, which is the light transit time at mean solar distance. Since the earth is about 3×10^6 miles closer to the sun in winter as compared to summer, this can produce an error in apparent solar position of about 2 seconds of arc.

The eccentricity of the earth's orbit is

$$e = 0.01675104 - 0.00004180T - 1.26 \times 10^{-7}T^2$$
$$= 0.01700254 \text{ on January 1.0, 1960.}$$

Also, the degrees of mean celestial longitude of the perigee of the sun's mean orbit about the earth as a reference is

$$L = 279.69668 + 0.9856473354d' + 0.000303T^2$$
$$= 282.25247 + 0.470684 \times 10^{-4}d.$$

Finally, the mean obliquity of the ecliptic in degrees is

$$\epsilon = 23.452294 - 0.0130125T - 1.64 \times 10^{-6}T^2 + 5.03 \times 10^{-7}T^3.$$

The apparent mean anomaly and eccentricity along with Kepler's equation permit the calculations of the apparent true anomaly of the sun by standard techniques.[6] This along with the apparent mean longitude of solar perigee and mean obliquity of the ecliptic allows a direct determination of the apparent right ascension and declination of the sun at any date by the simple geometry shown in Fig. 7.

The apparent right ascension and declination of the sun computed from its mean elements in this manner can depart from the tabular values (which include nutation) of *The American Ephemeris and Nautical Almanac* for the years 1962, 1963, and 1964 by about 5 seconds of arc as a maximum. This is well within the approximate 1° error of the measuring techniques for solar aspect as well as the 0.1° for the mirror technique.

3.3 *Construction of the Locus Circle*

As previously stated, the direction of the spin axis may be specified as a circle on the celestial sphere, centered on the right ascension and declination of the mirror normal. This locus may be generated by using the spherical triangle shown in Fig. 8. Here the dotted curve indicates the attitude locus of all possible positions of the spin axis. Let $\alpha_l\delta_l$ be any point on this circle making a fixed arc F with α_n , δ_n . In the case of either the Telstar I or II satellite, F will equal 68° or 95°, the angles

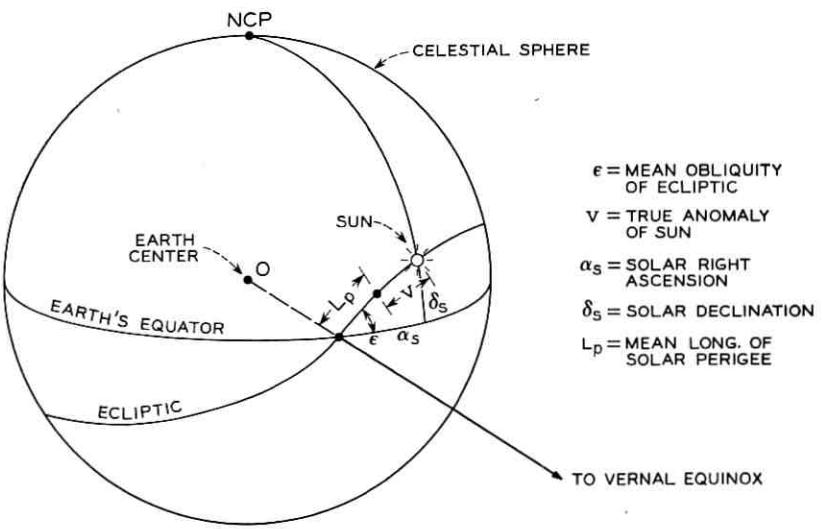


Fig. 7 — Geometry of the sun's position.

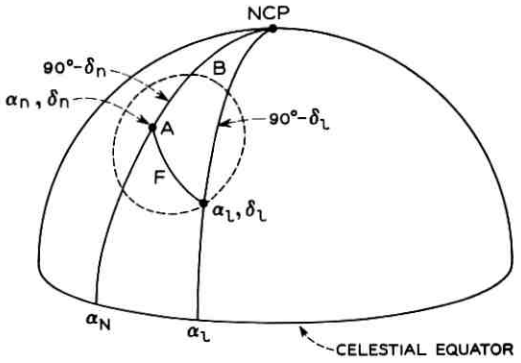


Fig. 8 — The attitude locus.

the mirror normals make with the spin axis. From Fig. 8,

$$\delta_l = \sin^{-1} [\sin \delta_n \cos F + \cos \delta_n \sin F \cos A] \tag{8}$$

where

A = a running variable which takes on values from 0 to 360° to generate the locus

δ_l = declination of a locus point which ranges from −90° to +90°, in general.

Also

$$\sin B = \frac{\sin F \sin A}{\cos \delta_l} \tag{9}$$

$$\cos B = \frac{(\cos F - \sin \delta_n) \sin \delta_l}{\cos \delta_n \cos \delta_l} \tag{9a}$$

$$B = \tan^{-1} \frac{\sin B}{\cos B} \tag{10}$$

and

$$\alpha_l = \alpha_n + B. \tag{10a}$$

In general, B may range from 0° to 360°. All quadrant ambiguities presented by (10) are resolved by noting the algebraic signs of (9) and (9a).

If there is a time uncertainty in measuring the midpoint of a flash series, this will result in more than one possible mirror normal, $\alpha_n \delta_n$, and hence a number of attitude loci, since the satellite and sun will occupy successive positions along their paths within the time error.

If during a single sky trajectory a ground station records two separate flash series (or a single series plus solar aspect data), each will generate a circular locus. The intersections of these loci determine attitude with an ambiguity of two. That is, there are two intersections and therefore two possible satellite attitudes which satisfy the two flash series. A priori knowledge of the attitude from previous measurements or initially from launch parameters together with some knowledge of the expected precession will permit selecting the proper intersection. Knowledge of expected precession is also needed to determine whether or not the flash series are close enough together in time to neglect incremental precession during the time between the series. Such precession can alter the position of the loci intersections.

Fig. 9 shows an example of the attitude determination of the Telstar II satellite for passes 135, 136, 199, 272, and 472. Solid lines indicate attitude loci determined from mirror flashes. The line of zeros is typical of those determined by solar sensor data. Two attitude determinations are shown by intersections on passes 135, 136, and on pass 272.

IV. ATTITUDE PREDICTION

Acting upon an orbiting spin-stabilized satellite to produce precession are certain disturbing torques. Those to be considered here are gravitational and magnetic torques. In the case of the Telstar I and II satellites, these are dominant over atmospheric drag torques, solar radiation torques, electrostatic torques, and others.

4.1 *Assumptions for Gravity Torque Calculations*

Differential gravity forces acting across the body of the satellite can produce torques which tend to rotate the body. These forces exist simply because the strength of the earth's gravity field is a function of the distance from earth. A body of finite size must, therefore, experience such torques.

For the purpose of calculating the mean gravity torques, the earth is assumed spherical with its radius equal to the equatorial radius. For the Telstar satellites, gravity torques are at least an order of magnitude less than magnetic torques, and therefore neglecting the earth's oblateness produces at worst only a second-order error.

It is also assumed that the moments of inertia of the satellites about all axes perpendicular to the spin axis and passing through the mass center are equal. The orbit is assumed elliptical and Keplerian, since the earth's figure produces but second-order effects over a single satel-

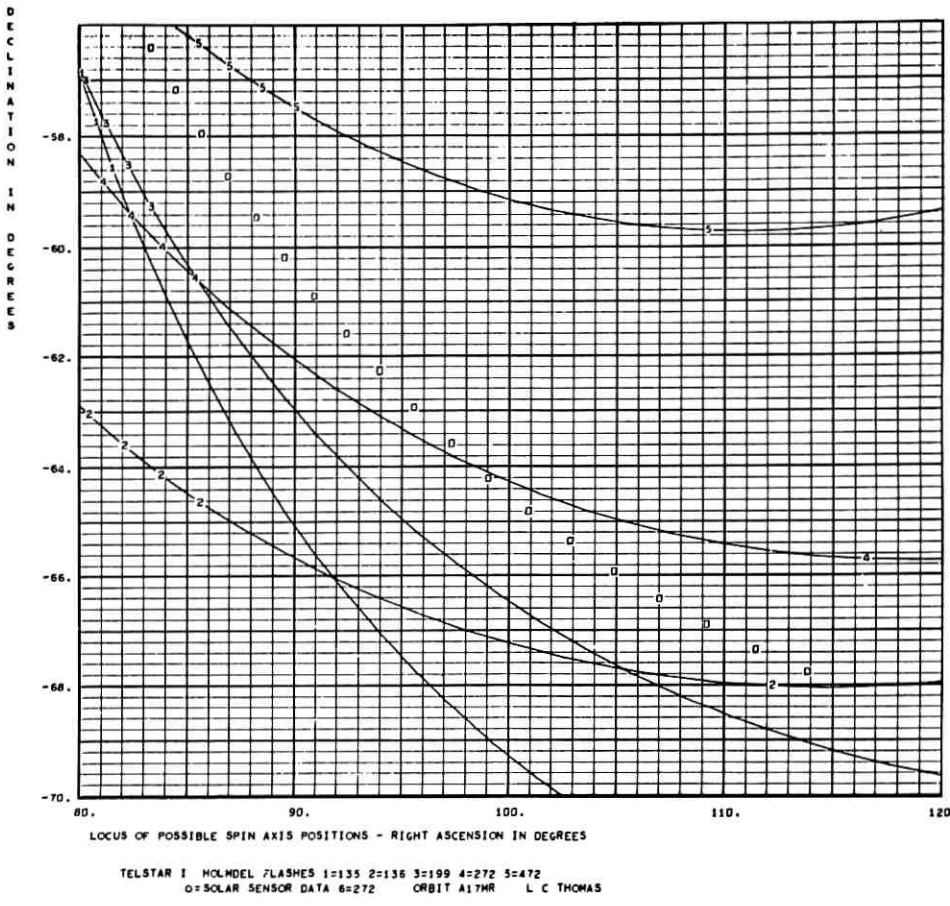


Fig. 9 — Computer plot of mirror and solar loci.

lite period. The orbit, however, for each succession of calculations will be updated using mean orbital elements, which include all secular perturbations, to produce the Keplerian orbit of the best fit at the beginning of each calculation.

4.2 *Assumptions for Magnetic Torque Calculations*

In order to construct a mathematical model which on one hand reasonably well represents the physics of the situation, but on the other hand does not, by its complexity, produce numerical equations costly to compute, the following simplifying assumptions are made:

(*i*) The magnetic field of the earth is represented by a dipole centered at the earth's center and inclined β degrees (equal 11.4°) to the earth's spin axis.[7,8]

(*ii*) Mass symmetry about the satellite's spin axis exists, as in Section 4.1.

(*iii*) The effects of magnetic moments transverse to the spin axis are either negligible or average out due to the satellite's spin.

(*iv*) For the calculation of mean or net magnetic torques over an orbital period, a Keplerian orbit is assumed as in Section 4.1.

4.3 *Assumptions for Gyroscopic Equations of Motion*

It will be useful in simplifying the gyroscopic equations of motion (see Section 4.7) to assume that the satellite angular momentum vector coincides with the spin axis and equals the spin rate times the moment of inertia about the spin axis. This is the same as assuming the entire angular momentum of the satellite is the result of its spin alone, and neglects that small amount provided by the precession of the spin axis itself. Precessional dampers[9] are provided on the Telstar satellites to prevent coning or rapid changes in attitude at rates comparable to the spin rate. This, more than ever, makes the assumption quite reasonable.

4.4 *Coordinate Systems*

It will be convenient to establish certain useful coordinate systems and their interrelationships prior to the torque calculations. These will be defined and related by Euler-type axis rotations expressed by matrices. All coordinate systems are rectangular and right-handed in the conventional sense such that the rotation of an X axis into a Y axis determines the positive direction of a Z axis as the direction of progress of a right-handed screw. Each system will be named, described, and interrelated in that order.

4.4.1 *The Earth-Centered Inertial System (IS)*

In this system, the three mutually perpendicular X, Y, and Z axes have their common origin at the earth's center. X contains the vernal equinox and increases positively from the earth's center in its direction. Y is perpendicular to X in the earth's equatorial plane. Z contains the earth's spin axis and increases positively toward the north celestial pole. This is the basic system to which others will be referred.

4.4.2 *Rigid-Body Systems*

Two rigid-body systems will be used. Both have their origin at the satellite mass center. The first of these is to be known as the SANOR or satellite nonrotating system. In this system, the z axis defines the satellite's spin axis, being positive in the direction of advance of a right-hand screw spinning with the satellite. The x axis is in the satellite

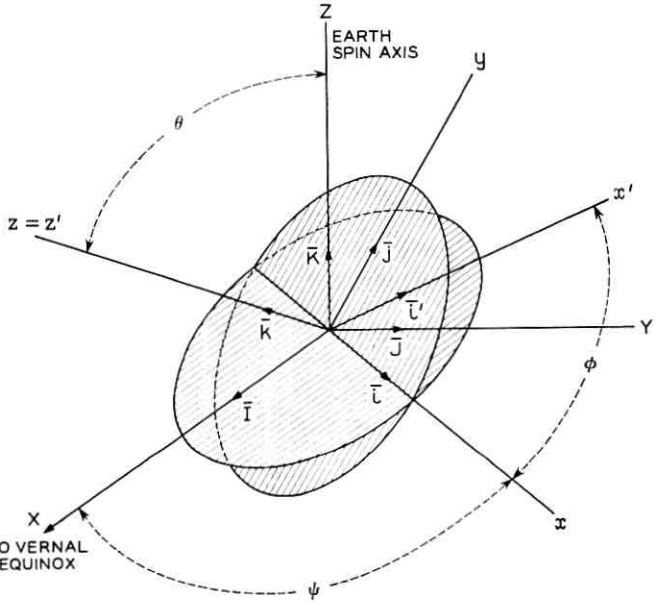


Fig. 10 — The IS, SANOR and SAR systems.

equatorial plane, ψ degrees rotated from X (see Fig. 10) about Z. This axis defines the xy-XY plane intersection. The y axis is in the satellite equatorial plane orthogonal to x. At this point, we note that the rotational equations of motion for the satellite will be developed in this SANOR system.

The SAR or satellite rotating system, as its name implies, differs from the SANOR system in that it rotates with the satellite. It has axes x', y', z', where z' coincides with z, and x' and y' are defined as being rotated φ degrees from x and y respectively about z.

To relate the SANOR and SAR systems to IS, we proceed as follows (refer to Fig. 10). The ψ rotation about Z yields the following matrix

$$D = \begin{vmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{vmatrix}. \tag{11}$$

The θ rotation about x yields

$$C = \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{vmatrix}. \tag{12}$$

The φ rotation about z gives

$$B = \begin{vmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{vmatrix}. \tag{13}$$

To transfer points in the IS to points in the SANOR system, we have

$$\begin{vmatrix} x \\ y \\ z \end{vmatrix} = (CD) \begin{vmatrix} X \\ Y \\ Z \end{vmatrix} \tag{14}$$

or, just to shorten the notation

$$x = (CD)X \tag{15}$$

where

(CD) = multiplication of the C and D matrices.

In like fashion

$$\mathrm{a}' = (BCD)X \tag{16}$$

$$= AX \tag{17}$$

where

$A \equiv (BCD)$ = multiplication of the B, C, D matrices.

Carrying out this multiplication, we may express A as the following

$$A = \begin{vmatrix} \cos\psi\cos\varphi & \cos\varphi\sin\psi & \sin\varphi\sin\theta \\ \quad -\cos\theta\sin\varphi\sin\psi & \quad +\cos\theta\cos\psi\sin\varphi & \\ -\sin\varphi\cos\psi & -\sin\varphi\sin\psi & \cos\varphi\sin\theta \\ \quad -\cos\theta\sin\psi\cos\varphi & \quad +\cos\theta\cos\psi\cos\varphi & \\ \quad \sin\theta\sin\psi & \quad -\sin\theta\cos\psi & \cos\theta \end{vmatrix}. \quad (18)$$

The matrix CD may be obtained from A by letting $\varphi = 0$,

$$(CD) = \begin{vmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi\cos\theta & \cos\psi\cos\theta & \sin\theta \\ \sin\psi\sin\theta & -\cos\psi\sin\theta & \cos\theta \end{vmatrix}. \quad (19)$$

Quite obviously the inverse operations apply by taking the transposes of the matrices. Indicating a transposed matrix by the symbol $\sim$, we have

$$X = (\tilde{D}\tilde{C})x \quad (20)$$

and

$$X = \tilde{A}x'. \quad (21)$$

4.4.3 *Orbital Coordinate Systems*

Let an orbit defining system (ORDEF) be described along the lines shown in Fig. 11. Here x_0 is the intersection of the orbit plane and the earth's equator plane with $+x_0$ drawn toward the ascending node of the orbit from the earth's center. Axis z_0 is normal to the orbit plane, positive in a direction a right-handed screw along oz_0 would advance if turned in the direction of the satellite's orbital motion. Axis y_0 completes a right-handed system by being mutually perpendicular to x_0 and z_0.

The satellite defining system (SADEF) has its z_s axis collinear with z_0, but x_s passes through the instantaneous satellite position, ω degrees from x_0. Axis x_s therefore will be referred to as the local vertical of the satellite.

The ORDEF and SADEF systems are related to the inertial system (IS) in a manner strictly analogous to the SANOR and the SAR matrices described in the previous section. The only change in the matrices is that

Ω replaces ψ

i replaces θ

ω replaces φ.

Thus we may write

$$x_g = (CD)_i X \tag{22}$$

$$X = (\tilde{D}\tilde{C})_i x_g \tag{23}$$

$$x_s = A_i X \tag{24}$$

$$X = \tilde{A}_i x_s \tag{25}$$

where the subscript i indicates the Ω, i, ω substitution.

4.4.4 *Magnetic Coordinate System*

A magnetic coordinate system (MAG) based on an inclined dipole model of the earth's magnetic field is constructed as follows. Let axis x_m define the intersection of the geomagnetic equatorial plane with the earth's geographical equatorial plane, positive toward the ascending node of the geomagnetic equator ($\eta°$ from X, see Fig. 12). Axis z_m is

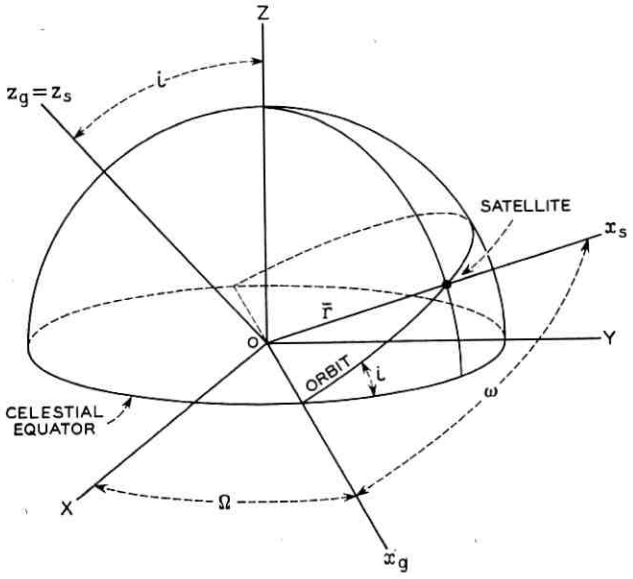


Fig. 11 — The ORDEF and SADEF coordinate systems.

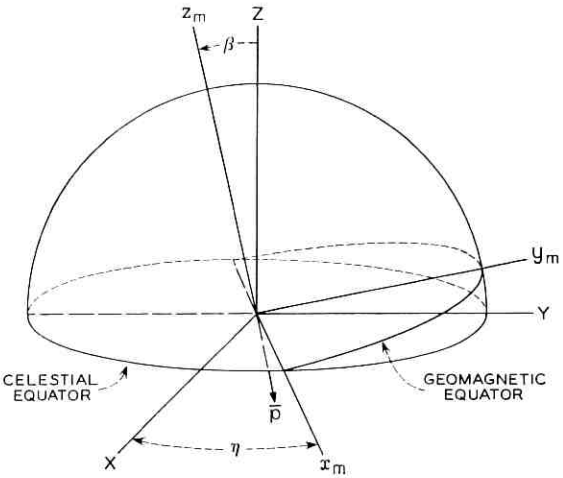


Fig. 12 — The geomagnetic coordinate system.

normal to the geomagnetic equator plane, positive toward the north geographical hemisphere, and $\beta°$ from Z. Here β is simply the magnetic dipole inclination to the earth's spin axis. Axis y_m is in the geomagnetic equatorial plane orthogonal to x_m.

Transformations from the MAG system into the IS system proceed as in (15) thru (20) with

$$\eta \text{ replacing } \psi$$

$$\beta \text{ replacing } \theta$$

in the matrices, so that

$$x_m = (CD)_\eta X \tag{26}$$

and

$$X = (\tilde{D}\tilde{C})_\eta x_m . \tag{27}$$

where the subscript η on the matrices indicates the above substitutions.

Fig. 13 shows the orientation of the earth-centered inclined dipole which produces a field of best fit[7,8] to a field based on all observations of field vectors made anywhere over the earth's surface. The anomalies, or differences between the actual and dipole field, decrease more rapidly with increasing heights above the earth's surface than does the dipole field itself, making the fit better and better as altitude increases.

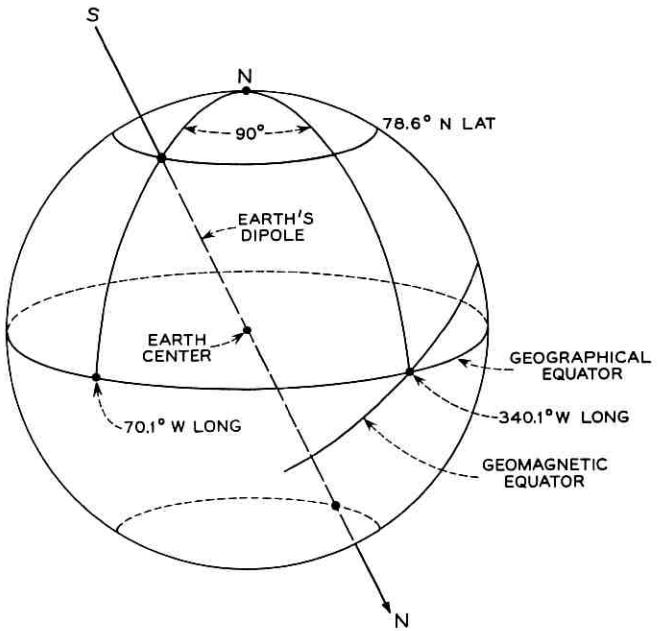


Fig. 13 — The earth-centered inclined dipole.

4.5 *The Mean Gravity Torque*

If the potential energy of an orbiting satellite is expanded about its center of mass in a Taylor series and differentiated with respect to the angles giving its orientation with respect to the IS coordinate system, the instantaneous gravitational torque acting on the satellite may be expressed as[10]

$$\mathbf{T}_g = 3\mu^2(I_3 - I)(\mathbf{i}_s \cdot \mathbf{k})(\mathbf{i}_s \times \mathbf{k}) \tag{28}$$

where

$$\mu^2 = GM/r^3 \tag{29}$$

GM = universal gravitational constant $\times$ the mass of the earth
r = geocentric distance of satellite
I_3 = moment of inertia about satellite's spin axis, z
I = moment of inertia transverse to z (assuming all transverse moments to be equal, i.e., $I_x = I_y$)
$\mathbf{k}$ = unit vector along satellite spin axis, z (more generally the

vector along the axis of cylindrical mass symmetry, which is the spin axis by assumptions in Section 4.1).

To determine the mean gravity torque, $\mathbf{T}_g$ will be integrated over a nodal period (T_n) assuming no precession during that period; i.e., the SANOR system stays fixed with respect to IS during the integration. Then

$$(\mathbf{T}_g)_{\text{mean}} = \frac{1}{T_n}\int_0^{T_n} 3\mu^2(I_3 - I_1)(\mathbf{i}_s\cdot\mathbf{k})(\mathbf{i}_s \times \mathbf{k})\, dt. \tag{30}$$

By interchanging the order of the dot product and taking $\mathbf{k}$ constant,

$$(\mathbf{T}_g)_{\text{mean}} = \frac{3GM(\Delta I)}{T_n}\left[\mathbf{k}\cdot\left\{\int_0^{T_n}\frac{\mathbf{i}_s\mathbf{i}_s}{r^3}\, dt\right\} \times \mathbf{k}\right]. \tag{31}$$

If a pure Keplerian orbit[6] is assumed, we have

$$r = \frac{a(1 - e^2)}{1 + e\cos(\omega - P)} \tag{32}$$

$$r^2\dot{\omega} = \frac{2\pi a^2(1 - e^2)^{\frac{1}{2}}}{T_n} \tag{33}$$

$$GM = \frac{4\pi^2 a^3}{T_n^2} \tag{34}$$

where

r = radius vector from the focus of the elliptical orbit to the satellite
a = semimajor axis of ellipse
e = eccentricity of the ellipse
P = argument of perigee
ω = argument of the satellite.

Hence the integral becomes

$$\int_0^{T_n}\frac{\mathbf{i}_s\mathbf{i}_s}{r^3}\, dt = \int_0^{\omega=2\pi}\frac{\mathbf{i}_s\mathbf{i}_s}{r^3\dot{\omega}}\, d\omega \tag{35}$$

and

$$(\mathbf{T}_g)_{\text{mean}} = \frac{6\pi(\Delta I)}{T_n^2(1 - e^2)^{\frac{1}{2}}} \times\left[\mathbf{k}\cdot\int_0^{2\pi}\mathbf{i}_s\mathbf{i}_s(1 + e\cos\omega\cos P + e\sin\omega\sin P)\, d\omega \times \mathbf{k}\right]. \tag{36}$$

Since in the ORDEF coordinates (see Fig. 11)

$$\mathbf{i}_s = \cos \omega \mathbf{i}_g + \sin \omega \mathbf{j}_g \tag{37}$$

the dyad, $\mathbf{i}_s\mathbf{i}_s$ becomes

$$\mathbf{i}_s\mathbf{i}_s = \cos^2 \omega\, \mathbf{i}_g\mathbf{i}_g + \sin^2 \omega\, \mathbf{j}_g\mathbf{j}_g + \sin \omega \cos \omega(\mathbf{i}_g\mathbf{j}_g + \mathbf{j}_g\mathbf{i}_g). \tag{38}$$

Substituting this into (36) results in three integrals. Those containing $\mathbf{i}_g\mathbf{i}_g$ and $\mathbf{j}_g\mathbf{j}_g$ yield π as a result of the integration. The other integral is 0. Therefore,

$$(T_g)_{\text{mean}} = \frac{6\pi^2(\Delta I)}{T_n^2(1 - e^2)^{\frac{3}{2}}} [\mathbf{k}\cdot(\mathbf{i}_g\mathbf{i}_g + \mathbf{j}_g\mathbf{j}_g) \times \mathbf{k}]. \tag{39}$$

By some vector maneuvering (see Appendix A) this reduces to

$$(T_g)_{\text{mean}} = -\frac{6\pi^2(\Delta I)}{T_n^2(1 - e^2)^{\frac{3}{2}}} [(\mathbf{k}\cdot\mathbf{k}_g)(\mathbf{k}_g \times \mathbf{k})]. \tag{40}$$

To transform the ORDEF vectors into the SANOR system, we write

$$x_g = (CD)_i X \tag{22}$$

$$X = (\tilde{D}\tilde{C})x \tag{20}$$

$$\therefore\ x_g = (CD)(\tilde{D}\tilde{C})x. \tag{41}$$

Performing the indicated operations, after a bit of labor we find that

$$\begin{aligned}\mathbf{k}_g = {} & (\sin \Omega \sin i \cos \psi - \cos \Omega \sin i \sin \psi)\mathbf{i} \\ & + (-\sin \Omega \sin i \sin \psi \cos \theta - \cos \Omega \sin i \cos \psi \cos \theta + \cos i \sin \theta)\mathbf{j} \\ & + (\sin \Omega \sin i \sin \psi \sin \theta + \cos \Omega \sin i \cos \psi \sin \theta + \cos i \cos \theta)\mathbf{k}\end{aligned} \tag{42}$$

where, referring to Figs. 10 and 11, we see that

ψ = an Euler rotational angle

i = orbital inclination

Ω = ascending node of orbit.

Substituting into (40), the mean gravity torque reduces to

$$\begin{aligned}(\mathbf{T}_g)_{\text{mean}} = {} & \frac{6\pi^2\Delta I}{T_n^2(1 - e^2)^{\frac{3}{2}}} \{\cos i \cos \theta + \sin i \sin \theta \cos (\Omega - \psi)\} \\ & \cdot [[(\sin i \cos \theta \cos (\Omega - \psi) - \cos i \sin \theta)\mathbf{i} + (\sin i \sin (\Omega - \psi))\mathbf{j}]].\end{aligned} \tag{43}$$

4.6 *The Mean Magnetic Torque*

It is well known[11] that the scalar potential (Φ_m) of a magnetic dipole

may be expressed as

$$\Phi_m = \frac{1}{4\pi\mu_0}\frac{\mathbf{p}\cdot\mathbf{i}_s}{r^2}$$

$$= \frac{1}{4\pi\mu_0}\frac{\mathbf{p}\cdot\mathbf{r}}{r^3} \tag{44}$$

$$= -\frac{p}{4\pi\mu_0}\frac{\mathbf{k}_m\cdot\mathbf{r}}{r^3} \tag{45}$$

where

μ_0 = permeability of free space
= $4\pi \times 10^{-7}$ webers/ampere-meter (Ref. 12)
= $4\pi \times 10^{-7}$ henry/meter (Ref. 12)
$\mathbf{p}$ = magnetic moment of earth's field, direction and magnitude
= 10^{17} weber-meters
= $10^{17}(10^{10}/4\pi)$ emu = 8.06×10^{25} emu
= 8.06×10^{25} erg/gauss (Ref. 12, p. 25)
$\mathbf{i}_s$ = unit vector along the x_s axis, local satellite vertical
r = distance from dipole center (earth's center) to satellite
$\mathbf{k}_m$ = unit vector describing direction of geomagnetic moment
p = magnitude of $\mathbf{p}$
$\mathbf{r}$ = vector form of r along i_s axis.

A satellite magnetic moment ($\mathbf{M}$) in the earth's magnetic field ($\mathbf{H}$) will produce a torque

$$\mathbf{T} = \mathbf{M} \times \mathbf{H}. \tag{46}$$

If the satellite spins rapidly, any magnetic moment components perpendicular to the spin axis will tend to produce torques which average to 0, while the component along the spin axis will produce a net torque of

$$\mathbf{T}_m = \mathbf{M}_s \times \mathbf{H} \tag{47}$$

$$= M_s\mathbf{k} \times \mathbf{H} \text{ ergs} \tag{48}$$

where

M_s = satellite magnetic moment along its spin axis (weber-meters or ergs/gauss)
$\mathbf{k}$ = unit vector along satellite spin axis (SANOR system)
$\mathbf{H}$ = geomagnetic field (ampere-turns/meters).

It is desired to integrate the instantaneous torque, $\mathbf{T}_m$, over one anomalistic period in order to calculate the mean torque. To do this,

we shall express **H** in the MAG system and eventually convert this expression into SANOR terms to be comparable with M_s.

We begin by setting

$$\mathbf{H} = -\nabla\Phi_m \tag{49}$$

from (45) and (49),

$$\mathbf{H} = -\frac{h}{\mu_0}\nabla\left[\frac{-\mathbf{k}_m\cdot(x_m\mathbf{i}_m + y_m\mathbf{j}_m + z_m\mathbf{k}_m)}{r^3}\right] \tag{50}$$

where

$$x_m, y_m, z_m = \text{components of } \mathbf{r} \text{ in MAG system}$$

and

$$h = p/4\pi. \tag{51}$$

Therefore,

$$\mathbf{H} = \frac{h}{\mu_0}\nabla(z_m/r^3) \tag{52}$$

and since

$$r^3 = (x_m^2 + y_m^2 + z_m^2)^{\frac{3}{2}}$$

we have

$$\mathbf{H} = \frac{-h}{\mu_0 r^5}[3x_m z_m\mathbf{i}_m + 3y_m z_m\mathbf{j}_m + (3z_m^2 - r^2)\mathbf{k}_m]. \tag{53}$$

Rewriting to spotlight **r** components, using the following relationships normalized to r,

$$x_m/r = r_x, \qquad y_m/r = r_y, \qquad z_m/r = r_z \tag{54}$$

one obtains

$$\mathbf{H} = -\frac{h}{\mu_0 r^3}[3r_x r_z\mathbf{i}_m + 3r_y r_z\mathbf{j}_m + (3r_z^2 - 1)\mathbf{k}_m] \tag{55}$$

or

$$\mathbf{H} = \frac{p}{4\pi\mu_0 r^3}[\mathbf{k}_m - 3(\mathbf{i}_s\cdot\mathbf{k}_m)\mathbf{i}_s] \tag{56}$$

since

$$\mathbf{r} = r\mathbf{i}_s. \tag{57}$$

By the following equations [(24) through (65)], the variables $\mathbf{i}_s$ and $\mathbf{k}_m$ will be expressed in forms useful in the ensuing mean magnetic torque calculations.

Let us first express $\mathbf{i}_s$ in the MAG system. As shown previously,

$$x_s = A_i X \tag{24}$$

and

$$X = (\tilde{D}\tilde{C})_\eta X_m \tag{27}$$

therefore

$$x_s = A_i(\tilde{D}\tilde{C})_\eta X_m \tag{58}$$

or

$$\begin{aligned}\mathbf{i}_s = {} & |(\cos\Omega\cos\omega - \cos i\sin\omega\sin\Omega) \\ & \cdot(\cos\omega\sin\Omega + \cos i\cos\Omega\sin\omega)(\sin\omega\sin i)| \\ & \times \begin{vmatrix} \cos\eta & -\sin\eta\cos\beta & \sin\eta\sin\beta \\ \sin\eta & \cos\eta\cos\beta & -\cos\eta\sin\beta \\ 0 & \sin\beta & \cos\beta \end{vmatrix} \begin{vmatrix} \mathbf{i}_m \\ \mathbf{j}_m \\ \mathbf{k}_m \end{vmatrix}.\end{aligned} \tag{59}$$

Expanding,

$$\begin{aligned}\mathbf{i}_s = {} & [\cos\eta(\cos\Omega\cos\omega - \cos i\sin\omega\sin\Omega) + \sin\eta(\cos\omega\sin\Omega \\ & + \cos i\cos\Omega\sin\omega)]\mathbf{i}_m + [-\sin\eta\cos\beta(\cos\Omega\cos\omega \\ & - \cos i\sin\omega\sin\Omega) + \cos\eta\cos\beta(\cos\omega\sin\Omega \\ & + \cos i\cos\Omega\sin\omega) + \sin\beta\sin\omega\sin i]\mathbf{j}_m + [\sin\eta\sin\beta \\ & (\cos\Omega\cos\omega - \cos i\sin\omega\sin\Omega) - \cos\eta\sin\beta(\cos\omega\sin\Omega \\ & + \cos i\cos\Omega\sin\omega) + \cos\beta\sin\omega\sin i]\mathbf{k}_m .\end{aligned} \tag{60}$$

This will be used later.

Also, $\mathbf{i}_s$ in the IS system may be written [see (24)] as

$$\mathbf{i}_s = A_i\mathbf{I} \tag{61}$$

$$\begin{aligned}\mathbf{i}_s = {} & [\cos\Omega\cos\omega - \cos i\sin\omega\sin\Omega]\mathbf{I} \\ & + [\cos\omega\sin\Omega + \cos i\cos\Omega\sin\omega]\mathbf{J} \\ & + [\sin\omega\sin i]\mathbf{K}.\end{aligned} \tag{62}$$

In a similar manner, $\mathbf{k}_m$ in the IS system is expanded as follows:

$$x_m = (CD)_\eta X \tag{63}$$

or

$$\mathbf{k}_m = \sin\eta \sin\beta\,\mathbf{I} - \cos\eta \sin\beta\,\mathbf{J} + \cos\beta\,\mathbf{K} \tag{64}$$

$$= \sin\beta(\sin\eta\,\mathbf{I} - \cos\eta\,\mathbf{J}) + \cos\beta\,\mathbf{K}. \tag{65}$$

The problem is now to integrate the instantaneous torque, $\mathbf{T}_m$, over one anomalistic period, T_a, to obtain the mean magnetic torque,

$$(\mathbf{T}_m)_{\text{mean}} = \frac{1}{T_a}\int_{t_i}^{t_i+T_a} \mathbf{T}_m dt = \frac{1}{T_a}\int_{v_i}^{v_i+2\pi} \mathbf{T}_m \frac{dt}{dv}\,dv = \frac{1}{T_a}\int_{v_i}^{v_i+2\pi} \mathbf{T}_m \frac{r^2}{h}\,dv \tag{66}$$

where

T_a = anomalistic period of the satellite
$\mathbf{T}_m$ = instantaneous magnetic torque
t = time
v = true anomaly of the satellite
r = geocentric satellite distance
h = defined by (68) below.

Substituting into (66) from (48) and (56)

$$(\mathbf{T}_m)_{\text{mean}} = \frac{M_a p}{4\pi\mu_0 T_a}\,\mathbf{k} \times \int_{v_i}^{v_i+2\pi} \frac{\mathbf{k}_m - 3(\mathbf{i}_s\cdot\mathbf{k}_m)\mathbf{i}_s}{r^3}\left(\frac{r^2}{h}\right) dv. \tag{67}$$

Since

$$h = \frac{2\pi a^2(1-e^2)^{\frac{1}{2}}}{T_a} \tag{68}$$ [13]

and

$$r = \frac{a(1-e^2)}{1+e\cos v} \tag{69}$$ [13]

where

a = semimajor axis of orbit
e = orbital eccentricity
T_a = anomalistic period of the satellite

(67) becomes

$$(\mathbf{T}_m)_{\text{mean}} = \frac{M_s p}{8\pi^2 \mu_0 a^3 (1 - e^2)^{\frac{3}{2}}} \cdot \mathbf{k} \times \int_{v_i}^{v_i + 2\pi} [\mathbf{k}_m - 3(\mathbf{i}_s \cdot \mathbf{k}_m)\mathbf{i}_s](1 + e \cos v) dv \tag{70}$$

$$= \mathfrak{M}\mathbf{k} \times \int_{v_i}^{v_i + 2\pi} [\mathbf{k}_m - 3(\mathbf{i}_s \cdot \mathbf{k}_m)\mathbf{i}_s](1 + e \cos v) dv. \tag{71}$$

First we shall evaluate the integral of (71) beginning by substituting the $\mathbf{k}_m$ value given in (65), to obtain

$$\frac{(\mathbf{T}_m)_{\text{mean}}}{\mathfrak{M}} = (\sin \beta)k \times \int_{v_i}^{v_i + 2\pi} [\sin \eta \mathbf{I} - \cos \eta \mathbf{J} - 3\{\mathbf{i}_s \cdot (\sin \eta \mathbf{I} - \cos \eta \mathbf{J})\}\mathbf{i}_s](1 + e \cos v) dv \tag{72}$$

$$+ (\cos \beta)\mathbf{k} \times \int_{v_i}^{v_i + 2\pi} \{\mathbf{K} - 3(\mathbf{i}_s \cdot \mathbf{K})\mathbf{i}_s\}(1 + e \cos v) dv$$

$$= \text{I} + \text{II}. \tag{73}$$

Integral II, the simpler of the two, is evaluated in Appendix B. The result is given below:

$$\text{II} = \pi \cos \beta [\{-2 \sin \theta + 3 \sin^2 i \sin \theta + 3 \sin i \cos i \cos \theta \cos (\Omega - \psi)\}\mathbf{i} + 3 \sin i \cos i \sin (\Omega - \psi)\mathbf{j}]. \tag{88}$$

Now we must evaluate the first integral, I, of (73) but first let us state that by examining (72) it is perfectly obvious that the I integral does not exist for a noninclined dipole. For this case, it follows from (73) with $\cos \beta = 1$ that

$$(\mathbf{T}_m)_{\text{mean}} = \frac{M_s p}{8\pi^2 \mu_0 a^3 (1 - e^2)^{\frac{3}{2}}} \text{II}. \tag{89}$$

By suitable variable substitution the first integral I of (73) may be compressed to yield

$$\text{I} = (\sin \beta)\mathbf{k} \times (\mathbf{I}A - \mathbf{J}B - 3\mathbf{I} \cdot \mathbf{C} + 3\mathbf{J} \cdot \mathbf{E}) \tag{90}$$

where

$$A = \int_{v_i}^{v_i + 2\pi} \sin \eta (1 + e \cos v) dv \tag{91}$$

$$B = \int_{v_i}^{v_i+2\pi} \cos \eta(1 + e \cos v)dv \tag{92}$$

$$\mathbf{C} = \int_{v_i}^{v_i+2\pi} \mathbf{i}_s\mathbf{i}_s \sin \eta(1 + e \cos v)dv \tag{93}$$

$$\mathbf{E} = \int_{v_i}^{v_i+2\pi} \mathbf{i}_s\mathbf{i}_s \cos \eta(1 + e \cos v)dv. \tag{94}$$

Now η must be expressed as a function of v in order to evaluate A, B, $\mathbf{C}$, and $\mathbf{E}$. For simplification of the integrals we shall also let the initial time for the integration be the time when the satellite passes through perigee. Then

$$\eta = \eta_0 + \omega_E t \tag{95}$$

$$= \eta_0 + \omega_E T_a M/2\pi \tag{96}$$

$$= \eta_0 + bM \tag{97}$$

where

η_0 = initial position of the ascending node of the geomagnetic equator at perigee passage with respect to the IS system (see Fig. 12)
ω_E = angular velocity of the earth in the IS system
t = time measured from passage of satellite through perigee
M = mean anomaly of the satellite (radians)
T_a = anomalistic period of the satellite (time units)
$b = (\omega_E/2\pi)T_a$ = number of turns of earth in time T_a.

Therefore

$$\sin \eta = \sin \eta_0 \cos (bM) + \cos \eta_0 \sin (bM) \tag{98}$$

$$\cos \eta = \cos \eta_0 \cos (bM) - \sin \eta_0 \sin (bM). \tag{99}$$

Unfortunately, M is related to v through Kepler's equation as

$$M = E - e \sin E \tag{100}$$[14]

where

E = the eccentric anomaly
e = eccentricity of the orbit

and

$$E = 2 \tan^{-1}\left[\left(\frac{1-e}{1+e}\right)^{\frac{1}{2}} \tan \frac{v}{2}\right] = 2 \tan^{-1}\left(q \tan \frac{v}{2}\right). \tag{101}$$[14]

Equations (100) and (101) certainly define M as a function of v, but in a most complicated manner. It can be shown, however, that a plot of v and M versus time normalized to T_a will look like Fig. 14. So as a reasonable approximation to M we might consider

$$M = v - (\lambda/b) \sin v \tag{102}$$

where

λ/b = maximum amplitude of the true anomaly "sine" wave of Fig. 14 = $2e$.[15]

Note that (102) resembles Kepler's equation with v replacing E. With less sophistication we might even let

$$M = v. \tag{103}$$

The only justification here is that we shall be dealing in mean torques averaged over an orbital period, and the v function makes one oscilla-

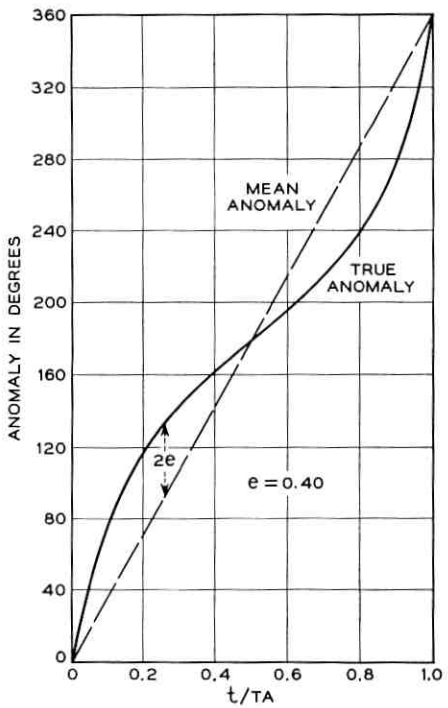


Fig. 14 — Mean and true anomaly comparison.

tion about the M function in that time. If (103) is assumed, (98) and (99) remain unaltered except that v replaces M. If (102) is used, (98) and (99) become

$$\begin{aligned}\sin \eta &= \sin \eta_0[\cos bv \cos (\lambda \sin v) + \sin bv \sin (\lambda \sin v)] + \\ &\quad \cos \eta_0[\sin bv \cos (\lambda \sin v) - \cos bv \sin (\lambda \sin v)] \qquad (104)\\ &= \sin \eta_0[A'] + \cos \eta_0[B'] \qquad (105).\end{aligned}$$

where A' and B' are defined by comparing (105) to (104). In like fashion,

$$\cos \eta = \cos \eta_0[A'] - \sin \eta_0[B']. \qquad (106)$$

We may express the sin $(\lambda \sin v)$ and cos $(\lambda \sin v)$ portions of A' and B' as

$$\sin (\lambda \sin v) \doteq \lambda \sin v - \frac{\lambda^3 \sin^3 v}{6} \qquad (107)$$

and

$$\cos (\lambda \sin v) \doteq 1 - \frac{\lambda^2 \sin^2 v}{2}. \qquad (108)$$

These approximations are reasonably valid for near earth satellites not having high eccentricities and avoid Bessel function complications. The neglect of higher-order terms in (107) and (108) for $e = 0.25$ and $T_a \doteq 150$ minutes results in errors of less than one part in 10^6.

As case I we shall evaluate the integrals A, B, $\mathbf{C}$, and $\mathbf{E}$ using (98) and (99) with v replacing M. For case II we shall return to evaluate these integrals again using (105), (106), (107), and (108). Case I details are given in Appendix C. Case II is outlined by Appendix D.

Using the expanded II integral of (88) and the evaluated A, B, $\mathbf{C}$, and $\mathbf{E}$ integrals from Appendices C and D, (73) may now be written as

$$\begin{aligned}(\mathbf{T}_m)_{\text{mean}} = \mathfrak{M}\Big[&\cos \beta\mathbf{k} \times \int_0^{2\pi} \{\mathbf{k} - 3(\mathbf{i}_a \cdot \mathbf{K})\mathbf{i}_a\}(1 + e \cos v)dv \\ &+ \sin \beta\mathbf{k} \times \{A\mathbf{I} - B\mathbf{J} - 3\mathbf{I}\cdot(\mathbf{i}_a\mathbf{i}_a C_1 + \mathbf{j}_a\mathbf{j}_a C_2 \\ &+ (\mathbf{i}_a\mathbf{j}_a + \mathbf{j}_a\mathbf{i}_a)C_3) + 3\mathbf{J}\cdot(\mathbf{i}_a\mathbf{i}_a E_1 + \mathbf{j}_a\mathbf{j}_a E_2 \\ &+ (\mathbf{i}_a\mathbf{j}_a + \mathbf{j}_a\mathbf{i}_a)E_3)\}\Big]\end{aligned} \qquad (130)$$

where

$A, B, C_1, C_2, C_3, E_1, E_2, E_3$ are all defined in Appendix C.

