THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLIV MAY-JUNE 1965 NUMBER 5

The N2 Carrier Terminal — Objectives and Analysis

By R. C. BOYD and F. J. HERR

(Manuscript received March 10, 1965)

The N2 carrier terminal, a 12-channel, double-sideband, amplitude modulated multiplex, is the first of a new family of short-haul carrier facilities. The new carrier family is designed to take advantage of improvements possible with solid-state technology and to meet changing needs of the growing Bell System toll network. Increased demands on the short-haul trunks have resulted from the growth of voice and nonvoice services on the DDD network, tightening requirements such as those for over-all net loss variations between channels and gain variations across the channel frequency band. Maintenance and operational features of short-haul facilities also are affected by the increase in complexity and size of the Bell System carrier network: newer equipment must have fewer adjustments, and the design must provide built-in margins to permit longer maintenance intervals and more effective alarm and trunk processing features in case of failure. The new carrier family must compete economically with existing short-haul systems, yet work with in-place carrier systems and facilities.

The new terminal meets the above objectives. In addition, improvements in the performance of N-carrier repeatered lines are indicated as a result of the systems analysis of short-haul carrier systems which led to the requirements for the N2 terminal.

I. INTRODUCTION

A comprehensive development program is in progress to provide a new family of solid-state terminal and repeatered line equipment for

short-haul carrier systems. Each unit will meet modern requirements imposed by the long distance Bell System network for the transmission of voice, data, and other special services. The new short-haul family includes:

— a 12-channel double-sideband N2 terminal to replace the N1 terminal
— a 24-channel single-sideband N3 terminal to replace the ON2 terminal
— an N2 repeatered line and its adjuncts to replace the N1 repeatered line.

The members of the family are being designed so that they will be compatible with each other and with their existing short-haul counterparts. Thus N1, ON, N2, or N3 terminals can provide the frequency division multiplexing for the N1 or N2 repeatered carrier lines.

This paper describes the N2 terminal, which is the first member of the new family. Included are the reasons for its development and its important objectives and requirements, system characteristics, and performance. Included also is a discussion of the over-all N system made up of the N2 terminal and the N repeatered line. Companion papers will cover the N2 circuitry in detail and discuss the terminal equipment features.[1,2] The other members of the new short-haul carrier family will be described in subsequent series of papers.

The operating telephone companies began using N2 terminals at the end of 1962 to provide telephone and special-service channels in a number of locations. In 1964 over 9600 N2 terminals were produced by the Western Electric Company. The first use of N3 terminals began in late 1964. The N2 repeatered line equipment is being developed on schedules that will make initial production units available in quantity from Western Electric early in 1966.

II. BACKGROUND FOR NEW SHORT-HAUL CARRIER FAMILY

The new solid-state short-haul carrier family is being developed to meet the changing needs of the growing Bell System network. In the late 1940's Bell Laboratories engineers developed the first short-haul system, the type N1 carrier.[3,4] The N1 channels were designed to provide short-haul toll and exchange trunks up to 200 miles long and the toll connecting trunks used as the end links in a switched, multilink connection. Initially the N1 short-haul carrier system was designed to provide only voice transmission channels. Other channel arrangements were developed in the years that followed. In addition, a number of related systems, including O, ON and ON/K carrier, and the ON radio multiplex, were developed subsequently to meet comparable needs.[5]

As use of direct distance dialing (DDD) has grown in the Bell System, the increasing complexity of the hierarchal switching network with automatic alternate routing has in turn increased the number of trunks that may be switched in tandem. The number of short-haul carrier trunks in a typical switched connection has increased to as many as 4 or 5. In addition, the trunk layout has grown more complex. A short-haul channel may be wired in tandem with a long-haul carrier channel to make a single trunk. Finally, short-haul carrier channels may be a part of one or more of the links in a switched connection spanning intercontinental distances. The growing use of submarine cables and the potential use of satellites increases the likelihood of such connections.