The following, expressed in the SANOR system, will now be substituted into (130) (refer to Fig. 11)

$$\mathbf{k} \times \mathbf{I} = \cos\psi\mathbf{j} + \sin\psi\cos\theta\mathbf{i} \tag{131}$$

$$\mathbf{k} \times (-\mathbf{J}) = \cos\psi\cos\theta\mathbf{i} - \sin\psi\mathbf{j} \tag{132}$$

$$\mathbf{J}\cdot\mathbf{i}_g = \sin\Omega \tag{133}$$

$$\mathbf{J}\cdot\mathbf{j}_g = \cos\Omega\cos i \tag{134}$$

$$\mathbf{I}\cdot\mathbf{i}_g = \cos\Omega \tag{135}$$

$$\mathbf{I}\cdot\mathbf{j}_g = -\sin\Omega\cos i \tag{136}$$

$$\mathbf{k} \times \mathbf{i}_g = -\cos\theta\sin(\Omega - \psi)\mathbf{i} + \cos(\Omega - \psi)\mathbf{j} \tag{137}$$

$$\mathbf{k} \times \mathbf{j}_g = (-\sin i\sin\theta - \cos i\cos\theta\cos(\Omega - \psi))\mathbf{i} - \cos i\sin(\Omega - \psi)\mathbf{j} \tag{138}$$

$$\mathbf{k} \times \mathbf{K} = -\sin\theta\mathbf{i}. \tag{139}$$

These yield the mean magnetic torque, which is

$$\begin{aligned}(\mathbf{T}_M)_{\text{mean}} = {} & \frac{M_s p}{8\pi^2\mu_0 a^3(1-e^2)^{\frac{3}{2}}}\Big\{\pi\cos\beta[\mathbf{i}\{-2\sin\theta + 3\sin^2 i\sin\theta \\ & + 3\sin i\cos i\cos\theta\cos(\Omega - \psi)\} \\ & + \mathbf{j}\{3\sin i\cos i\sin(\Omega - \psi)\}]\Big\} \\ & + \frac{M_s p\sin\beta}{8\pi^2\mu_0 a^3(1-e^2)^{\frac{3}{2}}}\Big\{A[\sin\psi\cos\theta\mathbf{i} + \cos\psi\mathbf{j}] \\ & + B[\cos\psi\cos\theta\mathbf{i} - \sin\psi\mathbf{j}] \\ & + 3[\{E_1\sin\Omega + E_3\cos\Omega\cos i - C_1\cos\Omega \\ & + C_3\sin\Omega\cos i\}\{-\cos\theta\sin(\Omega - \psi)\mathbf{i} \\ & + \cos(\Omega - \psi)\mathbf{j}\} + \{E_2\cos\Omega\cos i + E_3\sin\Omega \\ & + C_2\sin\Omega\cos i - C_3\cos\Omega\}\{-(\sin i\sin\theta \\ & + \cos i\cos\theta\cos(\Omega - \psi))\mathbf{i} - \cos i\sin(\Omega - \psi)\mathbf{j}\}]\Big\}\end{aligned} \tag{140}$$

or, collecting on $\mathfrak{M}\cos\beta$ and $\mathfrak{M}\sin\beta$,

$$\begin{aligned}(\mathbf{T}_M)_{\text{mean}} = {} & \pi\mathfrak{M}\cos\beta\Big[\mathbf{i}\{-2\sin\theta + 3\sin^2 i\sin\theta + 3\sin i\cos i \\ & \cdot\cos\theta\cos(\Omega - \psi)\} + \mathbf{j}\{3\sin i\cos i\sin(\Omega - \psi)\}\Big]\end{aligned} \tag{141}$$

$$+ \mathfrak{M} \sin \beta \Big[\mathbf{i} \Big\{ A \sin \psi \cos \theta + B \cos \psi \cos \theta - 3[\cos \theta \sin (\Omega - \psi)(E_1 \sin \Omega + E_3 \cos \Omega \cos i - C_1 \cos \Omega + C_3 \sin \Omega \cos i) + \{\sin i \sin \theta + \cos i \cos \theta \cos (\Omega - \psi)\}(E_2 \cos \Omega \cos i + E_3 \sin \Omega + C_2 \sin \Omega \cos i - C_3 \cos \Omega)] \Big\} + \mathbf{j} \Big\{ A \cos \psi - B \sin \psi + 3[\cos (\Omega - \psi)(E_1 \sin \Omega + E_3 \cos \Omega \cos i - C_1 \cos \Omega + C_3 \sin \Omega \cos i) - \cos i \sin (\Omega - \psi)(E_2 \cos \Omega \cos i + E_3 \sin \Omega + C_2 \sin \Omega \cos i - C_3 \cos \Omega)] \Big\} \Big].$$

Using the magnetic moment program, it has been shown in the case of the Telstar satellites that letting $M = v$ produces precessional results that follow the case II approximation over 1,000 orbits to within 0.01°. To document this result, the residual magnetic moment used was -0.9 microweber-meter, spin rate equaled 20 to 10 radians per second, spin axis and transverse moments of inertia were 4 slug-feet^2, orbit perigee was set to 4,500 miles, and eccentricity ranged from 0 to 0.95. Slight changes in the x-y torques of the order of thousandths of a microfoot pound were observed as the principal differences in the case I and II approximations for these eccentricity ranges. While these differences are negligible for the Telstar I and II satellites, other satellites in sufficiently lower orbits or having greater residual magnetic moments could require the case II approximation (no complete study has been made to date to bound the required ranges of the above mentioned variables for case I to achieve agreement to within 0.01° of case II).

4.7 *Equations of Motion of a Body Symmetrical about Its Spin Axis*[16]

The mean gravity and magnetic torque equations have been derived in the previous section. To analyze the motion of a spin-stabilized satellite responding to these torques, certain gyroscopic equations must now be developed. We begin by relating the vector angular momentum, $\mathfrak{C}$, for any rotating body to the external forces acting on the body as

$$\mathfrak{C}' = \mathbf{N} \tag{142}$$

where

$\mathbf{N}$ = resultant moment of all external forces acting on the body
$\mathfrak{C}'$ = the time derivative of $\mathfrak{C}$.

Using the assumptions in 4.3, we write in the SANOR system

$$\mathcal{H} = \dot{\varphi} I_3 \mathbf{k} \tag{143}$$

where

$\dot{\varphi}$ = the satellite spin rate.

From this, it follows that (see Fig. 10)

$$\mathcal{H}' = I_3[\ddot{\varphi}\mathbf{k} + \dot{\varphi}(\boldsymbol{\omega} \times \mathbf{k})] \tag{144}$$

where

$\boldsymbol{\omega}$ = angular velocity of the SANOR system referenced to IS coordinates.

We may express $\boldsymbol{\omega}$ in the SANOR system as (see Fig. 10)

$$\boldsymbol{\omega} = \omega_x \mathbf{i} + \omega_y \mathbf{j} + \omega_z \mathbf{k} \tag{145}$$

or

$$\boldsymbol{\omega} = \dot{\theta}\mathbf{i} + \dot{\psi} \sin \theta \mathbf{j} + \dot{\psi} \cos \theta \mathbf{k} \tag{146}$$

where

ω_x, ω_y, ω_z = angular velocity about the a, y, and z axes respectively.

Combining (142), (144), and (146), after performing the indicated operations

$$\mathbf{N} = I_3(\dot{\varphi}\dot{\psi} \sin \dot{\theta}\mathbf{i} - \varphi\dot{\theta}\mathbf{j} + \ddot{\varphi}\mathbf{k}). \tag{146}$$

Expressing (146) as x, y, and z torques in the SANOR system,

$$T_x = I_3\dot{\varphi}\dot{\psi} \sin \theta \tag{147}$$

$$T_y = -I_3\dot{\varphi}\dot{\theta} \tag{148}$$

$$T_z = I_3\ddot{\varphi}. \tag{149}$$

Quite obviously it is the x and y torques which produce precession. These torques, by the assumptions of Section IV, are the sums of the x and y components of the gravity and magnetic torques expressed by (43) and (141). T_z is zero as a result of assuming zero magnetic moment transverse to the spin axis and mass symmetry about that axis. We note that a nonzero T_z implies a change in the spin rate. For the Telstar I and II satellites this takes place principally because of induced eddy currents which produce transverse moments.*

* The general equations of motion referred to the center of mass for a rigid body spinning about the z axis, and symmetrical about this axis, are

4.8 *Alternate Inertial Coordinate System*

Clearly, (143) is valid only if the angular momentum vector, $\mathfrak{K}$, coincides with the satellite spin axis **k**. We note in (147), which was derived from (143), that the above assumption will be approximated only if $\dot{\psi} \sin \theta$ is small as compared to $\dot{\varphi}$. Otherwise the direction of $\mathfrak{K}$ shall certainly be influenced by that term as well as $\dot{\varphi}$. We note too that a valid condition within the bounds of the assumption is for $\dot{\psi} \sin \theta$ to be small even with large $\dot{\psi}$ provided only that θ itself be appropriately small. That is, solutions to the equation of motion exist for large $\dot{\psi}$, and those will occur only for small θ because of (143).* But if θ be near 0° or 180°, a singularity exists in (147), for $\dot{\psi}$, as a result, approaches infinity. An exit from this dilemma may be secured simply by transforming to a new reference set of inertial coordinates in place of the IS system whenever θ becomes small. Naturally, the new set should be chosen so that the equivalent θ then existing will be large. This is accomplished by redefining IS so that

$$X_2 \text{ corresponds to } Y$$

$$Y_2 \text{ corresponds to } Z$$

$$Z_2 \text{ corresponds to } X$$

as shown in Fig. 15. Let us call this new inertial frame the IS2 system.

Transformation equations to relate IS to IS2 are quite simple and are given below:

$$X_2 = QX \tag{150}$$

$$X = \bar{Q}X_2 \tag{151}$$

$$T_x = I\ddot{\theta} + (I_3 - I)\dot{\psi}^2 \sin \theta \cos \theta + I_3\dot{\varphi}\dot{\psi} \sin \theta$$

$$T_y = I\ddot{\psi} \sin \theta + (2I - I_3)\dot{\psi}\dot{\theta} \cos \theta - I_3\dot{\varphi}\dot{\theta}$$

$$T_z = I_3(\ddot{\psi} \cos \theta - \dot{\varphi}\dot{\theta} \sin \theta + \ddot{\varphi}).$$

For cases where $\dot{\varphi}$ dominates , $\dot{\theta}\dot{\psi} \sin \theta$, and $\dot{\psi} \cos \theta$, these torques reduce to those given by (147), (148) and (149). We note principally that a change in spin rate reflects a T_z processional torque, from the above equations, but this is generally small and is neglected in this paper. For Telstar I, this torque component is estimated at least two orders of magnitude below the magnetic torques considered herein. (See also Ref. 17.)

* This is the case for precession through or near the north or south celestial pole, where even small changes in attitude, or θ, can produce large changes in ψ. The situation is quite analogous to an azimuth-elevation antenna tracking a satellite that passes through or near the zenith, where the azimuth rates become extremely high even for small changes in satellite position.

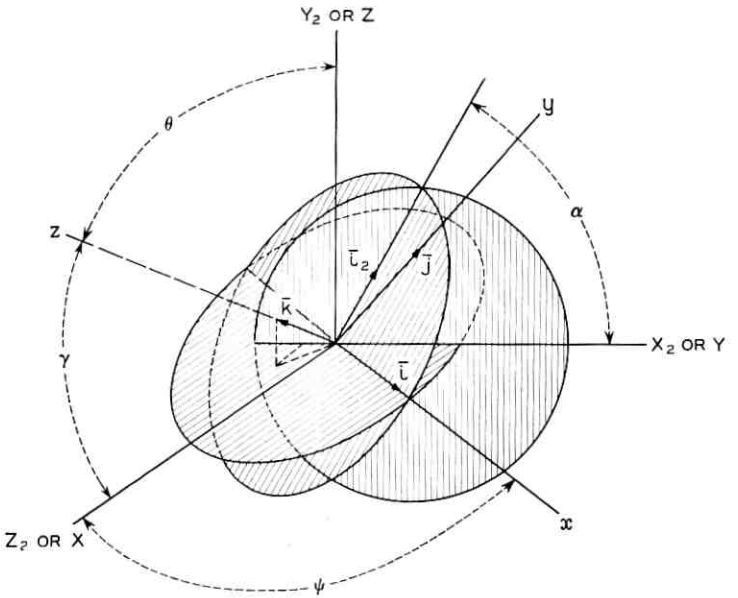


Fig. 15 — Euler rotations in the alpha-gamma system.

where

$$Q = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} \tag{152}$$

and

$$\tilde{Q} = \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}. \tag{153}$$

Relationships between the other coordinate systems and the IS2 system may be obtained using equations of Section 4.4 and applying the Q matrix as appropriate, except that in the expanded matrices

$$\alpha \text{ corresponds to } \psi$$

$$\gamma \text{ corresponds to } \theta$$

where

α = The Euler angle measured from X_2 to the intersection of the xy and X_2Y_2 plane, which defines a new x axis called the x_2 axis in a SANOR 2 system. (See Fig. 15.)

γ = The Euler angle measured from Z_2 to **k** in a plane perpendicular to the x_2 axis. (Rotation is about the x_2 axis.)

Quite obviously, the steps leading to the mean gravity and mean magnetic torques may be retraced using the Q matrix and the new α,γ Euler angles. This would lead to torque expressions as functions of these desired angles. The same result may be obtained by transforming the final torques expressed in the ψ-θ system to expressions in the α-γ system by using explicit relationships between these two systems. We shall choose this later route and the needed relationships will now be derived.

We note in Fig. 15 that the **k** (satellite spin) vector remains in the same spatial position whether expressed in the IS or IS2 system. It seems reasonable then to proceed to relate α-γ to ψ-θ by observing the projections of the **k** vector on the XY and X_2Y_2 planes, respectively.

Referring to Fig. 16, we see that

$$\tan \alpha = \frac{-\sin \theta \cos \psi}{-\cos \theta} \tag{157}$$

$$\cos \gamma = \sin \theta \sin \psi \tag{158}$$

$$\sin \gamma = (1 - \sin^2 \theta \sin^2 \psi)^{\frac{1}{2}}. \tag{159}$$

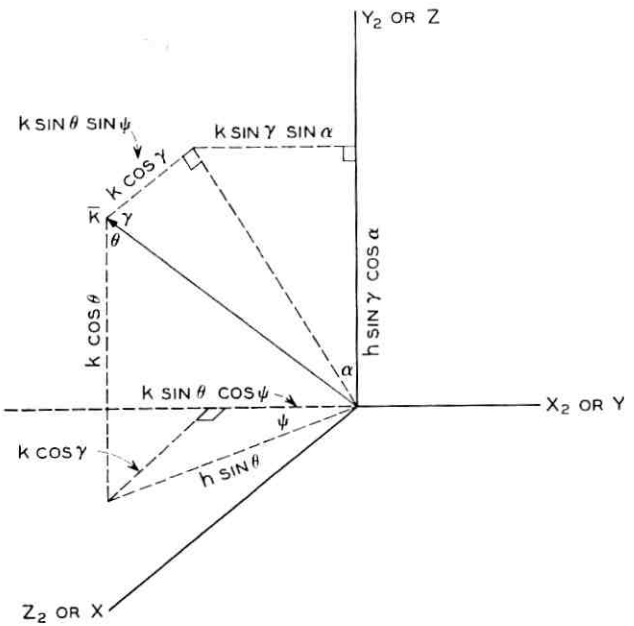


Fig. 16 — Theta-psi and alpha-gamma geometric relationships.

Quite obviously, inverse relationships may be developed. These are

$$\tan \psi = \frac{\cos \gamma}{-\sin \alpha \sin \gamma} \tag{160}$$

$$\cos \theta = -\cos \alpha \sin \gamma \tag{161}$$

$$\sin \theta = (1 - \cos^2 \alpha \sin^2 \gamma)^{\frac{1}{2}}. \tag{162}$$

4.9 *Mean Gravity and Magnetic Torques — Alternate Euler Rotations*

Using (157) through (159) it can be shown that the mean gravity torque expressed in the α-γ system is

$$\begin{aligned}(\mathbf{T}_G)_{\text{mean}} = \frac{6\pi^2 \Delta I}{T_n^2(1-e^2)^{\frac{3}{2}}} \Big\{ & (-\sin \Omega \sin i \cos \gamma \\ & + \cos \Omega \sin i \sin \alpha \sin \gamma + \cos i \cos \alpha \sin \gamma) \Big\} \\ & [\mathbf{i}_2(\sin \Omega \sin i \sin \gamma + \cos \Omega \sin i \sin \alpha \cos \gamma \\ & + \cos i \cos \alpha \cos \gamma) + \mathbf{j}_2(\cos \Omega \sin i \cos \alpha \\ & - \cos i \sin \alpha)] .\end{aligned} \tag{163}$$

Likewise, the mean magnetic torque can be expressed as

$$\begin{aligned}(\mathbf{T}_m)_{\text{mean}} = {} & \pi \mathfrak{M} \cos \beta \Big[\mathbf{i}_2 \Big\{ -2 \cos \alpha \cos \gamma + 3 \sin i(\sin i \cos \alpha \cos \gamma \\ & - \sin \Omega \cos i \sin \gamma - \cos \Omega \cos i \sin \alpha \cos \gamma) \Big\} \\ & + \mathbf{j}_2 \Big\{ 2 \sin \alpha - 3 \sin i(\sin i \sin \alpha \\ & + \cos \Omega \cos i \cos \alpha) \Big\} \Big] \\ & + \mathfrak{M} \sin \beta \Big[\mathbf{i}_2 \Big\{ -A \sin \gamma \\ & - B \sin \alpha \cos \gamma + 3G_1(\sin \Omega \sin \alpha \cos \gamma \\ & - \cos \Omega \sin \gamma) + 3G_2(-\sin i \cos \alpha \cos \gamma \\ & + \sin \Omega \cos i \sin \gamma + \cos \Omega \cos i \sin \alpha \sin \gamma) \Big\} \\ & + \mathbf{j}_2 \Big\{ -B \cos \alpha + 3G_1 \sin \Omega \cos \alpha \\ & + 3G_2(\sin i \sin \alpha + \cos i \cos \Omega \cos \alpha) \Big\} \Big]\end{aligned} \tag{164}$$

where

$$G_1 = E_1 \sin \Omega + E_3 \cos \Omega \cos i - C_1 \cos \Omega + C_3 \sin \Omega \cos i \tag{165}$$

$$G_2 = E_2 \cos \Omega \cos i + E_3 \sin \Omega + C_2 \sin \Omega \cos i - C_3 \cos \Omega. \tag{166}$$

All other terms have meanings described previously.

V. EXPERIMENTAL RESULTS — GENERAL

Using the methods described herein, the attitudes of the Telstar I and II satellites have been predicted, determined, and these two results compared. In the case of the Telstar I satellite, the maximum angular deviation between predicted and determined attitude is 0.90° with the special-case exception of pass 2154, which is covered in detail later.* The average angular deviation is 0.38°. The predicted attitude of the Telstar II satellite deviates from that determined by optical flashes and solar sensors by a maximum of 0.5° and an average of 0.09°. These results are tabulated in Tables II and III. Here the deviation in right ascension and declination is displayed by subtracting the predicted values from the determined values. The angular deviation between predicted and determined attitude is also given as the celestial great circle arc G. When this value is followed by N, this indicates that no attitude fix has been determined by loci intersections (see Section II), and so the shortest arc distance between the corresponding attitude locus and the prediction is quoted. The T preceding an attitude determination indicates the selection of that point as a target toward which a magnetic moment program embodying the above analysis attempts to converge.

5.1 *Experimental Results — The Telstar I Satellite*

The Telstar I satellite entered orbit on July 10, 1962. Its attitude, calculated by combining nominal third-stage burnout parameters and stage 1 and 2 telemetry data, is given in Table I. Between launch and pass 16, there were a number of instances when the orientation coil was inadvertently energized by misinterpreted command signals sent to the satellite at extreme ranges and/or low elevations. This orientation coil (often referred to as a "torque coil") is located in the equatorial plane of the Telstar satellite just under the outer skin. Its purpose is to enable attitude adjustments by the production of a magnetic moment whenever current is sent through the coil. A fixed amount of current may be caused to flow in either direction through the coil so as to produce a magnetic moment, in addition to the residual moment, of ± 7.8540 microweber-meters. Fig. 17 diagrams the sense of the magnetic vectors which are called positive, indicates the resulting north and south magnetic poles of the satellite, and shows the corresponding positive direction of the current flow in the orientation coil. The relative magnetic directions and current flow are in accord with established standards.

* There are apparent exceptions on passes 16, 72 and 1657, but these relate to antenna pattern techniques (see Section 5.1) which have an expected accuracy of $\pm 1°$ to $\pm 2°$.

TABLE II—TELSTAR I SATELLITE—ATTITUDE COMPARISON

Pass Number	Attitude Determination		Type Data for Determ.	Attitude Prediction		Deviation (Det.—Predict.)			MAG MOM (MWM)
	RA	DEC		RA	DEC	RA	DEC	G	
Burnout				83.7	−66.8				
16	84.2	−65.8	SM	84.2	−65.18	0	0	0	−0.86328
16	83.0	−66.0	A	84.2	−65.18	−1.2	−0.82	1.4	
72				87.78	−66.27			0.5N	
72	86.0	−67	A			−1.3	−0.73	1.2	
135-136	91.7	−66.0	M	92.19	−66.36	−.49	0.36	0.6	
199	single locus		M	96.49	−65.95			0.4N	
271	100	−65	A	100.80	−64.92	−0.8	−0.8	0.5	
272	single locus		M	100.90	−64.90			0.6N	
272	↓	↓	S	100.90	−64.90			0.2N	
472				107.50	−59.80			0.2N	
931	T100.5	−49.84	M	100.74	−49.93	−0.24	0.09	0.18	
									−0.86328
1051	single locus			97.25	−50.21			0.9N	+6.9907
									−0.86328
1069				98.42	−49.70			0.6N	−8.7173
1114				95.92	−51.01			0.1N	−0.71103
1430				92.62	−59.55			0.1N	
1567	T96.2	−64.2	M	96.71	−64.21	−0.51	−0.01	0.23	
									−0.71103
1657	99.2	−67.0	A	102.81	−66.64	−3.61	0.36	1.7	7.14297
									−0.59101
1695				78.58	−57.01				
1909				90.59	−59.71			0.7N	
2154-55	T104.0	−53.5	M	104.71	−54.81	−0.71	−1.3	1.5	
2200	single locus			106.03	−53.28			0.1N	
2264				107.20	−50.74			0.4N	
2464				107.09	−43.65			0 N	
2482				106.88	−43.04			0 N	
2509				106.57	−42.33			0 N	
2582-83	105.2	−40.2	M	105.44	−40.53	−0.24	0.38	0.4	
3340, 41, 42	106.3	−43.7	M	103.68	−61.54	−2.62	17.8	18.0	
3476-7	112.30	−46.32	M	120.21	−66.21	7.91	19.9	22.0	
3495				123.16	−66.41				
3476-7	T112.30	−46.32	M	112.30	−46.32†	0	0	0	−0.40945

M = Mirror flash data.
A = Antenna pattern data.
S = Solar sensor data.
† Connects 3340 to 3476.

Data relating to the inadvertent uses of the orientation coil are insufficient to reconstruct the detailed precessional motion of Telstar I from launch to pass 16. On pass 16, however, an attitude determination was made by reducing optical and solar sensor data. This along with an attitude fix on pass 135–136 made from mirror flash data indicated a magnetic moment of about −0.7 microweber-meter. This was later refined to −0.86328 microweber-meter by causing the magnetic moment program to connect, by the precessional theory herein described, the

TABLE III—TELSTAR II SATELLITE—ATTITUDE COMPARISON

Pass Number	Attitude Determination		Type Data	Attitude Prediction		Deviation (Det.-Predict.)			MAG MOM (MWM)
	RA	DEC		RA	DEC	ΔRA	ΔDEL	G	
Burnout				82.23	−57.31				
62, 63, 63	87.75	−55.08	M	87.75	−55.08	0	0	0	−0.48375
100				88.55	−55.05	0	0	0	−0.48375
132, 133, 133	88.75	−55.34	M	89.23	−54.97	−0.48	−0.37	−0.5	
281				92.17	−54.10			0N	
293, 293	92.3	−54.05	M	92.38	−53.99	−0.08	−0.06	0.1	
324, 25	92.40	−53.71	M	92.90	−53.71	0	0	0	
331				93.01	−53.66			0.08N	
490				94.95	−51.81				
496	T95.00	−51.73	M	95.00	−51.73			0N	
541				95.34	−51.10			0N	
573				95.49	−50.74			0N	
611				95.63	−50.20			0.13N	
643				95.71	−49.80			0.16N	

M = Mirror flash data.
S = Solar sensor data.

attitude on pass 16 to a point on the pass 931 attitude locus. The convergence is within 0.18°, as shown in Table II. Using this latter magnetic moment, the precessional motion from passes 16 to 1051 was established. This precessional history is given in Fig. 18, where both the predicted curve and the attitude loci data are displayed.

When an optical flash series is recorded at the Holmdel Laboratories, the midpoint of that series can be determined with a time accuracy of less than ±20 seconds when using visual observation through the on-site telescopes and within ±10 seconds photoelectrically. The early data recorded on the Telstar I satellite did not indicate the expected accuracy at the midpoints, so in Fig. 18 only central loci are plotted. We must therefore think of these loci as having possible tolerances up to ±10 seconds, since photoelectric data were reduced. A tolerance of ±10 seconds can cause the maximum locus limit to be displaced as much as ±0.7° on either side of the plotted central loci. Considering this, the predicted attitude curve passes through every loci and every loci intersection on Fig. 18 with the single exception of pass 1051. The prediction curve would miss the lower limit of this locus by 0.3°.

Extended attitude predictions were made on October 1, 1962 (around pass 800) and covered the period through January 22 (pass 1800). These indicated that attitude adjustments should be initiated during

January, 1963. This was needed to prevent the solar offset from exceeding about 15° as required (see Section I). In order to check the attitude correction procedures, a test maneuver was initiated on pass 1051. It began by turning on the orientation coil, so as to produce a magnetic moment of +7.8540 microweber-meters to oppose the −0.86328 residual moment and yield a resultant of +6.9907 microweber-meters. This was continued through pass 1058, whereupon the coil was turned off until pass 1069. On 1069 the coil was turned on in the opposite sense so as to produce a moment of −7.8540 microweber-meters, giving the satellite a net moment of −8.7173 microweber-meters. On pass 1075, the coil was turned off. All told, the coil was positive for about 18 hours and negative for about 16 hours, a fact which approximately nullified the

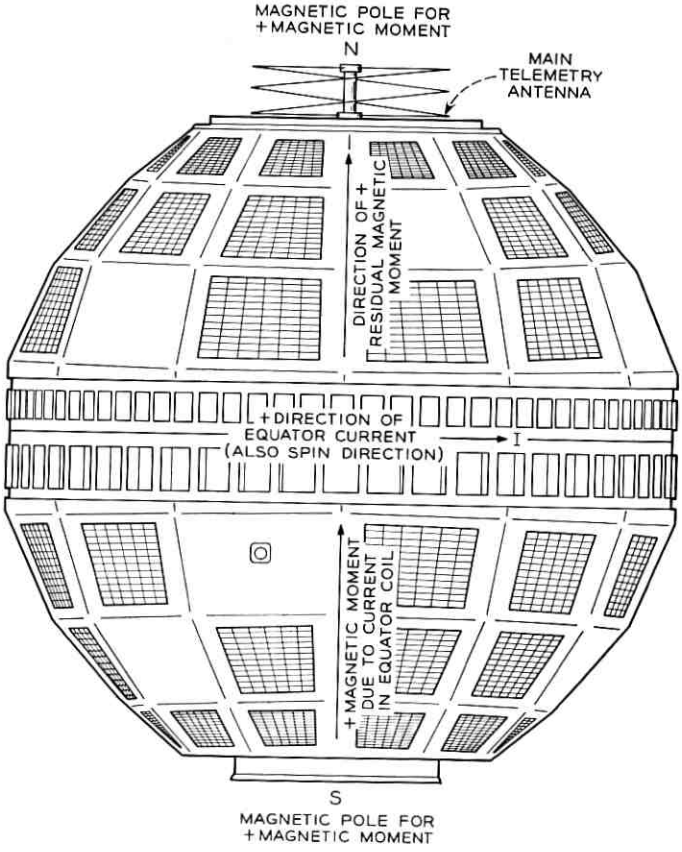


Fig. 17 — Direction of positive magnetic moments for Telstar satellites.

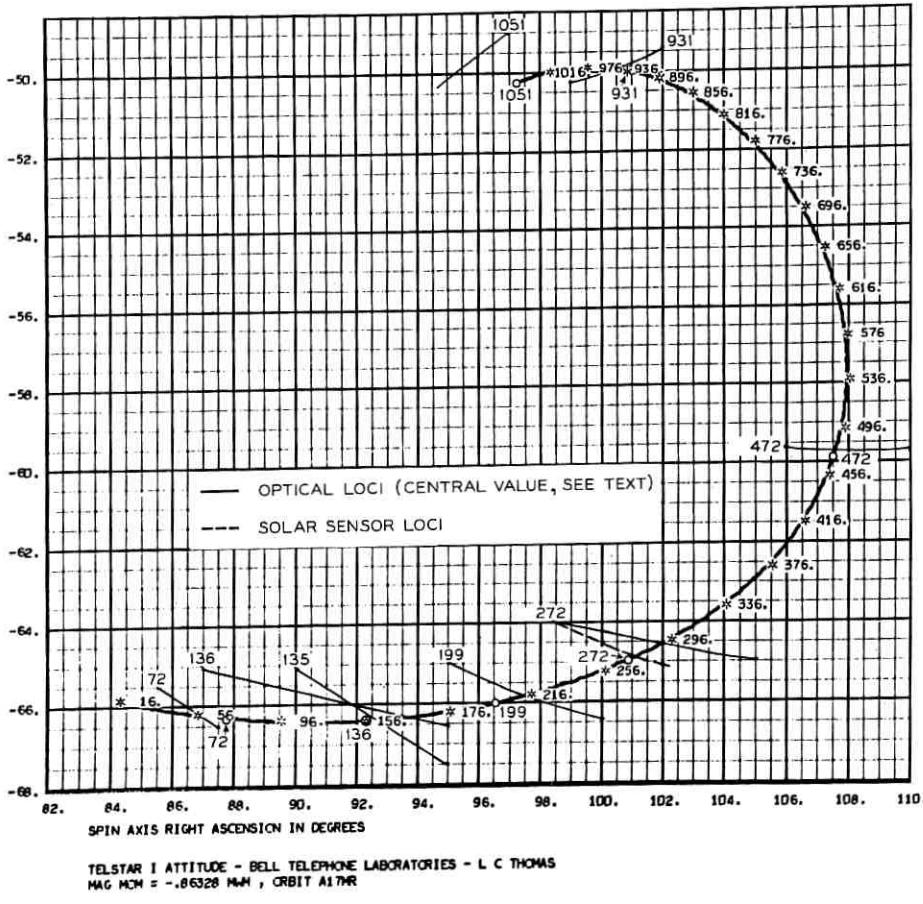


Fig. 18 — Telstar I precessional history — passes 16 through 1051.

precessional effect of the coil as intended for this test. Details of this maneuver are shown in Fig. 19.

Table II summarizes the agreement of loci to prediction during this period. Giving loci of unknown tolerances a $\pm 0.7°$ spread and using the known tolerance for pass 1114, the residuals shown in Table IV are produced.

The magnetic moment program connected the attitude on pass 1075 to that determined by mirror data on pass 1567 within 0.23°. In so doing, it calculated a residual magnetic moment of -0.71103 microweber-meter for that era. Using this calculated moment, predictions are extended to pass 1657 where, on January 7, 1963, the main orientation

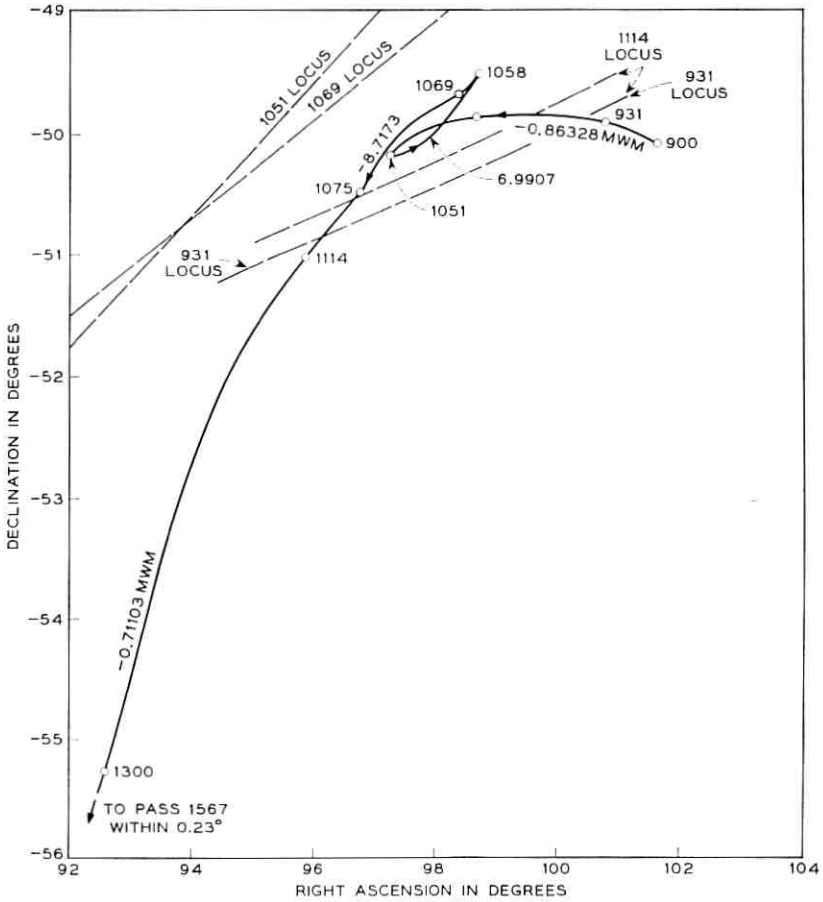


Fig. 19 — Details of the trial torqueing maneuver.

TABLE IV — TELSTAR I SATELLITE — TEST MANEUVER

Pass Number	Deviation of Prediction from Optical Locus
1051	0.3°
1069	0°
1114	0.1°

maneuver began. The region from pass 1075 to 1657 is shown in Fig. 20 along with the attitude loci. Predictions fit the determinations throughout within about 0.2°.

In addition to using mirror flash and solar sensor data to determine attitude, W. C. Jakes, Jr. and group at the Holmdel Laboratories deduced the attitude for a number of passes in this period by analyzing antenna pattern data.[18] Some of these determinations are included in Table II. Since their expected accuracy is ±1°, these data also serve to corroborate the predicted attitude.

The orientation coil was energized for the major attitude maneuver

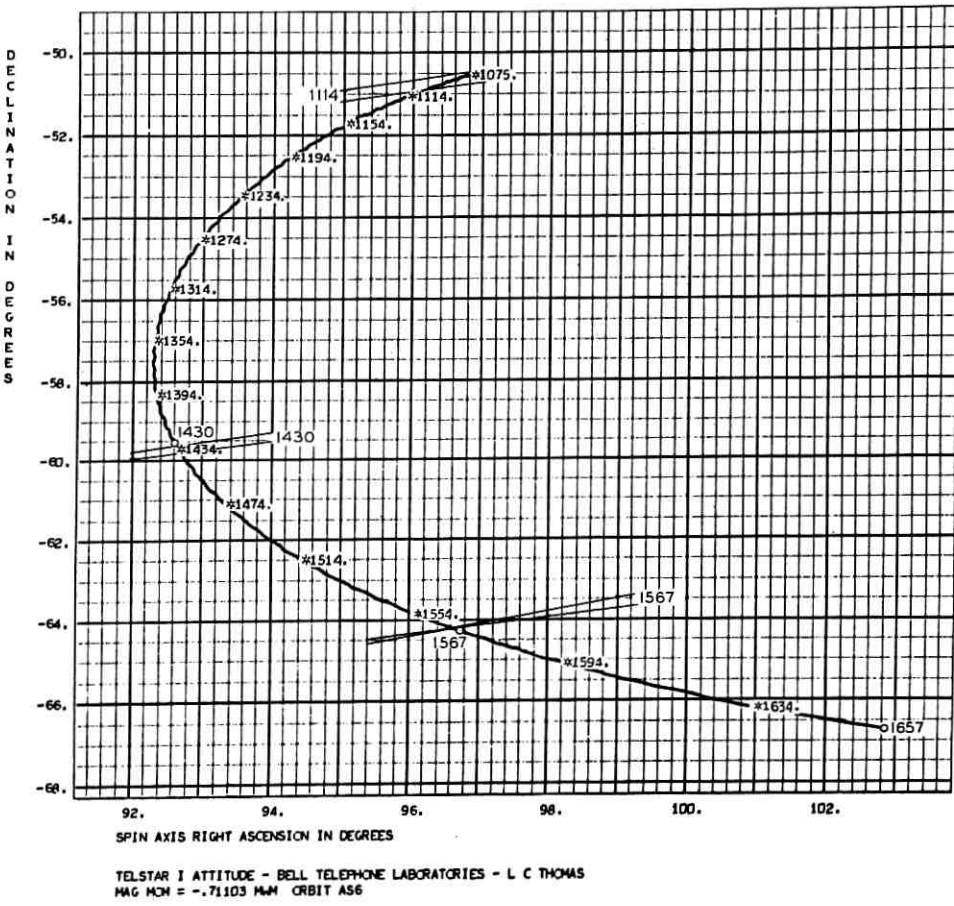


Fig. 20 — Telstar I precessional history — passes 1075 to 1657.

beginning on pass 1657, January 7, 1963, and ending on pass 1695, January 11, 1963. During this time the residual magnetic moment of −0.71103 microweber-meter had superimposed upon it a 7.8540 microweber-meter moment as a result of the coil. The attitude of the Telstar I satellite was changed by over 20° during this maneuver. The attitude and curve for this period are given in Fig. 21. The slight ripples in these curves are due to the inclination of the earth's dipole. Because of the small scale of this plot, these ripples are more noticeable than in the previous graphs.

At this point, we might well insert a brief discussion of the planning

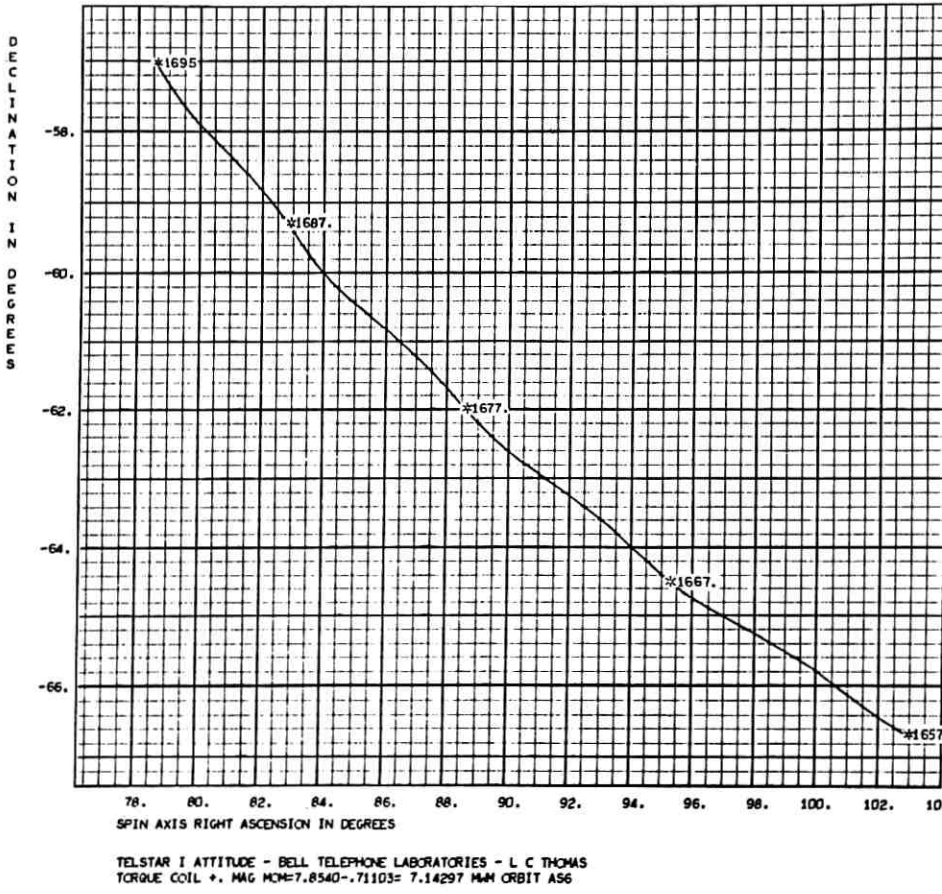


Fig. 21 — Telstar I precessional history — passes 1657 to 1695.

behind this attitude maneuver and why pass 1657 was chosen for its commencement. First of all, it can be shown that for any angular offset in the spin axis from a fixed, desired orientation there exists, for each instant in time, a determinable polarity of voltage for the equatorial coil which will tend to decrease the offset. To utilize this principle fully would require turning on the orientation coil on many occasions when the Telstar I satellite entered the Andover skies. While such a procedure would tend to decrease the angular offset generally in the most direct fashion, there are occasions in the lifetime of the Telstar satellite when leaving the coil on for a number of revolutions will produce faster corrections than pulsing the coil continually at Andover. This is because the instantaneous torque produced by the continuous magnetic moment of the coil in the earth's magnetic field varies during an orbital period in such a way that, in general, the average value exceeds the torque obtainable by energizing the coil only in the Andover skies. The optimum procedure then is for a period of continuous operation and this was undertaken. Fig. 22 shows the results of turning on the coil for steady torqueing at various dates. The heavy spiral shows the predicted position of the spin axis in right ascension and declination if the coil is never energized (the trial maneuver is omitted here for simplicity). The diverging curves indicate the motion due to the torqueing coil. Examination of Fig. 22 shows rapid precession as a result of the torque coil field, so that no mode of continuous coil operation will long permit the spin axis to point close to the south ecliptic pole. The solar offset also must ultimately increase with the enlarging ecliptic angle. It therefore follows that if a simplified mode of coil operation exists, it must begin with a relatively short period of coil operation.

Two factors concerning the precessional motions are noted in Fig. 22. They are:

(*i*) All precessional motions involving the orientation coil lie outside the spiral of residual precession for many orbits, because the net magnetic moment existing when the coil is actuated exceeds the residual magnetic moment.

(*ii*) Torque coil precession for both positive and negative coil polarity begins incrementally at the spiral in opposite directions.

We see that the torque coil should be used sparingly, because of the relatively rapid motion it produces. The problem that remains is to determine the times to turn on and turn off this coil and also to investigate whether or not future coil use is required. Because the spin rate of the satellite is decreasing, equal magnitudes of satellite magnetic moments will cause greater and greater precessional motion as time passes.

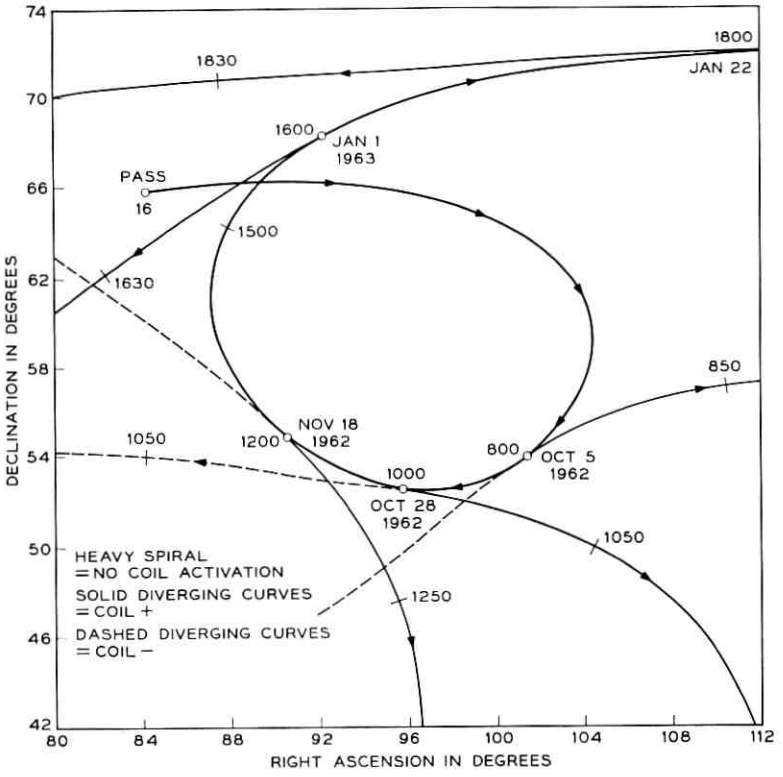


Fig. 22 — Telstar I predicted attitude showing steady torqueing.

It is to be expected that after the coil is actuated and then turned off a spiral of larger excursion than that occurring for passes 16 through 1655 will result strictly because of the lower spin rate existing at the later time. There is no way to avoid this except by repeated coil pulsing.

To avoid, or at least minimize, repeated coil usage, the procedure is to try to keep the maximum ecliptic angle as small as possible. If this value is expected to exceed the allowable solar offset, it is advantageous to time the operation so that the maximum ecliptic angle will occur at the time of minimum solar offset. This criterion governs both the selection of the day for coil turn-on as well as the duration of the attitude maneuver. As may be guessed from Fig. 22 and from the precessional motion factors stated above, only short torqueing operations during the month of January would result in the next precessional spiral (after coil turn-off) returning anywhere near the initial spiral shown in Fig. 22.

To cause the maximum ecliptic angle and the minimum solar offset to occur together, orientation coil use had to begin in the period from January 1 to January 7, 1963. As shown above, the maneuver which did begin on January 7 placed the attitude at right ascension 78.6° and declination at 57.0° by pass 1695.

The magnetic moment program, in connecting pass 1695 attitude to that determined by mirror data on passes 2154 and 2155, determined a residual magnetic moment of −0.59101 microweber-meter. This differs by about 0.1 microweber-meter from the residual moment going into the torqueing maneuver, but appears to be borne out by the closeness of fit to the mirror data from pass 1695 through pass 2583 (see Table II).* Fig. 23 plots this trajectory along with the mirror loci. The fit of attitude predictions to determinations is very good from passes 1695 through 2583 (see Table II). The largest deviation of 1.5° on passes 2154–2155 may be somewhat misleading. Referring to Fig. 23, we see that the predicted attitude on pass 2154 is easily within 0.1° of the 2154 locus. The interaction of the 2154 and 2155 loci however, produces a common area, or attitude box, 1.5° from the pass 2154 prediction. This box is produced by a somewhat grazing intersection of the loci, and its location is therefore affected by the precession which took place between these two passes. Antedating the 2155 locus to account for this would place the attitude box within 0.05° of the 2154 and 2155 prediction, so this fit is quite valid.

We are not in so comfortable a position for passes 3340 and thereafter. Attitude boxes for 3340–1 and 3476–7 are shown as dashed lines on Fig. 23. They are about 20° from the corresponding predicted attitudes, and the reason for this discrepancy is at present unknown. We note that the Telstar I satellite ceased its transmission on pass 2065 (February 20, 1963) because of radiation damage, but that the attitude remained predictable at least through pass 2583. The region from pass 2583 to 3340 is devoid of attitude data because no telemetry could be received to report solar aspect and no mirror flashes were recorded. This was due to the increased activity on Telstar II, a certain proportion of Telstar I passes occurring in daylight hours, and prevailing weather conditions at Holmdel.

Interestingly enough, it is noted that the attitude determinations on passes 3340–1 and 3476–7 can be connected by the magnetic moment program even though they cannot be sensibly joined to the preceding

* It is possible that coil usage can alter the residual moment of the satellite, which was made small in the first place by appropriately balancing much larger magnetic fields.[19]

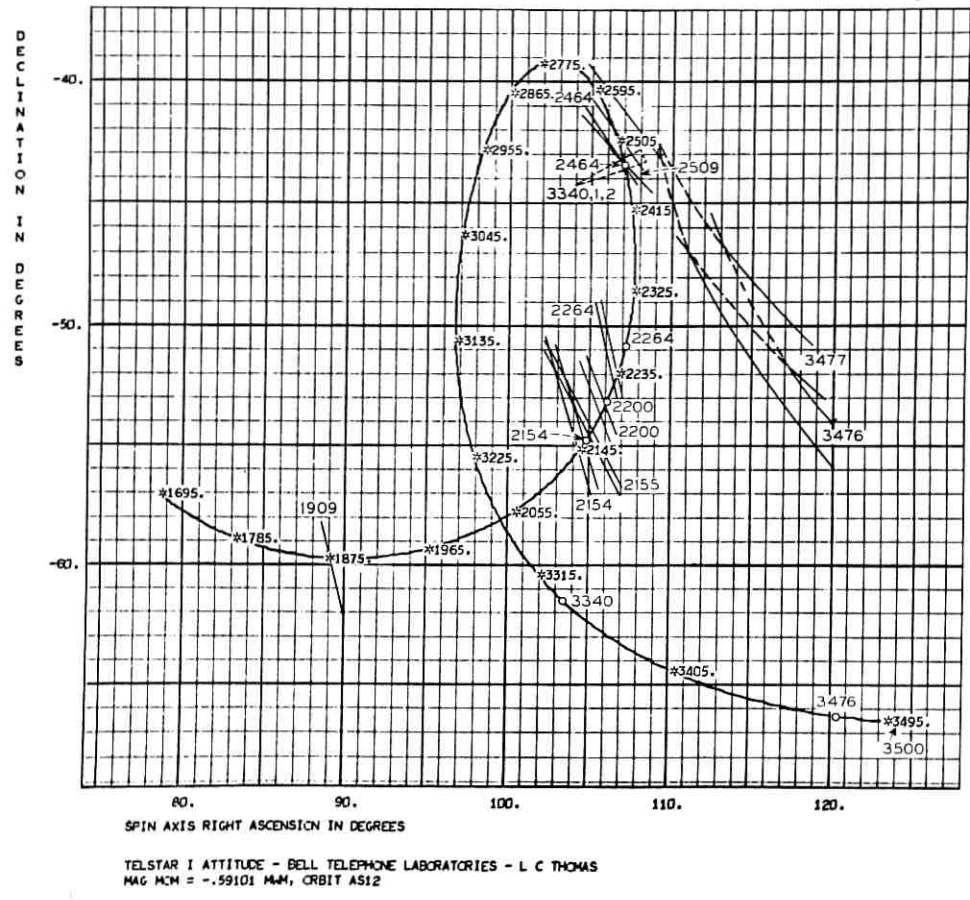


Fig. 23 — Telstar I precessional history — passes 1695 to 3500.

data. Fig. 24 shows this union. The magnetic moment of -0.59101 microweber-meter used in these last three plots cannot be considered too reliable, since it is one of a number which can connect the two attitude boxes shown in Fig. 24. However, it is the same as previously used in the 1695 to 2583 region and may tend to be valid because of this. In any event, more attitude data are required to resolve this issue.