In addition to its predominant use for voice channels, the Bell System network is being used to transmit a growing variety of non-voice services. One or several teletype or digital data circuits can be transmitted over a short-haul carrier channel of voice-frequency bandwidth (200–3000 cps). In addition, all or part of the 96-kc N-carrier band has been used to transmit wideband services, including high-speed digital data.

These changes in the Bell System environment have imposed more severe requirements on the individual short-haul channels than can be met by the performance capabilities designed into the original short-haul systems. In the switched message network, the possibility of more links in tandem has tightened the requirements on the performance of each link. The variations in over-all net loss of existing channels must be reduced materially. This applies both to long-term variations due to aging, changes in temperature, and office battery voltage variations and to short-term variations such as cyclic changes in net loss, called "beats." The bandwidth provided by a channel must be increased, and the variations in gain across the band reduced. The crosstalk coupling between channels within a system must be reduced by controlling the sources of coupling within the terminals and repeaters. The background noise contributions of the terminals and the repeatered line must be reduced.

The increasing number of links in tandem, carrying both voice and nonvoice services also has imposed new requirements on the control of impulse noise and envelope delay distortion. Consistent with economic limitations, both should be controlled in the basic terminals and also in the repeatered line. Where necessary, additional equalization may be provided as required for specific voice-band or wideband services.

These new requirements have been important in the design of the new N2 terminal. In addition, it will be necessary to correct deficiencies

in the performance of the existing short-haul carrier channels to meet as many of these up-to-date objectives as economically practical. At the end of 1961, the N1 and ON channels in service provided about 55 per cent of the carrier channels in the Bell System plant. Techniques developed for the N2 terminal have indicated improvements that are being incorporated into existing N1 and ON terminals, and this modernization program will continue.

The increase in complexity and sheer size of the Bell System carrier network has also focused attention on the maintenance and operational features of the N2 terminal. One important implication is that the design should have a minimum number of adjustments. Design margins have to be increased to permit longer intervals between adjustments. Also needed are more effective alarm and trunk processing features which operate when the system fails, including automatic restoral to service when the carrier system again has been made satisfactory for commercial service.

Finally, development of a modern family of short-haul carrier units will provide other benefits to Western Electric and the operating telephone companies. An up-to-date design taking advantage of new solid-state and ferrite components, modern design, and new manufacturing techniques will be easier to manufacture and operate in the long run than existing short-haul equipment.

III. BROAD OBJECTIVES

The basic objective for the new family of short-haul carrier systems is implied by the preceding section, which gave the reasons for undertaking the development program. That is, the new systems must meet the requirements imposed by the changing Bell System environment. In addition, however, there are certain other broad objectives which have had a major influence on the development of the new units.

One objective is that the members of the new short-haul family must be competitive in cost with their existing counterparts. Improvements in performance were expected to increase the basic costs of the equipment itself. These increases must be offset by savings in the first cost of dc power in central offices, due to the lower dc power drains of transistor circuitry vs electron tubes and the related use of lower, more economical central office battery voltages. Other savings can be expected from the smaller size of the equipment and from the substantially simpler engineering and installation and maintenance effort.

Another objective applying to all of the new carrier units is that each must be compatible with the transmission plan for existing short-

haul systems. The carrier frequencies, signal powers, and transmission levels must match. By doing so, it will be possible to operate the different terminals on repeatered lines in the same cable. The different repeaters must provide compatible carrier lines in a cable or work in tandem on a given line.

One important reservation has been applied to this desirable compatibility. An N2 terminal operated at one end of a repeatered line with an N1 terminal at the other end would restrict the performance improvements provided by the N2 terminal. Therefore, N1-N2 terminal compatibility was not taken as an objective for the project. This reservation was justified on two bases. One is the high rate of growth in new short haul carrier channels. The other is the anticipated low level of N1 terminal reassignments that might lead to the desire to have N1 and N2 terminals work together. The combined growth of short haul channels has been equal to or greater than 20 per cent per year, and in 1965 there will be equal numbers of N1 and N2 terminals in the plant.