As to what may have caused the anomalous behavior of Telstar I somewhere between pass 2583 and 3440, we offer the following, taken singularly or in combination, as possibilities:

(*i*) the orientation coil has been energized during this period (from

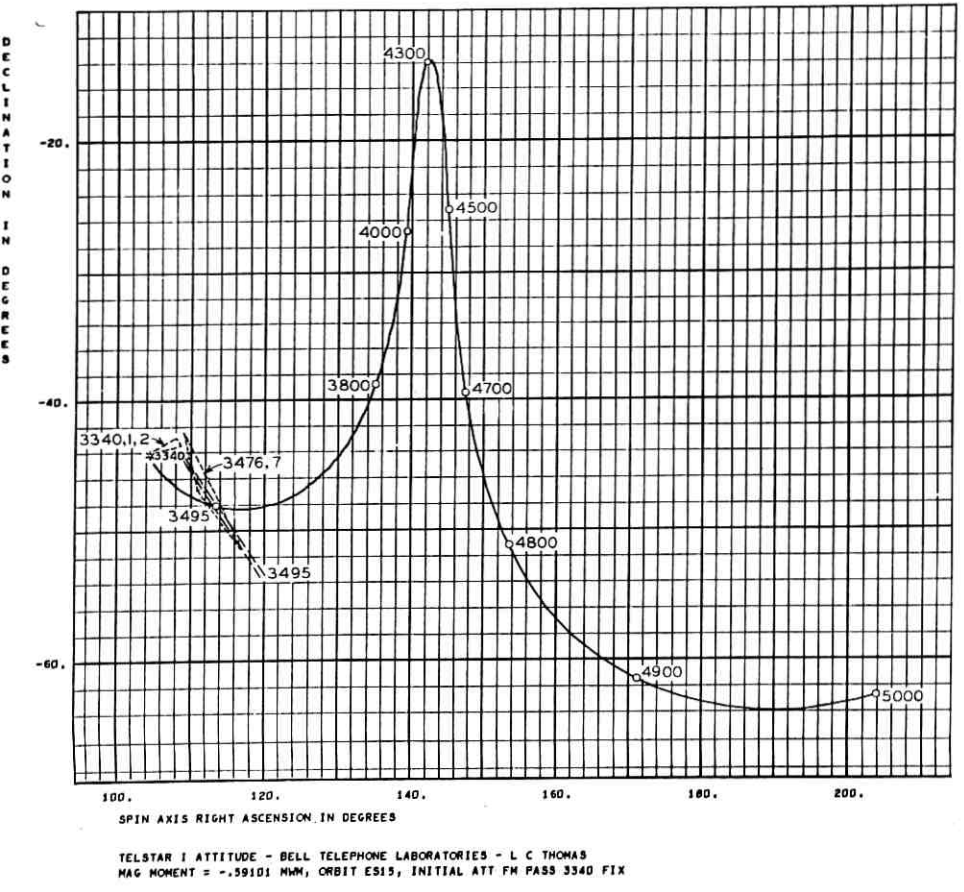


Fig. 24 — Telstar I precessional history and prediction — passes 3340 to 5000.

the present precessional motion, it appears to have been turned off between 3340 and 3477);

(*ii*) the residual magnetic moment has changed a multiplicity of times;

(*iii*) meteoric collision has taken place;

(*iv*) pressure leakage from instrument canister has occurred in a manner to alter the attitude by reaction forces.

5.2 *Experimental Results — The Telstar II Satellite*

The Telstar II satellite entered orbit on May 7, 1963. Its initial attitude calculated from the third-stage burnout parameters is shown

in Table III. The predicted burnout attitude proved inaccurate because the third stage of the Thor-Delta vehicle did not perform nominally. This is evidenced by the fact that the orbital period calculated from such assumed nominal performance differed from the actual period by as much as four minutes.

The first attitude fix of the Telstar II satellite occurred on passes 62 and 63, when a total of three flash series were observed. Connecting this determination to that of pass 496 requires a residual magnetic moment of -0.48375 microweber-meter. The fit through pass 643 is given in Table III, and some typical attitude loci are plotted along with the attitude prediction in Fig. 25. Corresponding ecliptic angle, solar aspect, X and Y torques appear in Figs. 26 through 29.

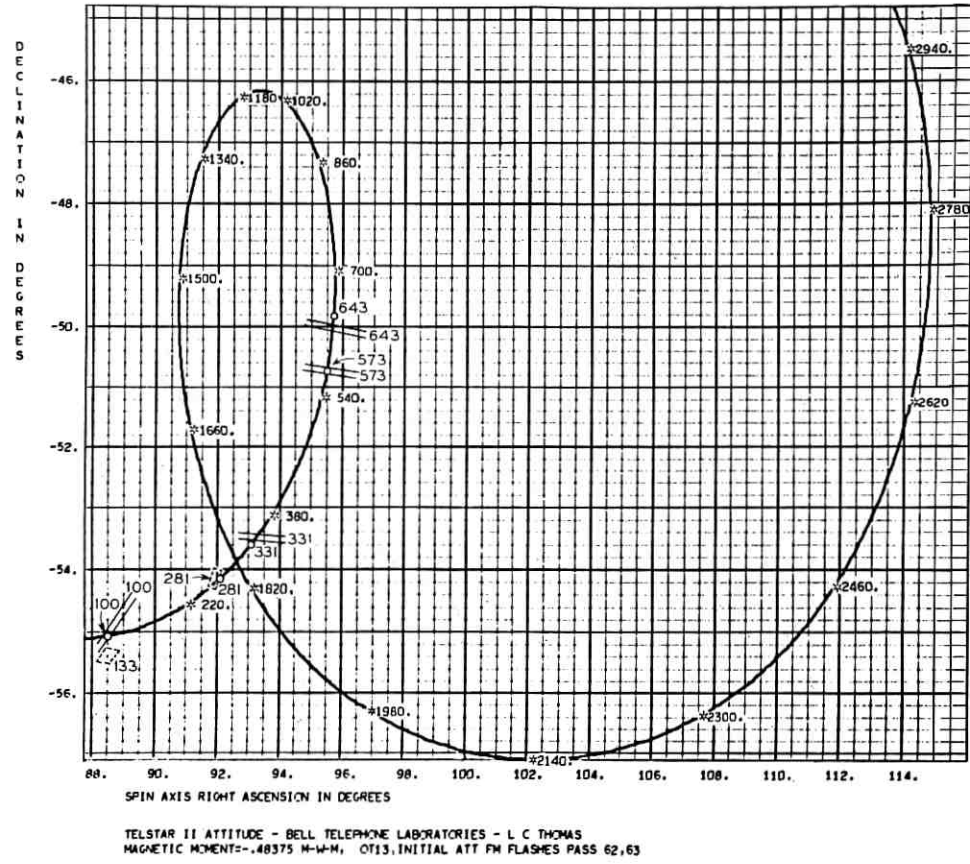


Fig. 25 — Telstar II precessional history and prediction — passes 62 to 3000. (See footnote, p. 1713.)

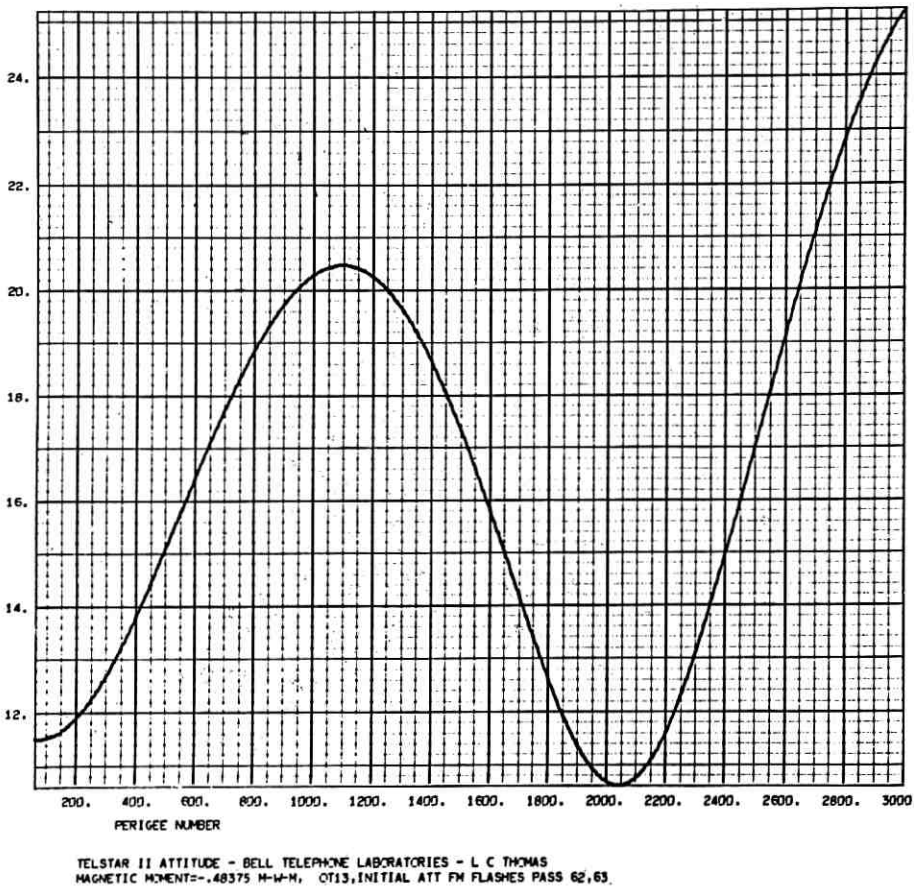


Fig. 26 — Telstar II ecliptic angle — passes 62 to 3000.

The solar offset is due to exceed the 15° limit dictated by temperature balance considerations (see Section I) by pass 1250 (see Fig. 27). Since the maximum offset which occurs on pass 1400 is only 18°, no plans are contemplated to reorient the Telstar II satellite strictly to prevent this mild excursion.

There is, however, another excursion beyond the 15° limit, around pass 2670. Besides this excursion being more serious than the former, some interesting differences between these two events exist, as Fig. 25 illustrates.

On pass 1400, the attitude will be returning to a region nearer the south ecliptic pole (located at right ascension 90°, declination 67.5°)

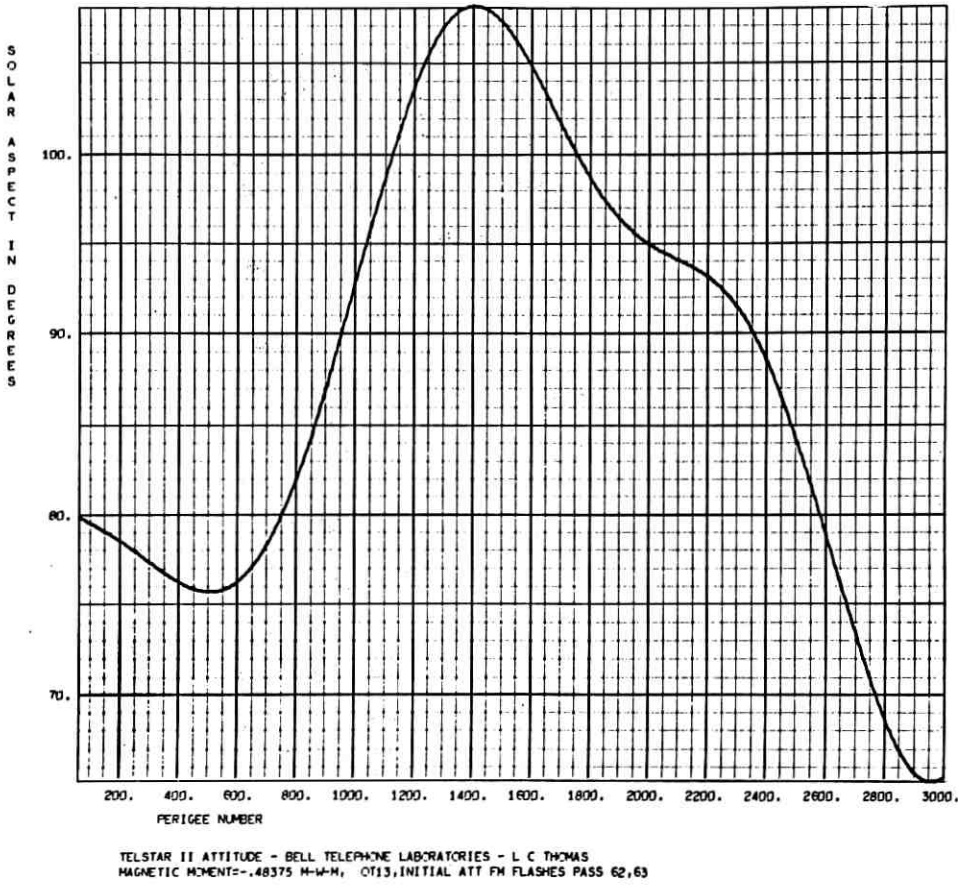


Fig. 27 — Telstar II solar aspect — passes 62 to 3000.

thereby limiting the solar offset to values below 15°. On pass 2670, however, the attitude is moving away from the south ecliptic pole, and, in fact, because of its now lower spin rate is entering a spiral of greater excursion than that previously traversed. It therefore behooves us to correct the attitude before that latter spiral occurs.

The question remaining is to determine the most profitable time for such a correction. Ideally, a well-chosen time would meet the following conditions:

(*i*) It would take place just after a good optical attitude fix was established to verify the predictions.

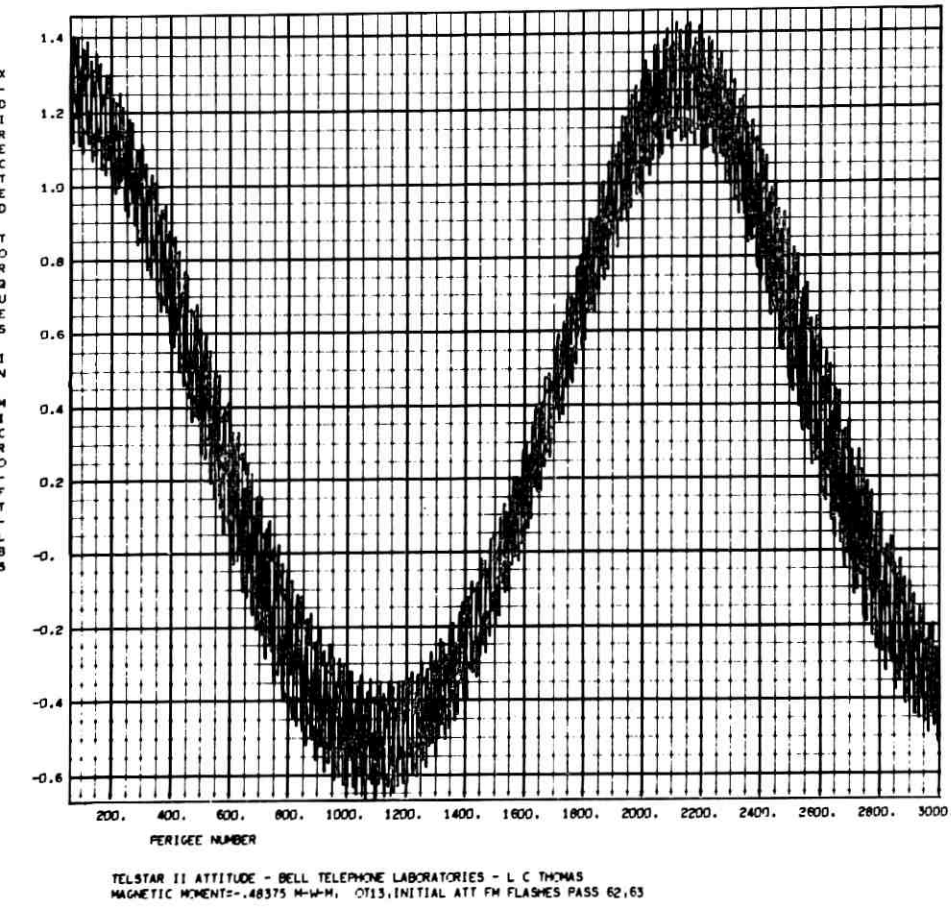


Fig. 28 — Telstar II X torques — passes 62 to 3000.

(*ii*) The attitude maneuver would result in as many mirror flashes as possible during correction to indicate its progress.

(*iii*) The attitude maneuver would end with the satellite in a position to guarantee a suitable attitude for as long a time in the future as is possible.

At the time of this writing, suitable attitude corrections are under study. Most probably the orientation coil will be energized sometime between passes 1300 to 1500 in a negative sense or in a positive sense near pass 2400 so as to drive the attitude downward and to the left in

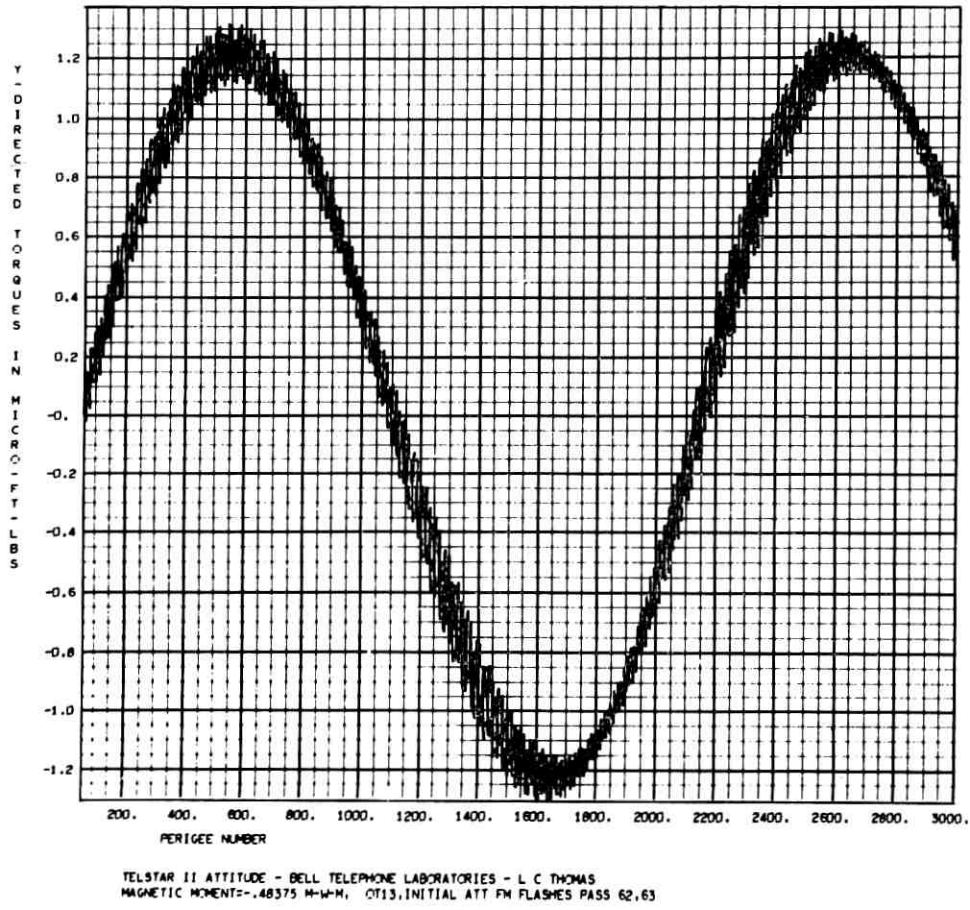


Fig. 29 — Telstar II Y torques — passes 62 to 3000.

the sense shown in Fig. 25. This permits the next attitude spiral to occur near the south ecliptic pole.*

VI. CONCLUSIONS

Techniques for both attitude determination and prediction for spin-stabilized satellites have been developed. Their use has been demon-

* The attitude of the Telstar II satellite was successfully reoriented by energizing the coil on pass 2402 (May 17, 1964) and leaving it in that state until pass 2421. Thus attitude predictions given in Fig. 19 beyond pass 2402 are no longer valid.

strated using Telstar I and II satellite data. It has been shown that an inclined dipole model of the earth's magnetic field and the method of averaging the gravitational and magnetic torques over each anomalistic period of the satellite permit attitude predictions to within a few tenths of a degree of determined values in most instances. In those few cases where departures are above one degree, explanations have been presented to show the reason for such discrepancies. The reasons are (1) unknown time errors in determining the midpoint of the optical flash series and (2) grazing intersections of the attitude determining loci. There remains but one anomaly in the precessional motion of the Telstar I satellite which, for the moment, is unexplained. Possible reasons for this anomaly are given in Section 5.1.

It has further been shown that the seemingly crude approximation of letting the mean anomaly of the satellite equal the true anomaly for the purposes of determining the mean torques produces attitudes in close agreement (0.01° over 1000 orbits) with more sophisticated approximations for orbit eccentricities up to 0.9 and perigee radii above about 4500 miles. (See Section 4.6, case II.)

The comparisons made herein of the precession given by the magnetic moment program and the attitude determinations substantiate the simplifying assumptions made in Sections 4.1, 4.2, and 4.3. These assumptions are most instrumental in producing a working technique for both attitude prediction and residual magnetic moment determination which is amenable to analytic solution and conservative of computer time.

Furthermore, the usefulness of combining optical flash and solar sensor data for attitude determination and their inherent accuracy is shown. Optical flash data can provide loci with a resolution of 0.1°. Solar sensor loci are resolved to within 1°. While it is clearly straightforward to determine analytically the boundaries of the attitude boxes from intersecting loci exhibiting estimated time tolerances, this paper indicates the decided advantage of graphing the individual loci to determine the angle of intersection and thereby gain an estimate of the validity of the boundaries in the presence of precession (see, for example, Section 4.1).

Finally, the techniques described have all been consolidated into working computer programs which follow closely the analysis presented. In addition, a number of important supporting calculations such as the solar position, sidereal time, orbit updating, etc. are developed. Because of the complexities of the mean torque and gyroscopic equations, the precessional analysis is most useful when embodied in suitable computer programs.

VII. ACKNOWLEDGMENTS

The author acknowledges his indebtedness to the following contributors as indicated: to J. F. Ossanna for orbital updating techniques specifically, and in general for helpful discussions concerning computer techniques useful for the calculations; to J. S. Courtney-Pratt, J. W. McLaughlin, and J. H. Hett for the development of techniques and equipment to record the optical flashes at the Holmdel Laboratories, and for their splendid cooperation throughout this project; to W. C. Jakes, Jr., L. R. Lowry, W. E. Legg, R. H. Brandt, P. E. Pheffer, J. W. Barto, and E. L. Frantsvog for providing and staffing Holmdel facilities to acquire and track the Telstar satellites and for useful work in ascertaining the attitude from antenna pattern measurements; to D. W. Hill for initially developing the concept of reducing attitude from mirror flashes and solar sensor data; to R. C. Chapman for patiently reading, correcting and rendering advice on this manuscript; to R. H. Shennum for close cooperation, advice and aid during the period of attitude correction; and to I. Welber for helpful integration of this work into the Telstar satellite communications experiment.

APPENDIX A

Derivation of Equation (40)[21]

Beginning with (39) of the main body of the paper, it is certainly evident that

$$\mathbf{k}\cdot(\mathbf{i}_g\mathbf{i}_g + \mathbf{j}_g\mathbf{j}_g) \times \mathbf{k} = (\mathbf{k}\cdot\mathbf{i}_g)(\mathbf{i}_g \times \mathbf{k}) + (\mathbf{k}\cdot\mathbf{j}_g)(\mathbf{j}_g \times \mathbf{k}). \quad (39a)$$

It will be shown that the following identity exists

$$(\mathbf{k}\cdot\mathbf{i}_g)(\mathbf{i}_g \times \mathbf{k}) + (\mathbf{k}\cdot\mathbf{j}_g)(\mathbf{j}_g \times \mathbf{k}) + (\mathbf{k}\cdot\mathbf{k}_g)(\mathbf{k}_g \times \mathbf{k}) \equiv 0. \quad (39b)$$

Rewriting (39b), one obtains

$$[(\mathbf{k}\cdot\mathbf{i}_g)\mathbf{i}_g + (\mathbf{k}\cdot\mathbf{j}_g)\mathbf{j}_g + (\mathbf{k}\cdot\mathbf{k}_g)\mathbf{k}_g] \times \mathbf{k} \equiv 0 \quad (39c)$$

and

$$(\mathbf{k}\cdot\mathbf{i}_g) = (x_g\mathbf{i}_g + y_g\mathbf{j}_g + z_g\mathbf{k}_g)\cdot\mathbf{i}_g \quad (39d)$$

$$= x_g \quad (39e)$$

where

x_g , y_g , z_g = components of $\mathbf{k}$ along the ORDEF axes.

In like fashion

$$(\mathbf{k}\cdot\mathbf{j}_g) = y_g \tag{39f}$$

$$(\mathbf{k}\cdot\mathbf{k}_g) = z_g\,. \tag{39g}$$

Substituting (39e,f,g) into (39c), we obtain

$$(x_g\mathbf{i}_g + y_g\mathbf{j}_g + z_g\mathbf{k}_g) \times \mathbf{k} \equiv 0 \tag{39h}$$

$$\mathbf{k} \times \mathbf{k} = 0 \tag{39i}$$

$$0 \equiv 0. \tag{39j}$$

Therefore

$$-(\mathbf{k}\cdot\mathbf{k}_g)(\mathbf{k}_g \times \mathbf{k}) \equiv \mathbf{k}\cdot(\mathbf{i}_g\mathbf{i}_g + \mathbf{j}_g\mathbf{j}_g) \times \mathbf{k}, \tag{39k}$$

and (40) follows from (39).

APPENDIX B

Evaluation of the II Integral of Equation (73)[21]

Copying the II integral from (73) of the main body of the paper, we have

$$\text{II} = (\cos\beta)\mathbf{k} \times \int_0^{2\pi} \{\mathbf{K} - 3(\mathbf{i}_s\cdot\mathbf{K})\mathbf{i}_s\}\,(1 + e\cos v)\,dv\,. \tag{74}$$

Split (74) into two parts by letting

$$\begin{aligned}\text{IIA} &= \cos\beta\mathbf{k} \times \mathbf{K} \int_0^{2\pi} (1 + \cos v)\,dv \\ &= \cos\beta\mathbf{K} \times \int_0^{2\pi} (1 + e\cos\omega\cos P + e\sin\omega\sin P)\,d\omega \\ &= (2\pi\cos\beta)\mathbf{k} \times \mathbf{K}\end{aligned} \tag{75}$$

and

$$\begin{aligned}\text{IIB} = (-3\cos\beta)\mathbf{k} \times \mathbf{K}\cdot \int_0^{2\pi} \mathbf{i}_s\mathbf{i}_s(1 + e\cos\omega\cos P \\ + e\sin\omega\sin P)\,d\omega\,.\end{aligned} \tag{76}$$

But

$$\mathbf{i}_s = \cos\omega\mathbf{i}_g + \sin\omega\mathbf{j}_g \tag{77}$$

therefore

$$\mathbf{i}_s\mathbf{i}_s = \cos^2 \omega\mathbf{i}_\theta\mathbf{i}_\theta + \sin^2 \omega\mathbf{j}_\theta\mathbf{j}_\theta + \sin \omega \cos \omega(\mathbf{i}_\theta\mathbf{j}_\theta + \mathbf{j}_\theta\mathbf{i}_\theta). \tag{78}$$

It follows that

$$\text{IIB} = (-3\pi \cos\beta)\mathbf{k} \times \mathbf{K}\cdot(\mathbf{i}_\theta\mathbf{i}_\theta + \mathbf{j}_\theta\mathbf{j}_\theta). \tag{79}$$

So that, combining IIA and IIB

$$\text{II} = (\pi \cos\beta)\mathbf{k} \times [2\mathbf{K} - 3\mathbf{K}\cdot(\mathbf{i}_\theta\mathbf{i}_\theta + \mathbf{j}_\theta\mathbf{j}_\theta)] \tag{80}$$

$$= (\pi \cos\beta)\mathbf{k} \times [2\mathbf{K} - 3(\mathbf{K}\cdot\mathbf{j}_\theta)\mathbf{j}_\theta] \tag{81}$$

$$= \pi \cos\beta[2\mathbf{k} \times \mathbf{K} - 3(\mathbf{K}\cdot\mathbf{j}_\theta)(\mathbf{k}X\mathbf{j}_\theta)]. \tag{82}$$

We shall now proceed to express (82) in the SANOR system. From Fig. 10,

$$\mathbf{K} = \sin\theta\mathbf{j} + \cos\theta\mathbf{k}. \tag{83}$$

From (22) and (20) we write

$$x_\theta = (CD)_i(\tilde{D}\tilde{C})x \tag{84}$$

or

$$\begin{aligned}\mathbf{j}_\theta = {} & (\sin\psi \cos\Omega \cos i - \sin\Omega \cos\psi \cos i)\mathbf{i} \\ & + (\sin i \sin\theta + \sin\Omega \cos i \sin\psi \cos\theta \\ & + \cos\Omega \cos i \cos\psi \cos\theta)\mathbf{j} \\ & + (\sin i \cos\theta - \sin\Omega \cos i \sin\psi \sin\theta \\ & - \cos\Omega \cos i \cos\psi \sin\theta)\mathbf{k}\end{aligned} \tag{85}$$

or, simplifying

$$\begin{aligned}\mathbf{j}_\theta = {} & [-\cos i \sin(\Omega - \psi)]\mathbf{i} \\ & + [\sin i \sin\theta + \cos i \cos\theta \cos(\Omega - \psi)]\mathbf{j} \\ & + [\sin i \cos\theta - \cos i \sin\theta \cos(\Omega - \psi)]\mathbf{k}.\end{aligned} \tag{86}$$

Substitute (83) and (86) into (82) to obtain

$$\begin{aligned}\text{II} = {} & \pi \cos\beta\Big[(-2 \sin\theta)\mathbf{i} - 3[\{\sin i \sin^2\theta \\ & + \cos i \sin\theta \cos\theta \cos(\Omega - \psi) + \sin i \cos^2\theta \\ & - \cos i \sin\theta \cos\theta \cos(\Omega - \psi)\}\{(-\sin i \sin\theta \\ & - \cos i \cos\theta \cos(\Omega - \psi))\mathbf{i} \\ & + (-\cos i \sin(\Omega - \psi))\mathbf{j}\}]\Big].\end{aligned} \tag{87}$$

Finally,

$$\begin{aligned}\text{II} = \pi \cos\beta[\{&-2\sin\theta + 3\sin^2 i \sin\theta \\ &+ 3\sin i \cos i \cos\theta \cos(\Omega - \psi)\}\mathbf{i} \\ &+ 3\sin i \cos i \sin(\Omega - \psi)\mathbf{j}].\end{aligned} \tag{88}$$

APPENDIX C

Case I Expansion

Refer to integrals of (91), (92), (93), and (94) of the main body of the paper:

$$A = \int_0^{2\pi} \sin\eta_0(1 + e\cos v)\,dv = \int_0^{2\pi} \sin\eta_0\,dv + e\int_0^{2\pi} \sin\eta_0 \cos v\,dv \tag{109}$$

$$\begin{aligned}&= \sin\eta_0 \int_0^{2\pi} \cos(bv)\,dv + \cos\eta_0 \int_0^{2\pi} \sin(bv)\,dv \\ &\quad + e\left[\sin\eta_0 \int_0^{2\pi} \cos(bv)\cos v\,dv\right. \\ &\quad \left. + \cos\eta_0 \int_0^{2\pi} \sin(bv)\cos v\,dv\right]\end{aligned} \tag{110}$$

$$= \sin\eta_0(D_1 + eD_3) + \cos\eta_0(D_2 + eD_4) \tag{111}$$

$$= \left(\frac{1}{b} - \frac{eb}{1 - b^2}\right)[\sin\eta_0 \sin 2\pi b + \cos\eta_0(1 - \cos 2\pi b)] \tag{112}$$

where the D factors are listed in Appendix E. In similar manner,

$$B = \cos\eta_0(D_1 + eD_3) - \sin\eta_0(D_2 + eD_4) \tag{113}$$

$$= \left[\frac{1}{b} - \frac{eb}{1 - b^2}\right][\cos\eta_0 \sin 2\pi b - \sin\eta_0(1 - \cos 2\pi b)]. \tag{114}$$

The **C** integral is expanded as follows

$$\mathbf{C} = \int_0^{2\pi} \mathbf{i}_s \mathbf{i}_s \sin\eta(1 + e\cos v)\,dv \tag{115}$$

but

$$\mathbf{i}_s = \cos\omega\mathbf{i}_g + \sin\omega\mathbf{j}_g\,. \tag{116}$$

Therefore

$$\mathbf{i}_s\mathbf{i}_s = \cos^2 \omega \mathbf{i}_0\mathbf{i}_0 + \sin^2 \omega \mathbf{j}_0\mathbf{j}_0 + \sin \omega \cos \omega(\mathbf{i}_0\mathbf{j}_0 + \mathbf{j}_0\mathbf{i}_0). \quad (117)$$

Let

$$\mathbf{C} = \mathbf{i}_0\mathbf{i}_0 C_1 + \mathbf{j}_0\mathbf{j}_0 C_2 + (\mathbf{i}_0\mathbf{j}_0 + \mathbf{j}_0\mathbf{i}_0)C_3 \quad (118)$$

and evaluate C_1, C_2, C_3 separately.

$$C_1 = \int_0^{2\pi} \cos^2 \omega \sin \eta(1 + e \cos v)\, dv. \quad (119)$$

Letting $\omega = v + P$, expanding $\cos^2 \omega$ and collecting terms, we have

$$\begin{aligned} C_1 = \sin \eta_0[&\cos^2 P(D_5 + eD_7) + \sin^2 P(D_9 + eD_{11}) \\ &- 2 \sin P \cos P(D_{13} + eD_{15})] \\ + \cos \eta_0[&\cos^2 P(D_6 + eD_8) + \sin^2 P(D_{10} + eD_{12}) \\ &- 2 \sin P \cos P(D_{14} + eD_{16})]. \end{aligned} \quad (120)$$

in like fashion

$$\begin{aligned} C_2 = \sin \eta_0[&\sin^2 P(D_5 + eD_7) + \cos^2 P(D_9 + eD_{11}) \\ &+ 2 \sin P \cos P(D_{13} + eD_{15})] \\ + \cos \eta_0[&\sin^2 P(D_6 + eD_8) + \cos^2 P(D_{10} + eD_{12}) \\ &+ 2 \sin P \cos P(D_{14} + eD_{16})] \end{aligned} \quad (121)$$

and

$$\begin{aligned} C_3 &= \int_0^{2\pi} \sin \omega \cos \omega \sin \eta(1 + e \cos v)\, dv \\ &= \sin \eta_0[(\cos^2 P - \sin^2 P)(D_{13} + eD_{15}) \\ &\quad + \sin P \cos P\,(D_5 - D_9 + eD_7 - eD_{11})] \\ &\quad + \cos \eta_0[(\cos^2 P - \sin^2 P)(D_{14} + eD_{16}) \\ &\quad + \sin P \cos P\,(D_6 - D_{10} + eD_8 - eD_{12})]. \end{aligned} \quad (122)$$

Since

$$\mathbf{E} = \int_0^{2\pi} \mathbf{i}_s\mathbf{i}_s \cos \eta(1 + e \cos v)\, dv \quad (94)$$

we expand as with $\mathbf{C}$, to yield

$$\mathbf{E} = \mathbf{i}_0\mathbf{i}_0 E_1 + \mathbf{j}_0\mathbf{j}_0 E_2 + (\mathbf{i}_0\mathbf{j}_0 + \mathbf{j}_0\mathbf{i}_0)E_3 \quad (123)$$

so that

$$E_1 = \int_0^{2\pi} \cos^2 \omega \cos \eta(1 + e \cos v)\, dv \tag{124}$$

$$\begin{aligned} &= \cos \eta_0[\cos^2 P(D_5 + eD_7) + \sin^2 P(D_9 + eD_{11}) \\ &\quad - 2 \sin P \cos P(D_{13} + eD_{15})] \\ &\quad - \sin \eta_0[\cos^2 P(D_6 + eD_8) + \sin^2 P(D_{10} + eD_{12}) \\ &\quad - 2 \sin P \cos P(D_{14} + eD_{16})] \end{aligned} \tag{125}$$

and

$$E_2 = \int_0^{2\pi} \sin^2 \omega \cos \eta(1 + e \cos v)\, dv$$

$$\begin{aligned} &= \cos \eta_0[\sin^2 P(D_5 + eD_7) + \cos^2 P(D_9 + eD_{11}) \\ &\quad + 2 \sin P \cos P(D_{13} + eD_{15})] \\ &\quad - \sin \eta_0[\sin^2 P(D_6 + eD_8) + \cos^2 P(D_{10} + eD_{12}) \\ &\quad + 2 \sin P \cos P(D_{14} + eD_{16})] \end{aligned} \tag{127}$$

and

$$E_3 = \int_0^{2\pi} \sin \omega \cos \omega \cos \eta(1 + e \cos v)\, dv \tag{128}$$

$$\begin{aligned} &= \cos \eta_0[(\cos^2 P - \sin^2 P)(D_{13} + eD_{15}) \\ &\quad + \sin P \cos P(D_5 - D_9 + eD_7 - eD_{11})] \\ &\quad - \sin \eta_0[(\cos^2 P - \sin^2 P)(D_{14} + eD_{16}) \\ &\quad + \sin P \cos P(D_6 - D_{10} + eD_8 - eD_{12})]. \end{aligned} \tag{129}$$

APPENDIX D

Case II Expansion

We will now evaluate (91), (92), (93) and (94) once again in a manner similar to Appendix C, but this time using the approximation

$$M = v - (\lambda/b) \sin v. \tag{102}$$

The same notation will be used for the integrals

$$A = \int_0^{2\pi} \sin \eta(1 + e \cos v)\, dv$$

$$\begin{aligned} = \sin \eta_0 \int_0^{2\pi} &\{\cos bv \cos (\lambda \sin v) + \sin bv \sin (\lambda \sin v)\} \\ &\cdot (1 + e \cos v)\, dv \\ + \cos \eta_0 \int_0^{2\pi} &\{\sin bv \cos (\lambda \sin v) - \cos bv \sin (\lambda \sin v)\} \\ &\cdot (1 + e \cos v)\, dv. \end{aligned} \tag{143}$$

Using the approximations of (107) and (108), we have

$$\begin{aligned} A = \sin \eta_0 &[D_1 + eD_3 + \lambda(D_{17} + eD_{14}) - (\lambda^2/2)(D_9 + eD_{11}) \\ &- (\lambda^3/6)(D_{18} + eD_{19})] \\ + \cos \eta_0 &[D_2 + eD_4 - \lambda(D_{20} + eD_{13}) - (\lambda^2/2) \\ &(D_{10} + eD_{12}) + (\lambda^3/6)(D_{21} + eD_{22})]. \end{aligned} \tag{144}$$

In like manner,

$$B = \int_0^{2\pi} \cos \eta(1 + e \cos v)\, dv \tag{145}$$

$$\begin{aligned} = \cos \eta_0 &[D_1 + eD_3 + \lambda(D_{17} + eD_{14}) - (\lambda^2/2)(D_9 + eD_{11}) \\ &- (\lambda^3/6)(D_{18} + eD_{19})] \\ - \sin \eta_0 &[D_2 + eD_4 - \lambda(D_{20} + eD_{13}) - (\lambda^2/2)(D_{10} \\ &+ eD_{12}) + (\lambda^3/6)(D_{21} + eD_{22})]. \end{aligned} \tag{146}$$

Similarly,

$$C_1 = \int_0^{2\pi} \cos^2 \omega \sin \eta(1 + e \cos v)\, dv \tag{147}$$

$$\begin{aligned} C_1 = \sin \eta_0 \Big[&\cos^2 P\{(D_5 + eD_7) + \lambda(D_{16} + eD_{26}) - (\lambda^2/2)(D_{23} \\ &+ eD_{25}) - (\lambda^3/6)(D_{31} + eD_{33})\} + \sin^2 P\{(D_9 + eD_{11}) \\ &+ \lambda(D_{18} + eD_{19}) - (\lambda^2/2)(D_{30} + eD_{38}) - (\lambda^3/6) \\ &\cdot (D_{35} + eD_{41})\} - 2 \sin P \cos P\{(D_{13} + eD_{15}) + \lambda(D_{12} \end{aligned}$$

$$+ eD_{24}) - (\lambda^2/2)(D_{22} + eD_{32}) - (\lambda^3/6)(D_{37} + eD_{39})\}\Big] + \cos\eta_0\Big[\cos^2 P\{(D_6 + eD_8) - \lambda(D_{15} + eD_{28}) - (\lambda^2/2) \cdot (D_{24} + eD_{27}) + (\lambda^3/6)(D_{32} + eD_{34})\} + \sin^2 P\{(D_{10} + eD_{12}) - \lambda(D_{21} + eD_{22}) - (\lambda^2/2)(D_{29} + eD_{37}) + (\lambda^3/6)(D_{36} + eD_{42})\} - 2\sin P\cos P\{(D_{14} + eD_{16}) - \lambda(D_{11} + eD_{23}) - (\lambda^2/2)(D_{19} + eD_{31}) + (\lambda^3/6)(D_{38} + eD_{40})\}\Big]. \tag{148}$$

To simplify the writing of the mean magnetic torque equation, let

$$F_1 = (D_5 + eD_7) + \lambda(D_{16} + eD_{26}) - (\lambda^2/2)(D_{23} + eD_{25}) - (\lambda^3/6)(D_{31} + eD_{33}) \tag{122}$$

$$F_2 = (D_9 + eD_{11}) + \lambda(D_{18} + eD_{19}) - (\lambda^2/2)(D_{30} + eD_{38}) - (\lambda^3/6)(D_{35} + eD_{41}) \tag{123}$$

$$F_3 = (D_{13} + eD_{15}) + \lambda(D_{12} + eD_{24}) - (\lambda^2/2)(D_{22} + eD_{32}) - (\lambda^3/6)(D_{36} + eD_{39}) \tag{124}$$

$$F_4 = (D_6 + eD_8) - \lambda(D_{15} + eD_{28}) - (\lambda^2/2)(D_{24} + eD_{27}) + (\lambda^3/6)(D_{32} + eD_{34}) \tag{125}$$

$$F_5 = (D_{10} + eD_{12}) - \lambda(D_{21} + eD_{22}) - (\lambda^2/2)(D_{29} + eD_{37}) + (\lambda^3/6)(D_{36} + eD_{42}) \tag{126}$$

$$F_6 = (D_{14} + eD_{16}) - \lambda(D_{11} + eD_{23}) - (\lambda^2/2)(D_{19} + eD_{31}) + (\lambda^3/6)(D_{38} + eD_{40}). \tag{127}$$

By the similarity of C_2 and C_3 integrals to C_1 and of E_1, E_2, E_3 to C_1, C_2, C_3 we immediately write and summarize the following:

$$C_1 = \sin\eta_0[F_1\cos^2 P + F_2\sin^2 P - 2F_3\sin P\cos P] + \cos\eta_0[F_4\cos^2 P + F_5\sin^2 P - 2F_6\sin P\cos P] \tag{128}$$

$$C_2 = \sin\eta_0[F_1\sin^2 P + F_2\cos^2 P + 2F_3\sin P\cos P] + \cos\eta_0[F_4\sin^2 P + F_5\cos^2 P + 2F_6\sin P\cos P] \tag{129}$$

$$C_3 = \sin\eta_0[(F_1 - F_2)\sin P\cos P + F_3(\cos^2 P - \sin^2 P)] + \cos\eta_0[(F_4 - F_5)\sin P\cos P + F_6(\cos^2 P - \sin^2 P)] \tag{130}$$

$$E_1 = \cos \eta_0[F_1 \cos^2 P + F_2 \sin^2 P - 2F_3 \sin P \cos P] - \sin \eta_0[F_4 \cos^2 P + F_5 \sin^2 P - 2F_6 \sin P \cos P] \tag{131}$$

$$E_2 = \cos \eta_0[F_1 \sin^2 P + F_2 \cos^2 P + 2F_3 \sin P \cos P] - \sin \eta_0[F_4 \sin^2 P + F_5 \cos^2 P + 2F_6 \sin P \cos P] \tag{132}$$

$$E_3 = \cos \eta_0[(F_1 - F_2) \sin P \cos P + F_3(\cos^2 P - \sin^2 P)] - \sin \eta_0[(F_4 - F_5) \sin P \cos P + F_6(\cos^2 P - \sin^2 P)]. \tag{133}$$

APPENDIX E

The D Integrals[22]

$$D_1 = \int_0^{2\pi} \cos (bv)\, dv = \frac{1}{b} \sin (2\pi b)$$

$$D_2 = \int_0^{2\pi} \sin (bv)\, dv = \frac{1}{b} [1 - \cos (2\pi b)]$$

$$D_3 = \int_0^{2\pi} \cos (bv) \cos v\, dv = -\frac{b \sin (2\pi b)}{1 - b^2}$$

$$D_4 = \int_0^{2\pi} \sin (bv) \cos v\, dv = -\frac{b}{1 - b^2} [1 - \cos (2\pi b)]$$

$$D_5 = \int_0^{2\pi} \cos (bv) \cos^2 v\, dv = \frac{2 - b^2}{b(4 - b^2)} \sin (2\pi b)$$

$$D_6 = \int_0^{2\pi} \sin (bv) \cos^2 v\, dv = \frac{2 - b^2}{b(4 - b^2)} [1 - \cos (2\pi b)]$$

$$D_7 = \int_0^{2\pi} \cos (bv) \cos^3 v\, dv = -\frac{b(7 - b)^2}{(1 - b^2)(9 - b^2)} \sin (2\pi b)$$

$$D_8 = \int_0^{2\pi} \sin (bv) \cos^3 v\, dv = -\frac{b(7 - b^2)}{(1 - b^2)(9 - b^2)} [1 - \cos (2\pi b)]$$

$$D_9 = \int_0^{2\pi} \cos (bv) \sin^2 v\, dv = \frac{2 \sin (2\pi b)}{b(4 - b^2)}$$

$$D_{10} = \int_0^{2\pi} \sin (bv) \sin^2 v\, dv = \frac{2}{b(4 - b^2)} [1 - \cos (2\pi b)]$$

$$D_{11} = \int_0^{2\pi} \sin^2 v \cos v \cos (bv)\, dv = -\frac{2b \sin (2\pi b)}{(1 - b^2)(9 - b^2)}$$

$$D_{12} = \int_0^{2\pi} \sin^2 v \cos v \sin (bv)\, dv = \frac{-2b}{(1 - b^2)(9 - b^2)} [1 - \cos (2\pi b)]$$

$$D_{13} = \int_0^{2\pi} \sin v \cos v \cos (bv)\, dv = \frac{1 - \cos (2\pi b)}{4 - b^2}$$

$$D_{14} = \int_0^{2\pi} \sin v \cos v \sin (bv)\, dv = -\frac{\sin (2\pi b)}{4 - b^2}$$

$$D_{15} = \int_0^{2\pi} \sin v \cos^2 v \cos (bv)\, dv = \frac{(3 - b^2)}{(9 - b^2)(1 - b^2)} [1 - \cos (2\pi b)]$$

$$D_{16} = \int_0^{2\pi} \sin v \cos^2 v \sin (bv)\, dv = -\frac{(3 - b^2) \sin (2\pi b)}{(9 - b^2)(1 - b^2)}$$

$$D_{17} = \int_0^{2\pi} \sin v \sin (bv)\, dv = -\frac{\sin (2\pi b)}{1 - b^2}$$

$$D_{18} = \int_0^{2\pi} \sin^3 v \sin (bv)\, dv = -\frac{6 \sin (2\pi b)}{(9 - b^2)(1 - b^2)}$$

$$D_{19} = \int_0^{2\pi} \sin^3 v \cos v \sin (bv)\, dv = -\frac{6 \sin (2\pi b)}{(16 - b^2)(4 - b^2)}$$

$$D_{20} = \int_0^{2\pi} \sin v \cos (bv)\, dv = \frac{1 - \cos (2\pi b)}{1 - b^2}$$

$$D_{21} = \int_0^{2\pi} \sin^3 v \cos (bv)\, dv = \frac{6[1 - \cos (2\pi b)]}{(9 - b^2)(1 - b^2)}$$

$$D_{22} = \int_0^{2\pi} \sin^3 v \cos v \cos (bv)\, dv = \frac{6[1 - \cos (2\pi b)]}{(16 - b^2)(4 - b^2)}$$

$$D_{23} = \int_0^{2\pi} \sin^2 v \cos^2 v \cos (bv)\, dv = \frac{2 \sin (2\pi b)}{b(16 - b^2)}$$

$$D_{24} = \int_0^{2\pi} \sin^2 v \cos^2 v \sin (bv)\, dv = \frac{2[1 - \cos (2\pi b)]}{b(16 - b^2)}$$

$$D_{25} = \int_0^{2\pi} \sin^2 v \cos^3 v \cos (bv)\, dv = -\frac{2b(13 - b^2) \sin (2\pi b)}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{26} = \int_0^{2\pi} \sin v \cos^3 v \sin (bv)\, dv = -\frac{(10 - b^2) \sin (2\pi b)}{(16 - b^2)(4 - b^2)}$$

$$D_{27} = \int_0^{2\pi} \sin^2 v \cos^3 v \sin (bv)\, dv = -\frac{2b(13 - b^2)[1 - \cos (2\pi b)]}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{28} = \int_0^{2\pi} \sin v \cos^3 v \cos (bv)\, dv = \frac{(10 - b^2)[1 - \cos (2\pi b)]}{(16 - b^2)(4 - b^2)}$$

$$D_{29} = \int_0^{2\pi} \sin^4 v \sin (bv)\, dv = \frac{24[1 - \cos (2\pi b)]}{b(16 - b^2)(4 - b^2)}$$

$$D_{30} = \int_0^{2\pi} \sin^4 v \cos (bv)\, dv = \frac{24 \sin (2\pi b)}{b(16 - b^2)(4 - b^2)}$$

$$D_{31} = \int_0^{2\pi} \sin^3 v \cos^2 v \sin (bv)\, dv = -\frac{6(5 - b^2) \sin (2\pi b)}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{32} = \int_0^{2\pi} \sin^3 v \cos^2 v \cos (bv)\, dv = \frac{6(5 - b^2)[1 - \cos (2\pi b)]}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{33} = \int_0^{2\pi} \sin^3 v \cos^3 v \sin (bv)\, dv = -\frac{6 \sin (2\pi b)}{(36 - b^2)(4 - b^2)}$$

$$D_{34} = \int_0^{2\pi} \sin^3 v \cos^3 v \cos (bv)\, dv = \frac{6[1 - \cos (2\pi b)]}{(36 - b^2)(4 - b^2)}$$

$$D_{35} = \int_0^{2\pi} \sin^5 v \sin (bv)\, dv = -\frac{120 \sin (2\pi b)}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{36} = \int_0^{2\pi} \sin^5 v \cos (bv)\, dv = \frac{120[1 - \cos (2\pi b)]}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{37} = \int_0^{2\pi} \sin^4 v \cos v \sin (bv)\, dv = -\frac{24b[1 - \cos (2\pi b)]}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{38} = \int_0^{2\pi} \sin^4 v \cos v \cos (bv)\, dv = -\frac{24b \sin (2\pi b)}{(25 - b^2)(9 - b^2)(1 - b^2)}$$

$$D_{39} = \int_0^{2\pi} \sin^4 v \cos^2 v \sin (bv)\, dv = \frac{24(6 - b^2)[1 - \cos (2\pi b)]}{b(36 - b^2)(16 - b^2)(4 - b^2)}$$

$$D_{40} = \int_0^{2\pi} \sin^4 v \cos^2 v \cos (bv)\, dv = \frac{24(6 - b^2) \sin (2\pi b)}{b(36 - b^2)(16 - b^2)(4 - b^2)}$$

$$D_{41} = \int_0^{2\pi} \sin^5 v \cos v \sin (bv)\, dv = -\frac{120 \sin (2\pi b)}{(36 - b^2)(16 - b^2)(4 - b^2)}$$

$$D_{42} = \int_0^{2\pi} \sin^5 v \cos v \cos (bv)\, dv = \frac{120[1 - \cos (2\pi b)]}{(36 - b^2)(16 - b^2)(4 - b^2)}.$$

REFERENCES

1. Bennett, S. B., and Thomas, L. C., The *Telstar* Communications Satellite Experiment Plan, AIEE Trans., No. 63-952, May 6, 1963, pp. 9–10. See also B.S.T.J., **42,** July, 1963, p. 765.
2. Hill, D. W., Calculation of the Spin-Axis Orientation of the *Telstar* Satellites from Optical Data, B.S.T.J., **42,** Nov., 1963, p. 2943.
3. Courtney-Pratt, J. S., Hett, J. H., and McLaughlin, J. W., Optical Meas-

urements on *Telstar* to Determine the Orientation of the Spin Axis and the Spin Rate, Jour. SMPTE, **72,** June, 1963, pp. 462–484.
4. *The American Ephemeris and Nautical Almanac for the Year 1962*, U. S. Government Printing Office, Washington, D. C., 1960, p. 482.
5. Ref. 4, pp. 466–471.
6. Smart, W. M., *Textbook on Spherical Astronomy*, Cambridge Press, 1956, pp. 116–117.
7. Vestline, E. H., et al., Description of the Earth's Main Magnetic Field and Its Secular Change, 1940–1945, Carnegie Institution of Washington, Publication 578, 1947.
8. Vestline, E. H., et al., The Geomagnetic Field, Its Description and Analysis, Carnegie Institution of Washington, Publication 580, 1959.
9. Hrycak, P., et al., The Spacecraft Structure and Thermal Design Considerations, B.S.T.J., **42,** July, 1963, p. 971.
10. Roberson, R. E., and Tatistcheff, D., The Potential Energy of a Small Rigid Body in the Gravitational Field of an Oblate Spheroid, J. Franklin Inst., **262,** 1956, pp. 209–214.
11. Harnwell, G. P., *Principles of Electricity and Electromagnetism*, 1st ed., McGraw-Hill, 1938, p. 377. For precessional theory similar to that of Section 4.6, but for circular orbits only, see: Bandeen, W. R., and Manger, W. P., Angular Motion of the Spin Axis of the *Tiros I* Meteorological Satellite Due to Magnetic and Gravitational Torques, NASA TN D-571, April, 1961.
12. Allen, C. W., *Astrophysical Quantities*, Athlone Press, 1955.
13. See Ref. 6, pp. 107–108.
14. See Ref. 6, p. 111–113.
15. Danby, J. M. A., *Fundamentals of Celestial Mechanics*, Macmillan, New York, 1962, pp. 128–129.
16. Goldstein, H., *Classical Mechanics*, Addison-Wesley, 1959.
17. Yu, E. Y., Spin Decay, Spin-Precession Damping, and Spin-Axis Drift of the *Telstar* Satellite, B.S.T.J., **42,** Sept., 1963, p. 2169.
18. Jakes, W. C., Participation of the Holmdel Station in the *Telstar* Project, B.S.T.J., **42,** July, 1963, p. 1421.
19. Harris, D. L., *Planets and Satellites*, ed. Kuiper and Middlehurst, Chicago, 1961.
20. Delchamps, T. B., et al., The Spacecraft Test and Evaluation Program, B.S.T.J., **42,** July, 1963, p. 1020 and p. 1024.
21. Phillips, H. B., *Vector Analysis*, John Wiley, 1933.
22. Pierce, B. O., *A Short Table of Integrals*, 4th ed., Ginn, 1957.