A further objective is that the N2 terminal would not use the built-in, 3700-cycle signaling system designed for N1, O, and ON terminals. In its place, the existing, in-band, single-frequency (SF) signaling system should be used.[6] With this change, all of the signaling options of the modern Type E SF signaling system can be used without any changes in the N2 channels. The need for this change has grown out of the increasing complexity of the Bell System trunk network. Many of the trunks made up of two or more carrier channels wired in tandem now include one short-haul carrier channel. For a trunk including two channels in tandem, with in-band tones used for signaling, only one set of signaling terminals is required, with none at the junction of the two channels. This saving of in-band equipment offsets the lower costs that might have been achieved with built-in, out-of-band signaling.

The use of SF signaling with N2 channels has led to the need to integrate the system failure indications provided by the N2 terminal and the appropriate signaling information provided by the SF units to condition properly the derived trunks in the event of carrier failure. System failure alarms in the N2 terminal should be similar to those now a part of N1 and ON terminals. In addition, during a system failure, the N2 terminal and SF units must have features like those in N1 or ON terminals that permit the customer to disconnect himself or be disconnected. The processing equipment must then make the trunks appear busy so they will not be seized when they are not usable. In addition, an objective for the N2 terminal is that its system failure processing equipment should be able to restore the trunks automatically to an operating condition when the N2 system is properly restored to a condition suitable

for commercial service. This objective poses the dual requirements of having the N system, including the associated SF units, first test itself to determine when transmission becomes satisfactory after a failure and then provide signals to process each of the trunks so that it will be ready to accept calls again. These objectives imply that a universal trunk processing arrangement be provided for all E-type signaling units connected to E and M, loop, or revertive trunk circuits.

IV. DESCRIPTION OF N2 CARRIER SYSTEM

Although the subject of this paper is the N2 carrier terminal, this section and the next will deal more generally with the N2 carrier system. That is, they will cover both the N2 terminal and the N repeatered line. This broader perspective will make it possible to point out similarities and differences between the N1 and N2 system philosophies and requirements. The perspective will also help to clarify the transition from over-all system objectives to requirements for the N2 terminal and N repeatered line.

As in N1 systems, the N2 terminal is a 12-channel, double-sideband, amplitude-modulated multiplex for N-carrier repeatered lines. The voice-frequency channels derived by the N systems are used by the operating telephone companies to provide direct, toll-connecting, and intertoll trunks up to about 200 miles long on paired cables. The cables are of both low-capacitance (0.066 mf/mile) toll and high-capacitance (0.085 mf/mile) exchange types. As an indication of the range of system lengths in the N family, Fig. 1 shows two length distributions: one for systems multiplexed with N1 or N2 terminals and one for those multiplexed with ON or N3 terminals. The curves are based on samples of system lengths existing and estimated through 1965 by the Bell System operating companies.

4.1 *Frequency Allocation*

Fig. 2 shows the frequency allocation for the N1 and N2 carrier terminals. Twelve N2 carrier channel modems amplitude modulate the voice-frequency input signals into double sidebands along with their modulating carrier in the high group, generally using the channel 2 through 13 assignments. Channel 1 can be used as a spare, or as a replacement in case unavoidable carrier-frequency interference is present in another channel assignment. However, the use of channel 1 is not preferred due to the relatively high delay distortion encountered on long N repeatered lines. The high group is applied to the repeatered line or