Directional Control in Light-Wave Guidance

By S. E. MILLER

(Manuscript received May 11, 1964)

The transmission of light waves for communication in a medium sheltered from atmospheric effects requires wave guidance providing frequent changes in direction of propagation. This paper shows that, in any electromagnetic waveguide having transverse planes in which the field is essentially equiphase, the transverse width of the field distribution 2a and wavelength λ *determine the order of magnitude of the direction-determining parameters,* R_{min}, *the minimum bending radius, and* δ_{max}, *the maximum abrupt angular changes, according to the relations*

$$R_{min} = 2(a^3/\lambda^2)$$

$$\delta_{max} = \tfrac{1}{2}(\lambda/a)$$

which are valid in the region $\lambda < a$. *The significance of* R_{min} *is apparent, with the note that in a system containing a multiplicity of bends, an appropriate way of summing the effects of the individual bends should be used to establish an over-all equivalent bend radius for the complete transmission path, which must be larger than* R_{min}. *The quantity* δ_{max} *may be regarded as the maximum value of the accumulated angular errors (rms sum, for example) in a transmission line including reflecting or refracting elements for directional control. For a light beam at* $\lambda = 0.6328$ *microns having a diameter of 1.0 mm,* $\delta_{max} = 0.036°$ *and* $R_{min} = 600$ *meters.*

Small-diameter beams ease the problem of directional control. There is no fundamental reason why small beams should not be achievable with low loss in the straight condition, but many guiding structures do have an inverse relation between beam diameter and straight-condition attenuation coefficient. To explore the direction-controlling properties of specific media and the interaction of R_{min} *and* δ_{max} *with straight attenuation coefficient, the following waveguides and associated criteria for establishing* R_{min} *and* δ_{max} *were studied:*

(1) sequence of lenses: criterion, beam deflection from nominal axis by one beam radius,

(2) hollow dielectric waveguide: criterion, added bend loss equal to straight condition loss,

(3) round metallic circular-electric waveguides: for helix guide, criterion is bend loss equal to straight loss; for simple metallic tube, criterion is a transmission ripple (due to mode conversion) of about 1.7 db.

In all cases the functional dependence on a and λ *for* R_{min} *or* δ_{max} *was the same (given above) as derived for the generalized electromagnetic waveguide, and the associated constants were in most cases of similar magnitude.*

I. INTRODUCTION

In research on techniques for transmitting light waves over appreciable distances for communication it has become evident that control of direction of propagation is an important and difficult problem. Electromagnetic waves in free space travel in a straight line. In a medium that is sheltered from atmospheric effects, frequent changes in direction are necessary to follow vertical terrain contours and to conform to a horizontal path avoiding physical obstacles and regions of high-cost installation. The wave guiding medium must provide these direction changes.

In this paper some simple relations are derived to give the order of magnitude of the direction-determining factors, bending radius and abrupt tilt angle, for any wave guiding structure as a function of wavelength and the transverse dimension of the guided electromagnetic wave beam. These simple relations are then compared to the corresponding more precisely defined quantities for specific waveguides: (1) a sequence of lenses,[1] (2) the hollow-dielectric waveguide,[5] and (3) round waveguides for circular electric waves.

II. DERIVATION OF GENERAL WAVEGUIDE DIRECTIONAL SENSITIVITY

In Fig. 1 we show a generalized waveguide for electromagnetic waves, with an abrupt open end radiating into free space. We assume the field at the aperture is essentially equiphase, which implies ending the guide

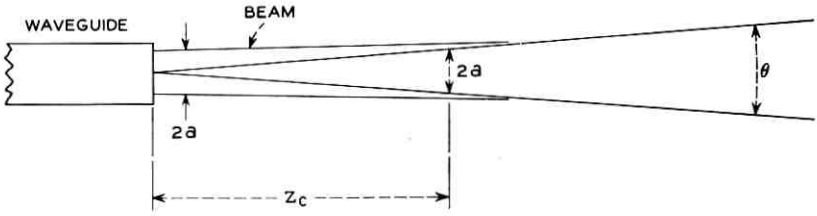


Fig. 1 — Waveguide with abrupt open end.

only at certain longitudinal locations if a periodic form of guidance (such as a sequence of lenses) is employed. Let the field strength variation across the aperture be approximately sinusoidal. Then, approximately, the far-field beam angle θ is

$$\theta = \lambda/a \text{ radians} \tag{1}$$

in which we require $a > \lambda$. Other aperture distributions would give the same order of magnitude for θ. In the near-field region the radiated beam remains collimated in a width approximately $2a$ out to a distance z_c from the aperture, where

$$z_c\theta = 2a \tag{2}$$

$$z_c = 2a^2/\lambda. \tag{2a}$$

The key inference on directional sensitivity is introduced here. Since in the absence of the guide the beam remains confined to essentially the same region as in the presence of the guide, it is concluded that the guide has little influence on the beam over the interval z_c. Thus any appreciable change in direction of wave propagation must not be made in a distance less than z_c.

With reference to Fig. 2, the departure of a circular arc from the tangent is

$$\Delta = \tfrac{1}{2}(l^2/R). \tag{3}$$

We now require that $\Delta = a$ when $l = z_c$. Using (3), (2) and (1), we obtain the minimum bend radius $R_{\min}$

$$R_{\min} = 2a^3/\lambda^2. \tag{4}$$

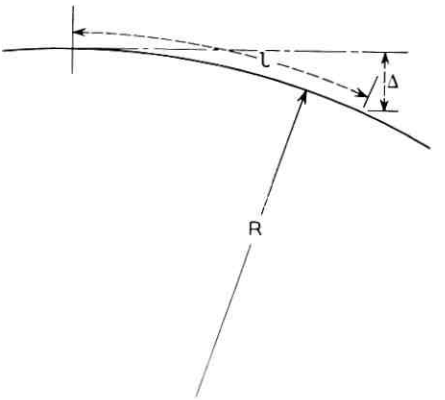


Fig. 2 — Departure of a circular arc from the tangent.

Alternatively, this same relation may be arrived at by specifying that the change in direction shall be one beamwidth θ after traveling a distance z_c in the minimum bending radius $R_{\min}$; i.e., $R_{\min}\theta = z_c$.

Equation (4) gives the order of magnitude of bend radius at which the wave propagation will change character. At longer bend radii the wave propagation will be essentially as in the straight guide, and at shorter bend radii something drastic will happen. Just what changes occur in the latter case depend on the nature of the medium in detail. If the medium is enclosed in a perfect conductor the change will be large mode conversion. If it consists of a sequence of infinitely wide lenses we will see that the change is a wide oscillation of the beam about the nominal axis of propagation. Note that in neither of these cases is energy lost due to the bend. Nonetheless, we regard either change as undesirable.

Consider an abrupt angular change in the guide direction, δ. Following a line of reasoning analogous to that given above, we can say that the character of wave propagation will change rapidly in the region where

$$\delta_{\max} = \theta/2 = \lambda/2a. \tag{5}$$

Smaller values of δ will cause progressively less change in wave character, whereas larger values of δ will cause violent changes.

If we consider the relation of these quantities to a wave guiding medium, it is apparent that $R_{\min}$ is intended as the smallest radius at which the otherwise uniform medium can be bent. When a multiplicity of bends is included in a single transmission link, some way of summing their effects is needed to form an equivalent bending radius which must be greater than $R_{\min}$.

The angle δ is somewhat different. In many media where $a \gg \lambda$ it is possible to insert a large plane reflector and introduce a change of direction of arbitrary size. As long as the guides at both approaches to the reflector are perfectly aligned according to geometric optics, the disturbance on wave propagation may be negligible. However, there will be an error in such angular alignment and $\delta_{\max}$ tells us how large that error may be. When many random angular errors are made, $\delta_{\max}$ is approximately the rms accumulation of such errors.

The numerical values of $R_{\min}$ and $\delta_{\max}$ have been plotted for λ from 0.6328 to 10 microns and beam radius a from 0.1 to 100 millimeters in Figs. 3 and 4, respectively. For example, at $\lambda = 0.6328$ microns and $2a = 1.0$ cm, $R_{\min} = 600{,}000$ meters and $\delta_{\max} = 3.6 \times 10^{-3}$ degrees. Dropping to $2a = 1.0$ mm, $R_{\min} = 600$ meters and $\delta_{\max} = 3.6 \times 10^{-2}$ degrees.

We consider next certain specific wave guiding structures to compare

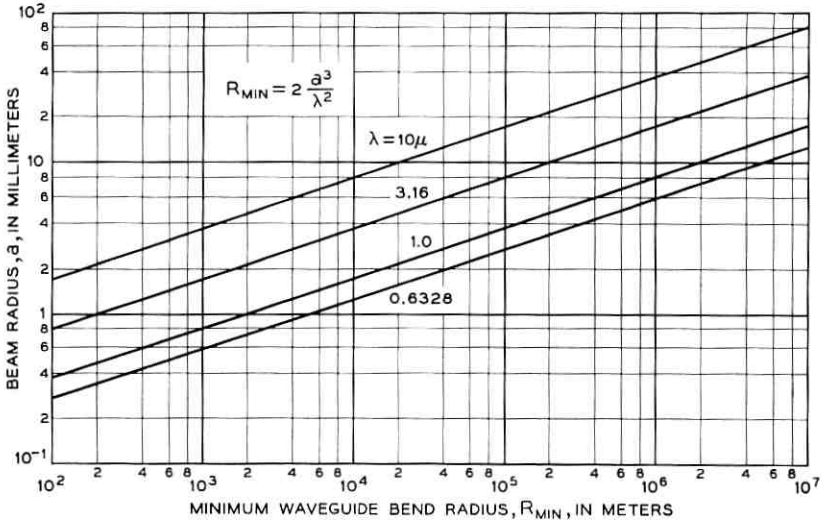


Fig. 3 — Beam radius vs minimum waveguide bend radius.

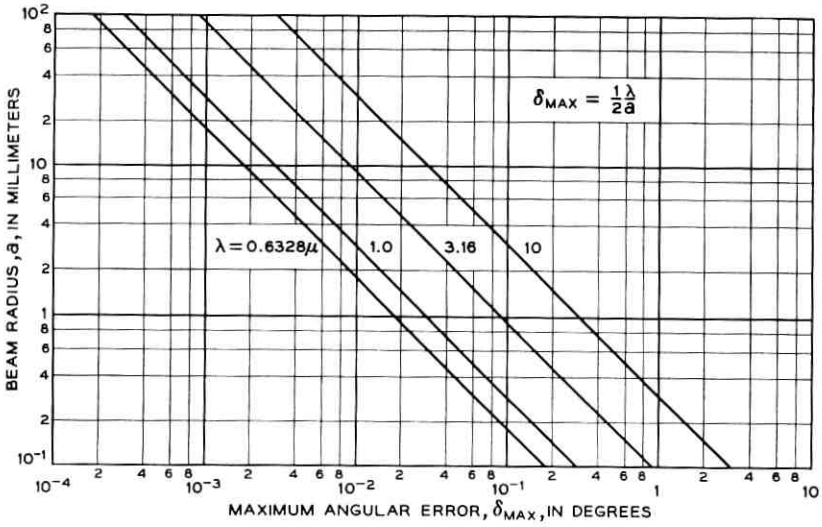


Fig. 4 — Beam radius vs maximum angular error.

the results of directional changes in those structures to the generalized conclusions drawn above.

III. SEQUENCE-OF-LENS WAVEGUIDE

G. Goubau has proposed[1] a waveguide for electromagnetic waves consisting of a series of lenses, and D. Marcuse has used geometric optics to determine the effects of bends in such a waveguide.[2]

If the input to a lens waveguide is a ray which is inclined at an angle δ to the longitudinal waveguide axis (Fig. 5) the departure of the ray from the longitudinal axis has a magnitude at successive lenses which is contained within an envelope which is a sinusoidal function of distance along the longitudinal axis. Starting with the work of Marcuse, one can show that there is an optimum strength of lens which minimizes the departure of a ray from the axis; the optimum focal length f is related to the lens spacing L by

$$2f = L \tag{6}$$

and under that condition the maximum deviation of the ray from the longitudinal axis is

$$r_{\max-1} = \delta L. \tag{7}$$

Consider a region of bend radius R following a straight region of lens waveguide. For a ray incident on the curved region from the axis of the straight region Marcuse has also calculated the ray's departure from the axis in the curved region; for the case $f = L/2$ the maximum departure is

$$r_{\max-2} = L^2/R. \tag{8}$$

We now relate these departures from the guide axis to the transverse dimension of the beam. It is convenient to consider the beam radius to be that value of radius beyond which a completely negligible amount of

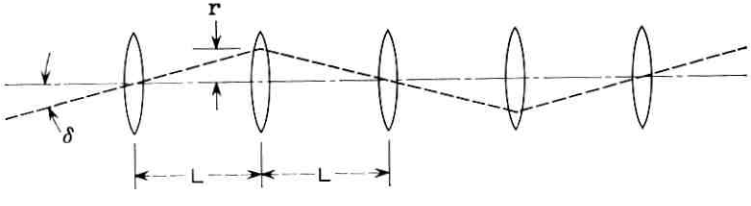


Fig. 5 — Sequence of lenses.

field exists. We define this beam radius as

$$r_b = (N_0 L \lambda)^{\frac{1}{2}}. \tag{9}$$

Here N_0 is the Fresnel number which previous work[3,9] shows is on the order of unity for negligibly small diffraction loss when all energy outside the radius r_b is absorbed at the lenses.*

As a criterion of the maximum permissible abrupt angular change, we somewhat arbitrarily set it to be that angle at which $r_{\max-1}$ is equal to the transverse beam radius r_b :

$$\delta_{\max} = (r_b/L) = (N_0\lambda/L)^{\frac{1}{2}} = N_0\lambda/r_b . \tag{10}$$

Since $N_0 \cong 1$, this specific guide and criterion gives a permissible angular change of twice that prescribed by (5). This may be considered an excellent agreement.

As a criterion of the minimum permissible bend radius, we set the resulting beam deflection $r_{\max-2}$ equal to the transverse beam radius r_b :

$$R_{\min} = L^2/r_b = r_b^3/(N_0\lambda)^2. \tag{11}$$

Since $N_0 \cong 1$, this specific guide and criterion gives a permissible bend radius of one-half that prescribed by (4), which again may be considered excellent agreement.

The most important aspect of the comparison between (4) and (5), (10), and (11) is that the corresponding equations have the identical dependence on λ and a, which determines in a broad way the magnitude of the direction determining parameters.

In this form of guide we can readily relate the beam radius to the associated lens spacing and the losses. Fig. 6 shows the lens spacing L versus beam radius in the 0.5- to 4-micron wavelength region. As before, N_0 will be about unity, but where extremely low losses per lens are required may have to be slightly greater than unity.[3] For the 1.0-mm beam diameter referred to above, the lens spacing is about 0.4 meter for $\lambda = 0.63$ microns.

In principle, vanishingly small transmission loss could be obtained by appropriate choice of lens diameter (i.e., choice of N_0), if the reflection, absorption, and scattering losses were negligible at the lenses. In practice, such losses may be very real. Fig. 7 shows the total losses per lens required as a function of lens spacing with net transmission loss as a parameter. For 3 db/mile net loss and the 0.4-meter lens spacing, a power loss per lens of about one part in 10^4 is required.

* Further discussion of N_0 and r_b is given in the Appendix.

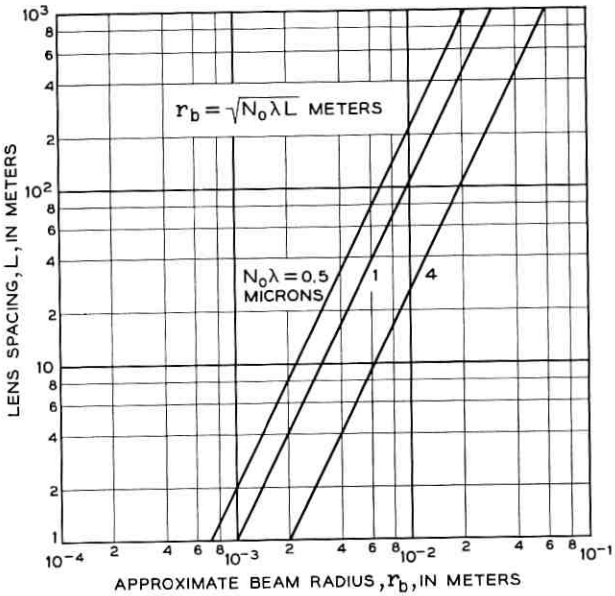


Fig 6. — Approximate beam radius vs lens spacing in a sequence of identical lenses.

In general, of course, the smaller beam diameters which permit rapid direction changes require more tight guidance (closer lens spacing) and tend to increase the losses. Note, however, that the only inherent losses associated with tight guidance for the lens guidance system are due to scattering or reflection at lens surfaces or bulk lens absorption loss, both of which may conceivably be made very small.

IV. HOLLOW DIELECTRIC WAVEGUIDE

E. A. J. Marcatili and R. A. Schmeltzer have proposed a waveguide for light waves consisting of a hollow dielectric tube in which the useful energy is entirely confined to the central hole.[5] When the guide is straight, loss takes place through very slow radiation into the dielectric, which is completely absorbing for the light energy.

The bending radius for such a guide which makes the extra loss due to bending (also a radiation loss) exactly equal to the straight-guide attenuation coefficient has also been determined.[5] They find, for the lowest-order mode (EH_{11}), in which the field varies roughly cosinusoidally from the axis to the inner wall of the tube,

$$R_{\min-D} = 9.5\, a^3/\lambda^2 \qquad (12)$$

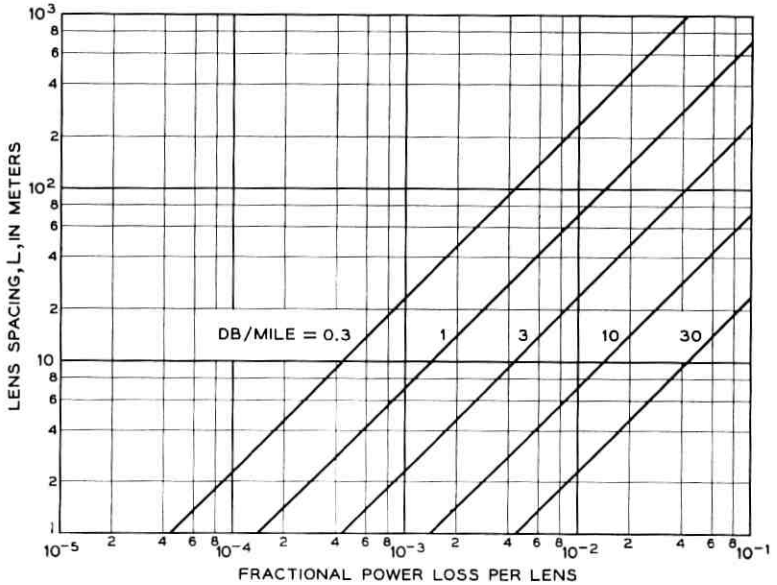


Fig. 7 — Lens loss vs spacing for prescribed total loss.

where a is the inner radius of the tube. Once again, the functional dependence of $R_{\min-D}$ on a and λ is identical to that in (4).

The straight-line attenuation coefficient for the lowest-order mode can be reduced to[5]

$$\alpha_s = 0.214\,\lambda^2/a^3 \tag{13}$$

for an index of refraction of the dielectric tube equal to 1.5. As Marcatili and Schmeltzer have pointed out, the dependence of α_s on a and λ is the exact inverse of that for $R_{\min-D}$; hence for any prescribed straight-guide loss there is a minimum permissible bending radius, which for the lowest-order mode is

$$R_{\min-D} = 2.03/\alpha_s\,. \tag{14}$$

For fixed α_s this is independent of λ. For $R_{\min-D} = 1000$ meters, $\alpha_s = 0.002$ nepers/meter or 27.9 db/mile, and at $\lambda = 0.6328$ microns, the inner radius of the tube $a = 0.35$ mm.

V. CIRCULAR ELECTRIC WAVEGUIDES

We consider now the directional control relations for round waveguides designed for the TE_{01} circular electric wave. Helix waveguide and smooth-

walled metallic waveguide involve quite different criteria for tolerable bending radius and will be discussed separately.

Consider first metallic guides in which the losses are negligible compared to the mode coupling coefficients. We assume the degeneracy between TE_{01} and TM_{11} is broken with a dielectric lining or other guide modification. Then the limit on bending radius or abrupt tilt is the interfering effect between the unperturbed TE_{01} energy and the energy which is converted to an undesired mode and reconverted back to the TE_{01} wave.

For an abrupt tilt, the amplitude conversion coefficient from TE_{01} to TE_{12} was found by S. P. Morgan[4] to be, approximately

$$p = 1.935\ (a/\lambda)\delta \tag{15}$$

where δ is the tilt angle in radians. When converted energy from one tilt strikes another tilt (presumed for simplicity here to be the same angle), energy is reconverted back to the TE_{01} wave with the same conversion coefficient given by (15). The amplitude of the reconverted wave compared to the unperturbed wave is then p^2. Depending on the relative phase of the reconverted vector, it may add at any phase angle to the unperturbed wave. Hence the amplitude transmission coefficient varies with wavelength between $(1 + p^2)$ and $(1 - p^2)$. Letting a transmission fluctuation of 1.7 db, corresponding to $p^2 = 0.1$, be the criterion of limiting tilt angle, we find

$$\delta_{\max-M} = p\lambda/1.935a = 0.164\ \lambda/a. \tag{16}$$

Comparing this to the generalized relation for $\delta_{\max}$ in (5), we note the identical dependence on λ and a and a somewhat smaller constant multiplier. In practice, the existence of coupling to other modes would tend to make somewhat smaller values of $\delta_{\max-M}$ needed, but the dependence on λ and a would not be affected.

Still considering guides with negligible losses, we examine the effect of a constant-radius bend. It may be shown that the reconverted vector has the magnitude

$$p_1^2 = k_t^2/(\Gamma_1 - \Gamma_2)^2 \tag{17}$$

where k_t is the distributed coupling coefficient between TE_{01} and an undesired mode and Γ_1 and Γ_2 are the propagation constants for TE_{01} and the undesired modes, respectively. For TE_{01} to TE_{12}

$$k_t \cong j(2a/\lambda_0 R) \tag{18}$$

where R is the bend radius and a is the radius of the guide. Also,

$$(\Gamma_1 - \Gamma_2) \cong j(\beta_1\beta_2) = j\,\lambda_0/a^2. \tag{19}$$

Hence

$$p_1^2 = 4a^6/R^2\lambda^4. \tag{20}$$

As a criterion for minimum bend radius we set p_1^2 equal to 0.1, giving the same ripple in transmission loss as noted above, with the resulting bending radius from (20)

$$R_{\min-M} = (4/p^2)^{\frac{1}{3}}(a^3/\lambda^2) = 6.3\ a^3/\lambda^2. \tag{21}$$

Comparing (21) to (4), we again find the identical dependence on a and λ with a slightly different constant.

Turning now to the case of helix waveguide in which very strong loss is introduced for the undesired mode, we find we do not have explicit forms for the coupling coefficient. We take advantage of some numerical evaluations carried out in the 30- to 100-kmc region on guide varying in diameter from 0.25 inch to 3 inches. It was found that the bend loss coefficient is given by the expression*

$$\alpha_B = 0.0726\ (a^3/R^2\lambda^{2.7}). \tag{22}$$

The TE_{01} loss of the guide when straight is very nearly that of a copper cylinder, given by S. A. Schelkunoff as[6]

$$\alpha_s = 4.46 \times 10^{-6}\ \lambda^{\frac{1}{2}}/a^3. \tag{23}$$

In (23) we have assumed $a^2 \gg (\lambda/2)^2$, so that the cutoff effect is negligible. As our criterion for minimum bending radius we equate the bend loss α_B and the straight-line loss α_s, yielding

$$R_{\min-H} = 128\ a^3/\lambda^{2.1}. \tag{24}$$

The functional dependence of $R_{\min-H}$ on λ and a is very nearly the same as in (4), but the constant multiplier is much greater. This is a consequence of the criterion $\alpha_s = \alpha_B$, which is thereby proven much more stringent than the rather lax transmission ripple criterion used above for the metallic tube guide in which dissipation was negligible. Since α_s of (23) and $R_{\min-H}$ of (24) have different dependence on λ, a change of wavelength will influence $R_{\min-H}$ even though α_s is held constant. We can express this by substituting (23) into (24), giving

$$R_{\min-H} = \frac{128}{\lambda^{0.6}} \times \frac{a^3}{\lambda^{\frac{1}{2}}} = \frac{5.71 \times 10^{-4}}{\lambda^{0.6}\alpha_s}. \tag{25}$$

At longer wavelengths, *smaller* bending radii are tolerable even though

* This is the result of unpublished calculations by the author, based on coupling coefficients derived by methods due to Unger[9] and using coupled-wave theory.[10]

α_s is held constant by increasing a. At a wavelength of 5 mm and a straight attenuation coefficient of 1 db/mile, $R_{\min-H} = 191$ meters. This result and the numerical constant in (22) are dependent to some extent on the wall impedance to the undesired modes used in the numerical evaluations referred to above, which was on the order of one-half the free-space intrinsic impedance.

VI. CONCLUSION

For any guided electromagnetic wave, the order of magnitude of the direction-determining parameters $R_{\min}$ (the minimum bending radius) and $\delta_{\max}$ (the maximum abrupt angular change) are uniquely determined by the wavelength and transverse beam dimension. Equations (4) and (5) were derived, determining $R_{\min}$ and $\delta_{\max}$ for a general guided electromagnetic wave by inferring the tightness of guidance from the behavior of a wave radiated from the open end of the waveguide. Investigation of specific forms of waveguide with precise criteria for setting limits on R and δ (as outlined in the abstract) lead to identical functional forms for $R_{\min}$ and $\delta_{\max}$, with similar constant multipliers.

APPENDIX

Previous workers[3,7] have calculated the diffraction loss at a reflector in a maser interferometer, and the same loss per lens would be expected in a sequence-of-lens waveguide if the entire plane outside the edge of the lens (of radius equal to that of the maser reflector) were absorbing. These losses are plotted versus $N = a^2/L\lambda$ (where a is the reflector radius) in Fig. 3 of Ref. 7 and in Fig. 15 of Ref. 3 for focal length $f = L/2$. We chose N_0 to be that value of N which gives satisfactorily low loss per lens; for example, for $N_0 = 1$, Ref. 7 gives a power loss per lens of one part in 10^4 for the lowest-order wave, and for $N_0 \cong 1.4$ the power loss is one part in 10^6. Fig. 3 of Ref. 1 shows that 99.8 per cent of the energy of the normal mode for infinite lenses lies within the radius $r = (L\lambda)^{\frac{1}{2}}$ at the lens. In practical cases, therefore, N_0 will differ little from unity.

Another item of interest is the relation between r_b of (9) and the field amplitude given by previous workers.[7,8] The field varies as a function of radius r from the axis of the guide according to

$$\exp\ (-r^2/w^2) \tag{26}$$

where w is the radius at which the field drops to e^{-1} of its maximum (on axis) value. The value of w varies with longitudinal position between

lenses; at midway between lenses $w = w_0$, where

$$w_0 = (L\lambda/2\pi)^{\frac{1}{2}}. \tag{27}$$

At the lenses, $w = w_s$, and for our cases of $f = L/2$

$$w_s = (L\lambda/\pi)^{\frac{1}{2}}. \tag{28}$$

It is apparent from (9) and (28) that

$$r_b = w_s (N_0\pi)^{\frac{1}{2}}. \tag{29}$$

In terms of w_s (10) becomes

$$\delta_{\max} = (N_0/\pi)^{\frac{1}{2}}(\lambda/w_s) \tag{30}$$

and (11) becomes

$$R_{\min} = (\pi^{\frac{3}{2}}N_0^{-\frac{1}{2}})\ (w_s^3/\lambda^2). \tag{31}$$

REFERENCES

1. Goubau, G., and Schwering, F., On the Guided Propagation of Electromagnetic Wave Beams, IRE Trans., **AP-9,** May, 1961, pp. 248–256.
2. Marcuse, D., Propagation of Light Rays through a Lens Waveguide with Curved Axis, B.S.T.J., **43,** March, 1964, pp. 741–753.
3. Fox, A. G., and Li, T., Resonant Modes in a Maser Interferometer, B.S.T.J., **40,** March, 1961, pp. 453–488.
4. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., **41,** May, 1962, pp. 1031–1170.
5. Marcatili, E. A. J., and Schmeltzer, R. A., Hollow Dielectric Waveguides for Long Distance Optical Transmission and Lasers, B.S.T.J., this issue, p. 1783.
6. Schelkunoff, S. A., *Electromagnetic Waves,* D. Van Nostrand Co., Princeton, 1945, p. 390.
7. Boyd, G. D., and Gordon, J. P., Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers, B.S.T.J., **40,** March, 1961, pp. 489–508.
8. Boyd, G. D., and Kogelnik, H., Generalized Confocal Resonator Theory, B.S.T.J., **41,** July, 1962, pp. 1347–1369.
9. Unger, H. G., Normal Modes and Mode Conversion in Helix Waveguide, B.S.T.J., **40,** January, 1961, pp. 255–280.
10. Miller, S. E., Coupled Wave Theory and Waveguide Applications, B.S.T.J., **33,** May, 1954, pp. 661–719.

Alternating-Gradient Focusing and Related Properties of Conventional Convergent Lens Focusing

By S. E. MILLER

(Manuscript received June 8, 1964)

A series of lenses whose focusing properties are alternately convergent and divergent (alternating-gradient focusing) is of potential interest in the guidance of light waves, and has previously been used to focus electron beams and high-energy particle streams. New information is provided herein on such focusing for the case of equal focal length f (but alternating-gradient) lenses equally spaced a distance L.

The alternating-gradient system formed by adding diverging lenses between the lenses of an all-converging sequence of lenses is found to have the same stability condition as the original system for $0 < L/f < 2$. A physical argument leads to the conclusion that weaker divergent lenses would also leave the stability criterion unchanged.

The focusing effect of the alternating-gradient system is surprisingly close to that of an all-converging lens system. After the focal length of each has been adjusted to an optimum value, the ray departure from the system axis is only 1.67 times as great for the alternating-gradient system as for an all-convergent lens system with the same spacing of convergent lenses.

For weak lenses (i.e. $2f/L \gg 1$) the output ray departure due to input ray displacement is independent of both the focal length and spacing of the lenses, and is independent of lens spacing but proportional to focal length for input ray slope.

Both the alternating-gradient system and all-convergent lens focusing arrangements exhibit discontinuities in the maximum ray displacement versus focal strength relation.

Viewed over-all, alternating-gradient focusing for light guidance does a surprisingly efficient job and may be advantageous over all-convergent lens systems if the alternating-gradient arrangement has structural or economic advantages.

I. INTRODUCTION

In research on guidance of light waves for communication we are considering use of a sequence of lenses of alternately convergent and divergent types. For example, a guidance system using tubular thermal gas lenses might employ continuous flow of gas through a tube whose walls are alternately warmer and cooler than the gas within. Thus, the mechanism used to cause the focusing may have the alternating character, and the question comes to the fore — how well can one focus with such a structure as compared to the use of a sequence of all-convergent lenses?

Alternating-gradient focusing has previously been used on electron beams[1] and on particle accelerators.[2] The present study discovered an error in the previous determination of stability conditions and revealed some little known but interesting properties of alternating-gradient focusing. A comparison is made with conventional focusing using all-convergent lenses.

II. ANALYSIS OF ALTERNATING-GRADIENT SYSTEMS

One might wonder whether a series of equal-focal-length and alternately converging and diverging lenses would give any net focusing at all, since the average dielectric constant along all paths parallel to the axis would be the same. It is well known, however, that a divergent lens followed by a convergent lens of equal focal length, spaced a finite distance less than the focal length, gives a net *converging* lens, and the same is true if the order of the lenses is reversed. Thus, a net focusing is to be expected for an infinite series of such lens pairs.[1]

We consider a sequence of alternating convergent and divergent lenses equally spaced a distance L and of equal focal lengths, f. We follow the method of analysis used by Pierce.[1] There are two cases to cover, one in which the first lens of the array is a divergent lens (obtained by starting at $n = 0$ in Fig. 1) and the other in which the first lens of the array is a convergent lens (obtained by starting at $N = 0$ in Fig. 4, below).

With reference to Fig. 2, and taking the input ray at plane a to have a slope r_a' and a displacement r_a from the longitudinal axis, the output ray from the lens at plane b will be

$$r_b = r_a + Lr_a' \tag{1}$$

$$r_b' = r_a' + (1/f)r_b\,. \tag{2}$$

At plane c this ray will be described by

$$r_c = r_b + Lr_b' \tag{3}$$

$$r_c' = r_b'. \tag{4}$$

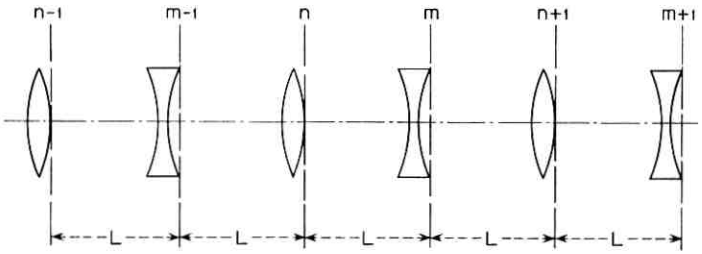


Fig. 1 — Sequence of lenses — first case.

Eliminating r_b and r_b' from (3) and (4) gives

$$r_c = \left(1 + \frac{L}{f}\right) r_a + L\left(2 + \frac{L}{f}\right) r_a' \tag{5}$$

$$r_c' = \left(1 + \frac{L}{f}\right) r_a' + \frac{r_a}{f}. \tag{6}$$

Hence, with reference to Fig. 1, we can write

$$r_{n+1} = \left(1 + \frac{L}{f}\right) r_n + L\left(2 + \frac{L}{f}\right) r_n' \tag{7}$$

$$r_{n+1}' = \left(1 + \frac{L}{f}\right) r_n' + \frac{r_n}{f} - \frac{r_{n+1}}{f}. \tag{8}$$

These two equations lead to

$$r_{n+2} - \left[2 - \left(\frac{L}{f}\right)^2\right] r_{n+1} + r_n = 0. \tag{9}$$

The solution to (9) is

$$r_n = A \cos n\theta + B \sin n\theta \tag{10}$$

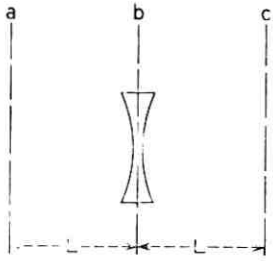


Fig. 2 — Lens subsection for Fig. 1.

where

$$\theta = \cos^{-1}\left[1 - \frac{1}{2}\left(\frac{L}{f}\right)^2\right] \tag{11}$$

and where A and B are constants to be determined.

Equation (11) differs from the corresponding equation on page 200 of Ref. (1), which is believed to be in error. The correct condition for stability, from (11), is

$$0 < \tfrac{1}{2}(L/f)^2 < 2 \tag{12}$$

or

$$0 < L/f < 2.$$

We put in the boundary conditions, at $n = 0$

$$r_n' = r_0' \tag{13}$$

$$r_n = r_0 . \tag{14}$$

We make use of a general theorem* stating the orthogonality of the effects of r_0 and r_0' and seek a solution with those quantities as factors. This leads to the following form for r_n, using (10), (7), (13) and (14):

$$r_n = r_0 k_1 \cos(n\theta - \varphi_1) + r_0' L\, k_2 \sin n\theta \tag{15}$$

where

$$k_1 = \left[\frac{2}{1 - (L/2f)}\right]^{\frac{1}{2}} \tag{16}$$

$$\varphi_1 = |\cos^{-1} k_1^{-1}| \tag{17}$$

$$k_2 = \frac{[(2f/L) + 1]}{[1 - \frac{1}{4}(L/f)^2]^{\frac{1}{2}}} = \frac{[2 + (L/f)]}{\sin\theta} . \tag{18}$$

The general form of k_1 and k_2 versus L/f is shown in Fig. 3. Further discussion will be postponed to a later point in this paper.

We are also interested in the displacement r_m at the output of the mth diverging lens (Fig. 1). Using the relation

$$r_m = r_n + L r_n' \tag{19}$$

and using (7) for r_n', (15) for r_n with appropriate trigonometric relations, it can be shown that

$$r_m = r_0 k_3 \cos(m\theta - \varphi_3) + r_0' L\, k_4 \cos(m\theta - \varphi_4) \tag{20}$$

* See Appendix B.

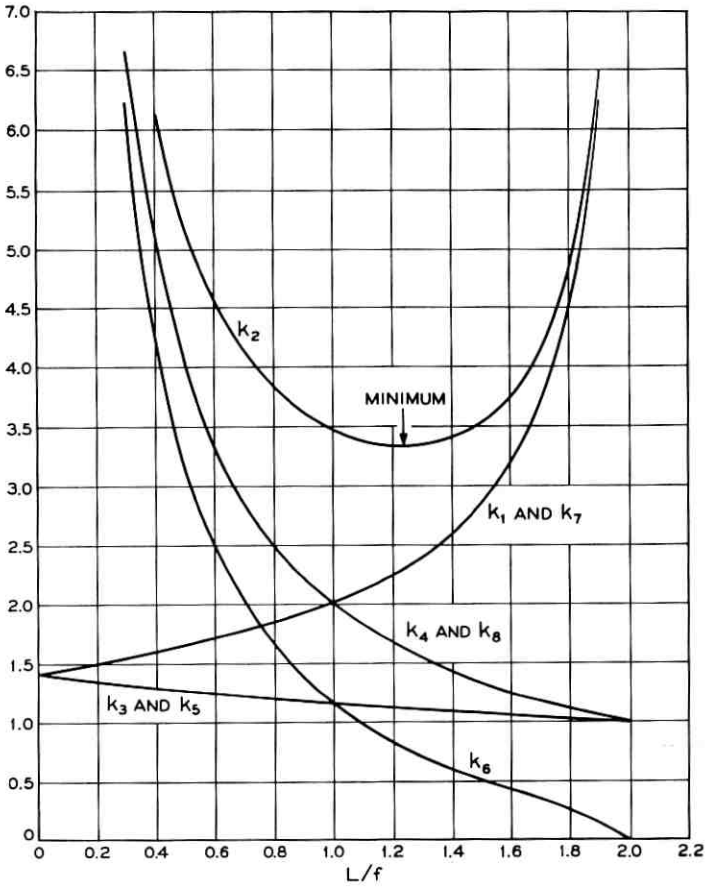


Fig. 3 — Coefficients relating input ray slope and displacement to ray displacement at the nth, mth, Nth, and Mth lenses of the alternating-gradient lens systems of Figs. 1 and 4.

where

$$k_3 = \left[\frac{2}{1 + (L/2f)}\right]^{\frac{1}{2}} \tag{21}$$

$$\varphi_3 = |\cos^{-1} k_3^{-1}| \tag{22}$$

$$k_4 = 2f/L \tag{23}$$

$$\varphi_4 = |\cos^{-1} k_4^{-1}| \tag{24}$$

and θ is again defined by (11). Plots of k_3 and k_4 are given in Fig. 3.

Equations (15) and (20) give the ray displacements at any lens in the system when the input is at a plane adjacent to a converging lens (i.e. at $n = 0$, Fig. 1). Before discussing interesting features of such ray propagation we will give the corresponding solutions for the case where the input is adjacent to a diverging lens (i.e. at $N = 0$, Fig. 4).

With reference to Figs. 4 and 5, and following a derivation similar to that carried out in connection with equations (1) to (9), it is found that

$$r_{N+2} - [2 - (L/f)^2]\, r_{N+1} + r_N = 0. \tag{25}$$

Note that (25) is identical to (9) and with the change of n into N the solution for (25) is again (10) and (11). When the initial conditions are put in, at $N = 0$, $r_N = r_0$ and $r_N' = r_0'$, we get

$$r_N = r_0 k_5 \cos(N\theta + \varphi_5) + r_0' L k_6 \sin N\theta \tag{26}$$

where

$$k_5 = \frac{(L/f)[2 - (L/f)]^{\frac{1}{2}}}{\sin\theta} = \left[\frac{2}{1 + (L/2f)}\right]^{\frac{1}{2}} \tag{27}$$

$$\varphi_5 = |\cos^{-1} k_5^{-1}| \tag{28}$$

$$k_6 = \frac{[(2f/L) - 1]}{[1 - \frac{1}{4}(L/f)^2]^{\frac{1}{2}}} = \frac{[2 - (L/f)]}{\sin\theta}. \tag{29}$$

Note that k_5 is identical to k_3 in (21).

For r_M we find

$$r_M = r_0 k_7 \cos(M\theta + \varphi_7) + r_0' L k_8 \sin(M\theta - \varphi_8) \tag{30}$$

where

$$k_7 = \left[\frac{2}{1 - (L/2f)}\right]^{\frac{1}{2}} \tag{31}$$

$$\varphi_7 = |\cos^{-1} k_7^{-1}| \tag{32}$$

$$k_8 = 2f/L \tag{33}$$

$$\varphi_8 = |\cos^{-1} k_8^{-1}|. \tag{34}$$

Note that k_8 and φ_8 are identical to k_4 and φ_4, and that k_7 is identical to k_1. Plots of k_5, k_6, k_7, k_8 are given in Fig. 3.

III. RELATIONS FOR A SEQUENCE OF CONVERGING LENSES

For comparison purposes we will want to refer to the case of a sequence of identical convergent lenses equally spaced. The analysis is similar to that above for the alternating gradient lenses. The results are

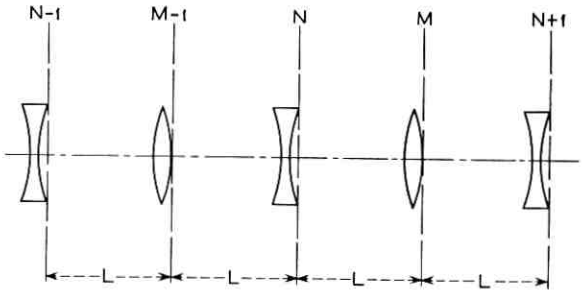


Fig. 4 — Sequence of lenses — second case.

as follows. Let the lens spacing be s, and the focal length be f. Then the displacement r_p at the pth lens is

$$r_p = r_0 k_9 \cos(p\delta - \varphi_9) + r_0' s\, k_{10} \sin p\delta \tag{35}$$

where

$$k_9 = \left[\frac{4f/s}{(4f/s) - 1}\right]^{\frac{1}{2}} \tag{36}$$

$$\varphi_9 = |\cos^{-1} k_9^{-1}| \tag{37}$$

$$k_{10} = \frac{(f/s)^{\frac{1}{2}}}{[1 - (s/4f)]^{\frac{1}{2}}} = \frac{1}{\sin\delta} \tag{38}$$

$$\delta = \cos^{-1}[1 - (s/2f)] \tag{39}$$

$$s/f = 2(1 - \cos\delta). \tag{40}$$

The system is stable in the sense that an input displacement r_0 or slope r_0' will remain bounded as p is increased if

$$0 < s/f < 4. \tag{41}$$

Fig. 6 shows the values of k_9 and k_{10} for comparison to Fig. 3.

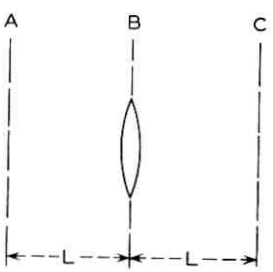


Fig. 5 — Lens subsection for Fig. 4.

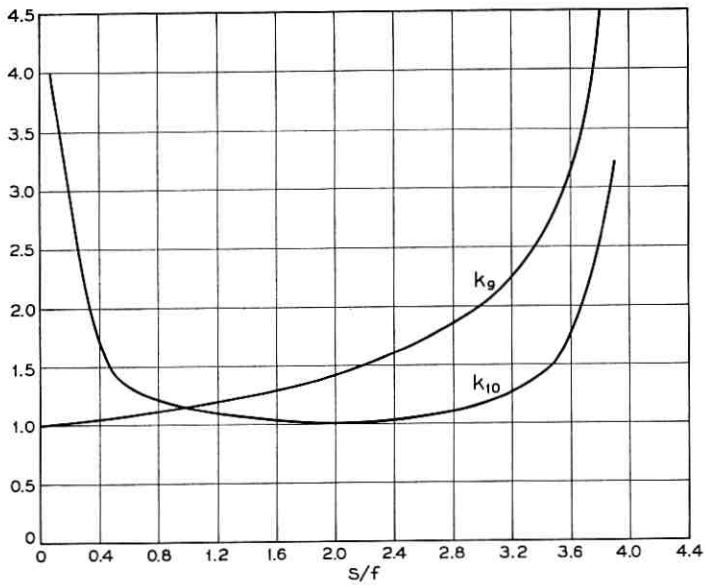


Fig. 6 — Coefficients relating ray slope and displacement to ray displacement at the pth lens of an all-convergent lens system; f = focal length, s = lens spacing.

IV. STABILITY COMPARISON

Since the converging lenses in Fig. 1 or Fig. 4 are spaced a distance $2L$, a comparison of (41) and (12) shows that the alternating-gradient system formed by adding a divergent lens halfway between the convergent lenses of an all-convergent lens system has the same stability condition as the original system. At a later point it will be shown that this is reasonable physically.

V. THE WEAK LENS CASE

When the lenses are weak, i.e., when $2f/L \gg 1$, the general expressions may be simplified to show some remarkable properties of alternating-gradient focusing. When the first lens is a diverging one, (15) and (20) yield

$$r_m = r_n = r_0 \sqrt{2} \cos [n\theta - (\pi/4)] + r_0' 2f \sin n\theta \qquad (42)$$

and when the first lens is a converging one, (26) and (30) yield

$$r_N = r_M = r_0 \sqrt{2} \cos [N\theta + (\pi/4)] + r_0' 2f \sin N\theta. \qquad (43)$$

These expressions show that for an input ray displacement without

slope the maximum displacement of the transmitted ray (as n or N varies) is $\sqrt{2}$ times the input ray displacement, independent of both focal length and lens spacing! Also, for an input ray of zero displacement but finite slope r_0', the maximum displacement of the transmitted ray is $2fr_0'$, independent of lens spacing L. The angle θ is dependent on f and L and goes to zero as $f \to \infty$.

In an all-convergent lens system the similar condition $4f/s \gg 1$ leads to [from (35)]

$$r_p = r_0 \cos p\delta + r_0' \sqrt{fs} \sin p\delta. \tag{44}$$

In comparing the alternating-gradient system to the all-convergent lens system for weak lenses, we see that for an input ray with zero slope r_0' but finite displacement, r_0, the maximum output displacement for the alternating-gradient system is $\sqrt{2}$ times that of the all-convergent lens system. For input ray displacement $r_0 = 0$ but finite r_0', we see that the maximum output displacement for the alternating-gradient system is larger than for the all-converging lens system by the factor [see (42) and (44)]:

$$\frac{2fr_0'}{(fs)^{\frac{1}{2}}r_0'} = (4f/s)^{\frac{1}{2}}. \tag{45}$$

Our assumption of weak lenses made $4f/s \gg 1$, so (45) is a factor of two or more.

In this weak lens case both θ and δ are small angles, and from (11) and (39)

$$\theta \cong L/f \tag{46}$$

$$\delta \cong (s/f)^{\frac{1}{2}}. \tag{47}$$

Using the case of $s = 2L$, which is the alternating-gradient system formed by adding a diverging lens between the lenses of an all-convergent lens system

$$\theta/\delta = (s/4f)^{\frac{1}{2}}. \tag{48}$$

Since $4f/s \gg 1$ by our weak lens definition, θ/δ is less than unity and the period of the alternating-gradient system encompasses a great many more convergent lenses than does the all-convergent lens system with the same spacing of convergent lenses. This is as would be expected.

VI. OPTIMUM FOCAL LENGTHS

We now inquire as to whether there is a best value for the lens strength in order to minimize output ray displacement. On the assumption that

the sine and cosine terms of (15), (20), (26), (30) and (35) go through unity for some number of lenses, the question is whether or not the coefficients k_1, $k_2 \cdots k_{10}$ have any minima.

For the all-convergent lens system Fig. 6 illustrates that k_9 has no useful minimum but that k_{10}, relating input ray slope to output ray displacement as in (35), does have a minimum. By setting

$$\frac{d}{df}(k_{10}) = 0 \tag{49}$$

we find

$$\left.\frac{s}{f}\right|_{\text{opt}} = 2 \tag{50}$$

at which condition $k_{10} = 1.0$. We note that the displacement r_p due to r_0' is $r_0' s k_{10} \sin p\delta$, so we have a minimum in this displacement when k_{10} is a minimum provided $\sin p\delta$ goes through unity for some number of lenses p. This is the most typical case, but there are notable exceptions. Suppose, for example, that δ of (39) is $\pi/3$, corresponding to $s/f = 1$; then $p\delta = \pi/3, 2\pi/3, \pi, 4\pi/3$, etc., as illustrated in Fig. 7, and $|\sin p\delta|$ never exceeds $\sin \delta$. Hence the maximum value of $r_0' s\, k_{10} \sin p\delta$ is $r_0' s$ for $s/f = 1$. It is shown in Appendix A that there is an infinite series of such discrete values, but the largest departure of the maximum value of $r_0' s k_{10} \sin p\delta$ from k_{10} is 15 per cent, occurring at s/f values of 1 and 3, as illustrated in Fig. 14 (see Appendix A).

Turning now to the alternating-gradient system, the only coefficient having a useful minimum is k_2 of (15), relating input ray slope to ray displacement at the converging lenses of Fig. 1. We find the minimum in k_2 by setting

$$\frac{d}{df}(k_2) = 0 \tag{51}$$

which leads to the equation

$$(L/f)^3 + 4(L/f)^2 - 8 = 0. \tag{52}$$

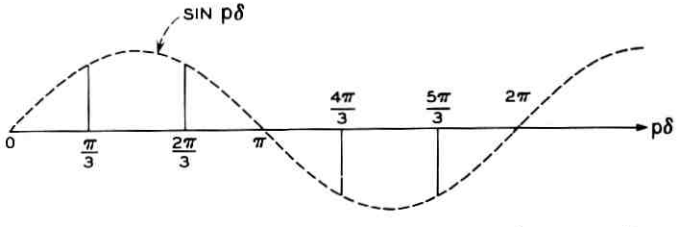


Fig. 7 — Diagram of $\sin p\delta$, $p = 1,2,3 \cdots$, when $\delta = \pi/3$.

The appropriate root of this equation is

$$\left.\frac{L}{f}\right|_{\text{opt}} = 1.237 \tag{53}$$

at which we calculate $\theta = 76.4°$, and $k_2 = 3.33$. Fig. 3 shows this minimum is sharper than the corresponding one for k_{10} of the all-converging lens system, Fig. 6. We note that $k_2 \sin n\theta$ of (15) contains the $\sin n\theta/\sin \theta$ factor, so once again (as described in Appendix A) for $\theta = \pi/3$ and other values, the maximum value of $k_2 \sin n\theta$ will be somewhat less than the value of k_2.

It is important to compare the optimized focusing effect of the alternating-gradient system to that for the all-convergent lens system. We make the comparison on the alternating-gradient system formed by adding a divergent lens of equal focal length in between the lenses of an all-convergent lens system; then we have $s = 2L$. The optimized maximum displacement due to input ray slope is

$$r_0' s = 2 r_0' L$$

for the convergent lens system, and is

$$3.33\, r_0' L$$

for the alternating-gradient system. It is remarkable that the focusing effect of the alternating-gradient system is so nearly the same as that of the all-convergent lens system. In practice it may be advantageous to get the focusing action in a manner that inherently reverses itself periodically. This analysis shows that such structures are nearly as effective as those wherein the focusing effect is always convergent.

Fig. 3 shows that the focal length which is optimum with respect to the input ray slope (k_2) is also an acceptable region with respect to input ray displacement (k_1 and k_3).

VII. RAY PATHS

One can get a useful physical feel for the wave propagation by tracing the rays in a few of the important cases.

For the all-convergent lens system optimized according to (50), Fig. 8 shows the ray paths for a zero-slope finite-displacement input ray and for a zero-displacement finite-slope input ray.* Here $\delta = 90°$ [see (35)] and a period is completed in 4 lenses. As proved in Appendix B, the

* Note that $k_9 = \sqrt{2}$, with $s = 2f$ in (36), but $k_9 \cos(p\delta - \varphi_9)$ is always ± 1 for any p. (The angle $\varphi_9 = 45°$.) This is an example of the caution that must be exercised in regarding the k's as maximum values of the various terms in r_n, r_m, r_p, etc.

response to an arbitrary input ray can be obtained by a linear superposition of the responses shown in Fig. 8.

For the alternating-gradient system the optimum according to (53) corresponds to an angle θ in (15), (20), (26) and (30) of 76.4°, which makes the ray path periodic only at a very large number of lenses. However, a very useful feel can be obtained from the ray paths for $\theta = 90°$, corresponding to $L = \sqrt{2}f$ and giving a value of k_2 only slightly larger than the minimum value (3.414 compared to 3.33). These ray plots are shown in Figs. 9 and 10 for the two types of input rays at the two possible points in the alternating-gradient system. We note that input ray displacement causes the same maximum displacement in the response regardless of where it occurs. Input ray slope is much more serious when

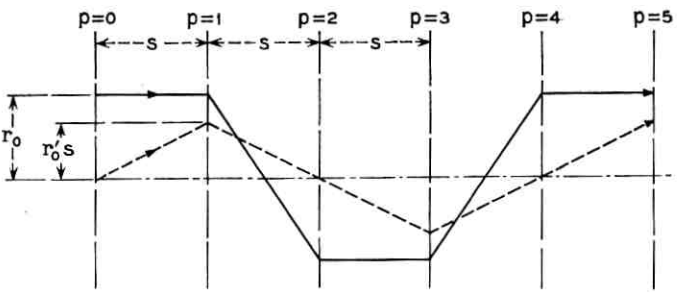


Fig. 8 — Ray paths in the confocal all-convergent lens system, $s = 2f$.

it occurs in front of a diverging lens than when it occurs in front of a converging lens. Again, the response to arbitrary input rays can be obtained by adding the plotted responses.

One can gain a little feel for the stability comparison made previously by looking at Figs. 11 and 12. Even though the r_0 term and the r_0' term of (35) for the all-convergent lens system go to infinity individually when $s = 4f$, a suitable combination of input ray slope and displacement remains bounded and this is illustrated in Fig. 11.* Any reduction in focal length f causes instability, and any increase in f leaves the system completely stable. It is clear that adding a lens of any kind at the midpoint between lenses in Fig. 11 will not alter the propagation of that ray. In Fig. 12 we see that adding a divergent lens in between the converging

* One can obtain these values of r_p from (35) by a suitable limiting process. It is helpful to start with the alternative form of (35):

$$r_p = r_0 \left\{ \cos p\delta + \frac{1}{[(4f/s) - 1]^{\frac{1}{2}}} \sin p\delta \right\} + r_0' s\, k_{10} \sin p\delta.$$

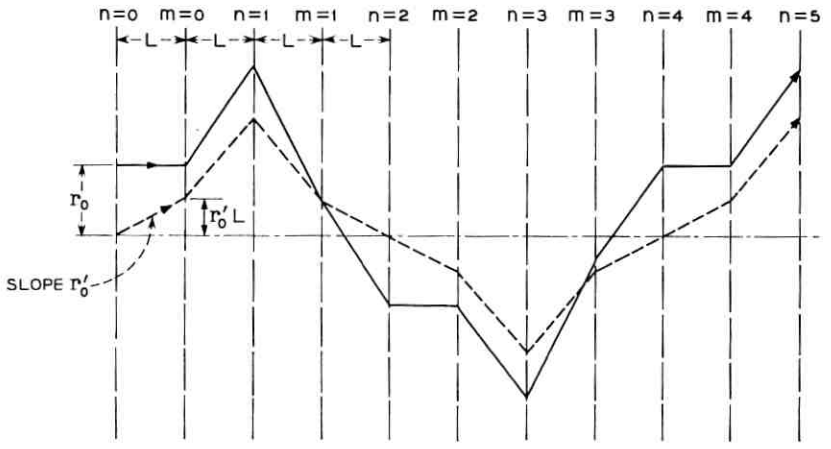


Fig. 9 — Ray paths for system of Fig. 1, $L = \sqrt{2f}$.

lenses will cause reductions in the focal length of the $p = 1$ converging lens to make the ray sent on to the $p = 2$ lens diverge even more; for increases in the focal length of the $p = 1$ lens, the divergent lens reduces the angle of the ray sent on to the $p = 2$ lens. Hence, it is plausible that the addition of the divergent lens between the convergent lenses does not alter the stability requirement on the focal lengths.

Given the mathematically-derived condition that divergent lenses of

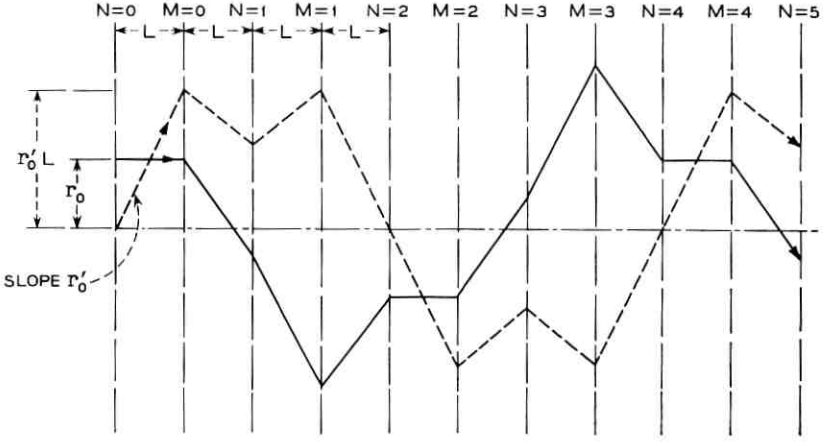


Fig. 10 — Ray paths for system of Fig. 3, $L = \sqrt{2f}$.

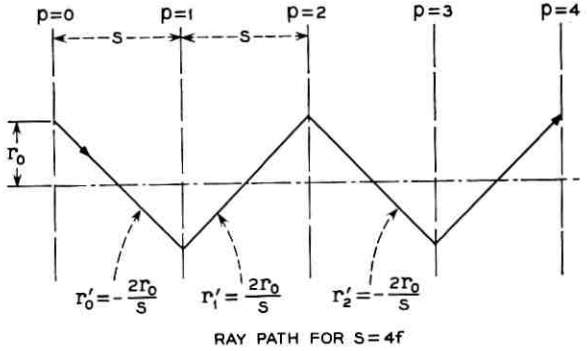


Fig. 11 — Ray path for $s = 4f$ in all-convergent lens system.

focal length f added to a chain of convergent lenses of focal length f do not change the stability criterion as described above, the physical argument just outlined leads to the conclusion that weaker divergent lenses would also leave the stability criterion unchanged.*

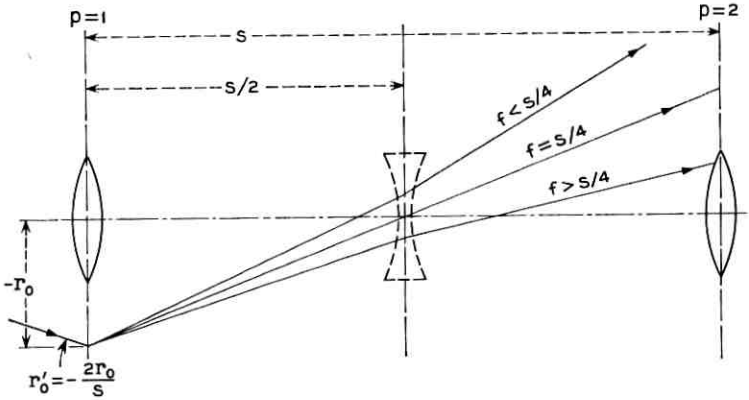


Fig. 12 — Ray path for f near $s/4$ to illustrate effect of divergent lens.

VIII. CONCLUSION

Alternating-gradient focusing is surprisingly close to an all-convergent lens system in focusing ability, and may be preferred if practical matters such as structural features or cost favor the alternating-gradient system.

* When reading this manuscript, Mr. J. P. Gordon commented that this conclusion is in agreement with the work of Boyd and Kogelnik[3] which can be shown to yield the relation $f_2 > f_1 - L/2$ for the required focal length f_2 of the diverging lens in terms of the focal length f_1 of the converging lens and the spacing L.

APPENDIX A

We examine here the maximum value that the term $r_0's$ $k_{10} \sin p\delta$ of (35) can take as a function of lens number p when our objective is to minimize the term through appropriate choice of focal length. In the body of the article it has been shown that k_{10} has a minimum at $s = 2f$. This corresponds to a value of $\delta = \pi/2$ from (39) and it is evident that $\sin p\delta = \sin p\,\pi/2$ is either zero or unity for all integral values of p.

In the more general case we want to know the value of

$$k_{10} \sin p\delta = \sin p\delta/\sin \delta. \tag{54}$$

When it is recognized that p may take on all integral values greater than zero it follows that the maximum value of $[(\sin p\delta)/\sin \delta]$ as p varies can never be less than unity for any fixed δ.

It is possible for $[(\sin p\delta)/\sin \delta]$ to have a maximum value which is smaller than $k_{10} = 1/\sin \delta$. That is to say, $\sin p\delta$ does not necessarily go through unity even though p ranges from 0 to ∞ in integral steps.

Referring to Fig. 13, the maximum value of $\sin p\delta$ will be less than unity if

$$\delta(q + \tfrac{1}{2}) = \pi/2 \tag{55}$$

or

$$\delta = \pi/(2q + 1) \tag{56}$$

where $q = 1, 2, 3, \cdots$. It also is true that $\sin p\delta$ will have a maximum value less than unity for

$$\delta = r[\pi/(2q + 1)] \tag{57}$$

where $r = 1, 2, 3, 4, \cdots$.

The values of s/f corresponding to these values of δ and the resultant values of maximum $k_{10} \sin p\delta$ are given in Table I. Column 5 shows the ratio of k_{10} to the maximum of $k_{10} \sin p\delta$, and is a measure of the error

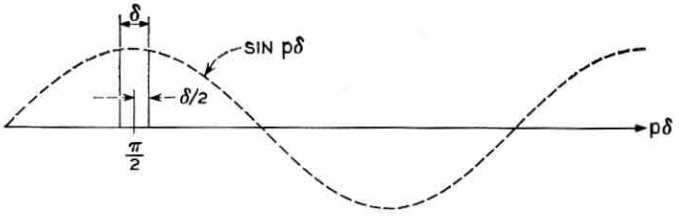


Fig. 13 — Diagram of $\sin p\delta$ yielding $|\sin p\delta| < 1$ for all p.

TABLE I

1 δ	2 s/f	3 k_{10}	4 Max. Value of $k_{10} \sin p\delta$	5 Column 3 ÷ Column 4
$\pi/2$	2	1.0	1.0	1.0
$\pi/3$	1	1.15	1.0	1.15
$2\pi/3$	3	1.15	1.0	1.15
$\pi/5$	0.38	1.70	1.62	1.05
$2\pi/5$	1.38	1.05	1.0	1.05
$3\pi/5$	2.62	1.05	1.0	1.05
$4\pi/5$	3.62	1.70	1.62	1.05
$\pi/7$	0.194	2.31	2.255	1.023
$2\pi/7$	0.750	1.28	1.25	1.023
$3\pi/7$	1.554	1.023	1.0	1.023
$4\pi/7$	2.444	1.023	1.0	1.023
$5\pi/7$	3.226	1.28	1.25	1.023
$6\pi/7$	3.806	2.31	2.255	1.023

made in assuming sin $p\delta$ goes through unity. That ratio is $1/\cos(\delta/2)$ where δ is given by (56). All values for a given q in (57) result in the same error, but the various values of r indicate the values of δ and s/f at which that error will appear. Fig. 14 summarizes the data of Table I; for p ranging up to infinity it is only at the discrete values of δ given by (57) that the maximum of $k_{10} \sin p\delta$ differs from k_{10}.

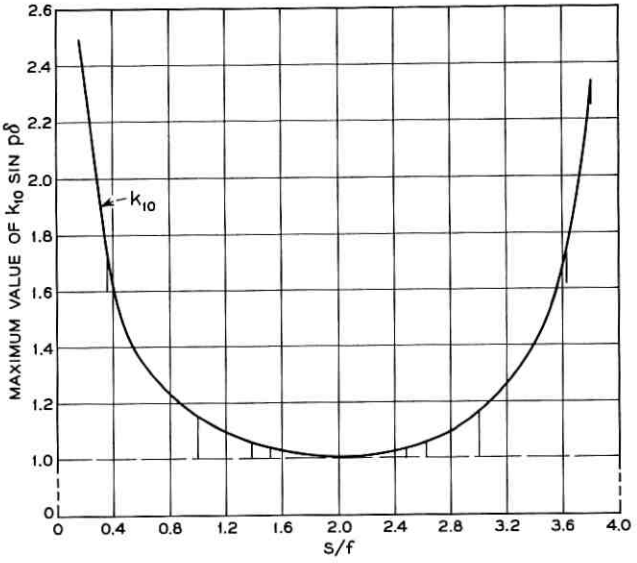


Fig. 14 — Maximum value of $k_{10} \sin p\delta$ vs s/f.

If p were finite and the ratio s/f was varied, the plot of Fig. 14 would presumably show finite-width dips of the same over-all depth as those plotted.

APPENDIX B

It is the purpose of this appendix to point out that the "thin lens" description of light ray propagation leads to the following conclusion (see Fig. 15):

The slope and displacement of the output ray of an arbitrary sequence of lenses for an input ray of slope r_0' and displacement r_1 is exactly the algebraic sum of the slopes and displacements of the output ray found (*i*) for an input ray of slope r_0' with zero displacement from the axis, and assuming all lens displacements $d_n = 0$, (*ii*) for an input ray of displacement r_1 with zero slope and assuming all lens displacements $d_n = 0$, and (*iii*) for an input ray of zero slope and zero displacement and assuming one lens displacement at a time is nonzero, summing the ray output slopes and displacements thus found over all lens displacements.

The proof is as follows: For the nth lens in a sequence of lenses (see Fig. 1):

$$r_n' = r_{n-1}' - \left(\frac{r_n - d_n}{f_n}\right) \tag{58}$$

where r_n' is the slope of the ray immediately following the nth lens and r_n is the displacement at the nth lens. We note that the angular lens rotation φ_n does not affect the ray propagation, an approximation which implies that

$$(r_n - d_n) \cos \varphi_n \cong (r_n - d_n) \tag{59}$$

or

$$\varphi_n \ll 1.$$

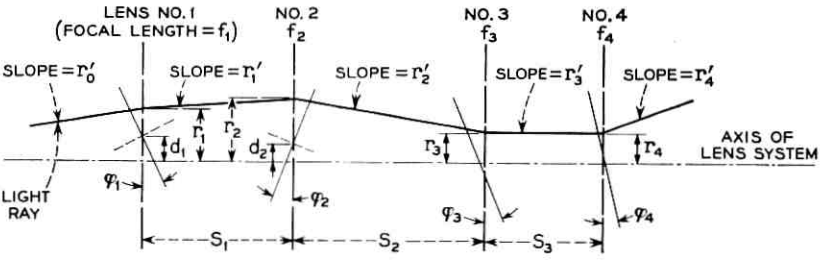


Fig. 15 — Light-ray path in an arbitrary lens system.

We may expand (58) to form

$$r_n' = r_0' - \frac{(r_1 - d_1)}{f_1} - \frac{(r_2 - d_2)}{f_2} \cdots - \frac{(r_n - d_n)}{f_n}. \tag{60}$$

Similarly, the displacement r_n at the nth lens is

$$r_n = r_{n-1} + s_{n-1} r_{n-1}' \tag{61}$$

which may be expanded to

$$r_n = r_{n-1} + s_{n-1} \left\{ r_0' - \frac{(r_1 - d_1)}{f_1} - \frac{(r_2 - d_2)}{f_2} \cdots - \frac{(r_{n-1} - d_{n-1})}{f_{n-1}} \right\} \tag{62}$$

which is valid for $n \geqq 2$.

We may examine each term of (60) and (62) and find that

$$r_n' = A_n r_0' + B_n r_1 + \sum_{m=1}^{m=n} \alpha_m \frac{d_m}{f_m} \tag{63}$$

$$r_n = C_n r_0' + D_n r_1 + \sum_{m=1}^{m=n} \beta_m \frac{d_m}{f_m} \tag{64}$$

in which A_n, B_n, C_n, D_n and the α_m and β_m are all independent of r_0', r_1 and the d_m.

We have thus proven the above-stated conclusion.

REFERENCES

1. Pierce, J. R., *Theory and Design of Electron Beams*, 2nd ed., D. Van Nostrand, Princeton, 1954.
2. Courant, E. D., Livingston, M. S., and Snyder, H. S., The Strong-Focusing Synchrotron — A New High-Energy Accelerator, Phys. Rev., **88**, Dec. 1, 1952, pp. 1190–1196.
3. Boyd, G. D., and Kogelnik, H., Generalized Confocal Resonator Theory, B.S.T.J., **41**, July, 1962, pp. 1347–1369.

Analysis of a Tubular Gas Lens

By D. MARCUSE and S. E. MILLER

(Manuscript received June 8, 1964)

If a cool gas is blown into a hot tube, it acts as a positive lens which will focus a light beam passing through the tube.

Using a theory presented in Ref. 3, we give curves which show the temperature distribution in the tube as a function of the distance from the tube axis and also as a function of the distance along the axis.

The focusing power of the lens is described by the difference in phase angle between a ray on the tube axis minus a ray at arbitrary distances from this axis and also as the second derivative of the phase angle on the axis of the tube. The phase curves, as a function of distance r from the tube axis, follow very closely an r^2 dependence. Expressions are given for the focal length of the lens.

The power consumption of the lens is discussed, and a figure of merit is defined as focusing power per watt. The gas used for this lens should be selected such that $(n - 1)/k$ is as large as possible (where n is the index of refraction, k the heat conductivity of the gas).

Using CO_2 and a $\frac{1}{4}$-inch ID tube 5 inches long heated 20°C above the incoming gas, a focal length of 5 feet with a power consumption of 0.325 watt is calculated; the focal length is inversely proportional to power consumption within certain limits.

I. INTRODUCTION

A communications system using light as the carrier of intelligence needs an efficient medium to propagate light from transmitter to receiver. Among the several alternatives, the idea of Goubau and Schwering[1] of confining and propagating an electromagnetic wave with a system of lenses appears promising. However, in a lens-waveguide system there is a wide range of possibilities as to what types of lenses to use. Conventional glass lenses present problems, since they may not only absorb light in the glass medium itself, but furthermore present important reflection losses which can be only partially avoided by special techniques such as coating the lens surfaces or making use of the Brewster angle.

Even if such corrective measures are used, there is still residual reflection and scattering of light due to unavoidable surface irregularities.

It appears that most of the problems connected with glass lenses could be overcome if, instead of a high-index medium such as glass, a very low-index focusing medium were used. If the transition from air into a dense medium could be avoided, the problem of light reflection would not exist. Gases present themselves as an obvious choice of a low-index dielectric medium. Their dielectric constant can be influenced by changing their density. A change of density is most easily effected by varying the gas temperature.

D. W. Berreman[2] built a successful gas lens by maintaining a temperature gradient between a hot helix and a cold cylindrical enclosure. Alternatively, D. W. Berreman and S. E. Miller proposed a gas lens formed by blowing a cool gas into a hot tube (Fig. 1). Since the gas heats up first at the wall of the tube and remains cool longer at its center, it has a density distribution of higher-density gas in the center of the tube and decreasing density towards the wall. Since an increase in density is accompanied by an increase in dielectric constant, it is easy to understand that the cool gas flowing through the hot tube acts as a positive lens and tends to focus a light beam traveling along the axis of the tube.

We present in this article some theoretical results of the temperature distribution in the gas and the difference in phase angle between two light beams, one traveling along the tube axis and the other traveling closer to the wall of the tube. This phase difference is a measure of the focusing power of the lens. For an economical lens we want maximum phase shift with a minimum of thermal power. We present curves showing the power consumption of the lens as well as the ratio of phase difference to power consumption. These data allow the construction of an optimum lens. Different gases give different lens properties. A gas is most efficient if the ratio $(n - 1)/k$ is large, where n is the index of refraction of the gas and k is its heat conductivity.

II. TEMPERATURE DISTRIBUTION

The theory of temperature distribution in a cool gas which is blown into a hot tube of constant temperature is presented in Ref. 3.

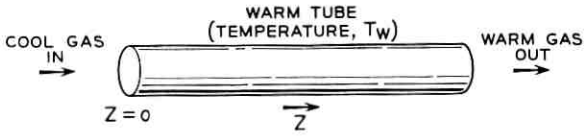


Fig. 1 — Gas lens using forced flow in a tube.

It is assumed that the gas flow is laminar and has the radial velocity distribution of a viscous fluid

$$v(r) = v_0[1 - (r/a)^2] \tag{1}$$

where

r = distance from tube axis
a = radius of tube
v_0 = gas velocity at $r = 0$.

The temperature T of the gas is given as

$$\theta = T_w - T$$

where T_w is the temperature of the wall of the tube. It is normalized with respect to

$$\theta_0 = T_w - T_0$$

with T_0 being the temperature of the cool gas before it enters the hot tube. θ/θ_0 is expanded in terms of functions $R_n(r/a)$, which are shown in Fig. 2 for $n = 0$, 1 and 2. Values of $R_n(r/a)$ are listed in Table I. The temperature depends on the distance z measured from the beginning of the hot tube, the gas velocity v_0, and the following material parameters

k = heat conductivity measured in cal/cm sec deg)
(deg = degrees Kelvin)
ρ = gas density in gram/cm^3
c_p = specific heat at constant pressure in cal/gram.

All these parameters depend somewhat on the temperature but are considered constant in the derivation of the theory. They enter the equations in the combination

$$\sigma = k/av_0\rho c_p . \tag{2}$$

TABLE I — R FUNCTIONS OF FIG. 2

x	$R_0(x)$	$F(x)$	$R_1(x)$	$R_2(x)$
0	1	1	1	1
0.1	0.9819	0.9805	0.8923	0.753
0.2	0.9290	0.9261	0.6067	0.206
0.3	0.8456	0.8432	0.2367	−0.290
0.4	0.7382	0.7382	−0.1062	−0.407
0.5	0.6147	0.6175	−0.3399	−0.204
0.6	0.4833	0.4880	−0.4317	0.104
0.7	0.3506	0.3535	−0.3985	0.278
0.8	0.2244	0.2244	−0.3051	0.278
0.9	0.1069	0.1041	−0.1637	0.144
1.0	0	0	0	0

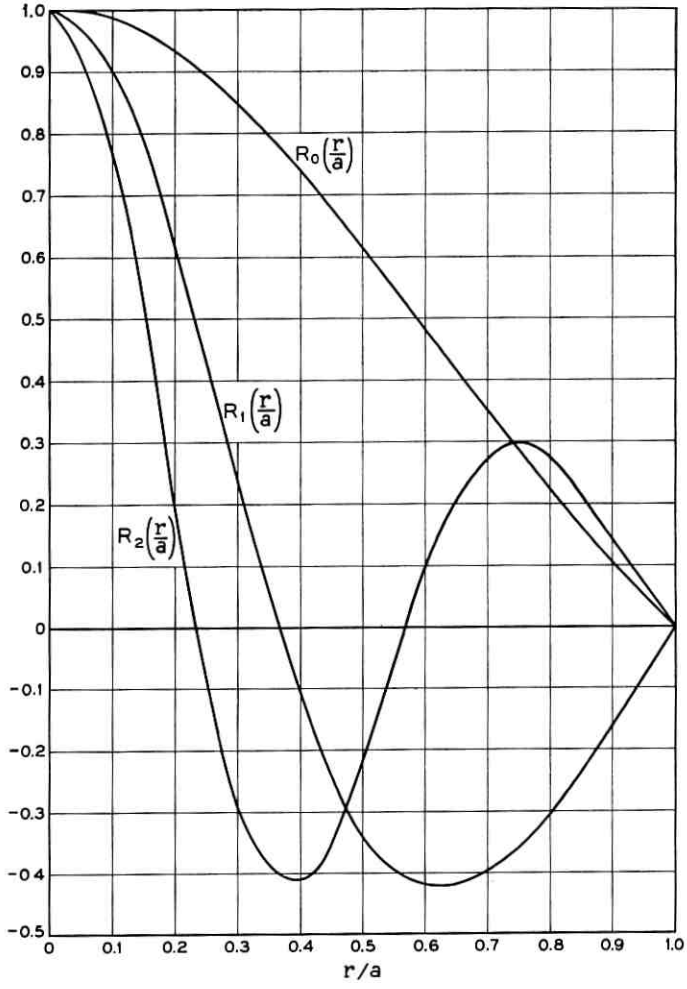


Fig. 2 — Functions R_0, R_1 and R_2 vs r/a.

The first three terms of the infinite series describing the temperature distribution in the tube are

$$\frac{\theta}{\theta_0} = 1.477 \exp\left(-7.316\sigma \frac{z}{a}\right) R_0\left(\frac{r}{a}\right) - 0.810 \exp\left(-44.36\sigma \frac{z}{a}\right) \cdot R_1\left(\frac{r}{a}\right) + 0.385 \exp\left(-106\sigma \frac{z}{a}\right) R_2\left(\frac{r}{a}\right) \pm \cdots. \tag{3}$$

The approximation is fairly poor at $z = 0$. However, the exponential factors in the higher terms of the series drop off very rapidly as z increases so that the approximation is already very good for values of

$$\sigma(z/a) > 0.01.$$

Fig. 3 shows θ/θ_0 at $r = 0$ as a function of $\sigma(z/a)$. It is apparent that for $\sigma(z/a) > 0.05$, the distribution of θ/θ_0 drops off exponentially. Fig. 4 shows the r/a dependence of θ/θ_0 for different values of $\sigma(z/a)$.

To make Figs. 3 and 4 more meaningful, we list in Table II the material parameters for several gases at 20°C and a pressure of 760 mm Hg.

III. POWER CONSUMPTION

The principle of operation of our gas flow lens requires that we heat the cool gas inside the hot tube. Even if we neglect all power losses to the environment, we have to spend a certain amount of heat power to operate our lens. For a subsequent study of lens efficiency we need to know this basic power consumption. At any given length z of the tube we obtain the power absorbed by the gas as

$$\begin{aligned} P(z) &= \int_0^a [T(r,z) - T_0]\rho c_p v(r) 2\pi r \, dr \\ &= 2\pi\rho c_p v_0 \theta_0 \int_0^a r \left[1 - \left(\frac{r}{a}\right)^2\right]\left[1 - \frac{\theta(r,z)}{\theta_0}\right] dr. \end{aligned} \tag{4}$$

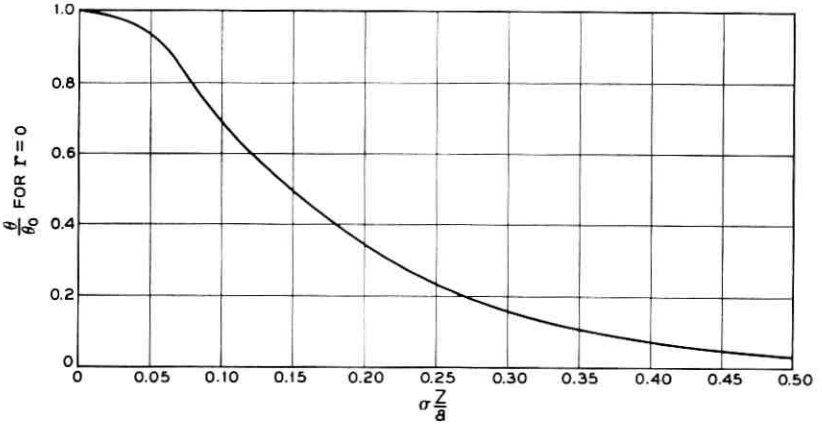


Fig. 3 — Normalized gas temperature on tube axis vs normalized distance along tube.

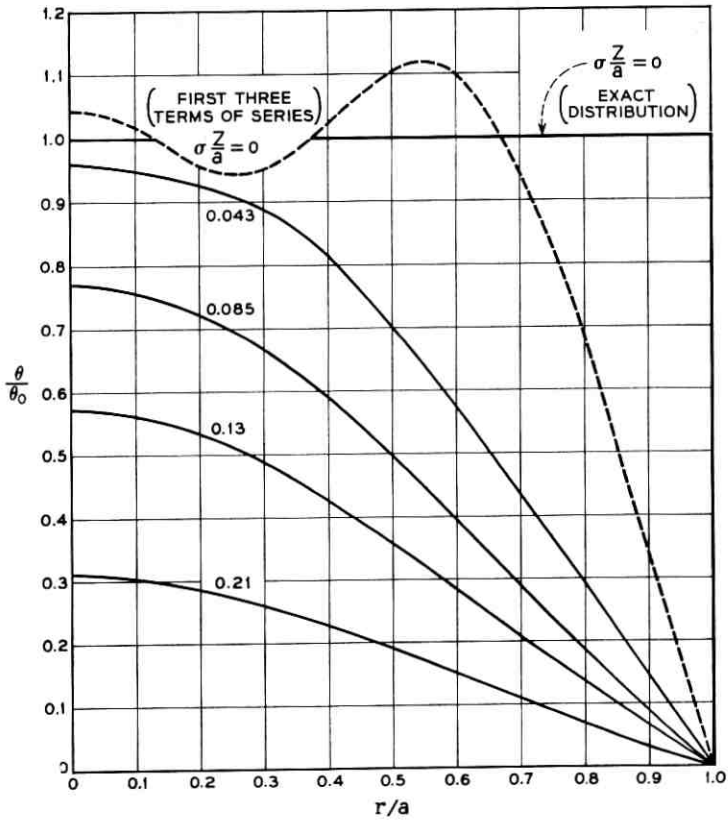


Fig. 4 — Normalized gas temperature vs radial position with longitudinal position as a parameter.

In order to be able to perform the integration easily, we restrict ourselves to values of

$$\sigma(z/a) > 0.05$$

which allows us to express θ/θ_0 by the first term of (3). In addition, we replace $R_0(x)$ by

$$R_0(x) \cong F(x) = 1 - 2.06\,x^2 + 1.06\,x^3. \tag{5}$$

This approximation deviates no more than 2.5 percent from the actual value of $R_0(x)$. The values of $F(x)$ can be compared to those of $R_0(x)$ in Table I. The integration can now be performed easily, and we obtain:

$$P = \frac{\pi}{2}\,a^2\rho c_p v_0\theta_0\left[1 - 0.820\exp\left(-7.316\sigma\frac{z}{a}\right)\right]. \tag{6}$$

TABLE II — GAS PARAMETERS VS TEMPERATURE

Gas	k (cal/cm sec deg)	ρ (gram/cm³)	c_p (cal/gram deg)	$a_{v\sigma} = k/\rho c_p$ (cm²/sec)	$n - 1$	$(n-1)/k$ (cm sec deg/cal)
CO_2	$3.93\ 10^{-5}$	$1.84\ 10^{-3}$	0.199	0.107	$4.20\ 10^{-4}$	10.7
NH_3	$5.90\ 10^{-5}$	$0.72\ 10^{-3}$	0.523	0.157	$3.48\ 10^{-4}$	5.9
CH_4	$7.80\ 10^{-5}$	$0.67\ 10^{-3}$	0.528	0.220	$4.13\ 10^{-4}$	5.3
Air	$6.28\ 10^{-5}$	$1.21\ 10^{-3}$	0.240	0.216	$2.73\ 10^{-4}$	4.35
H_2	$41.0\ 10^{-5}$	$0.084\ 10^{-3}$	3.39	1.44	$1.23\ 10^{-4}$	0.30
He	$35.0\ 10^{-5}$	$0.166\ 10^{-3}$	1.25	1.69	$0.34\ 10^{-4}$	0.097

Using ρ and c_p from Table II, P is in calories/sec. Fig. 5(a) shows the power consumption as a function of normalized lens length, $\sigma z/a$. An alternative form of (6) brings out the dependence of P on the flow velocity v_0 more clearly:

$$P = \frac{\pi}{2} kz\theta_0 \frac{v_0}{V}\left[1 - 0.820 \exp\left(-7.316 \frac{V}{v_0}\right)\right]. \tag{7}$$

The quantity

$$V = kz/a^2\rho c_p = \sigma(z/a)v_0 \tag{8}$$

has the dimension of velocity and is characteristic of the gas and the tube geometry. Fig. 5(b) shows the power consumption as a function of normalized gas velocity, v_0/V.

The ratio of V/v_0 can be related to the time $t_0 = z/v_0$ which it takes the gas particles on the axis to traverse the tube of length z with the velocity v_0 and to a time τ which is defined by

$$\frac{1}{\tau} = \frac{\dfrac{dT(0,t)}{dt}}{T_w - T(0,t)}. \tag{9}$$

τ is characteristic of the heat diffusion rate in a gas which rests in a tube whose wall temperature is T_w. At $t = 0$ the gas has the uniform temperature $T_0 < T_w$. Its temperature at a given radius r and time t is $T(r,t)$, so that $T(0,t)$ is the gas temperature at the tube axis at time t. $1/\tau$ is the time rate of temperature rise on the axis per degree of temperature difference between wall and axis.

It is shown in Appendix A that

$$V/v_0 = \sigma z/a = 0.173(t_0/\tau). \tag{10}$$

This shows that V/v_0 expresses the ratio of the time it takes the gas particles (on the tube axis) to flow through the tube of length z to the heat diffusion rate on the tube axis. Equation (10) may be used to replace

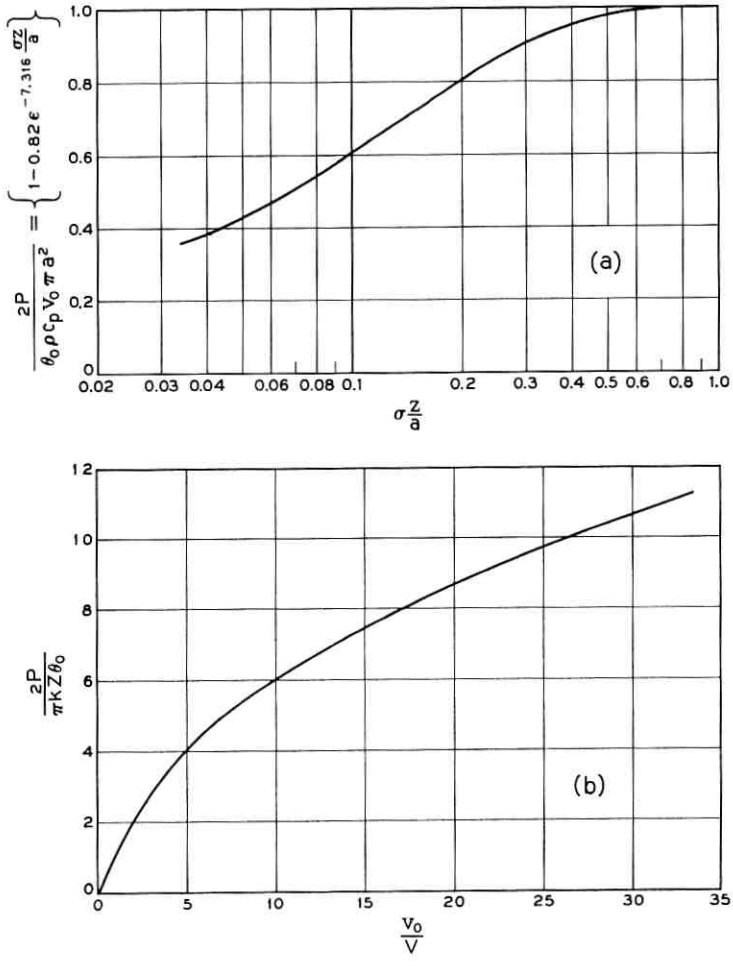


Fig. 5 — (a) Normalized power flow to the gas vs longitudinal position. (b) Normalized power flow to the gas vs normalized gas velocity.

V/v_0 in (7) and in the other places where V/v_0 or $\sigma z/a$ appears, resulting (for example) in

$$\exp\left(-7.316\,\frac{\sigma z}{a}\right) = \exp\left(-7.316\,\frac{V}{v_0}\right) = \exp\left(-1.265\,\frac{t_0}{\tau}\right).$$

The parameters lens length z and gas velocity v_0 are important variables for other reasons, and the first two forms of the exponential may be pre-

ferred. We may note, however,

$$\tau = 0.173(a^2\rho c_p/k)$$

and typical values are 0.161 second and 0.08 second for CO_2 and air, respectively, when $a = 0.125$ inch. It is surprising that this time constant is so short.

Another physical meaning one can give V is that it represents that velocity of gas flow along the pipe axis which assures that θ/θ_0 drops from its initial value of one at the beginning of the tube to

$$\theta/\theta_0 = 9.910^{-4} \approx 10^{-3}$$

on the axis at its end.

IV. FOCUSING ACTION

A lens focuses because the optical path length varies for rays traveling at different distances from its axis.

We describe the focusing action of our lens by the phase angle of a ray traveling parallel to the axis of the structure. The phase angle is given by

$$\Phi(r,z) = \beta_0 \int_0^z n(r,x)\, dx. \tag{11}$$

Here, $\beta_0 = 2\pi/\lambda_0$ is the free-space propagation constant of the light beam, and n is the index of refraction of the gas.

$$n(r,x) = 1 + (n_0 - 1)\frac{T_0}{T(r,x)} \tag{12}$$

$$\frac{T_0}{T} = \frac{1}{1 + \frac{\theta_0}{T_0}\left(1 - \frac{\theta}{\theta_0}\right)} \approx 1 - \frac{\theta_0}{T_0}\left(1 - \frac{\theta}{\theta_0}\right). \tag{13}$$

The last step is an approximation for $\theta_0/T_0 \ll 1$. The temperature in (12) has to be expressed in degrees Kelvin.

We decompose $\Phi(r,z)$ into two parts:

$$\Phi(r,z) = \varphi + \psi(r,z). \tag{14}$$

The first part

$$\varphi = \beta_0\left[1 + (n_0 - 1)\left(1 - \frac{\theta_0}{T_0}\right)\right]z \tag{15}$$

is independent of the position r of the ray in the gas lens, while the sec-

ond part

$$\begin{aligned}\psi &= \beta_0(n_0 - 1)\frac{\theta_0}{T_0}\int_0^z \frac{\theta(r,x)}{\theta_0}\,dx \\ &= \beta_0 z(n_0 - 1)\frac{\theta_0}{T_0}\cdot\frac{v_0}{V}\left\{0.202R_0\left(\frac{r}{a}\right)\left[1 - \exp\left(-7.316\,\frac{V}{v_0}\right)\right]\right. \\ &\quad - 0.0183R_1\left(\frac{r}{a}\right)\left[1 - \exp\left(-44.3\,\frac{V}{v_0}\right)\right] \\ &\quad \left. + 0.00363R_2\left(\frac{r}{a}\right)\left[1 - \exp\left(-106\,\frac{V}{v_0}\right)\right] + \cdots\right\}\end{aligned} \tag{16}$$

accounts for the different amounts of phase shift in different parts of the lens.