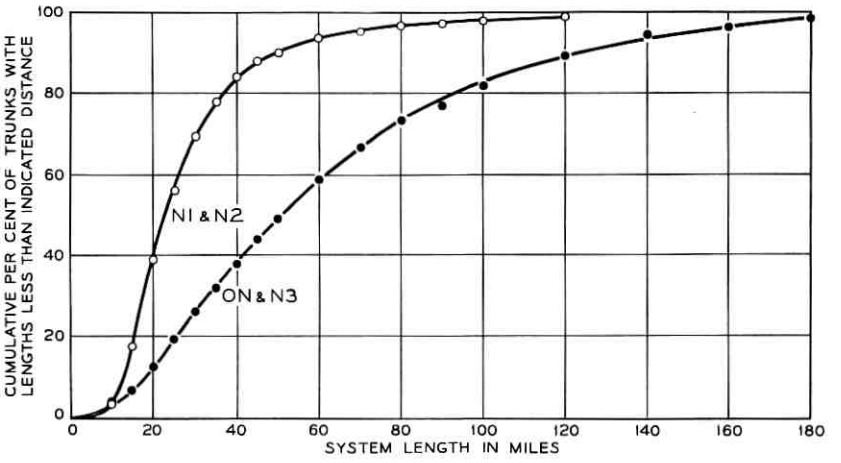

Fig. 1 — Distribution of lengths of typical short-haul carrier systems.

modulated into the low-group allocation, which is then applied to the line.

4.2 *N System Transmission Plan*

Fig. 3 shows one combination of transmitting and receiving groups. The figure illustrates that both groups are used in each repeater section, one for each direction of transmission. Fig. 3 also shows the frequency inversion, or "frogging," used at each repeater. The group received at the repeater is modulated with 304 kc into the other group and transmitted to the next repeater section. The frequency separation between the input and output groups at a terminal or repeater greatly simplifies the control of the effects of near-end crosstalk and of near-end interaction crosstalk between cable pairs adjacent to the repeaters. The frogging of the two groups also improves far-end crosstalk. For a given direction of transmission for the high-group frequencies where crosstalk is poorer, the total length of exposure is cut in half between parallel systems in the same cable.

Frequency frogging of the two groups also provides first-order equalization for the variation of line loss with frequency. The slope of line loss encountered, for example, across the high group band in one line section is nearly offset by the opposite slope in the loss of the low group in the next line section, after the group is inverted by the frogging repeater. This compensation and the slope equalization provided in the repeaters and terminals are discussed further later in this section.