The difference between the phase angle of a ray traveling along the lens axis and the phase angle of a ray traveling at a distance r from the axis is

$$\begin{aligned}\Delta\Phi &= \beta_0 z(n_0 - 1)\frac{\theta_0}{T_0}\frac{v_0}{V}\left\{0.202(1 - R_0)\left[1 - \exp\left(-7.316\,\frac{V}{v_0}\right)\right]\right. \\ &\quad - 0.0183(1 - R_1)\left[1 - \exp\left(-44.3\,\frac{V}{v_0}\right)\right] \\ &\quad \left. + 0.00363(1 - R_2)\left[1 - \exp\left(-106\,\frac{V}{v_0}\right)\right] + \cdots\right\}.\end{aligned} \tag{17}$$

This form of $\Delta\Phi$ shows clearly its dependence on flow velocity for a fixed tube length z. To study the dependence of $\Delta\Phi$ for fixed flow rate and varying length, the following form is preferable.

$$\begin{aligned}\Delta\Phi &= \beta_0(a/\sigma)\,(n_0 - 1)\,(\theta_0/T_0)\,\{0.202(1 - R_0) \\ &\cdot(1 - \exp(-7.316\sigma z/a)) - 0.0183(1 - R_1)\,(1 - \exp(-44.3\sigma z/a)) \\ &\quad + 0.00363(1 - R_2)\,(1 - \exp(-106\sigma z/a)) + \cdots\}.\end{aligned} \tag{18}$$

Fig. 6 shows a plot of $\Delta\Phi$ versus length of lens for $r/a = 0.4$. That means that we compare the phase difference between a ray on the axis and another at distance $r = 0.4a$ from the axis of the lens.

Fig. 7 shows the phase difference at a fixed length z as a function of gas velocity. In this case, $\Delta\Phi$ goes through a maximum which for $r/a = 0.4$ occurs at $v_0/V = 6.9$. The position of this maximum depends somewhat on the radius r of the ray used for phase comparison with the axial ray. Appendix B gives the theory and Table III gives the values of

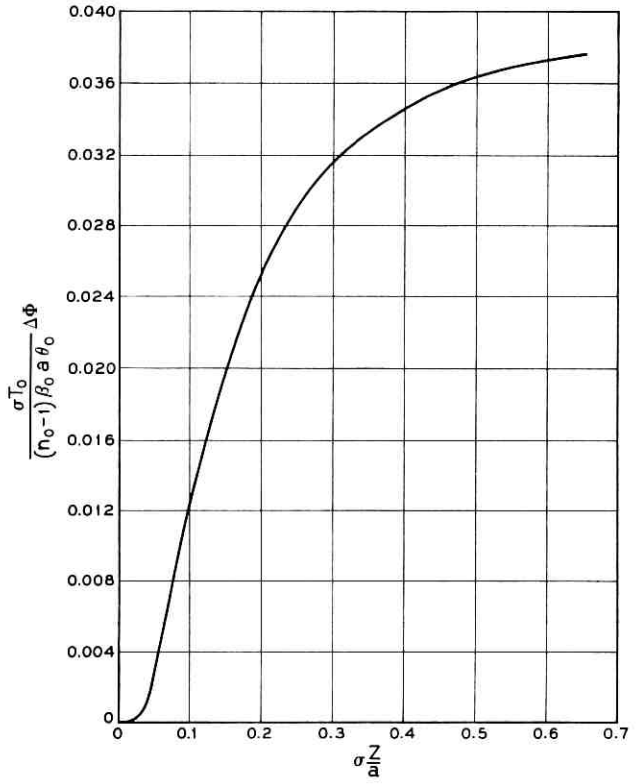


Fig. 6 — Phase difference between ray at $r = 0$ and ray at $r = 0.4a$ vs longitudinal position.

the position of the maximum v_0/V for different values of r/a. Table III shows that the position of the maximum does not change much with the radial position of the reference ray.

We can explain physically this maximum in $\Delta\Phi$ versus v_0 as follows. At very low gas velocity the majority of the gas in the tube is at the same temperature — the temperature of the walls, T_w . It is heated up in a time τ [see (9)] after entering. With little temperature difference between the gas at $r = 0$ and at $r > 0$ there is little $\Delta\Phi$. As v_0 increases, the gas on the axis remains at or near T_0 , but that nearer the walls is heated because it flows more slowly [see (1)] and larger $\Delta\Phi$ develops. Beyond some velocity, further increases in velocity cause the gas at $r = 0.4a$ (for example) to leave the tube at lower and lower temperatures — i.e., less temperature difference will exist between $r = 0$ and $r = 0.4a$ because the

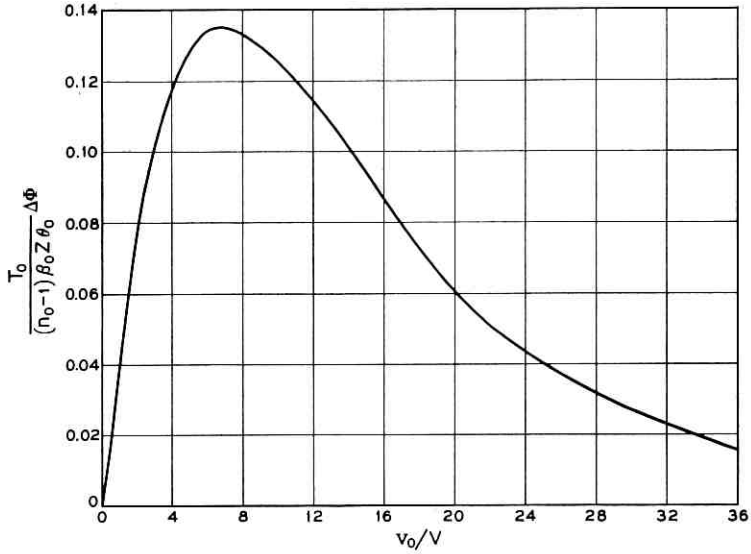


Fig. 7 — Phase difference between ray at $r = 0$ and ray at $r = 0.4a$ vs gas velocity.

transit time of the gas through the tube becomes less than the thermal diffusion time constant τ. Thus, at high velocities the $\Delta\Phi$ decreases. Note that this explanation (and the theory) depends upon laminar flow of the gas. If there is radial mixing of the gas, less $\Delta\Phi$ would be expected than predicted above, and the maximum in $\Delta\Phi$ versus v_0 might not occur.

The second derivative $d^2\Phi/d(r/a)^2$ is a good measure of the effectiveness of the lens for light rays close to its axis. For a glass lens

$$\Phi = n\beta_0 d(r)$$

where $d(r)$ is the thickness of the lens as a function of the distance from its axis. The radius of curvature R of the glass lens is given by

$$\frac{1}{R} = \frac{1}{n\beta_0}\frac{d^2\Phi}{dr^2}.$$

In order to be able to take the second derivative of Φ we have to express the functions R_0, R_1, and R_2 by power series with respect to r/a. For

TABLE III — MAXIMUM v_0/V FOR VALUES OF r/a

r/a	0.2	0.4	0.6
v_0/V	6.73	6.9	8.26

the second derivative on the axis at $r = 0$, it is sufficient to know the coefficient of $(r/a)^2$ in the expansion. Jakob[3] gives a series expansion of the R-functions.

$$R_n = 1 - \tfrac{1}{4}\beta_n^2(r/a)^2 \pm \cdots \tag{19}$$

with $\beta_0 = 2.705$, $\beta_1 = 6.66$, and $\beta_2 = 10.3$.

We get these values

$$\left(\frac{d^2\Phi}{d\left(\frac{r}{a}\right)^2}\right)_{r=0} = \beta_0 z(n_0 - 1)\frac{\theta_0}{T_0}\frac{v_0}{V}\left\{0.738\left[1 - \exp\left(-7.316\,\frac{V}{v_0}\right)\right]\right.$$
$$- 0.405\left[1 - \exp\left(-44.3\,\frac{V}{v_0}\right)\right] \tag{20}$$
$$\left. + 0.192\left[1 - \exp\left(-106\,\frac{V}{v_0}\right)\right] + \cdots\right\}.$$

The maximum of this curve as a function of v_0/V appears at $v_0/V = 6.75$.

Fig. 8(a) is a plot of a normalized value of $d^2\Phi/d(r/a)^2$ as a function of v_0/V, while Fig. 8(b) shows it (with a different normalization) as a function of $\sigma(z/a)$.

The r/a dependence of $\Delta\Phi$ is shown in Fig. 9.

In order to show what values the phase difference might actually assume, and also to compare different gases, we have plotted $\Delta\Phi$ in Fig. 10 for several gases and the following geometry and flow rate:

$2a = 0.25$ inch
$v_0 = 212$ cm/sec, corresponding to 2 liters/minute or 4.77 miles/hr.
$\beta_0 = 1.07 \times 10^5$ cm^{-1}, corresponding to $\lambda_0 = 5890$Å
$T_w = 343$°K
$T_0 = 293$°K

and with the values of $n_0 - 1$ and σ as listed in Table II. These curves assume a tube length z so long that no further $\Delta\Phi$ would be realized with a longer z (i.e., outgoing gas at uniform temperature).

It is interesting to compare the r/a dependence of $\Delta\Phi$ to the simple function $c(r/a)^2$. For this purpose we use (17), which is written so that $\Delta\Phi = 0$ at $r = 0$. Fig. 11 gives the normalized value of $\Delta\Phi$ as a function of r/a for several values of the gas velocity v_0/V. For comparison the function $c(r/a)^2$ is shown by dotted lines. The constant c is adjusted so that both curves coincide at $r/a = 0.4$. The actual curves of $\Delta\Phi$ are surprisingly close to the simple square law dependence in all cases. If the

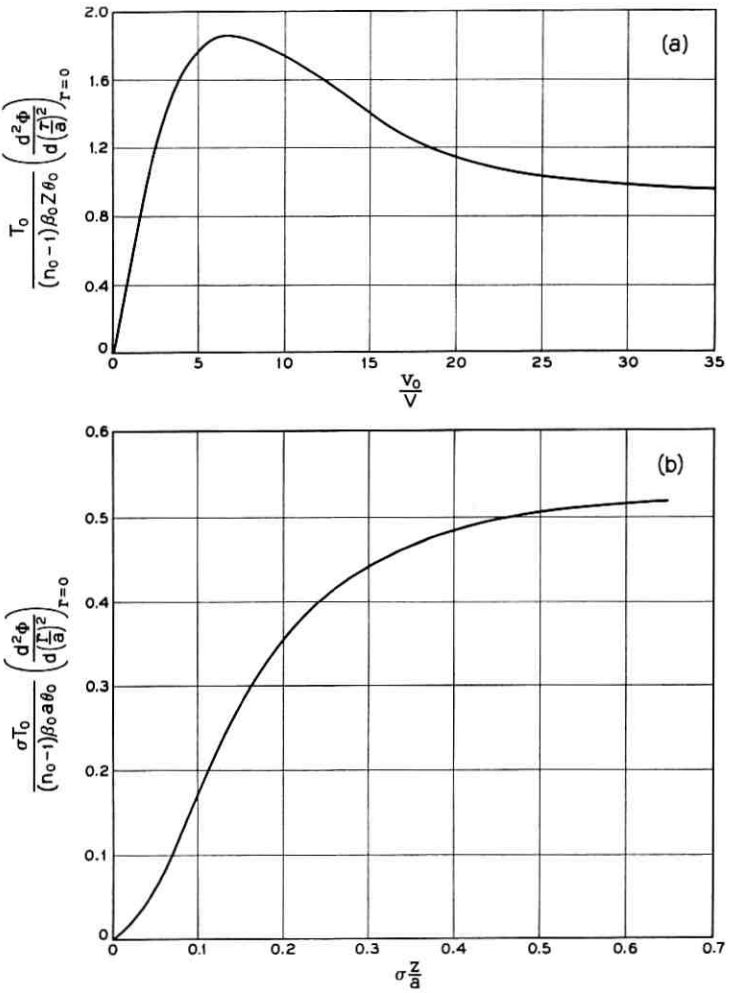


Fig. 8 — (a) Normalized $d^2\Phi/d(r/a)^2$ vs gas velocity; (b) normalized $d^2\Phi/d(r/a)^2$ vs longitudinal position.

gas lens could be treated as a thin lens, it would act very similar to a glass lens with spherically curved surfaces.

However, the gas lens is not thin and the question presents itself: how do different sections of the lens contribute to the over-all focusing effect? Fig. 12 shows the phase difference $\Delta\Phi$ between a ray on the axis at $r = 0$ and a ray at r for a fixed value $v_0/V = 6.9$ for different sections

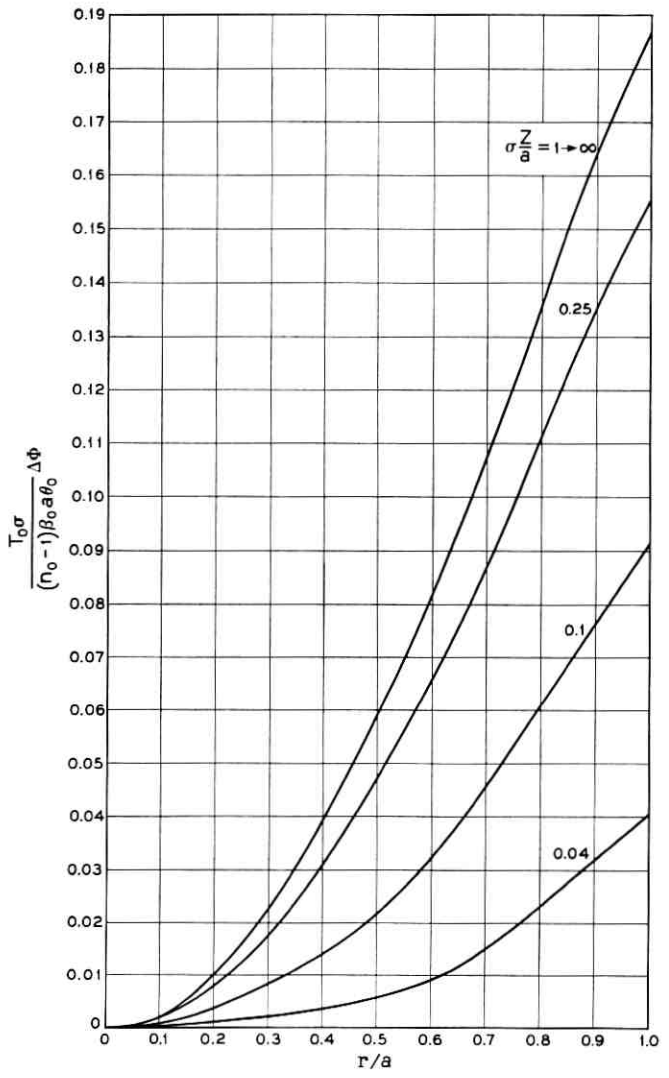


Fig. 9 — Normalized $\Delta\Phi$ vs r/a with longitudinal position as a parameter.

of the lens. The curve showing $\Delta\Phi$ for the section $0 \rightarrow z$ has already been shown in Fig. 11. The other curves show $\Delta\Phi$ in the first $\frac{1}{3}$ of the lens [curve $0 \rightarrow (\frac{1}{3})z$], the second $\frac{1}{3}$[curve $(\frac{1}{3})z \rightarrow (\frac{2}{3})z$], and the last $\frac{1}{3}$[curve $(\frac{2}{3})z \rightarrow z$]. The contributions are surprisingly different at different radii and do not resemble simple $(r/a)^2$ dependences. However, they all add

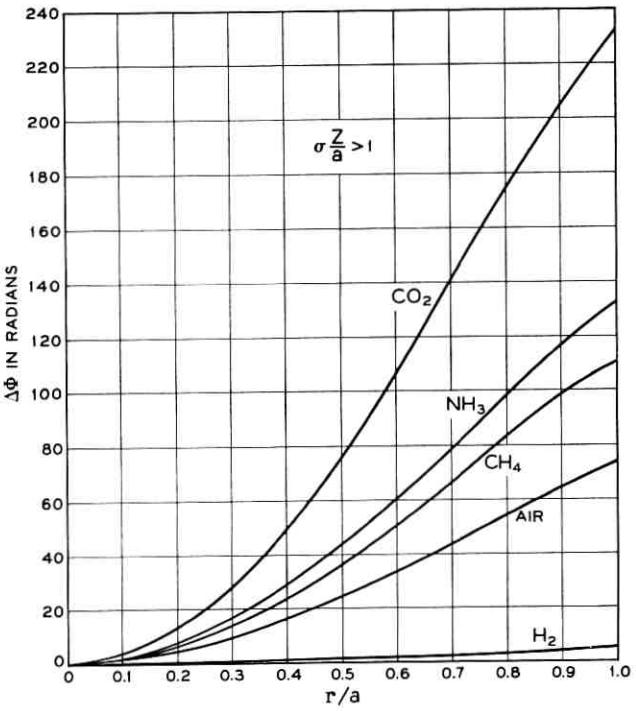


Fig. 10 — $\Delta\Phi$ vs r/a when $a = 0.125$ inch, $v_0 = 212$ cm/sec (corresponding to 2 liters/min), $\lambda_0 = 5890$ Å, $T_W = 343°K$, $T_0 = 293°K$ and $\sigma Z/a > 1$.

up to the $0 \rightarrow z$ curve which does resemble the $(r/a)^2$ dependence very closely, as was shown in Fig. 11.

V. FIGURE OF MERIT

The focusing action of the gas lens becomes independent of the length of the lens if $\sigma(z/a) > 1$, as Figs. 6 and 8(b) show. We also know that, for a fixed length of the lens, there is an optimum flow velocity, as shown by Figs. 7 and 8.

For practical applications one would like not only to obtain an effective lens but also to do so with a minimum expenditure of power. It is, therefore, interesting to study the lens action, that is, $d^2\Phi/d(r/a)^2$, per unit of applied power. We may introduce the ratio

$$M = \frac{1}{P}\left(\frac{d^2\Phi}{d(r/a)^2}\right)_{r=0} \tag{21}$$

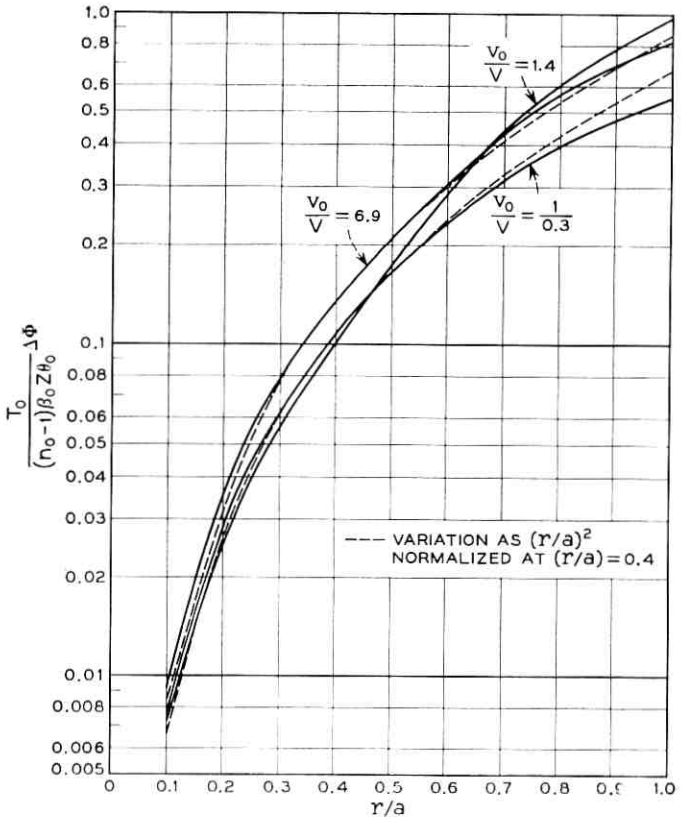


Fig. 11 — Normalized $\Delta\Phi$ vs r/a with gas velocity as a parameter; dotted curve represents variation as $(r/a)^2$, normalized to ordinate at $r/a = 0.4$.

as the figure of merit of the lens. From (6) and (20) we obtain

$$M = \frac{2\beta_0}{\pi T_0} \cdot \frac{n_0 - 1}{k} \left\{ \frac{F}{1 - 0.820 \exp\left(-7.316 \frac{V}{v_0}\right)} \right\} \tag{22a}$$

with

$$F = 0.738 \left[1 - \exp\left(-7.316 \frac{V}{v_0}\right)\right] - 0.405 \cdot \left[1 - \exp\left(-44.3 \frac{V}{v_0}\right)\right] + 0.192 \left[1 - \exp\left(-106 \frac{V}{v_0}\right)\right]. \tag{22b}$$

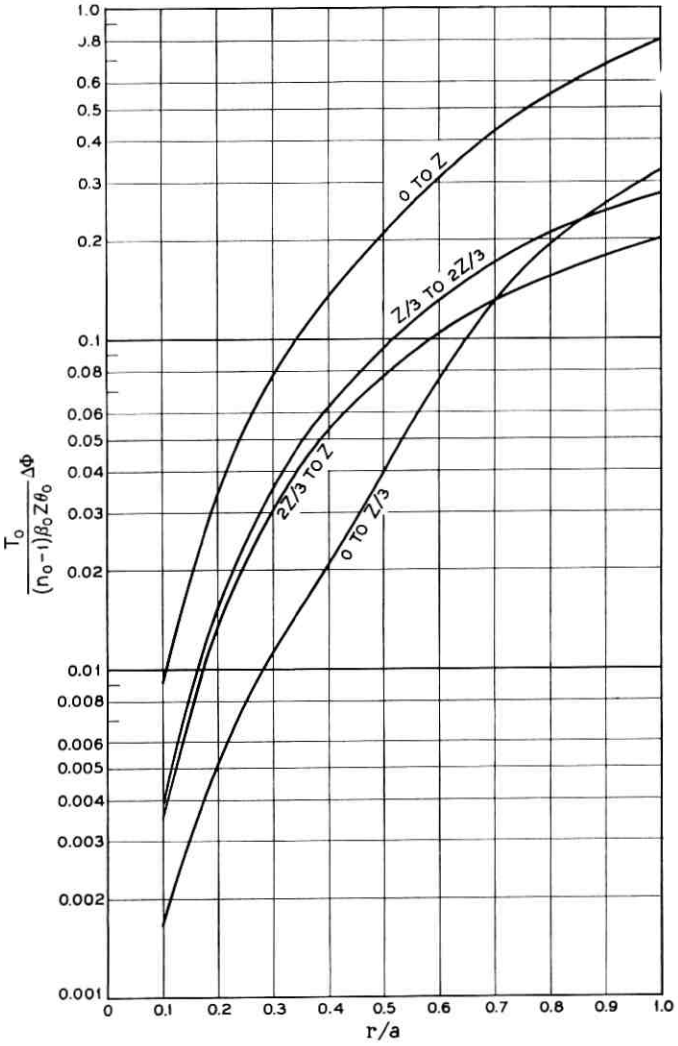


Fig. 12 — $\Delta\Phi$ contributed by first third, second third, and output third of gas lens for $v_0/V = 6.9$.

For a given value of v_0/V, the figure of merit is proportional to $(n_0 - 1)/k$. It is advantageous to make this number as large as possible.

The figure of merit M, given in (22b), is plotted in Fig. 13.

The gases in Table II are arranged in decreasing order of $(n_0 - 1)/k$. Of all the gases listed in that table, carbon dioxide is best suited for a gas

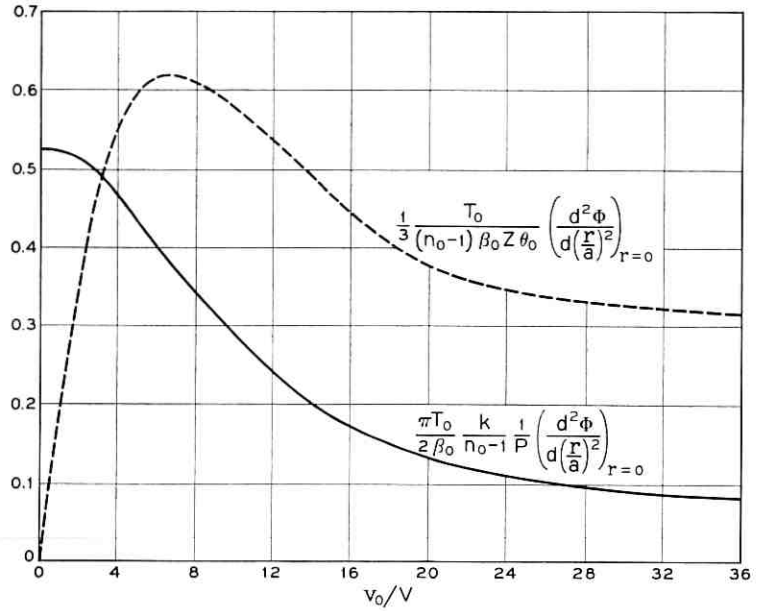


Fig. 13 — Figure of merit given by (22a) and focusing power (as in Fig. 8) vs gas velocity.

lens. This does not mean, however, that gases with larger values of $(n_0 - 1)/k$ cannot be found.

VI. FOCAL LENGTH

We have seen that $\Delta\Phi$ varies nearly as $(r/a)^2$. In the region where the lens is weak, we may treat it as a thin lens and obtain directly a simple expression for focal length.

For any thin lens it may be shown that the focal length f is given by

$$f = \tfrac{1}{2}\beta_0(r^2/\Delta\Phi) \tag{23}$$

where

$\Delta\Phi$ is the phase shift added on axis as compared to that for a ray at radius r

β_0 = phase constant of the region surrounding the lens.

We obtain $\Delta\Phi$ from (17), which is given by the following for the gas velocity set to maximize $\Delta\Phi$ at $(r/a) = 0.4$ — i.e., at $(v_0/V) = 6.9$:

$$\Delta\Phi = 0.839(r/a)^2(\theta_0/T_0)\beta_0 z(n_0 - 1). \tag{24}$$

Putting (24) into (23) we obtain the following expression for the focal length of a weak gas lens:

$$f = 0.596 \frac{\lambda^2}{z} \frac{T_0}{\theta_0(n_0 - 1)}. \tag{25}$$

If $a = 0.125$ inch, $z = 5$ inches, $T_0 = 293°K$, and $\theta_0 = 20°C$, we find $f \cong 5$ feet using CO_2 as the gas and $f \cong 8$ feet using air as the gas; the power transferred to the gas with CO_2 would be 0.0775 cal/sec = 0.325 watt.

When the gas lens is not weak, one should take into account that the refractive index varies both with radial position and with longitudinal position. Work is under way to analyze this very difficult situation. A simpler approach, and one which should give a first-order answer for gas lenses operated near the velocity producing maximum $\Delta\Phi$ [see (17)], is to assume a medium within the lens

$$n(r,x) = n_a(1 - \tfrac{1}{2}a_2r^2) \tag{26}$$

where

x = distance (within lens)from start of lens
r = radius
n_a = index of refraction on the axis.

For the gas velocity $(v_0/V) = 6.9$ it may be shown that

$$a_2 = \frac{1.68}{n_a a^2} \frac{\theta_0}{T_0} (n_0 - 1). \tag{27}$$

In other unpublished work the authors have shown that the radial position of a ray (or of the axis of a Gaussian beam mode) is

$$r = r_i \cos \sqrt{a_2}x + \frac{r_i'}{\sqrt{a_2}} \sin \sqrt{a_2}x \tag{28}$$

where

r_i = displacement of ray at lens input
r_i' = slope of ray at lens input.

We can use this general result to specify the focal length of a strong (or weak) gas lens with reference to Fig. 14. All input rays with zero slope will converge to a point on the axis a distance d beyond the output face of the lens (Fig. 14), where

$$d = \frac{1}{\sqrt{a_2}} \cot (\sqrt{a_2}t), \tag{29}$$

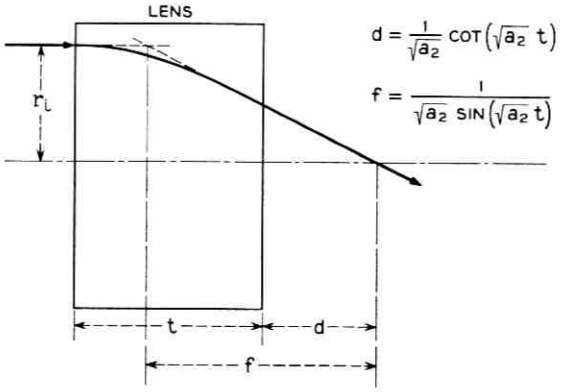


Fig. 14 — Diagram defining focal length and position of the equivalent thin lens for a thick gas lens.

t is the length of the lens, and n_a has been taken as unity. An equivalent thin lens may be located a distance f back from the focal point, where

$$f = \frac{1}{\sqrt{a_2} \sin (\sqrt{a_2} t)} . \tag{30}$$

This expression for focal length is valid up to $(\sqrt{a_2} t) = \pi/2$, at which point $d = 0$. For $(\sqrt{a_2} t) > \pi/2$ the rays cross within the lens, per (28). For $(\sqrt{a_2}\, t) \ll 1$ it may be shown that (30) passes into (25) and the location of the equivalent thin lens is in the center of the distributed lens.

E. A. J. Marcatili[4] has solved Maxwell's equations for a medium characterized by (26) and has found the normal modes for a sequence of lenses composed of segments of such a medium. This work relates closely to a sequence of gas lenses described in this paper. Experiments with tubular thermal gas lenses are reported by A. C. Beck.[5]

VII. CONCLUSION

When a cool gas is blown through a warmer tube, the gas at the axis has a lower temperature than that near the walls. Thus the density and refractive index is larger at the axis and a converging lens is formed. If the tube were at a lower temperature than the input gas, a diverging lens would be formed.

There is an optimum gas velocity for maximizing the focusing power of such lenses, and expressions are given for this velocity. It turns out that the optimum transit time for gas through the tube is approximately the time constant for temperature changes in a gas at rest in the tube, which

for typical gases (air and carbon dioxide) is about 0.1 second in a $\frac{1}{4}$-inch ID tube. Although not discussed in this paper, it is found that a $\frac{1}{4}$-inch tube 6 inches long yields (at the optimum velocity for focusing power) a Reynolds number well below that at which turbulence is expected.

Expressions are given for focal length and a figure of merit expressed as focusing power per watt of power transferred to the moving gas.

The best gas is one with a maximum $(n - 1)/k$, where n is the refractive index and k is the heat conductivity.

APPENDIX A

Derivation of (10)

The relation (10)

$$V/v_0 = 0.173(t_0/\tau) \tag{31}$$

can be derived as follows.

The time development of a cool gas resting in a tube of wall temperature T_w can be described as follows

$$T = T_w - 2(T_w - T_0) \sum_{n=1}^{\infty} e^{-\lambda_n t} \frac{J_0\left(w_n \frac{r}{a}\right)}{w_n J_1(w_n)} \tag{32}$$

with

$$\lambda_n = (k/a^2\rho c_p)w_n^2 \quad \text{and} \quad J_0(w_n) = 0.$$

At $t = 0$, (32) becomes $T(r,0) = T_0$, which is constant throughout the tube's cross section.

As time progresses the exponents $\lambda_n t$ become large, so that very soon the first term of the series is the only contributing factor. Neglecting all the terms except the first, we get

$$\frac{1}{\tau} = \frac{\frac{dT(0,t)}{dt}}{T_w - T(0,t)} \cong \lambda_1 = \frac{k}{a^2\rho c_p} w_1^2 = 5.79 \frac{V}{z} \tag{33}$$

and substituting $z = v_0 t_0$ we get (31). Since we neglected all but the first term in the series, (33) represents the asymptotic value which $1/\tau$ assumes after the initial transients have died down.

To obtain a feeling for the accuracy of the approximation involved in deriving (33), we write down the ratio of the second to the first term in

the sum (32) for $r = 0$:

$$\frac{w_1 J_1(w_1)}{w_2 J_1(w_2)} \exp[-(\lambda_2 - \lambda_1)t] = 0.666 \exp\left(-4.26 \frac{t}{\tau}\right).$$

This ratio is 10^{-2} for $t/\tau = 0.986$ and is 10^{-1} for $t/\tau = 0.45$. The approximation is excellent for times $t \geqq \tau$ and is quite good for $t > 0.5\tau$.

For the special example used in Fig. 10 we get for

$$\begin{aligned} &CO_2 : \tau = 0.161 \text{ sec} \\ &\text{air:} \quad \tau = 0.08 \text{ sec}. \end{aligned}$$

APPENDIX B

The Maximum of (17)

We seek an expression for the value of v_0/V which brings $\Delta\Phi$, equation (17), to a maximum:

$$\begin{aligned} \frac{d(\Delta\Phi)}{d(v_0/V)} = \beta_0 z(n_0 - 1) \frac{\theta_0}{T_0} \\ \cdot\Bigg[0.202(1 - R_0)\left\{1 - \exp\left(-7.316 \frac{V}{v_0}\right)\left(1 + 7.316 \frac{V}{v_0}\right)\right\} \\ - 0.0183(1 - R_1)\left\{1 - \exp\left(-44.3 \frac{V}{v_0}\right)\left(1 + 44.3 \frac{V}{v_0}\right)\right\} \\ + 0.00363(1 - R_2)\left\{1 - \exp\left(-106 \frac{V}{v_0}\right)\left(1 + 106 \frac{V}{v_0}\right)\right\}\Bigg]. \end{aligned} \tag{34}$$

We set (34) equal to zero, and noting that the second and third exponentials are small, we neglect them (to be justified by the solutions thus obtained) yielding

$$\begin{aligned} \exp\left(-7.316 \frac{V}{v_0}\right)\left(1 + 7.316 \frac{V}{v_0}\right) \\ = 1 - 0.0906 \frac{(1 - R_1)}{(1 - R_0)} + 0.018 \frac{(1 - R_2)}{(1 - R_0)}. \end{aligned} \tag{35}$$

Equation (35) gives the approximate value of v_0/V at the maximum of $\Delta\Phi$ for any chosen radius, r/a. Additional terms in (34) can be taken if more accuracy is desired, which was done in computing the v_0/V for $r/a = 0.2$, as given in the body of the paper.

REFERENCES

1. Goubau, G., and Schwering, F., On the Guided Propagation of Electromagnetic Wave Beams, I.R.E. Trans. **AP-9,** May, 1961, pp. 248–256.
2. Berreman, D. W., B.S.T.J., this issue, pp. 1469 and 1476.
3. Jakob, M., *Heat Transfer*, Vol. 1, John Wiley, New York, 1949, pp. 451–464.
4. Marcatili, E. A. J., Wave Optics in a Periodic Sequence of Arbitrarily Thick Lens-Like Focusers, to be published.
5. Beck, A. C., Thermal Gas Lens Measurements, B.S.T.J., this issue, p. 1818.

Hollow Metallic and Dielectric Waveguides for Long Distance Optical Transmission and Lasers

By E. A. J. MARCATILI and R. A. SCHMELTZER

(Manuscript received June 12, 1964)

The field configurations and propagation constants of the normal modes are determined for a hollow circular waveguide made of dielectric material or metal for application as an optical waveguide. The increase of attenuation due to curvature of the axis is also determined.

The attenuation of each mode is found to be proportional to the square of the free-space wavelength λ and inversely proportional to the cube of the cylinder radius a. For a hollow dielectric waveguide made of glass with $\nu = 1.50$, $\lambda = 1\mu$, and $a = 1$ mm, an attenuation of 1.85 db/km is predicted for the minimum-loss mode, EH_{11}. This loss is doubled for a radius of curvature of the guide axis $R \approx 10$ km. Hence, dielectric materials do not seem suitable for use in hollow circular waveguides for long distance optical transmission because of the high loss introduced by even mild curvature of the guide axis. Nevertheless, dielectric materials are shown to be very attractive as guiding media for gaseous amplifiers and oscillators, not only because of the low attenuation but also because the gain per unit length of a dielectric tube containing He-Ne "masing" mixture at the right pressure can be considerably enhanced by reducing the tube diameter. In this application, a small guide radius is desirable, thereby making the curvature of the guide axis not critical. For $\lambda = 0.6328\mu$ and optimum radius $a = 0.058$ mm, a maximum theoretical gain of 7.6 db/m is predicted.

It is shown that the hollow metallic *circular waveguide is far less sensitive to curvature of the guide axis. This is due to the comparatively large complex dielectric constant exhibited by metals at optical frequencies. For a wavelength $\lambda = 1\mu$ and a radius $a = 0.25$ mm, the attenuation for the minimum loss TE_{01} mode in an aluminum waveguide is only 1.8 db/km. This loss is doubled for a radius of curvature as short as $R \approx 48$ meters. For $\lambda = 3\mu$ and $a = 0.6$ mm, the attenuation of the TE_{01} mode is also 1.8 db/km. The radius of curvature which doubles this loss is approximately 75 meters. The*

straight guide loss for the EH_{11} mode for $\lambda = 1\mu$ and $a = 0.25$ mm is 57 db/km and is increased to 320 db/km for $\lambda = 3\mu$ and $a = 0.6$ mm.

In view of the low-loss characteristic of the TE_{01} mode in metallic waveguides, the high-loss discrimination of noncircular electric modes, and the relative insensitivity to axis curvature, the hollow metallic *circular waveguide appears to be very attractive as a transmission medium for long distance optical communication.*

I. INTRODUCTION

During recent years the potentially large frequency range made available to communications by the development of the optical maser has stimulated much interest in efficient methods for long distance transmission of light. The most promising contenders for long distance optical transmission media consist of sequences of lenses or mirrors, highly reflective hollow metallic pipes, and dielectric waveguides.[1-10]

In this paper we present an analysis of the field configurations and propagation constants of the normal modes in a hollow circular waveguide which, because of its simplicity and low loss, may become an important competitor. The guiding structure considered here may consist of an ordinary metallic pipe of precision bore whose inner surface is highly reflective, or of a hollow dielectric pipe — i.e., one in which the metal is replaced with dielectric. Although the transmission characteristics of metallic waveguides are well known for microwave frequencies, this theory is invalidated for operation at optical wavelengths, because the metal no longer acts as a good conductor but rather as a dielectric having a large dielectric constant. In the subsequent analysis, therefore, both the dielectric and metallic guide are considered as special cases of a general hollow circular waveguide having an external medium made of arbitrary isotropic material whose optical properties are characterized by a finite complex refractive index. If the free-space wavelength is much smaller than the internal radius of the tube, the energy propagates not in the external medium but essentially within the tube, bouncing at grazing angles against the wall. Consequently, there is little energy loss due to refraction. The refracted field is partially reflected by the external surface of the tube and may, in general, interfere constructively or destructively with the field inside the tube, decreasing or increasing the attenuation. Because of the difficulty of controlling the interference paths, it seems more convenient to eliminate the effect completely by introducing sufficient loss in the dielectric or, in the case of a glass dielectric, by frosting the external surface. The field in the hole of the

tube is then unaffected by wall thickness. We shall therefore simplify the analysis of the hollow circular waveguide by assuming infinite wall thickness, as depicted in Fig. 1.

This structure will be shown to be attractive as a low-loss transmission medium for long distance optical communication as well as for optical gaseous amplifiers and oscillators. It is known, for example, that in a tube containing a He-Ne mixture such that the product of radius and pressure is roughly a constant, the gain per unit length is inversely proportional to the radius of the tube.[11] On the other hand, we find in this paper that the attenuation of the normal modes is inversely proportional to the cube of the radius. Hence there is an optimum tube radius for which the net gain per unit length is a maximum. Furthermore, because the guidance is continuous, there is no need for periodic focusing. Consequently, no restriction need be imposed on the length of the amplifying or oscillating tube.

We begin by analyzing an idealized guide having a straight axis and a cylindrical wall. The results are then extended to include the effects of mild curvature of the guide axis by finding a perturbation correction for field configurations and propagation constants of the idealized straight guide.

II. MODAL ANALYSIS OF THE GENERAL STRAIGHT CIRCULAR WAVEGUIDE

Consider a waveguide consisting of a circular cylinder of radius a and free-space dielectric constant ϵ_0 embedded in another medium of dielectric or metal having a complex dielectric constant ϵ. The magnetic permeability μ_0 is assumed to be that of free space for both media. We are interested in finding the field components of the normal modes of the waveguide and in determining the complex propagation constants of these modes.

The problem is substantially simplified if it is assumed that

$$ka = 2\pi a/\lambda \gg |\nu| u_{nm} \tag{1}$$

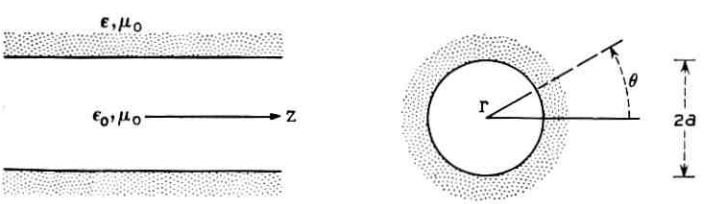


Fig. 1 — Hollow dielectric waveguide.

and

$$|\,(\gamma/k) - 1\,| \ll 1 \tag{1}$$

where $k = \omega\sqrt{\epsilon_0\mu_0} = 2\pi/\lambda$ is the free-space propagation constant; u_{nm} is the mth root of the equation $J_{n-1}(u_{nm}) = 0$, and n and m are integers that characterize the propagating mode; $\nu = \sqrt{\epsilon/\epsilon_0}$ is the complex refractive index of the external medium; and γ is the axial propagation constant of the mode under consideration. The first inequality states that the radius a is much larger than the free-space wavelength λ. In the case of metalization of the external medium, $|\,\nu\,|$ may be quite large but is finite at optical frequencies. The second inequality restricts our analysis to low-loss modes, which are those whose propagation constants γ are nearly equal to that of free space.

The field components of the natural modes of the most general circular cylindrical structure with arbitrary isotropic internal and external media have been determined by Stratton.[12] This structure supports three types of modes: first, transverse circular electric modes whose only field components are E_θ, H_r and H_z; second, transverse circular magnetic modes whose components are H_θ, E_r and E_z; and third, hybrid modes with all the electric and magnetic components present. The approximate field components of these modes are written below. They have been derived using the inequalities (1) and neglecting terms with powers of λ/a larger than one. The superscripts i and e refer to the internal and external media, respectively.

1. Circular electric modes TE_{0m} $(n = 0)$

$$\left.\begin{aligned}
E_{\theta 0m}{}^{i} &= J_1(k_i r) \\
H_{r0m}{}^{i} &= -\sqrt{\frac{\epsilon_0}{\mu_0}}\, J_1(k_i r) \\
H_{z0m}{}^{i} &= -i\sqrt{\frac{\epsilon_0}{\mu_0}}\,\frac{u_{0m}}{ka}\, J_0(k_i r)
\end{aligned}\right\} \exp i(\gamma z - \omega t)$$

$$\left.\begin{aligned}
E_{\theta 0m}{}^{e} &= -1 \\
H_{r0m}{}^{e} &= \sqrt{\frac{\epsilon_0}{\mu_0}} \\
H_{z0m}{}^{e} &= -i\sqrt{\nu^2 - 1}\,\sqrt{\frac{\epsilon_0}{\mu_0}}
\end{aligned}\right\} i\,\frac{u_{0m}}{k\sqrt{ar(\nu^2 - 1)}}\, J_0(u_{0m})\, \exp i[k_e(r - a) + \gamma z - \omega t] \tag{2}$$

2. Circular magnetic modes TM_{0m} $(n = 0)$

$$\left.\begin{aligned}
E_{r0m}{}^{i} &= J_1(k_i r) \\
E_{z0m}{}^{i} &= i\,\frac{u_{0m}}{ka}\,J_0(k_i r) \\
H_{\theta 0m}{}^{i} &= \sqrt{\frac{\epsilon_0}{\mu_0}}\,J_1(k_i r)
\end{aligned}\right\} \exp i(\gamma z - \omega t)$$

$$\left.\begin{aligned}
E_{r0m}{}^{e} &= -\frac{1}{\nu^2} \\
E_{z0m}{}^{e} &= \sqrt{\nu^2 - 1} \\
H_{\theta 0m}{}^{e} &= -\sqrt{\frac{\epsilon_0}{\mu_0}}
\end{aligned}\right\} i\,\frac{u_{0m} J_0(u_{0m})}{k\sqrt{ar(\nu^2 - 1)}} \exp i\,[k_e(r - a) + \gamma z - \omega t] \qquad (3)$$

3. Hybrid modes EH_{nm} $(n \neq 0)$

$$\left.\begin{aligned}
E_{\theta nm}{}^{i} &= \left[J_{n-1}(k_i r) + \frac{i u_{nm}{}^2}{2nka}\sqrt{\nu^2 - 1}\,J'_n(k_i r)\right] \cdot \cos n(\theta + \theta_0) \\
E_{rnm}{}^{i} &= \left[J_{n-1}(k_i r) + \frac{i u_{nm}}{2kr}\sqrt{\nu^2 - 1}\,J_n(k_i r)\right] \cdot \sin n(\theta + \theta_0) \\
E_{znm}{}^{i} &= -i\,\frac{u_{nm}}{ka}\,J_n(k_i r)\,\sin n(\theta + \theta_0)
\end{aligned}\right\} \exp i(\gamma z - \omega t)$$

$$H_{\theta nm}{}^{i} = \sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{rnm}{}^{i}$$

$$H_{rnm}{}^{i} = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{\theta nm}{}^{i}$$

$$H_{znm}{}^{i} = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{znm}{}^{i}\,\operatorname{ctn} n(\theta + \theta_0)$$

$$\left.\begin{aligned}
E_{\theta nm}{}^{e} &= \cos n(\theta + \theta_0) \\
E_{rnm}{}^{e} &= \sin n(\theta + \theta_0) \\
E_{znm}{}^{e} &= -\sqrt{\nu^2 - 1}\,\sin n(\theta + \theta_0)
\end{aligned}\right\} i\,\frac{u_{nm}}{k\sqrt{ar(\nu^2 - 1)}}\,J_n(u_{nm}) \cdot \exp i[k_e(r - a) + \gamma z - \omega t]$$

$$H_{\theta nm}{}^{e} = \nu^2\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{rnm}{}^{e}$$

$$H_{rnm}{}^{e} = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{\theta nm}{}^{e}$$

$$H_{znm}{}^{e} = -\sqrt{\frac{\epsilon_0}{\mu_0}}\,E_{znm}{}^{e}\,\operatorname{ctn} n(\theta + \theta_0) \qquad (4)$$

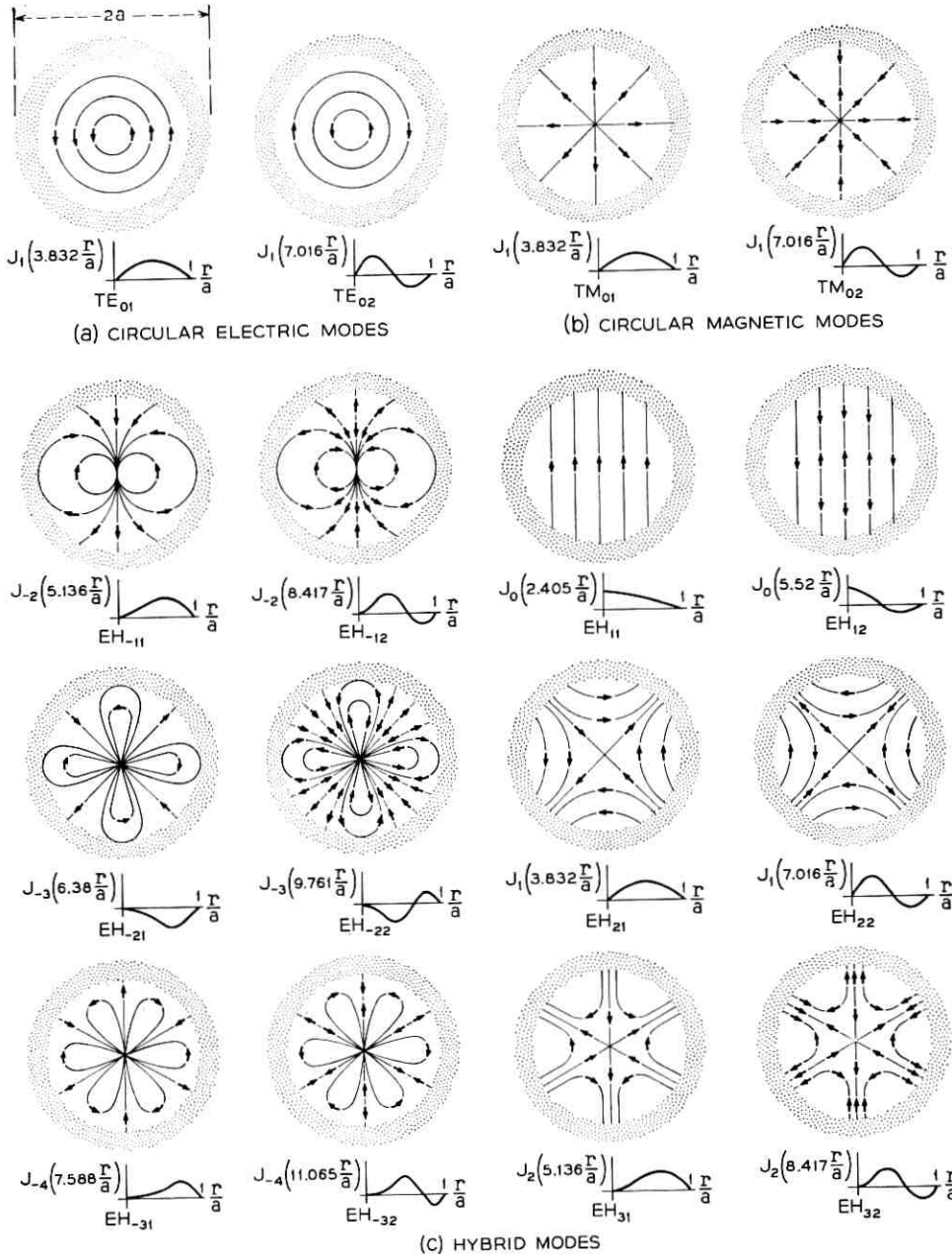


Fig. 2 — Electric field lines of modes in hollow dielectric waveguides: (a) circular electric modes, (b) circular magnetic modes, (c) hybrid modes.

where the complex propagation constant γ satisfies the relationships

$$\begin{aligned} k_i^2 &= k^2 - \gamma^2 \\ k_e^2 &= \nu^2 k^2 - \gamma^2 \end{aligned} \tag{5}$$

and u_{nm} is the mth root of the equation

$$J_{n-1}(u_{nm}) = 0. \tag{6}$$

As usual, $| n |$ is the number of periods of each field component in the θ direction, and m is both the order of the root of (6) and the number of maxima and minima of each component counted in the radial direction within the internal medium. The constant θ_0 appearing in (4) will become of interest later on when we study the waveguide with curved axis, because it will admit any orientation of the transverse electric field relative to the plane of curvature of the guide axis.