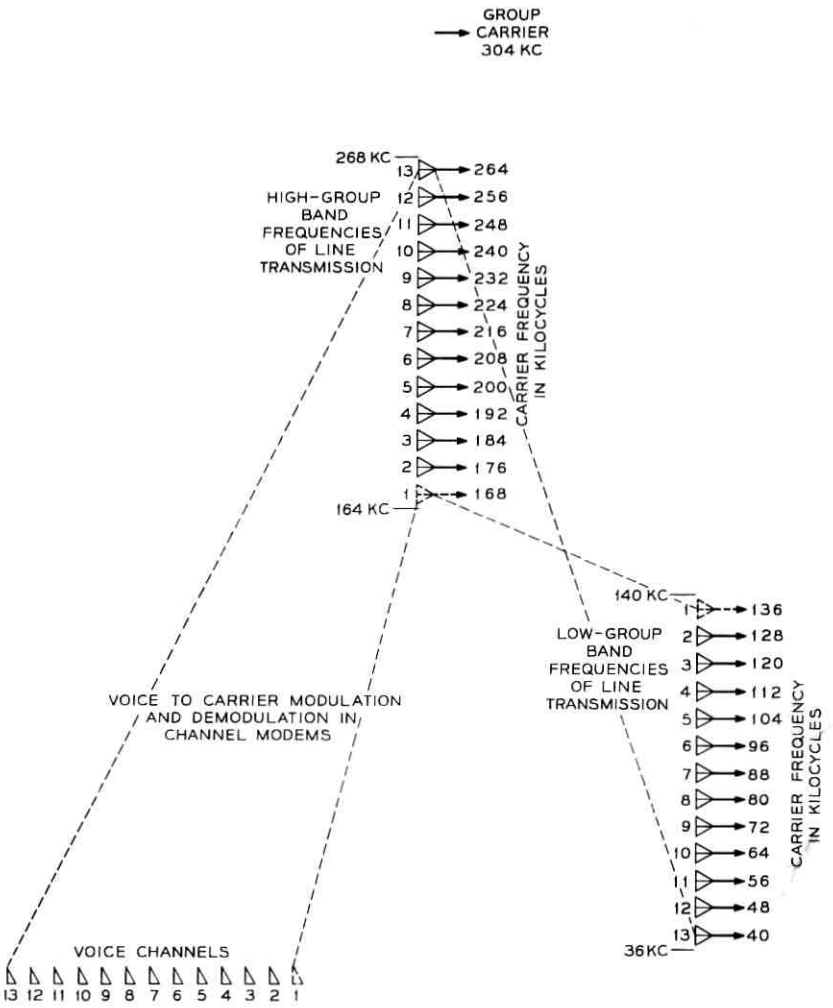

Fig. 2 — Frequency allocation for N terminals.

Fig. 4 summarizes the important transmission level points for the major functional transmission units in an N2 system. The values for message power on the line side of the channel modulator and demodulator are the power which would be measured in one of the sidebands of a carrier channel if a 0-dbm one thousand-cycle sine wave were applied at a 0-db system level (SL) point (see Appendix of Ref. 1) on the voice-frequency input leads. The message power on the line for that (compressed) tone is 12.5 db below the channel carrier power.