For $n = 0$, the modes are either transverse electric TE_{0m} (2), or transverse magnetic TM_{0m} (3). The lines of electric field of the TE_{0m} modes are transverse concentric circles centered on the z axis. The lines of magnetic field are in planes containing the z axis. Similarly, the lines of magnetic field of the TM_{0m} modes are transverse concentric circles centered on the z axis with the electric field contained in radial planes. The electric field lines of the modes TE_{01}, TE_{02}, TM_{01} and TM_{02} are shown in Figs. 2(a) and 2(b); each vector represents qualitatively the intensity and direction of the local field.

For $n \neq 0$, the modes are hybrid, EH_{nm} (4); therefore, the magnetic and electric field are three-dimensional with relatively small axial field components in the internal medium. Thus the hybrid modes are almost transverse.

Let us examine the projection of these three-dimensional field lines on planes perpendicular to the axis z of the waveguide. The differential equations for the projected lines of electric field in both media are

$$\begin{aligned} \frac{1}{r}\frac{dr}{d\theta} &= \frac{E_{rnm}{}^{i}}{E_{\theta nm}{}^{i}} \\ \frac{1}{r}\frac{dr}{d\theta} &= \frac{E_{rnm}{}^{e}}{E_{\theta nm}{}^{e}}. \end{aligned} \tag{7}$$

$E_{rnm}{}^{i}$ as well as $E_{\theta nm}{}^{i}$ contain two terms as given in (4). Both are necessary to satisfy the boundary conditions. If we neglect the second term, however, no substantial error is introduced except very close (a few wavelengths) to the boundary, where the second term dominates as

the first tends to zero. With this simplification, the differential equations (7) in both media become identical

$$(1/r)(dr/d\theta) = \tan n\theta.$$

Upon integrating, one obtains an equation for the locus of the projected electric field lines

$$(r/r_0)^n \cos n\theta = 1 \tag{8}$$

where r_0 is a constant of integration that individualizes the member of the family of lines. The electric field of an EH_{nm} mode is different from that of EH_{-nm} mode.

The projection of the magnetic field lines is determined in a similar way. These equations are

$$(r/r_0)^n \sin n\theta = 1 \tag{9}$$

for the internal medium and

$$(r/r_0)^{n\nu^2} \sin n\theta = 1$$

for the external medium.

The projections of the internal electric (8) and magnetic (9) field lines are identical for any given mode except for a rotation of $\pi/(2n)$ radians around the z axis. In Fig. 2(c) the lines of the electric field in the internal medium are depicted for the first few hybrid modes. Again the vectors represent qualitatively the field intensities and directions.

What happens at the boundary? Consider, for example, the projected electric lines of mode EH_{11}, as shown in Fig. 3(a). These field lines satisfy (8), an equation which is valid everywhere except near the boundary. The boundary conditions are violated in Fig. 3(a) because there is continuity not only of the tangential electric component but also of the normal component. The internal normal component must be ν^2 times larger than the external one. Consequently, the electric field line must be discontinuous. This result is shown qualitatively in Fig. 3(b).

A three-dimensional representation of the field lines is far more complicated than the two-dimensional one depicted in Fig. 2. As a typical example, the electric field lines of the EH_{22} mode are shown in Fig. 4 in a three-dimensional perspective.

The propagation constants of the TE_{0m}, TM_{0m} and EH_{nm} ($n \neq 0$) modes are determined below (21). It is found that the hybrid mode $EH_{-|n|,m}$ is degenerate (same propagation constant) with the $EH_{|n|+2,m}$; i.e., for every hybrid mode with negative azimuthal index there is a degenerate hybrid mode with positive aximuthal index. The

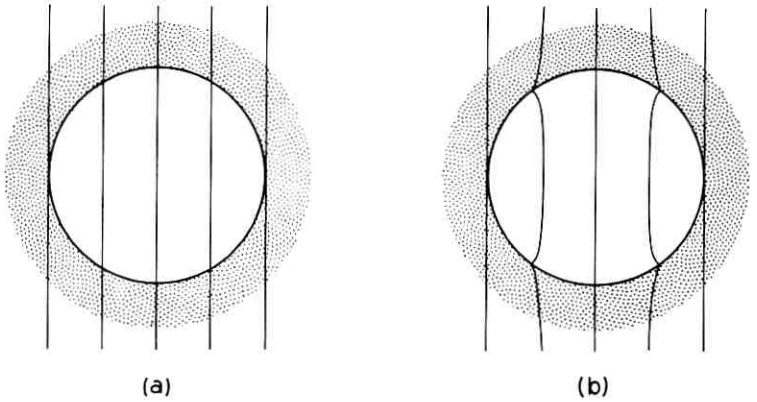


Fig. 3 — (a) Electric field lines of EH_{11} mode violating boundary conditions; (b) same EH_{11} mode with electric field lines qualitatively corrected.

transverse modes TE_{0m} and TM_{0m} and the hybrid modes EH_{1m} and EH_{2m} have no degenerate counterpart.

If the field components of the degenerate $EH_{-|n|,m}$ and $EH_{|n|+2,m}$ modes (4) are added, we obtain new composite modes whose electric and magnetic field lines project as straight lines on a plane perpendicular to the z axis. Some of those composite modes are shown in Fig. 5.

It should be noted that if the refractive index of the external medium,

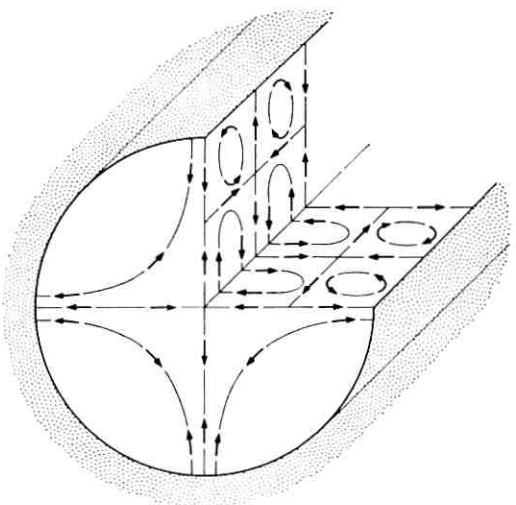


Fig. 4 — Cutaway view of electric field lines of EH_{22} mode. The axial period is grossly exaggerated.

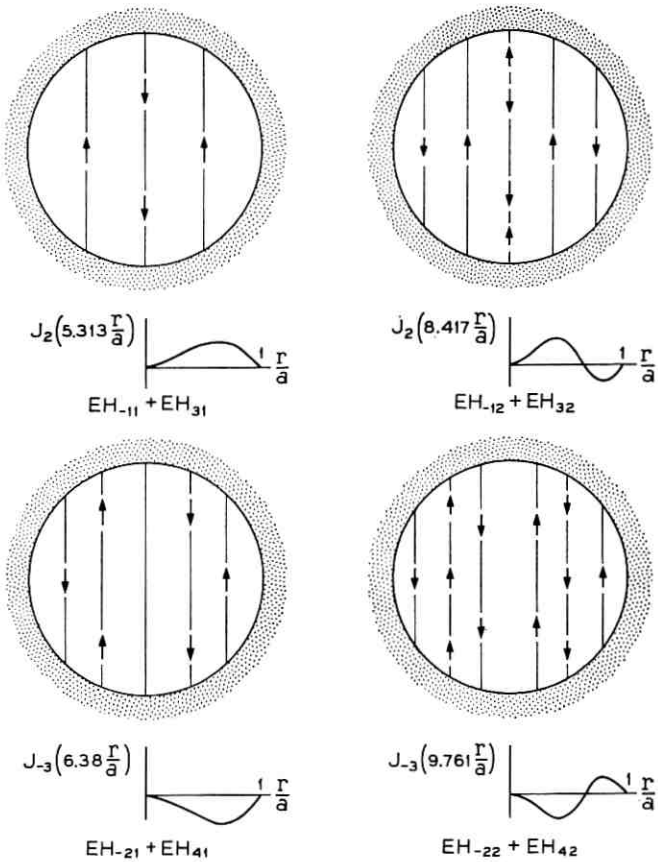


Fig. 5 — Electric field lines of composite modes $EH_{-|n|,m} + EH_{|n|+2,m}$.

ν, is very close to unity, then for each value of m, the TE_{0m}, TM_{0m} and EH_{2m} modes also become degenerate (17), (21) and the sum of the components of TE_{0m} (2) and EH_{2m} (4) yields a new composite mode, as shown in Fig. 6. This mode, together with those in Fig. 5 and the EH_{1m} of Fig. 2(c), form a complete set that closely resembles the set found for interferometers with plane circular mirrors or for sequences of circular irises.[1]

Let us now consider the field intensity distribution outside and inside the hollow dielectric waveguide. The external field (2), (3) and (4) has the radial dependence

$$\frac{\exp\,[ik_e(r - a)]}{\sqrt{r}}.$$

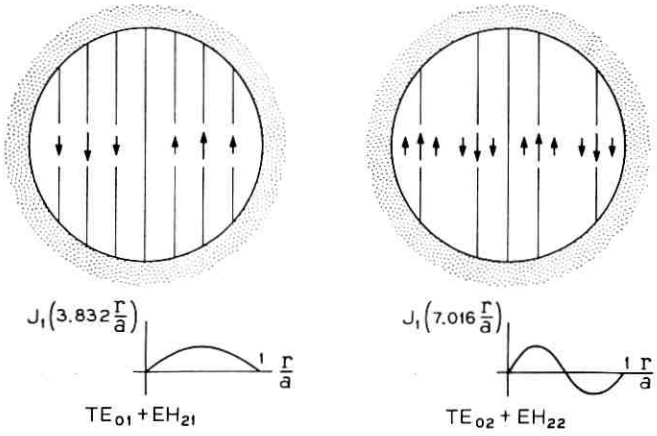


Fig. 6 — Electric field lines of composite modes $TE_{0m} + EH_{2m}$.

From (5) and (20) we obtain, neglecting terms of order $(\lambda/a)^2$ and higher, $k_c = k\sqrt{\nu^2 - 1}$. The radial dependence is then

$$\frac{\exp\left[ik\sqrt{\nu^2 - 1}(r - a)\right]}{\sqrt{r}}.$$

If the dielectric is lossy, the refractive index ν has a positive imaginary part. The external electric and magnetic fields then oscillate with period of the order of $\lambda/|\sqrt{\nu^2 - 1}|$ and decay exponentially in the radial direction. The maximum field intensities in the external medium occur at the boundary $r = a$. Being proportional to λ/a, these maxima are small.

The field intensity inside the hollow waveguide is more interesting. Again if we substitute γ (20) into (2), (3) and (4) and neglect terms of the order λ/a, only the internal transverse components remain.

For TE_{0m} modes

$$E_{\theta 0m}{}^i = -\sqrt{\frac{\mu_0}{\epsilon_0}}\, H_{r0m}{}^i = J_1\left(u_{0m}\frac{r}{a}\right). \tag{10}$$

For TM_{0m} modes,

$$E_{r0m}{}^i = \sqrt{\frac{\mu_0}{\epsilon_0}}\, H_{\theta 0m}{}^i = J_1\left(u_{0m}\frac{r}{a}\right). \tag{11}$$

For EH_{nm} modes,

$$E_{\theta nm}{}^{i} = -\sqrt{\frac{\mu_0}{\epsilon_0}}\, H_{rnm}{}^{i} = J_{n-1}\left(u_{nm}\frac{r}{a}\right)\cos n\theta$$
$$E_{rnm}{}^{i} = \sqrt{\frac{\mu_0}{\epsilon_0}}\, H_{\theta nm}{}^{i} = J_{n-1}\left(u_{nm}\frac{r}{a}\right)\sin n\theta. \tag{12}$$

The field components of each mode have approximately the same radial dependence, varying as Bessel functions of the first kind, and tending to negligibly small values at the boundary (6). This approximate radial dependence (10), (11) and (12) is reproduced under each mode pattern in Figs. 2(a), 2(b) and 2(c).

III. PROPAGATION CONSTANTS FOR THE GENERAL CIRCULAR CYLINDRICAL GUIDE

In this section we shall determine the propagation constants γ, of the TE_{0m}, TM_{0m} and EH_{nm} modes in the straight hollow guide at optical wavelengths. The propagation constants are the roots of the following characteristic equation for the general circular cylindrical structure.[12] They are related to k_i and k_e by expressions (5).

$$\left[\frac{J_n'(k_ia)}{J_n(k_ia)} - \frac{k_i}{k_e}\frac{H_n^{(1)'}(k_ea)}{H_n^{(1)}(k_ea)}\right]\left[\frac{J_n'(k_ia)}{J_n(k_ia)} - \frac{\nu^2 k_i}{k_e}\frac{H_n^{(1)'}(k_ea)}{H_n^{(1)}(k_ea)}\right] = \left[\frac{n\lambda}{kk_ia}\right]^2\left[1 - \left(\frac{k_i}{k_e}\right)\right]^2. \tag{13}$$

This equation is simplified substantially when the approximations in (1) are introduced. Since $k_ea \gg 1$, the asymptotic value of the Hankel functions may be used

$$\frac{H_n^{(1)'}(k_ea)}{H_n^{(1)}(k_ea)} \approx i + 0(1/k_ea), \qquad k_ea \gg 1. \tag{14}$$

Since

$$\frac{\nu^2}{k_ea} \approx \frac{\nu^2}{(\nu^2-1)^{\frac{1}{2}}}\left(\frac{\lambda}{2\pi a}\right) \ll 1 \tag{15}$$

powers of ν^2/k_ea larger than one shall be neglected. The characteristic equation then simplifies to

$$J_{n-1}(k_ia) = i\nu_n(k_i/k)J_n(k_ia) \tag{16}$$

where

$$\nu_n = \begin{cases} \dfrac{1}{\sqrt{\nu^2 - 1}} & \text{for } TE_{0m} \text{ modes } (n = 0) \\ \dfrac{\nu^2}{\sqrt{\nu^2 - 1}} & \text{for } TM_{0m} \text{ modes } (n = 0) \\ \dfrac{\frac{1}{2}(\nu^2 + 1)}{\sqrt{\nu^2 - 1}} & \text{for } EH_{nm} \text{ modes } (n \neq 0). \end{cases} \tag{17}$$

To solve the characteristic equation for $k_i a$ we notice that because of (1) and (5), the right-hand side of (16) is close to zero. Using a perturbation technique and keeping only the first term of the perturbation,

$$k_i a \approx u_{nm}(1 - i\nu_n/ka) \tag{18}$$

where u_{nm} as before is the mth root of the equation

$$J_{n-1}(u_{nm}) = 0. \tag{19}$$

The validity of (18) is assured provided that the order of the mode is low enough so that $| \nu_n | u_{nm} \ll ka$. The propagation constants γ can then be obtained from (5)

$$\gamma \approx k\left[1 - \frac{1}{2}\left(\frac{u_{nm}\lambda}{2\pi a}\right)^2\left(1 - \frac{i\nu_n\lambda}{\pi a}\right)\right]. \tag{20}$$

The phase constant and attenuation constant of each mode are the real and imaginary parts of γ, respectively,

$$\begin{aligned} \beta_{nm} &= \text{Re } (\gamma) = \frac{2\pi}{\lambda}\left\{1 - \frac{1}{2}\left[\frac{u_{nm}\lambda}{2\pi a}\right]^2\left[1 + \text{Im}\left(\frac{\nu_n\lambda}{\pi a}\right)\right]\right\} \\ \alpha_{nm} &= \text{Im } (\gamma) = \left(\frac{u_{nm}}{2\pi}\right)^2 \frac{\lambda^2}{a^3} \text{Re } (\nu_n). \end{aligned} \tag{21}$$

IV. PROPAGATION CONSTANTS FOR STRAIGHT DIELECTRIC GUIDES

For guides made of dielectric material, ν_n is usually real and independent of λ, so that the phase and attenuation constants are

$$\left.\begin{aligned} \beta_{nm} &= \frac{2\pi}{\lambda}\left\{1 - \frac{1}{2}\left(\frac{u_{nm}\lambda}{2\pi a}\right)^2\right\} \\ \alpha_{nm} &= \left(\frac{u_{nm}}{2\pi}\right)^2 \frac{\lambda^2}{a^3} \begin{cases} \dfrac{1}{\sqrt{\nu^2 - 1}}, & \text{for } TE_{0m} \text{ modes } (n = 0) \\ \dfrac{\nu^2}{\sqrt{\nu^2 - 1}}, & \text{for } TM_{0m} \text{ modes } (n = 0) \\ \dfrac{\frac{1}{2}(\nu^2 + 1)}{\sqrt{\nu^2 - 1}}, & \text{for } EH_{nm} \text{ modes } (n \neq 0) \end{cases} \end{aligned}\right\}. \quad (22)$$

The phase constant of modes in hollow dielectric waveguides have the same frequency dependence as modes in perfectly conducting metallic waveguides when operating far from cutoff; both transmission media are then similarly dispersive.

The attenuation constants are proportional to λ^2/a^3. Consequently, the losses can be made arbitrarily small by choosing the radius of the tube a sufficiently large relative to the wavelength λ.

The refractive index ν affects the attenuation of each of the three types of modes (22) in different ways. This fact is reasonable on physical grounds. TE_{0m} modes can be considered to be composed of plane wavelets, each impinging at grazing angle on the interface between the two media with polarization perpendicular to the plane of incidence. It is known from the laws of refraction that the larger the value of ν, the smaller the refracted power.

TM_{0m} modes may also be thought of as consisting of plane wavelets, but with the electric field of each now contained in the plane of incidence. For ν very close to unity, there is little reflection and the refracted loss is high; as the value of ν is allowed to become large, each wavelet gets close to the Brewster angle of incidence and again the refracted loss is high. The minimum occurs for $\nu = \sqrt{2}$.

EH_{nm} modes are composed of both types of plane wavelets. Therefore, as is reasonable from the above argument, the attenuation constant α_{nm} has a ν dependence which is an average of those of TE_{0m} and TM_{0m} modes. The value of ν that minimizes α_{nm} is $\nu = \sqrt{3} = 1.73$.

The attenuation constants (22) are proportional to u_{nm}^2. Some values of u_{nm} (19) are presented in Table I. For a fixed value of n the attenuation constant increases with m. This statement is not true for m fixed and n variable.

Comparing the attenuation constants (22) of the different modes, we find that the mode with lowest attenuation is TE_{01} if $\nu > 2.02$ and EH_{11} if $\nu < 2.02$. Most glasses have a refractive index $\nu \approx 1.5$, and

TABLE I— SOME VALUES OF u_{nm}

n/m	1	2	3	4
1	2.405	5.52	8.654	11.796
2 or 0	3.832	7.016	10.173	13.324
3 or −1	5.136	8.417	11.62	14.796
4 or −2	6.380	9.761	13.015	16.223

consequently for hollow glass tube EH_{11} should be preferred. The attenuation of this mode ($8686\alpha_{11}$ in db/km) has been plotted in Fig. 7 as a function of λ/a for $\nu = 1.50$ using λ as a parameter. Typically, for a wavelength $\lambda = 1\mu$ and radius $a = 1$ mm, the attenuation of the EH_{11} mode is 1.85 db/km (≈3 db/mile). If the radius of the guide is doubled, the attenuation is reduced to 0.231 db/km.

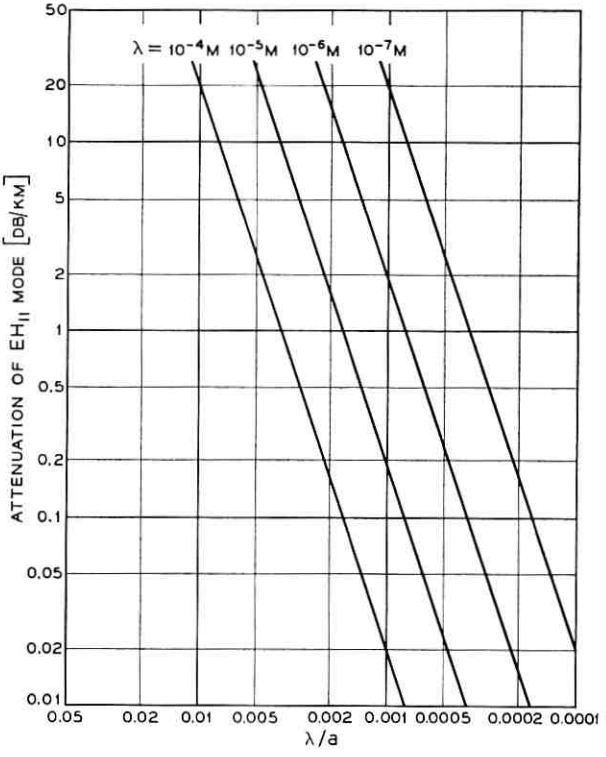


Fig. 7 — Attenuation of EH_{11} modes ($1.85\ \lambda^2/a^3$) versus wavelength/radius ($\nu = 1.5$).

V. HOLLOW DIELECTRIC WAVEGUIDE FOR OPTICAL MASER AMPLIFIERS AND OSCILLATORS

A mode traveling in a hollow dielectric waveguide filled with "masing" material experiences a net gain which is given by the difference between the amplification due to the active medium and the loss due to leakage through the walls. It has been shown[11] that in a tube filled with the right mixture of He and Ne at the proper pressure, the gain G is inversely proportional to the radius a of the tube. Then

$$G = (A/a) \text{ db/m} \tag{23}$$

where the radius a is measured in meters and the constant A is

$$A = 0.00066 \text{ db.}$$

On the other hand, we have found that the transmission loss of the EH_{11} mode in the hollow waveguide with a refractive index $\nu = 1.50$ is $L = 8.686\alpha_{11}$. From (22)

$$L = B(\lambda^2/a^3) \text{ db/m} \tag{24}$$

where the constant B is

$$B = 1.85 \text{ db.}$$

The net gain per unit length is then

$$G - L = (A/a) - B(\lambda^2/a^3) \tag{25}$$

passing through a maximum at the value of the radius for which

$$\partial(G - L)/\partial a = 0.$$

The optimum radius and the maximum net gain are respectively

$$\begin{aligned} a_{\text{opt}} &= \sqrt{3\frac{B}{A}}\,\lambda = 91.7\lambda \\ (G - L)_{\max} &= \frac{2}{3^{\frac{3}{2}}}\frac{A^{\frac{3}{2}}}{B^{\frac{1}{2}}}\frac{1}{\lambda} = 4.81\,\frac{10^{-6}}{\lambda} \text{ db/m.} \end{aligned} \tag{26}$$

For the He-Ne mixture, $\lambda = 0.6328\ 10^{-6}$ m. Consequently

$$\begin{aligned} a_{\text{opt}} &= 0.058 \text{ mm} \\ (G - L)_{\max} &= 7.6 \text{ db/m.} \end{aligned} \tag{27}$$

Although the diameter of the tube is quite small, the gain per unit length is sufficiently large as to make hollow dielectric amplifiers and oscillators attractive for experimentation.

Present-day confocal He-Ne masers employ tubes whose approximate length and radius are 1 m and 3 mm respectively. The gain per passage (23) is 0.22 db ($\approx$5 per cent). If a hollow dielectric waveguide with an optimum radius 0.058 mm were used, the same gain would be achieved with a length of only 0.22/7.6 = 29 mm. This presents an excellent possibility for a very compact maser.

Even for radii larger than the optimum, the hollow dielectric waveguide is still attractive. For example with $a = 0.25$ mm, the gain is 2.6 db/m, a value far larger than the gain 0.22 db/m obtained for the 3-mm radius tube commonly used for masers.

Nevertheless, for long-wavelength masers the optimum values (26) are not practical. Consider for example a tube containing an active material which amplifies at $\lambda = 10^{-4}$ m. Let us assume that the constant A is still 0.00066. Then from (26), the optimum radius and maximum gain are

$$\begin{aligned} a_{\text{opt}} &= 9.14 \text{ mm} \\ (G - L)_{\max} &= 0.0481 \text{ db/m}. \end{aligned} \tag{28}$$

The gain is very small. It could be enhanced by reducing the radius and by increasing the refractive index ν of the walls to a value much larger than 1.5. This can be accomplished if metal is used instead of dielectric, as is shown in the next section.

VI. ATTENUATION CONSTANTS FOR THE STRAIGHT METALLIC GUIDE

In order to discuss the attenuation characteristics of metallic waveguides, we shall need to have some quantitative information about the behavior of metals at optical frequencies. We examine as a typical example the optical properties of aluminum, even though this may not be the most suitable metal. The dispersion characteristics of the conductivity and relative dielectric constants of aluminum have been studied extensively by Hodgson,[13] Beattie and Conn,[14] and Schulz.[15] The data used below have been taken from a compilation of the results of these studies,[16] and is presented graphically in Fig. 8. It is evident from these dispersion curves that the dielectric constant for aluminum is much larger than for ordinary dielectrics and increases monotonically with wavelength in the range $0.3\mu < \lambda < 4.0\mu$.

The circular electric modes have the lowest loss in metallic waveguides, while the circular magnetic and hybrid modes are rapidly attenuated even for a wavelength as short as 0.3μ. The attenuation constant α_{01} for the lowest-loss TE_{01} mode is plotted in Fig. 9 for wave-

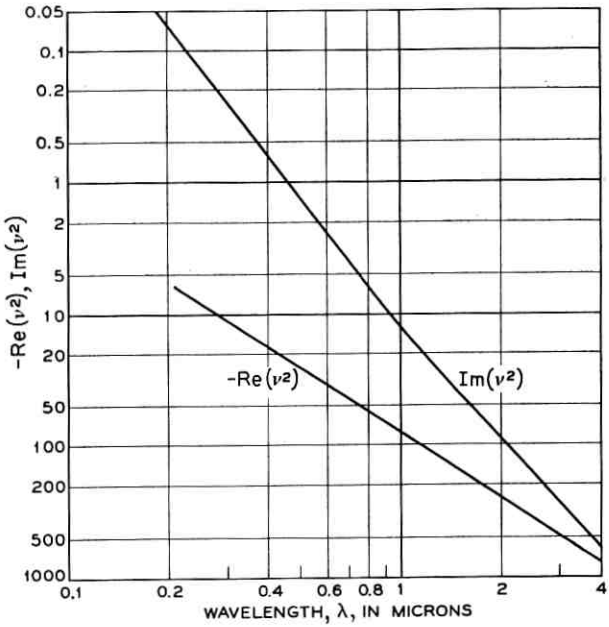


Fig. 8 — Dispersion curve for aluminum $\nu^2 = \epsilon/\epsilon_0 = \mathrm{Re}(\nu^2) + i\,\mathrm{Im}(\nu^2)$ versus wavelength $\lambda(\mu)$.

lengths in the range $0.3\mu < \lambda < 4.0\mu$ for $a = 0.25$ mm, 0.50 mm and 1 mm. These data show a considerable improvement over that corresponding to the lowest-loss mode EH_{11} for the dielectric guide. We saw that for a hollow glass dielectric waveguide, the EH_{11} mode has a loss of 1.8 db/km for a radius $a = 1$ mm and wavelength $\lambda = 1\mu$. The attenuation for the TE_{01} mode for the aluminum guide with the same radius and wavelength is only 0.028 db/km. For a wavelength $\lambda = 1\mu$ and a radius $a = 0.25$ mm, the minimum-loss TE_{01} mode for the aluminum waveguide has an attenuation constant $\alpha_{01} = 1.8$ db/km. The same attenuation is achieved for $\lambda = 3\mu$ and $a = 0.6$ mm. The attenuation constant for the TE_{02} mode under the last two conditions is $\alpha_{02} = 6.05$ db/km. For a wavelength $\lambda = 1\mu$ and $a = 0.25$ mm, the straight guide losses for the TM_{01} and EH_{11} modes are approximately 145 db/km and 57 db/km, respectively.

VII. FIELD CONFIGURATION AND ATTENUATION OF MODES IN THE CURVED GUIDE

In order to achieve a more realistic evaluation of the hollow circular waveguide for long distance optical transmission, it is necessary to

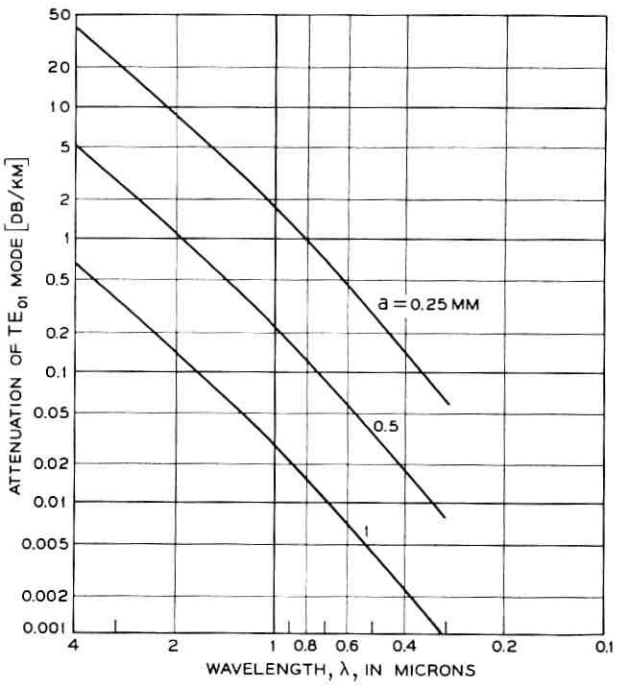


Fig. 9 — Attenuation of TE_{01} mode, α_{01} versus wavelength $\lambda(\mu)$, for aluminum guide.

evaluate the effects of mild curvature of the guide axis. This is most easily accomplished by determining a perturbation correction for both the field configuration and the attenuation constants for the idealized straight guide whose characteristics have been described above.

VIII. FORMULATION OF THE PROBLEM

Consider the toroidal system (r, θ, z) with metric coefficients

$$\begin{aligned} e_r &= 1 \\ e_\theta &= r \\ e_z &= 1 + r/R \sin \theta \end{aligned} \tag{29}$$

as depicted in Fig. 10. In this system of coordinates, a differential length is given by

$$ds = (e_r^2 dr^2 + e_\theta^2 d\theta^2 + e_z^2 dz^2)^{\frac{1}{2}} \tag{30}$$

where R is the radius of curvature of the toroidal system and is chosen

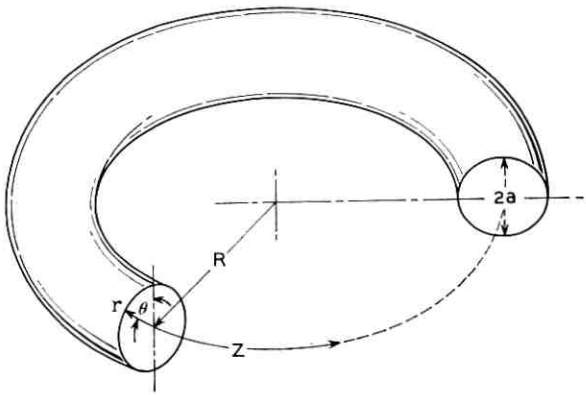


Fig. 10 — The curved hollow dielectric waveguide and the associated toroidal coordinate system (r, θ, Z).

equal to the radius of curvature of the guide axis, so that the guide wall is located at $r = a$, and the axis of the guide coincides with the curved z-axis. In this toroidal coordinate system, Maxwell's equations are

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\{(1 + r/R \sin\theta)\mathcal{H}_z\} - i\gamma_c r\mathcal{H}_\theta + i\omega\epsilon r(1 + r/R \sin\theta)\mathcal{E}_e = 0 \\
&i\gamma_c\mathcal{H}_r - \frac{\partial}{\partial r}\{(1 + r/R \sin\theta)\mathcal{H}_z\} + i\omega\epsilon(1 + r/R \sin\theta)\mathcal{E}_\theta = 0 \\
&\frac{\partial}{\partial r}(r\mathcal{H}_\theta) - \frac{\partial}{\partial\theta}\mathcal{H}_r + i\omega\epsilon r\mathcal{E}_z = 0 \\
&\frac{\partial}{\partial\theta}\{(1 + r/R \sin\theta)\mathcal{E}_z\} - i\gamma_c\mathcal{E}_\theta - i\omega\mu r(1 + r/R \sin\theta)\mathcal{H}_r = 0 \\
&i\gamma_c\mathcal{E}_r - \frac{\partial}{\partial r}\{(1 + r/R \sin\theta)\mathcal{E}_z\} - i\omega\mu(1 + r/R \sin\theta)\mathcal{H}_\theta = 0 \\
&\frac{\partial}{\partial r}(r\mathcal{E}_\theta) - \frac{\partial}{\partial\theta}\mathcal{E}_r - i\omega\mu r\mathcal{H}_z = 0
\end{aligned} \tag{31}$$

where we have omitted the common factor

$$\exp i(\gamma_c z - \omega t)$$

in which γ_c is the propagation constant along the curved z-axis.

The toroidal system (r, θ, z) and the curved waveguide degenerate into a cylindrical system and a straight guide, respectively, as R approaches infinity. Maxwell's equations for the straight guide are therefore obtained from (31) by letting $R \to \infty$.

$$\begin{aligned}
\frac{\partial}{\partial\theta} H_z - i\gamma r H_\theta + i\omega\epsilon r E_r &= 0 \\
i\gamma H_r - \frac{\partial}{\partial r} H_z + i\omega\epsilon E_r &= 0 \\
\frac{\partial}{\partial r}(rH_\theta) - \frac{\partial}{\partial\theta} H_r + i\omega\epsilon r E_r &= 0 \\
\frac{\partial}{\partial\theta} E_z - i\gamma r E_\theta - i\omega\epsilon r H_r &= 0 \\
i\gamma E_r - \frac{\partial}{\partial r} E_z - i\omega\mu H_\theta &= 0 \\
\frac{\partial}{\partial r}(rE_\theta) - \frac{\partial}{\partial\theta} E_r - i\omega\mu r H_z &= 0
\end{aligned} \tag{32}$$

where γ is the propagation constant for the straight guide, and the superscript i and subscripts nm are suppressed.

IX. SOLUTION FOR THE CURVED GUIDE

We proceed to solve (31) for the field vectors $\vec{\mathcal{E}}$, $\vec{\mathcal{H}}$ and obtain the propagation constant γ_c for the curved guide as functions of the field vectors $\vec{E}$, $\vec{H}$ and the propagation constant γ of the straight guide. The latter quantities are known [(2), (3), (4) and (20)]. We introduce a parameter

$$\sigma = \frac{k}{\gamma - k} \cdot \frac{a}{R} \approx 2\left(\frac{2\pi a}{u_{nm}\lambda}\right)^2 \frac{a}{R}. \tag{33}$$

The range of interest is that for which the radius of curvature R is so large that $\sigma \ll 1$.

Using a first-order perturbation technique, the solution of (31) is

$$\begin{aligned}
\mathcal{E}_\theta &= (1 + \sigma r/a \sin\theta)E_\theta \\
\mathcal{E}_r &= (1 + \sigma r/a \sin\theta)E_r \\
\mathcal{E}_z &= (1 + \sigma r/a \sin\theta)E_z + (i\sigma/ka)(E_r \sin\theta + E_\theta \cos\theta) \\
\mathcal{H}_\theta &= (1 + \sigma r/a \sin\theta)H_\theta \\
\mathcal{H}_r &= (1 + \sigma r/a \sin\theta)H_r \\
\mathcal{H}_z &= (1 + \sigma r/a \sin\theta)H_z + (i\sigma/ka)(H_r \sin\theta + H_\theta \cos\theta).
\end{aligned} \tag{34}$$

The effect of curvature of the guide axis is to make unsymmetrical the

transverse field configuration of the straight guide. Each transverse component is enhanced in the half cross section farthest from the center of curvature.

To a first-order perturbation of σ, the propagation constants of the curved and straight guide are identical; i.e., $\gamma_c \approx \gamma$. Nevertheless, knowing the field components of the mildly curved structure, it is possible to calculate its attenuation constants $\alpha_{nm}(R) = \text{Re}\ \gamma_c$.

X. ATTENUATION CONSTANTS $\alpha_{nm}(R)$

The mean radial power flowing into the dielectric per unit length at the surface of the guide is

$$P_r = \frac{1}{2}\int_0^{2\pi} \text{Re}\,[\mathcal{E}_\theta \mathcal{H}_z^* - \mathcal{E}_z \mathcal{H}_\theta^*]_{r=a}\,[1 + a/R \sin\theta] a\, d\theta. \tag{35}$$

The power flow in the axial z direction within the internal medium $r < a$ is

$$P_z = \frac{1}{2}\int_0^a \int_0^{2\pi} \text{Re}\,[\mathcal{E}_r \mathcal{H}_\theta^* - \mathcal{E}_\theta \mathcal{H}_r^*] r\, d\theta\, dr \tag{36}$$

and decreases along z at a rate equal to the radial flow per unit length P_r; i.e.,

$$\frac{dP_z}{dz} = -2\alpha_{nm}(R) P_z = -P_r \tag{37}$$

where $\alpha_{nm}(R)$ is the attenuation constant of the mode under consideration for the curved hollow dielectric waveguide. Consequently

$$\alpha_{nm}(R) = \tfrac{1}{2}(P_r/P_z). \tag{38}$$

To compute P_r we substitute the known field quantities into (35). This yields

$$P_r = \text{Re}\sqrt{\frac{\epsilon_0}{\mu_0}}\,\frac{u_{nm}^2 J_n^2(u_{nm})}{2k^2 a\sqrt{\nu^2 - 1}} \int_0^{2\pi} |1 + \sigma \sin\theta|^2 (1 + a/R \sin\theta) \begin{cases} 1 & \text{for } TE_{0m} \text{ modes} \\ \nu^2 & \text{for } TM_{0m} \text{ modes} \\ \nu^2 \sin^2 n(\theta + \theta_0) + \cos^2 n(\theta + \theta_0) & \text{for } EH_{nm} \text{ modes.} \end{cases} d\theta \tag{39}$$

Terms with powers of $\lambda/(2\pi a)$ larger than two have been neglected. Upon integrating,

$$P_r = \pi\ \text{Re}\sqrt{\frac{\epsilon_0}{\mu_0}}\,\frac{u_{nm}^2 J_n^2(u_{nm})}{k^2 a\sqrt{\nu^2 - 1}}$$

$$\begin{cases} (1 + \tfrac{1}{2}\sigma^2) & \text{for } TE_{0m} \text{ modes} \\ \nu^2(1 + \tfrac{1}{2}\sigma^2) & \text{for } TM_{0m} \text{ modes} \\ \dfrac{1}{2}(\nu^2 + 1)\left[1 + \dfrac{1}{2}\sigma^2\left(1 + \dfrac{\delta_n(\pm 1)}{2}\dfrac{\nu^2 - 1}{\nu^2 + 1}\cdot\cos 2\theta_0\right)\right] & \text{for } EH_{nm} \text{ modes} \end{cases} \tag{40}$$

where

$$\delta_n(\pm 1) = \begin{cases} 1, & n = \pm 1 \\ 0, & n \neq \pm 1. \end{cases} \tag{41}$$

The power P_z flowing radially in the guide is obtained by substituting (34) into (36) and integrating

$$P_z = \frac{\pi a^2}{2}\sqrt{\frac{\epsilon_0}{\mu_0}}\, J_n^2(u_{nm})\left\{1 + \frac{\sigma^2}{6}\left[1 + 2n(n-2)/u_{nm}^2\right]\right\}. \tag{42}$$

Hence

$$\alpha_{nm}(R) = \alpha_{nm}(\infty)\left\{1 + \frac{4}{3}\left(\frac{2\pi a}{u_{nm}\lambda}\right)^4\left(\frac{a}{R}\right)^2 \cdot\left[1 - \frac{n(n-2)}{u_{nm}^2} + \frac{3}{4}\delta_n(\pm 1)\frac{\operatorname{Re}\sqrt{\nu^2 - 1}}{\operatorname{Re}\dfrac{\nu^2 + 1}{\sqrt{\nu^2 - 1}}}\cos 2\theta_0\right]\right\} \tag{43}$$

where $\alpha_{nm}(\infty) = \alpha_{nm}$ is the attenuation constant for modes in the straight guide ($R = \infty$) given by (21). The attenuation constant $\alpha_{nm}(R)$ can also be written in the following form

$$\alpha_{nm}(R) = \alpha_{nm}(\infty) + (a^3/\lambda^2 R^2)\operatorname{Re} V_{nm}(\nu) \tag{44}$$

where

$$V_{nm}(\nu) = \frac{4}{3}\begin{Bmatrix} \dfrac{1}{\sqrt{\nu^2 - 1}} \\ \dfrac{\nu^2}{\sqrt{\nu^2 - 1}} \\ \dfrac{\frac{1}{2}(\nu^2 + 1)}{\sqrt{\nu^2 - 1}} \end{Bmatrix}\left(\frac{2\pi}{u_{mn}}\right)^2 \cdot\left\{1 - \frac{n(n-2)}{u_{nm}^2} + \frac{3}{4}\delta_n(\pm 1)\left(\frac{\nu^2 - 1}{\nu^2 + 1}\right)\cos 2\theta_0\right\}. \tag{45}$$

The values of $\mathrm{Re}V_{nm}(\nu)$ are always positive. Some of them have been calculated in Table II for a refractive index $\nu = 1.50$.

The attenuation constant of any mode consists of two terms (44). The first coincides with that of the straight guide and is proportional to $u_{nm}{}^2\lambda^2/a^3$; the second term represents an increase in attenuation due to curvature of the guide axis and is proportional to $a^3/\lambda^2R^2u_{nm}{}^2$. Therefore the lower the straight guide attenuation constant (small $u_{nm}{}^2\lambda^2/a^3$), the larger the loss due to bends and vice versa. From (43) or (45) we find that only for the $EH_{\pm1,m}$ modes, the orientation of the field with respect to the plane of curvature influences the attenuation. If $\theta_0 = 0$, the electric field in the center of the guide is in the plane of curvature and the attenuation is a maximum. For $\theta_0 = \pm\pi/2$, the electric field is normal to the plane of curvature and the attenuation is a minimum. The ratio of maximum to minimum is mild, however. For the lowest attenuation mode EH_{11} and $\nu = 1.50$, it is

$$\frac{V_{nm}(\theta_0 = 0)}{V_{nm}(\theta_0 = \pi/2)} = 1.65. \tag{46}$$

If $|\nu| \gg 1$, that ratio is

$$\frac{V_{nm}(\theta_0 = 0)}{V_{nm}(\theta = \pi/2)} = 4.6.$$

From equation (43) we find that the radius of curvature which doubles the straight guide attenuation is

$$R_0 = \frac{2}{\sqrt{3}}\left(\frac{2\pi}{u_{mn}}\right)^2\frac{a^3}{\lambda^2} \cdot\left[1 - \frac{n(n-2)}{u_{nm}{}^2} + \frac{3}{4}\delta_n(\pm1)\ \mathrm{Re}\frac{\mathrm{Re}\sqrt{\nu^2-1}}{\nu^2+1}\cos 2\theta_0\right]^{-\frac{1}{2}}. \tag{47}$$

This value of R_0 is only approximate since (43) was derived by assuming $\sigma \ll 1$.

XI. EFFECT OF CURVATURE ON ATTENUATION OF MODES IN THE HOLLOW DIELECTRIC WAVEGUIDE

For a straight hollow glass waveguide with $\nu = 1.5$ and a radius $a = 1$ mm operating typically at a wavelength $\lambda = 1\mu$, the attenuation of the lowest-loss mode EH_{11} is $\alpha_{11} = 1.85$ db/km. This loss is doubled for a radius of curvature $R_0 \approx 10$ km. For long distance optical transmission a radius of curvature of at least a few hundred meters would

TABLE II — SOME VALUES OF $V_{nm}(\nu)$

n/m		1	2	3	4
-1		2.57 $(1 + 0.326 \cos 2\theta_0)$	1.034 $(1 + 0.301 \cos 2\theta_0)$	0.553 $(1 + 0.295 \cos 2\theta_0)$	0.347 $(1 + 0.293 \cos 2\theta_0)$
0	TE	3.22	0.955	0.455	0.265
	TM	7.22	2.145	1.022	0.596
1		15.5 $(1 + 0.246 \cos 2\theta_0)$	2.60 $(1 + 0.279 \cos 2\theta_0)$	1.034 $(1 + 0.284 \cos 2\theta_0)$	0.554 $(1 + 0.286 \cos 2\theta_0)$
2		5.22	1.55	0.735	0.432
3		2.57	1.034	0.553	0.347
4 or -2		1.51	0.737	0.430	0.287

be tolerable. Therefore hollow dielectric waveguides do not seem suitable for long distance optical transmission.

On the other hand, the curvature in hollow dielectric waveguides for application in gaseous amplifiers and oscillators is not critical. For example, if $a = 0.25$ mm and $\lambda = 1\mu$, the straight guide attenuation is 0.12 db/meter. The radius of curvature which doubles this quantity for the lossiest polarization — i.e., with the electric field at the center of the guide contained in the plane of curvature — is approximately 150 meters, a value well within the limits of laboratory precision. Consequently, the hollow dielectric waveguide does remain very attractive as a guiding medium for optical amplifiers and oscillators where a small guide radius is desirable, thereby making the guide less sensitive to curvature of the axis.

XII. EFFECT OF CURVATURE ON ATTENUATION OF MODES IN THE METALLIC GUIDE

The attenuation constants $\alpha_{0m}(R)$ for the lowest-loss TE_{0m} modes in the curved metallic guide are given by

$$\alpha_{0m}(R) \approx \alpha_{0m}(\infty)\left\{1 + \frac{4}{3}\left(\frac{2\pi a}{\lambda u_{0m}}\right)^4\left(\frac{a}{R}\right)^2\right\} \tag{48}$$

where $\alpha_{0m}(\infty)$ is the attenuation constant for the TE_{0m} mode in the straight guide, $R = \infty$. For a radius $a = 0.25$ mm and wavelength $\lambda = 1\mu$, the straight guide loss for the lowest-loss TE_{01} mode, $\alpha_{01}(\infty) = 1.8$ db/km, is doubled for a radius of curvature of only $R_0 \approx 48$ meters.

For $\lambda = 3\mu$ and $a = 0.6$ mm, the straight TE_{01} loss is also 1.8 db/km and the radius of curvature that doubles that loss is 75 m.

XIII. CONCLUSIONS

The hollow dielectric waveguide at optical wavelengths supports a complete set of normal modes that are either circular electric, circular magnetic or hybrid. They resemble the modes found in a sequence of circular irises not only in field configuration but also in loss discrimination among them. For hollow metallic waveguides the mode discrimination is far larger.

The field configuration and propagation constants have been determined. The attenuation is practically independent of the loss tangent of the dielectric but depends essentially on the refraction mechanism at the wall. Assuming refractive index of the dielectric, 1.5 for hollow dielectric waveguides, the EH_{11} mode exhibits the lowest power attenuation, viz., 1.85 (λ^2/a^3) db/m. For a wavelength $\lambda = 1\mu$ and a tube radius $a = 1$ mm, the attenuation is only 1.85 db/km.