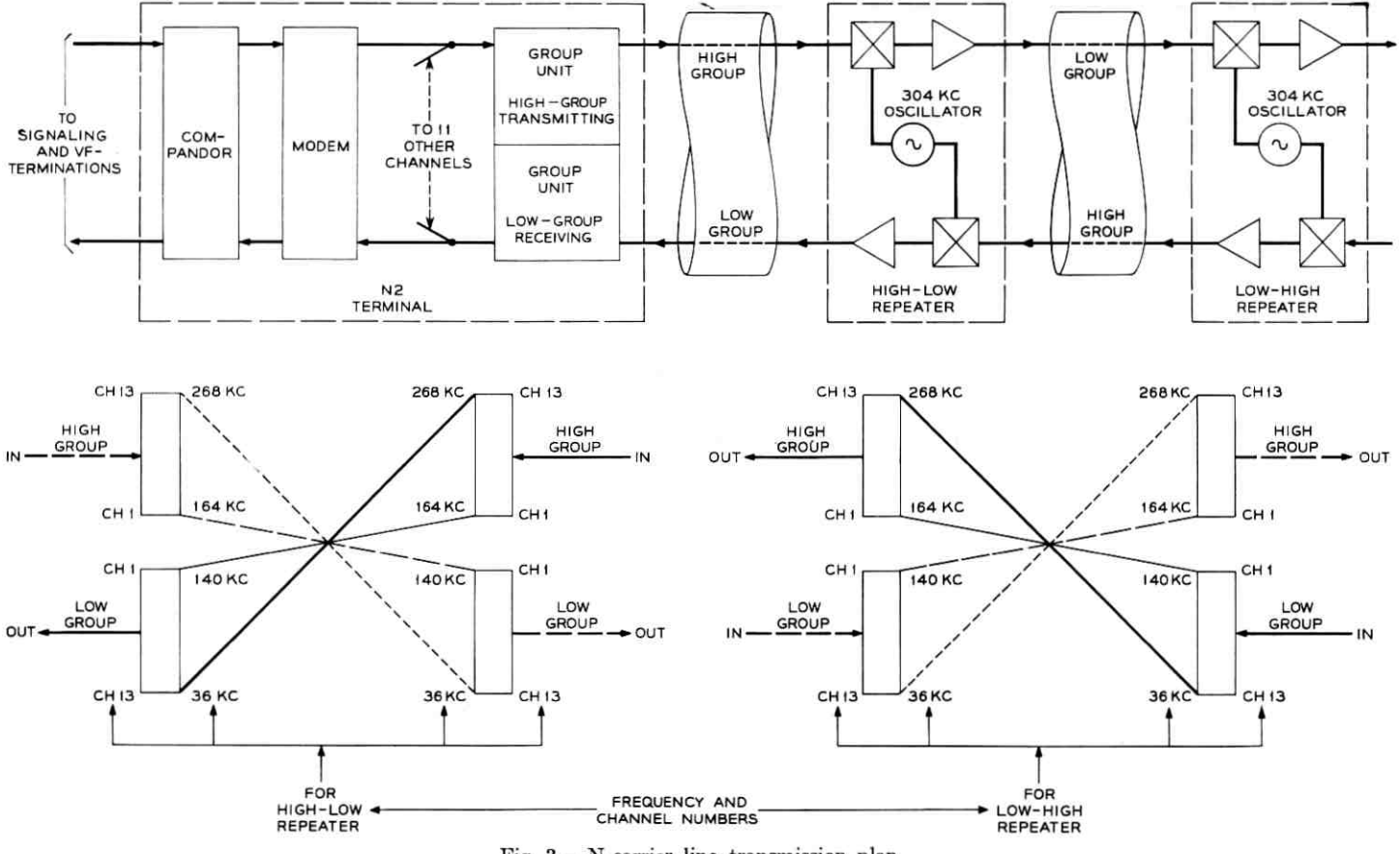

Fig. 3 — N-carrier line transmission plan.

In the earlier paper on the N1 system,[3] values were also given for "message level," which represented a fictitious set of levels in the terminal and on the line. They were powers for a 0-dbm at 0-SL one thousand-cycle input wave if it were not compressed in the compressor. The message levels could be used along with an appropriate expandor advantage to calculate the output power of noise or other interferences on N lines. The message level for an uncompressed 0-dbm at 0-SL tone in a single sideband in the carrier portion of the N system would be 15 db below the carrier. Message levels have been omitted from Fig. 4 for simplification.

The carrier-to-sideband ratios mentioned above are those used in both the N1 and N2 systems. This is necessary to simplify operation of N1 and N2 terminals in a telephone office and of N1 and N2 systems in the same cable. The ratios were chosen so that, for even the highest speech volumes, the index of modulation would not exceed 100 per cent.

The power of the channel carrier is used to define the transmission level of the channel signals in the carrier portion of the N system. Thus, the carrier must represent that level without significant interference from the sidebands. The over-all net gain of the channel is regulated by the receiving channel regulator, indicated as "REG" in the channel unit in Fig. 4, which operates on the amplitude of the received carrier signal. This is necessary because the receiving amplifiers in the receiving group units and their counterparts in the repeater all regulate the total group power at their outputs, which essentially consists of the channel carriers.

4.3 *N-Carrier Line Engineering and Equalization*

The carrier powers shown on the carrier line in Fig. 4 were set by the practical need to have N2 terminals work on N1 and N1A repeatered lines. The total powers of the high-group or low-group carriers transmitted to the carrier line are the same as those for the comparable groups at N1 terminals or the appropriate N repeaters. However, the slope of the carrier powers is used as a flexible parameter in engineering modern N lines, and the N2 terminal provides slope equalization in a much wider range than does the N1 terminal. Slope is defined for a terminal or repeater as the difference in power output, or gain, for Channel 13 with respect to Channel 2, being positive when Channel 13 has greater power or gain. The block shown in Fig. 4 for slope equalization in the N2 terminal is provided by plug-in networks available in 3-db steps from +9 to −9 db supplemented by mop-up switched equalization in three 1-db steps in the receiving terminal.

Fig. 4 — N2 carrier power level.

The slope of the carrier powers — and indeed the layout of the entire carrier line including line losses, repeater and terminal gains, and the use of equalizers — is aimed at controlling the slope and signal magnitudes of the channel carriers and the accompanying sidebands. That control has several objectives:

(1) to control over-all channel noise by keeping the channel carriers and sidebands above the noise present on the line

(2) to control crosstalk between systems by keeping the powers of each carrier in one system close to the powers of the corresponding carriers of other systems in the same cable

(3) to keep the slope linear across both sidebands of a given channel so that the two sidebands will add in the channel demodulator without distorting the channel gain-frequency response.