The hollow dielectric waveguide does not, however, seem suitable for long distance optical transmission because of the high loss introduced by even mild curvature of the guide axis. Nevertheless it remains very attractive as a guiding medium for optical amplifiers and oscillators, since here a small radius of the guide is desirable. Consequently, curvature of the guide axis is not critical. Filled with "masing" material, the hollow dielectric waveguide provides not only guidance but also gain which is almost inversely proportional to the radius. For the right He-Ne mixture, the maximum theoretical gain attainable is 7.6 db/m provided that the radius is 0.058 mm. But even if the radius is 0.25 mm, the predicted gain is still large, viz. 2.6 db/m.

The metallic waveguide is superior to the hollow dielectric waveguide for use in long distance optical transmission. Because of the relatively large dielectric constant exhibited by aluminum at optical frequencies, the attenuation constant for the lowest-loss mode TE_{01} is comparatively small and less sensitive to curvature of the guide axis. For a radius $a = 0.25$ mm and a wavelength $\lambda = 1\mu$, the attenuation constant for TE_{01} modes in the straight aluminum guide is only 1.8 db/km, which is doubled for a radius of curvature of about 48 meters. For $a = 0.6$ mm and $\lambda = 3\mu$, the TE_{01} straight guide loss is also 1.8 db/km but is doubled if the radius of curvature of the waveguide axis is 75m.

We have considered some of the theoretical problems of the hollow dielectric or metallic waveguide. The results are promising. Nevertheless, the usefulness of these guides has yet to be proven experimentally,

and furthermore the attenuation constants discussed here do not include scattering losses due to surface imperfections.

XIV. ACKNOWLEDGMENTS

It is a pleasure to thank R. Kompfner and S. E. Miller for their suggestions.

REFERENCES

1. Fox, A. G., and Li, Tingye, Resonant Modes in a Maser Interferometer, B.S.T.J., **40**, March, 1961, p. 453.
2. Boyd, G. D., and Gordon, J. P., Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers, B.S.T.J., **40**, March, 1961, p. 489.
3. Boyd, G. D., and Kogelnik, H., Generalized Confocal Resonator Theory, B.S.T.J., **41**, July, 1962, p. 1347.
4. Goubau, G., and Schwering, F., On the Guided Propagation of Electromagnetic Wave Beams, Trans. I.R.E., **AP-9**, May 1961, p. 248.
5. Eaglesfield, C. C., Optical Pipeline: A Tentative Assessment, The Inst. of Elect. Engineers, January, 1962, p. 26.
6. Simon, J. C., and Spitz, E., Propagation Guidée de Lumière Cohérente, J. Phys. Radium, **24**, February, 1963, p. 147.
7. Goubau, G., and Christian, J. R., Some Aspects of Beam Waveguides for Long Distance Transmission at Optical Frequencies, IEEE Trans. on Microwave Theory and Techniques, **MTT-12**, March, 1964, pp. B.S.T.J., 212–220.
8. Marcuse, D., and Miller, S. E., Analysis of a Tubular Gas Lens, B.S.T.J., this issue, p. 1759.
9. Berreman, D. W., A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases, B.S.T.J., this issue, p. 1469.
10. Berreman, D. W., A Gas Lens using Unlike, Counter-Flowing Gases, B.S.T.J., this issue, p. 1476.
11. Gordon, E. I., and White, A. D., Similarity Laws for the He-Ne Gas Maser, Appl. Phys. Letters, **3**, December 1, 1963, p. 199.
12. Stratton, J. A., *Electromagnetic Theory*, McGraw-Hill Book Co., New York and London, 1941, p. 524.
13. Hodgson, J. N., Proc. Phys. Soc. (London), **B68**, 1955, p. 593.
14. Beattie, J. R., and Conn, G. K. T., Phil. Mag., **7**, 1955, pp. 46, 222, and 989.
15. Schulz, L. G., J. Opt. Soc. Am., **41**, 1951, p. 1047; **44**, 1954 p. 357.
16. Givens, M. Parker, Optical Properties of Metals, *Solid State Physics*, **6**, Acad. Press Inc., New York, 1958, p. 313.

Contributors to This Issue

VACLAV E. BENEŠ, A.B., 1950, Harvard College; M.A. and Ph.D., 1953, Princeton University; Bell Telephone Laboratories, 1953—. Mr. Beneš has been engaged in mathematical research on stochastic processes, traffic theory, and servomechanisms. In 1959–60 he was visiting lecturer in mathematics at Dartmouth College. He is the author of *General Stochastic Processes in the Theory of Queues* (Addison-Wesley, 1963). Member, American Mathematical Society, Association for Symbolic Logic, Institute of Mathematical Statistics, Society for Industrial and Applied Mathematics, Mind Association and Phi Beta Kappa.

MARTIN R. BIAZZO, Bell Telephone Laboratories, 1951—. He attended Rutgers University College. His first assignment at Bell Laboratories was on PCM systems, primarily in the area of ferrite core transformers and PCM repeaters. His present interest is in the measurement of optical losses with the application of electronic techniques on optical lasers.

JAMES O. CAPPELLARI, JR., B.S., 1953, Marshall University; B.S.A.E., 1956, M.S.A.E., 1957, Ph.D., 1961, Purdue University; Instructor, School of Aeronautical and Engineering Sciences, Purdue, 1957–1961; Bell Telephone Laboratories, 1961–1962, Bellcomm, Inc., 1962—. While at Bell Laboratories he was engaged in attitude control, rigid body dynamics, and orbital mechanics problems connected with Project Telstar. Since going to Bellcomm, Inc., he has been concerned with a wide variety of trajectory analysis studies connected with Project Apollo. Member, A.I.A.A. and Sigma Xi.

G. F. FOXHALL, B.S. (Physics), Worcester Polytechnic Institute, 1961, M.S. (Physics), University of Illinois, 1962; Bell Telephone Laboratories, 1961—. He has investigated the effects of gaseous ambients on silicon surfaces, worked on the development of a germanium bridge rectifier polarity guard, and studied metal-semiconductor contacts. Member, Sigma Xi.

ROGER M. GOLDEN, B.S., 1954, M.S., 1955, Ph.D., 1959, California Institute of Technology; Fulbright student Technical Institute at Eind-

hoven, 1959–1960; Bell Telephone Laboratories, 1960—. Since joining Bell Laboratories, he has been working on speech bandwidth compression devices, vocoders, and speech analysis-synthesis systems for telephone communications. He is presently studying such systems by means of newly developed digital computer simulations. Member, Acoustical Society of America, IEEE, Sigma Xi and Tau Beta Pi.

JOHN P. HYDE, A.B., Princeton University, 1959; M.S., Northwestern University, 1960; Bell Telephone Laboratories, 1960—. He has worked on machine aids to design, and especially computer aids to sequential circuit synthesis. He is presently engaged in further development of the ALPAK system and other aspects of computer algebra.

JAMES F. KAISER, E.E., University of Cincinnati, 1952; S.M., 1954, and Sc.D., 1959, Massachusetts Institute of Technology; faculty of the Massachusetts Institute of Technology, 1956–1960; Bell Telephone Laboratories, 1959—. He has been concerned with problems of data processing, digital filter design, and system simulation. Member, IEEE, Association for Computing Machinery, Society for Industrial and Applied Mathematics, Eta Kappa Nu, Sigma Xi and Tau Beta Pi.

B. K. KINARIWALA, B.S., 1951, Benares University (India); M.S., 1954, and Ph.D., 1957, University of California; Bell Telephone Laboratories, 1957—. He was first engaged in research in circuit theory involving, in particular, active and time-varying networks. More recently, he has been concerned with problems in digital communication systems. Member, IEEE and Sigma Xi.

DAVID A. KLEINMAN, S.B., 1946, and S.M., 1947, Massachusetts Institute of Technology; Ph.D., 1952, Brown University; Brookhaven National Laboratory, 1949–53; Bell Telephone Laboratories, 1953—. Mr. Kleinman has worked in the areas of neutron scattering in solids, semiconductor electronics, electron energy bands, and the infrared properties of crystals, and is currently working on problems related to the optical maser. Member, American Physical Society.

JOHN A. LEWIS, B.S., 1944, Worcester Polytechnic Institute, Sc. M., 1948, Brown University, Ph.D., 1950, Brown University, Bell Telephone Laboratories, May, 1951—. A member of the mathematical physics department, he has been engaged in theoretical investigation of problems of fluid dynamics, piezoelectric vibrations, heat transfer, and

satellite attitude control. He is currently studying problems of hypersonic flow. Member, American Mathematical Society, Society for Industrial and Applied Mathematics, l'Unione Matematica Italiana and the Society for Natural Philosophy.

E. A. J. Marcatili, Aeronautical Engineer, 1947, and E.E., 1948, University of Córdoba (Argentina); Research staff, University of Córdoba, 1947–54; Bell Telephone Laboratories, 1954—. He has been engaged in theory and design of filters in multimode waveguides and in waveguide systems research. More recently he has concentrated on the study of optical transmission media. Member, IEEE.

Dietrich Marcuse, Diplom Vorpruefung, 1952 and Dipl. Phys., 1954; Berlin Free University; D.E.E., 1962, Technische Hochschule, Karlsruhe, Germany; Siemens and Halske (Germany), 1954–1957; Bell Telephone Laboratories, 1957—. At Siemens and Halske Mr. Marcuse was engaged in transmission research, studying coaxial cable and circular waveguide transmission. At Bell Laboratories he has been engaged in studies of circular electric waveguides and work on gaseous masers. Member, IEEE.

Stewart E. Miller, B.S. and M.S., 1941, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1941—. He first worked on coaxial carrier repeaters and later shifted to microwave radar systems development. At the close of World War II he returned to coaxial carrier repeater development until 1949, when he joined the radio research department. There his work has been in the fields of circular electric waveguide communication, microwave ferrite devices, and other components for microwave radio systems. As Director, Guided Wave Research Laboratory, he heads a group engaged in research on communication techniques for the millimeter wave and optical regions. Fellow, IEEE.

A. J. Rack, B.S., 1930, University of Illinois; M.A., 1935, Columbia University; Bell Telephone Laboratories, 1930—. He has been engaged in the application of circuits in the communication field, including studies of tube noise, feedback amplifiers, transistor circuits and PCM. At present he is investigating the field of optical loss measurement.

Irwin W. Sandberg, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1958—.

He has been concerned with analysis of military systems, particularly radar systems, and with synthesis and analysis of active and time-varying networks. He is currently involved in a study of the signal-theoretic properties of nonlinear systems. Member, IEEE, Society for Industrial and Applied Mathematics, Eta Kappa Nu, Sigma Xi and Tau Beta Pi.

Robert A. Schmeltzer, B.S., 1955, The Cooper Union; M.S., 1958, The Stevens Institute of Technology; Eng.Sc.D. 1962, New York University; Bell Telephone Laboratories, 1963—. He has been engaged in theoretical research in low-loss guided wave systems for long distance optical transmission and lasers. Member, American Mathematical Society, IEEE and Eta Kappa Nu.

Lewis C. Thomas, B.E.E., Cornell University, 1949, M.S. in E.E., c.s.l., 1958, Newark College of Engineering; Bell Telephone Laboratories, 1949—. He has worked on the Nike missile systems, pulse code modulation systems, and data transmission systems. His recent work has included attitude and orbital mechanics studies for Project Telstar and communication satellite system studies. He is on the lecture staff of the American Museum–Hayden Planetarium. Member, IEEE, A.I.A.A., Royal Astronomical Society of Canada and Eta Kappa Nu; honorary member, Epsilon Pi Tau.

B. S. T. J. BRIEFS

A Condition for the $\mathfrak{L}_\infty$-Stability of Feedback Systems Containing a Single Time-Varying Nonlinear Element

By I. W. SANDBERG

(Manuscript received May 14, 1964)

In the automatic control literature, a feedback system is frequently said to be stable if, regardless of the initial state of the system, each bounded input applied at $t = 0$ produces a bounded output. The purpose of this brief is to present a sufficient condition for the feedback system of Fig. 1 to be stable in this sense.

Our discussion is restricted to cases in which g_1, f, u, and v (in Fig. 1) denote real-valued measurable functions of t defined for $t \geqq 0$. The block labeled ψ is assumed to represent a memoryless time-varying element that introduces the constraint $u(t) = \psi[f(t),t]$, in which $\psi(x,t)$ is a function of x and t with the properties that $\psi(0,t) = 0$ for $t \geqq 0$ and there exist a positive constant β and a real constant α such that

$$\alpha \leqq \frac{\psi(x,t)}{x} \leqq \beta, \qquad t \geqq 0$$

for all real $x \neq 0$.

The block labeled **K** represents the linear time-invariant portion of the forward path, and is assumed to introduce the constraint

$$v(t) = \int_0^t k(t - \tau)u(\tau)d\tau - g_2(t), \qquad t \geqq 0 \tag{1}$$

in which k and g_2 are real-valued measurable functions such that

$$\int_0^\infty |k(t)|\,dt < \infty, \qquad \sup_{t \geqq 0} |g_2(t)| < \infty.$$

The function g_2 takes into account the initial conditions at $t = 0$. We do not require that u and v be related by a differential equation (or by a system of differential equations).

Assumption: We shall assume that the response $v(t)$ is well defined and such that for all finite $y > 0$

$$\sup_{0 \leqq t \leqq y} |v(t)| < \infty,$$

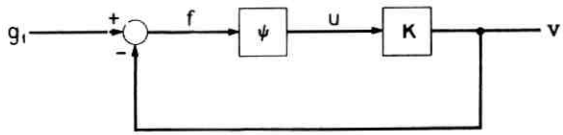


Fig. 1 — Nonlinear feedback system.

for each initial-condition function g_2 that meets the conditions stated above and each input g_1 such that

$$\sup_{t \geqq 0} | g_1(t) | < \infty .$$

Definition: We shall say that the feedback system of Fig. 1 is "$\mathfrak{L}_\infty$-stable" if and only if there exists a positive constant ρ with the property that the response v satisfies

$$\sup_{t \geqq 0} | v(t) | \leqq \rho \sup_{t \geqq 0} | g_1(t) + g_2(t) | + \sup_{t \geqq 0} | g_2(t) |$$

for every initial-condition function g_2 that meets the conditions stated above and every input g_1 such that

$$\sup_{t \geqq 0} | g_1(t) | < \infty .$$

Clearly, if the system is $\mathfrak{L}_\infty$-stable, then the response is bounded whenever the input is bounded.

Theorem: Suppose that

(*i*) *with* $K(s) = \int_0^\infty k(t)e^{-st}dt$ *for* $Re[s] \geqq 0$,

$1 + \frac{1}{2}(\alpha + \beta)K(s) \neq 0$ *for* $Re[s] \geqq 0$, *and*

(*ii*) *with* $h(\cdot)$ *the inverse Laplace transform of*

$$\frac{K(s)}{1 + \frac{1}{2}(\alpha + \beta)K(s)},^*$$

$$\tfrac{1}{2}(\beta - \alpha) \int_0^\infty | h(t) | dt < 1.$$

Then the feedback system of Fig. 1 is $\mathfrak{L}_\infty$*-stable.*

It is of interest to note that condition (*ii*) is satisfied whenever condition (*i*) is met, $\alpha > 0$, $h(t) \geqq 0$ for $t \geqq 0$, and

$$\int_0^\infty k(t)\, dt > 0,$$

* Condition (*i*) and our assumption regarding $k(\cdot)$ imply that $h(\cdot)$ exists, and that its modulus is integrable on $[0,\infty)$ [see Ref. 1].

for then

$$\tfrac{1}{2}(\beta - \alpha) \int_0^\infty |h(t)|\, dt = \frac{\tfrac{1}{2}(\beta - \alpha)K(0)}{1 + \tfrac{1}{2}(\alpha + \beta)K(0)} < 1.$$

The theorem can be proved with the techniques discussed in Refs. 2 and 3. More specifically, consider the relation between f and $(g_1 + g_2)$:

$$g_1(t) + g_2(t) = f(t) + \int_0^t k(t - \tau)\psi[f(\tau),\tau]d\tau, \qquad t \geqq 0 \qquad (2)$$

and suppose that

$$\sup_{t \geqq 0} |g_1(t) + g_2(t)| < \infty.$$

Arguments very similar to those used to prove Theorem 1 of Ref. 2 and Theorem I of Ref. 3 show that if (a) f satisfies (2), (b)

$$\sup_{0 \leqq t \leqq y} |f(t)| < \infty$$

for all finite $y > 0$, and (c) conditions (i) and (ii) of our theorem are satisfied, then there exists a positive constant ρ_1 (which does not depend upon g_1 or g_2) such that

$$\sup_{t \geqq 0} |f(t)| \leqq \rho_1 \sup_{t \geqq 0} |g_1(t) + g_2(t)|.$$

Our theorem is a direct consequence of this fact, in view of (1) and the relation between u and f.

REFERENCES

1. Paley, R. E., and Wiener, N., Fourier Transforms in the Complex Domain, Publ. Am. Math. Soc., Providence, R. I., 1934, pp. 60–61.
2. Sandberg, I. W., On the $\mathcal{L}_2$-Boundedness of Solutions of Nonlinear Functional Equations, B.S.T.J., this issue, p. 1581.
3. Sandberg, I. W., Signal Distortion in Nonlinear Feedback Systems, B.S.T.J., **42**, Nov., 1963, p. 2533.

Thermal Gas Lens Measurements

By A. C. BECK

(Manuscript received June 12, 1964)

The refractive index of a gas is an inverse function of its temperature under isobaric conditions. Therefore a cool gas flowing into a heated tube will have a lower refractive index near the heated walls than in the center, and the combination becomes a convex lens which will focus a light beam transmitted through the tube.[1,2]

A simple arrangement was built to get some information about the behavior of such a device. It is sketched in Fig. 1. The gas flowed through a 5-inch long electrically heated brass tube with a $\frac{1}{4}$-inch inside diameter. The tube was mounted in a large polyfoam cylinder to reduce external heat losses. A single-mode light beam from a helium-neon gas laser oscillating at $\lambda = 0.63$ micron was collimated at a diameter of about $\frac{1}{8}$ inch with glass lenses and then sent along the axis of the heated tube, through which a gas was flowing. The light was intercepted on a screen about 10 to 20 feet away. When this system, acting like a convex lens, is placed in such a collimated beam, the light goes through a focus at the lens focal length, and then expands to form a much larger area of light on the screen. A Foucault knife edge cutting the beam at the focus was used for measuring the focal length of the lens.

The reciprocal of the lens focal length in meters (often called the focusing power or convergence of the lens, and usually expressed in diopters) is plotted as a function of wall temperature rise on Fig. 2. Measurements were made with air and with carbon dioxide at the indicated flow rates. Helium was found to give very small effects, as expected.[2]

It will be noted from Fig. 2 that the lens power increases more slowly with flow rates at the higher values. At somewhat higher flow rates it flattens off and stops increasing, but at these very high flow rates, gas turbulence sets in and laminar flow is no longer present, so accurate measurements become impossible. At the flow rates shown on Fig. 2, constant and steady light patterns and focal lengths are observed when heated gas emerging from the tube is kept out of the light beam. This indicates that when laminar flow exists, the gas is very stable in a tube as small as this one; therefore steady optical lens effects are obtained. At higher temperatures the lens power continues to increase, at least up to the temperatures obtainable with this equipment. At the highest lens powers there is an effect that appears like spherical aberration,

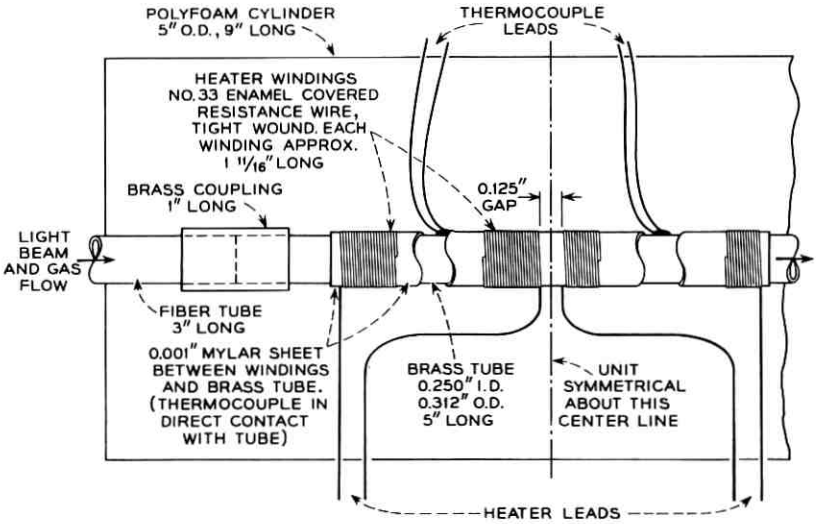


Fig. 1 — Experimental arrangement for thermal gas lens.

shown by a concentric pattern of light rings beginning to appear on the screen,[3,4] but this is very slight in the ranges shown on Fig. 2.

The brass tube was heated with direct current, and the input electrical power was measured for all points. The electrical power required with no gas flow was also measured as a function of temperature. The heat power taken away by the flowing gas was the difference between these two values at the measured temperature rise. From this information, a "figure of merit," defined as the reciprocal of the focal length in meters divided by the watts of heat extracted by the flowing gas (diopters per watt), can be calculated.

Representative data, and some comparisons with theoretical values, are summarized in Table I. The theory, notation, and gas constants are given by Marcuse and Miller.[2]

Agreement between calculated and measured values is fair for heat power consumption of the flowing gas, but not as good for focusing power, particularly at large flow rates of the heavy gas.

The focusing power and figure of merit of these lenses can be improved by using longer sections of tubing, since this one was not optimized. They may also be improved by using other gases having a higher $(n - 1)/k$ ratio, where n is the refractive index and k is the heat conductivity of the gas, and by operating at higher gas pressures.

Thermal gas lenses have short focal lengths with low power consumptions, and the effects are remarkably steady. The focusing power can

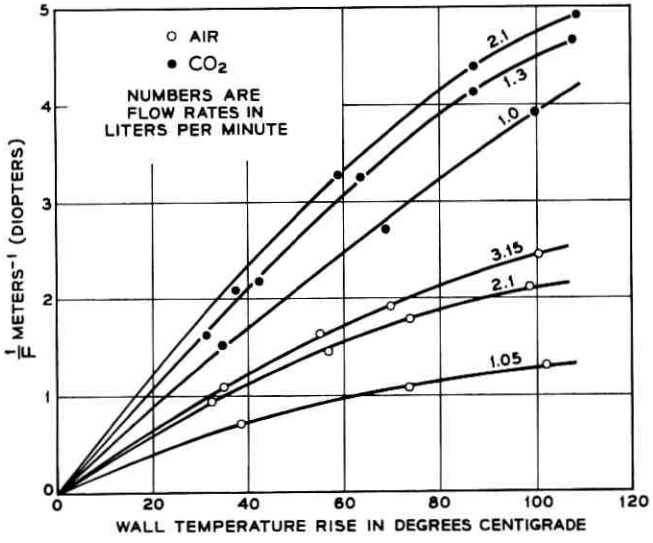


Fig. 2 — Lens focusing power versus wall temperature rise.

be adjusted easily by changing the gas flow rate. Such devices may find applications to long distance optical communication systems.

The assistance of W. E. Whitacre in constructing and setting up the equipment is gratefully acknowledged.

TABLE I — REPRESENTATIVE DATA AND THEORETICAL VALUES

Gas	Flow Rate (liters/min)	V (cm/sec)	v_0 (cm/sec)	v_0/V	1/F (diopters)		Power (watts)		Figure of merit (diopters/watt)	
					calc.	meas.	calc.	meas.	calc.	meas.
Air	1	27.24	105	3.87	1.67	1.3	1.74	1.8	0.96	0.72
Air	2	27.24	211	7.74	1.95	2.1	2.75	2.9	0.71	0.72
CO_2	1	13.52	105	7.80	3.00	3.8	1.73	2.0	1.73	1.90
CO_2	2	13.52	211	15.6	2.02	4.7	2.49	2.9	0.81	1.62

REFERENCES

1. Berreman, D. W., A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases, B.S.T.J., **43**, July, 1964, Part 1, p. 1469.
2. Marcuse, D., and Miller, S. E., Analysis of a Tubular Gas Lens, B.S.T.J., this issue, p. 1759.
3. Born, Max, and Wolf, Emil, *Principles of Optics*, Pergamon Press, 1959, pp. 472–477.
4. Beck, A. C., Gas Mixture Lens Measurements, B.S.T.J., this issue, p. 1821.

Gas Mixture Lens Measurements

By A. C. BECK

(Manuscript received June 15, 1964)

Different gases have different refractive indexes. If two different gases are introduced into different parts of a straight hollow tube and allowed to continue in laminar flow, they diffuse into each other slowly but remain sufficiently separated so that a refractive index gradient is maintained for some distance. It has been suggested by A. R. Hutson, G. E. Conklin, and the author that various optical components such as prisms and lenses can be obtained by using such gaseous structures. These components have very low losses and reflections, with no solid surfaces to cause matching or cleaning problems. For these reasons they may have important applications in optical communications systems.[1,2,3]

A simple structure, sketched in Fig. 1, was built to test this principle and to get some information about the performance of a gas mixture lens. A cylindrical porous-walled tube with an inside diameter of $\frac{1}{4}$ inch and an exposed length of 3 inches was mounted inside a larger phenolic cylinder with its ends blocked to form a gas reservoir. A single-mode light beam from a helium-neon gas laser oscillating at $\lambda = 0.63$ micron was collimated at a diameter of about $\frac{1}{8}$ inch with glass lenses and then sent along the axis of the porous tube through which one gas was flowing. The second gas was introduced into the tube through the porous walls. The light beam was then intercepted on a screen about 10 to 20 feet away. If there were no optical refractive effects on this beam, a spot about $\frac{1}{8}$ inch in diameter was seen at these distances, since the beam was collimated. A diverging lens increases the spot size without the beam going through a focus between the lens and the screen. A converging lens causes the beam to become smaller, go through a focus, then expand, forming a much larger spot on the screen at distances of many focal lengths. When measuring converging lenses, a Foucault knife edge cutting the beam was used to determine the focal point and thus the focal length of the lens.

With this system, diverging lenses are obtained if the gas of higher refractive index is introduced through the walls of the porous tube while the gas with the lower index flows down the tube parallel to the light beam. Conversely, sending the less refractive gas through the porous wall of the tube through which the more refractive gas and the light beam are passing gives a converging or positive lens. The reciprocal of the lens focal length in meters (often called the focusing power or con-

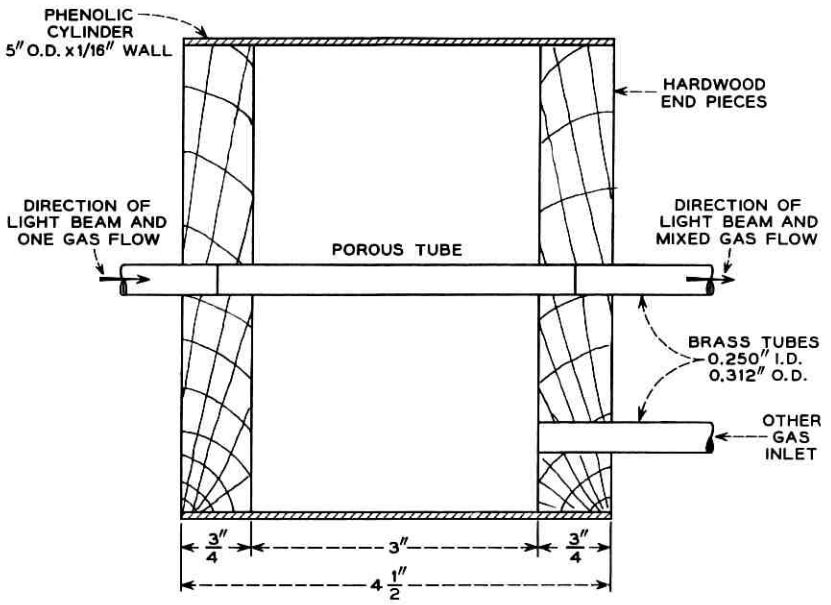


Fig. 1 — Experimental arrangement for gas mixture lens.

vergence of the lens, and usually expressed in diopters) is plotted for various gas flow rates in Figs. 2 and 3. Three different gas mixtures are shown, although other gas mixtures might have even greater focusing power. With carbon dioxide-helium and air-helium, focusing power increased with increased flow rate in the range of Fig. 2, but Fig. 3 shows that this is not the case for the carbon dioxide-air mixture in the range shown there, since focusing power at first increased, reached a maximum, then decreased as the carbon dioxide flow rate was increased. For these curves, Dextilose paper, No. 17V, was rolled up with about five thicknesses to form the porous tube. Other materials work equally well or better, porous ceramics, sintered metals, fritted glass and fine copper screen showing similar results.

At the indicated flow rates, the patterns and focal lengths are constant and steady when the mixed gases emerging from the tube are kept out of the light beam. This shows that the gas mixture is quite stable when laminar flow exists in the tube. The tube continued beyond the porous region exposed to the second gas for about $1\frac{1}{2}$ inches, but greater focusing power than that shown here was obtained with longer exit tube lengths. This indicates that the gases were still in laminar flow and were not completely mixed in this short length, so that focusing continued with

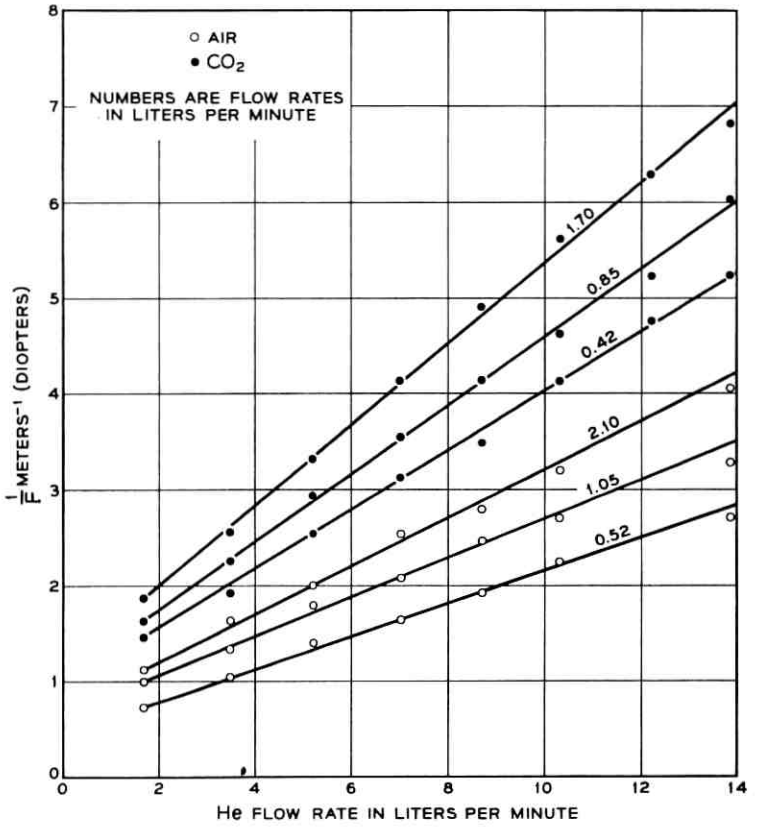


Fig. 2 — Lens focusing power versus gas flow rates, air-helium and carbon dioxide-helium.

more length. At higher flow rates, particularly of the gas first introduced into the system, turbulence appeared. This broke up the light beam and gave unsteady and flickering patterns on the screen.

Some photographs of patterns observed with gas mixture arrangements are reproduced on Fig. 4. At the upper left, the image of the beam is shown with no gases flowing in the tube. The diameter of this spot on a screen at any distance was about $\frac{1}{8}$ inch. At the upper right is the enlarged spot formed by carbon dioxide flowing down the tube and helium introduced through the porous tube walls at rather high flow rates so that a short focal length (approximately 6 inches) was obtained. This pattern, of course, increased in size as the screen was moved farther away. At lower flow rates, giving longer focal lengths, the light pattern is quite uniform, but as the focal length is made short, concentric rings

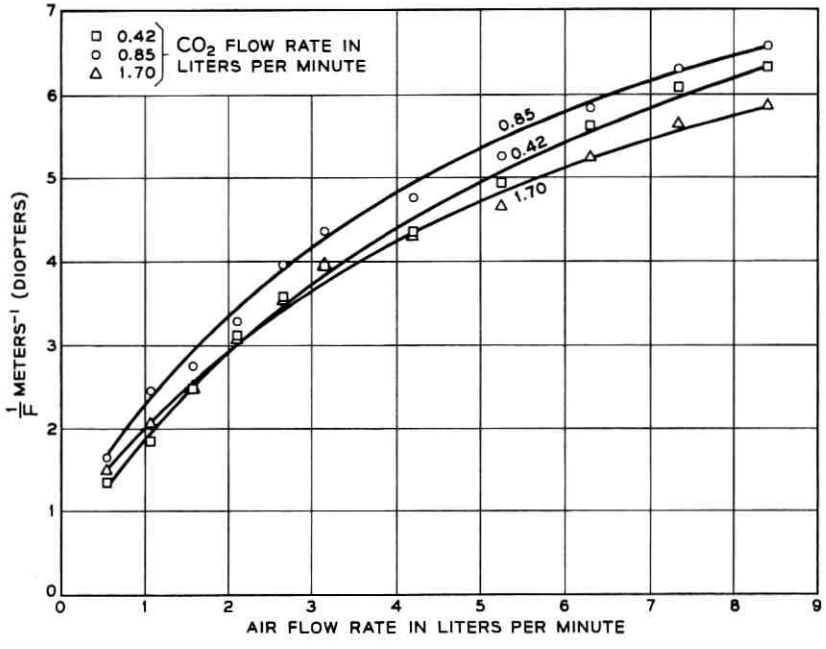


Fig. 3 — Lens focusing power versus gas flow rates, carbon dioxide-air.

of light begin to appear, as can just be seen here. These rings are similar to those which appear when spherical aberration forms a light interference pattern.[4] With a 3-inch porous tube, this effect is slight until very short focal lengths are produced. However, when the porous tube was replaced by two solid brass tubes in line with a gap of about $\frac{1}{10}$ inch between them at the center of the second gas reservoir, the light pattern shown in the photograph at the lower left was obtained. When the porous tube was replaced by a solid brass tube having six $\frac{1}{10}$-inch diameter holes drilled symmetrically every 60 degrees around its circumference in a plane located at the center of the second gas reservoir, the pattern at the lower right was observed. The use of tubes with different numbers of equally spaced holes produced patterns similar to this, having an n-fold symmetry where n is the number of holes. The steadiness and stability of these patterns of light interference are graphic indications of the stability of the gas flow, and the stability of gas lenses.

There are ways of separating the mixed gases, if desired, by permeable membranes or other means so that they can be mixed again in a later lens farther along a transmission system. Complete separation is not necessary to give lens action.

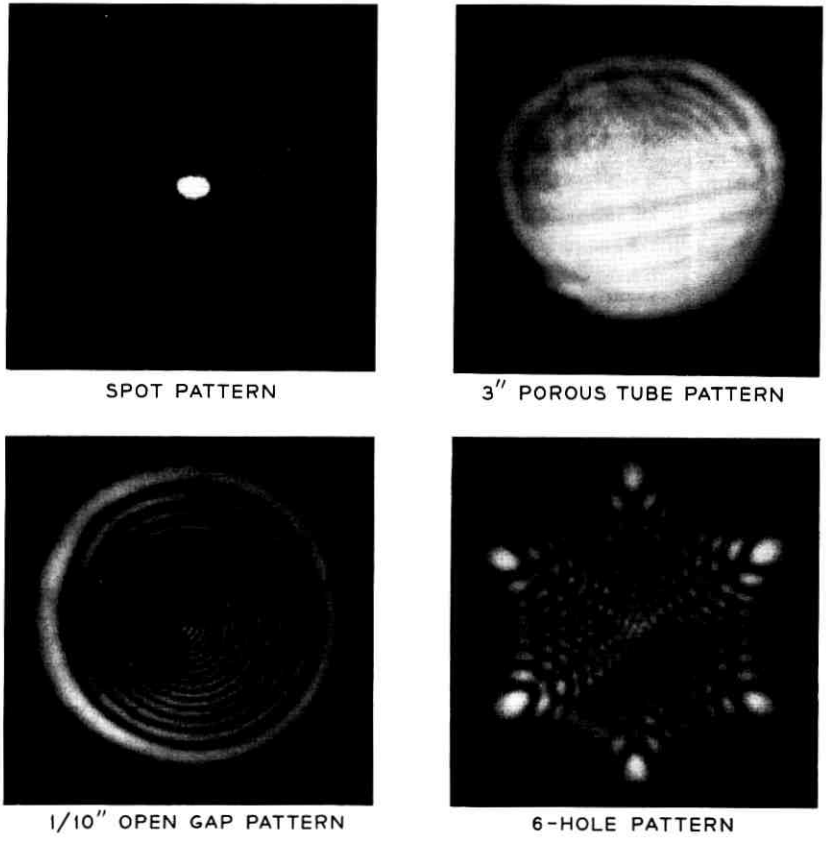


Fig. 4 — Observed patterns.

Gas mixture lenses can have short focal lengths with moderate gas flows. The effects are remarkably steady and constant. The focal length is easily changed by changing the gas flow rates. Such lenses may have possible uses in optical transmission systems.

The assistance of W. E. Whitacre in constructing and setting up the equipment is gratefully acknowledged.

REFERENCES

1. Berreman, D. W., A Lens or Light Guide Using Convectively Distorted Thermal Gradients in Gases, B.S.T.J., **43**, July, 1964, Part 1, p. 1469.
2. Berreman, D. W., A Gas Lens Using Unlike, Counter-Flowing Gases, B.S.T.J., **43**, July, 1964, Part 1, p. 1476.
3. Marcuse, D., and Miller, S. E., Analysis of a Tubular Gas Lens, B.S.T.J., this issue, p. 1759.
4. Born, Max, and Wolf, Emil, *Principles of Optics*, Pergamon Press, 1959, pp. 472–477.

A Circle Diagram for Optical Resonators

By J. P. GORDON

(Manuscript received June 16, 1964)

A graphical representation of the relationships between the parameters of Hermite Gaussian light beams has been introduced recently[1,2] by S. A. Collins, Jr. T. Li has pointed out[3] that Collins' chart has two equivalent forms. Collins' chart relates the spot radius and phase front curvature at any position on the beam to the position and spot radius of the beam waist. In this note we point out that a similar chart can be made which relates the curvature parameters[4,5] of any two phase fronts along a Gaussian beam and the spot radii on those phase fronts. The curvature parameters are defined as $g_1 = 1 - d/R_1$, $g_2 = 1 - d/R_2$, where R_1 and R_2 are the radii of curvature of the phase fronts, and d is their separation. This new chart directly relates mirror curvatures and spot radii in a spherical mirror resonator.

The equations on which the chart is based are[5,6]

$$\frac{g_1}{g_2} = \frac{w_2^2}{w_1^2} \tag{1}$$

$$w_1 w_2 = \frac{\lambda d}{\pi} (1 - g_1 g_2)^{-\frac{1}{2}}, \tag{2}$$

where w_1 and w_2 are the spot radii on the two phase fronts. If we eliminate w_2 and g_2, respectively, from these two equations we get

$$\left(g_1 - \frac{1}{2g_2}\right)^2 + \left(\frac{\lambda d}{\pi w_1^2}\right)^2 = \left(\frac{1}{2g_2}\right)^2 \tag{3}$$

and

$$\left(\frac{\lambda d}{\pi w_1^2} - \frac{\pi w_2^2}{2\lambda d}\right)^2 + g_1^2 = \left(\frac{\pi w_2^2}{2\lambda d}\right)^2. \tag{4}$$

On a graph whose Cartesian coordinates are $\lambda d/\pi w_1^2$ and g_1, these two equations represent circles of diameters $1/g_2$ and $(\pi w_2^2/\lambda d)$, respectively, as shown in Fig. 1. The point of intersection of the two circles gives the values of w_1 and g_1 which satisfy both (3) and (4). A more complete chart, similar to Collins' chart, can be used to read off spot radii in a resonator whose mirror curvatures are known, or to find mirror curvatures from measurements of the beam spots. As does Collins' chart, this new chart has two equivalent forms, which in our case differ only in having the subscripts 1 and 2 interchanged.

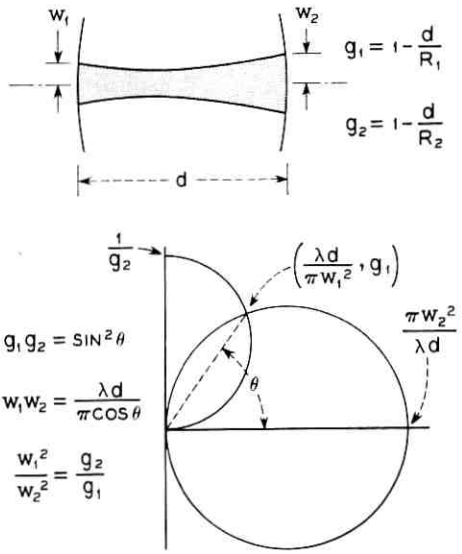


FIG. 1 — Circle diagram for spherical mirror resonators. Cartesian coordinates of the important points are indicated, along with some geometrical relationships.

REFERENCES

1. Collins, S. A., Jr., J. Opt. Soc. Am., **53,** 1963, p. 1339.
2. Collins, S. A., Jr., Appl. Opt., to be published.
3. Li, T., Appl. Opt., to be published.
4. Fox, A. G., and Li, T., Proc. IEEE, **51,** 1963, p. 80.
5. Boyd, G. D., and Kogelnik, H., B.S.T.J., **41,** July, 1962, p. 1347.
6. Gordon, J. P., and Kogelnik, H., B.S.T.J., to be published.

Gas Pumping in Continuously Operated Ion Lasers

By E. I. GORDON and E. F. LABUDA

(Manuscript received June 26, 1964)

Gas ion lasers[1] operate at discharge currents of several amperes in small-bore tubing. Under these conditions the discharge acts to pump gas from the cathode to the anode[2] and pressure differences in excess of 10:1 can be established in less than one minute of discharge operation. Since the optimum pressure range for laser operation is narrowly defined relative to the range of pressures existing in the discharge tube (see Fig. 1), laser action usually deteriorates or goes out shortly after turn-on.

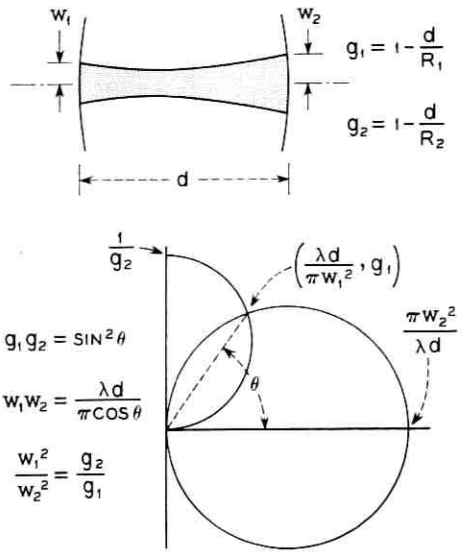


FIG. 1 — Circle diagram for spherical mirror resonators. Cartesian coordinates of the important points are indicated, along with some geometrical relationships.

REFERENCES

1. Collins, S. A., Jr., J. Opt. Soc. Am., **53**, 1963, p. 1339.
2. Collins, S. A., Jr., Appl. Opt., to be published.
3. Li, T., Appl. Opt., to be published.
4. Fox, A. G., and Li, T., Proc. IEEE, **51**, 1963, p. 80.
5. Boyd, G. D., and Kogelnik, H., B.S.T.J., **41**, July, 1962, p. 1347.
6. Gordon, J. P., and Kogelnik, H., B.S.T.J., to be published.

Gas Pumping in Continuously Operated Ion Lasers

By E. I. GORDON and E. F. LABUDA

(Manuscript received June 26, 1964)

Gas ion lasers[1] operate at discharge currents of several amperes in small-bore tubing. Under these conditions the discharge acts to pump gas from the cathode to the anode[2] and pressure differences in excess of 10:1 can be established in less than one minute of discharge operation. Since the optimum pressure range for laser operation is narrowly defined relative to the range of pressures existing in the discharge tube (see Fig. 1), laser action usually deteriorates or goes out shortly after turn-on.

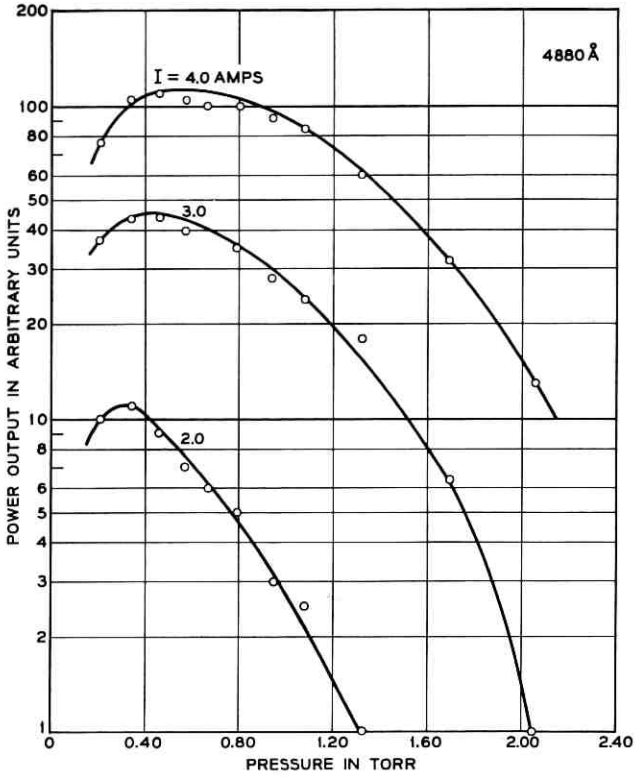


Fig. 1 — Laser output vs pressure at various discharge currents.

When the discharge is turned off for several seconds to allow the pressure to equalize and then turned on again, the output returns to its initial value only to deteriorate again.

By placing a connecting tube of high gas flow conductance between the anode and cathode as shown in Fig. 2, the pressure difference between the anode and cathode can be virtually eliminated. Tubes operated with the connecting tube show no deterioration over long periods of time. The tube length and bore are chosen so that the sustaining voltages for the two paths are comparable.

The connecting tube is fashioned in the form of a helix to relieve any strains that might develop from differential expansion and because a helix provides a convenient means of getting a long length of tubing into a small volume.

The connecting tube also serves to eliminate gas separation or cata-

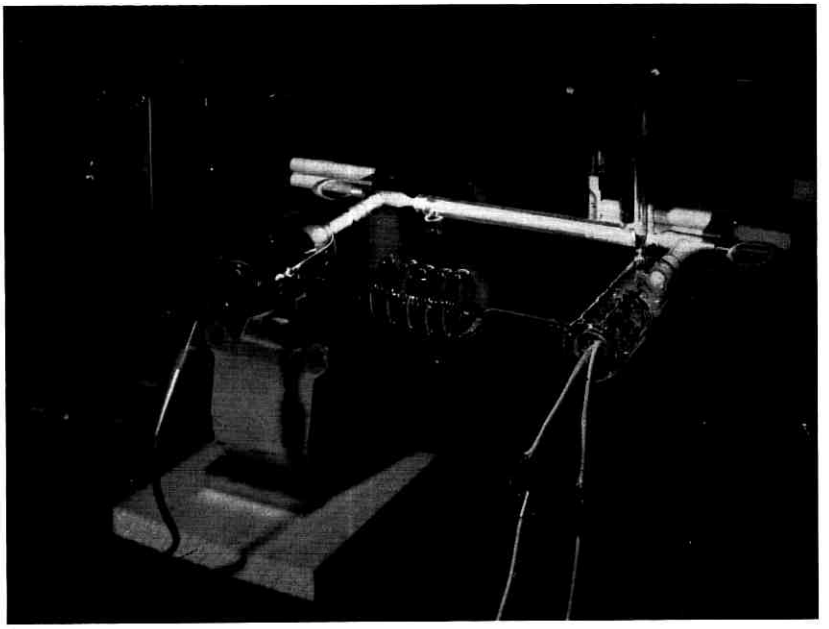


Fig. 2 — Photograph of laser illustrating the connecting tube.

phoresis that occurs when using mixed gases. This is of particular importance in dc-excited laser discharges such as helium-xenon.[3] Cataphoresis is no special problem in the helium-neon discharges in the current range at which they are operated, and the side tube is unnecessary.

We are indebted to J. T. Bannon, who constructed the experimental tubes.

REFERENCES

1. Gordon, E. I., Labuda, E. F., and Bridges, W. B., Continuous Visible Laser Action in Singly Ionized Argon, Krypton and Xenon, Appl. Phys. Lett., **4**, May 15, 1964, p. 178.
2. Francis, Gordon, *Handbuch Der Physik*, XXII, ed. S. Flugge, p. 198.
3. Bridges, W. B., High Optical Gain at 3.5μ in Pure Xenon, Appl. Phys. Lett., **3**, Aug. 1, 1963, p. 45. See also Faust, W. L., McFarlane, R. A., Patel, C. K. N., and Garrett, C. G. B., Gas Maser Spectroscopy in the Infrared, Appl. Phys. Lett., **1**, Dec., 1962, p. 4.