In meeting the above objectives, the philosophy used in engineering the N line has changed over the years. Originally, all N lines were laid out so that there would be a nominal 14-db total slope in the cable loss between repeaters. For shorter repeater sections, the loss was built out with flat loss span pads and the slope made up by one of the three steps of slope equalization built into the N1 and N1A repeaters. For quite short repeater sections, the pads were supplemented by artificial cable networks roughly simulating the loss and slope of one or two miles of 19-gauge low-capacitance cable. For any length repeater section, the transmitted signal was pre-equalized to a 7-db slope, one-half of the expected nominal slope of the cable loss in a repeater section. The pre-equalization was necessary so that the frequency-frogging at the repeaters could equalize adjacent line slopes. However, the pre- and post-equalization also helped minimize the noise produced by misalignment and imperfections in equalization.

The original layout philosophy has been changed for several reasons that arise from the need to control the interference that external impulse noise produces in both voice and nonvoice signals which must be transmitted over N systems. N repeater spacings have been shortened for repeater sections adjacent to telephone central offices which are sources of impulse noise. For 19-gauge cable, sections of 3 to 5 miles are now prevalent, compared to the 7- to 8-mile sections used in early N systems. Use of the original N line engineering philosophy for short sections required use of large numbers of sloped line buildout units. The only such buildout units generally available until recently were the one- or two-mile artificial cable networks. These networks do not provide a good match for the loss of modern polyethylene insulated cable (PIC). The mismatch shows up as a bulge in the carrier frequency characteristic

across an N group. The result penalizes the channels in the center of the group band. As a means of controlling impulse noise in sections with severe exposures, it is advantageous to have all carriers received at the same power at a terminal or repeater rather than with a slope across the group, thus not forcing a few of the channels farther down into the noise than the others.

All of the above influences resulted in a change to a new philosophy of N line engineering termed "natural-slope" engineering. The basis of the philosophy is that any value of slope within wide limits is permitted at repeater or terminal outputs. That is, the loss slope "naturally" contributed by the line section is acceptable and should not be supplemented by artificial lines to obtain the fixed 7-db slope previously required at all repeater outputs. This degree of freedom facilitates obtaining zero slope when required at terminals at locations of severe noise induction. Occasionally, when the range of slope adjustments within the repeater is inadequate, fixed slope equalizers may be required.

The examples of slopes shown in Fig. 4 illustrate a variety of slopes of line loss and carrier powers at terminal or repeater outputs. Actual values on practical N lines may differ very widely from those shown.

4.4 *N2 Terminal Levels*

Most of the signal levels in the N2 terminal shown in Fig. 4 were determined by noise and crosstalk considerations. The minimum levels were chosen as a compromise between levels high enough to keep above noise sources and low enough to keep from producing distortion in nonlinear circuits or circuit elements.

The noise sources in the terminal include first circuit noise in the amplifiers and external interference. An example of the latter which posed a problem was magnetic pick-up of 60-cps energy within the expandor. As shown in Fig. 4, a very low level for voice frequencies exists in the expandor at the output of the variolosser and input of the amplifier. Until proper shielding was provided for the expandor amplifier input transformer, the 60-cps interference substantially exceeded the noise objective for program channels.

The primary sources of nonlinear distortion within the voice-frequency portion of the terminal are the variable loss ("variolosser") elements in the compressor and expandor. In the carrier portion of the terminal, nonlinear distortion occurs in the group amplifier and group modulator and demodulator resulting in the generation of unwanted modulation products. Such distortion shows up as intermodulation crosstalk between channels in the group. The carrier levels were chosen so

that the intermodulation crosstalk arising in an N2 terminal does not contribute materially to the total interchannel crosstalk in an N system. Interchannel crosstalk in the terminal is primarily leakover from adjacent channels through the receiving channel band filters.

The maximum levels in the terminal were determined by two considerations. In the receiving channel portion of the terminal, enough carrier power was needed to ensure that the envelope detector operates properly as a demodulator. High carrier power at that point also minimized the dc gain needed in the channel regulator control loop. The carrier power needed at the regulator and demodulator, taking into account the in-band loss of the channel band filter, determined the output power of the amplifier in the receiving group unit.

The maximum terminal power is at the output of the transmitting group amplifier. It is desirable to transmit as much power as feasible from the terminal to the line to permit the maximum cable loss to be spanned in the face of external noise and interference. That power was limited by the transmission plan of the N line. However, even within that constraint, an amplifier transmitting a total of +12 dbm (rms) with the tight intermodulation requirements imposed by crosstalk objectives posed a design challenge. The slope equalizer was put ahead of the amplifier so that the amplifier output power would not be reduced before it was applied to the line.

4.5 *Terminal Net-Loss Stability*

As discussed in earlier sections, the N2 terminal must meet exacting requirements on its net-loss stability. The carrier-to-sideband ratio established in the transmitting terminal must be determined precisely and held constant over the life of the channel modem. In addition, the voice-frequency transmission levels must not vary significantly. Both precision and stability are provided by large amounts of feedback in the voice-frequency amplifiers and by a precision channel modulator. The carrier-to-sideband ratio, i.e., the index of modulation, is carefully controlled by fixed voice-frequency gain ahead of the modulator and by built-in bias on the modulator. The receiving terminal channel regulator provides very tight regulation of the carrier, unaffected by sideband power.

4.6 *Special-Service N2 Terminal Units*

In addition to the terminal units which provide the message channels described thus far, other channel options are available for use in the N2

terminal. A VF amplifier can replace the compandor unit when non-compandored operation is desired for voice-band data transmission on private lines. Two VF amplifier units can be used at the junction of two N2 compandored channels which are wired together to form one derived circuit. By that means, only one compressor and one expandor are needed at the ends of the circuit rather than using two complete compandors in tandem. A Schedule C and D program channel modem can be used to replace any of the normal message channel modems for channels 3 through 7 to meet the more stringent program requirements for the channel gain-frequency characteristic. Finally, plug-in modems can replace the plug-in units for channels 5 through 11 to provide a wideband channel to transmit 40.8-kilobit synchronous data over part of the N carrier frequency band.

4.7 *N2 Terminal Alarm Arrangements*

In the N2 terminal, a number of alarm functions are carried out automatically if either the carrier line or the terminal common equipment fail. The alarm arrangements:

(1) keep customers from being charged falsely for a call in progress when service has failed,

(2) keep the failed trunks from being seized by the switching machine, and

(3) restore the system automatically when the failed portion is made good so that the derived trunks can be used without delay.

The functional alarm units include the terminal alarm unit and an external carrier group alarm (CGA) panel through which pass all the derived voice channels as well as the supervisory signaling leads. The operation of the alarm system is described in a companion paper.[2]

A special feature of the alarm system is necessary for automatic restoration. When an N line fails, the many N repeaters in tandem will increase their gain as their built-in regulators regulate on the power of the noise and crosstalk present in the failed line. The receiving channel regulators do likewise. This regulation provides enough extra gain so that the total interference power is usually equal to the normal total carrier power. The excess gain is enough so that when carrier transmission is restored the net gain of the individual channels will be many db too high, often to the point where the trunks will sing. In addition, the channels will be very noisy. The many regulators in tandem will gradually return to their proper operating points, but only after that time has elapsed will the channels be suitable for service. Therefore, the alarm

system includes a means of assuring that transmission is satisfactory on the restored channels before they are returned to service.

4.8 *Summary of N System Features*

Table I summarizes a number of the features of N systems discussed in Section IV. It highlights features common to both N1 and N2 terminals. In addition, it points out those features of the N2 terminal which are an improvement over the N1 terminal. The quantitative value of some of those improvements are discussed further in Section VI.

TABLE I — SUMMARY OF N SYSTEM FEATURES

Features *Common* to N1 and N2 Terminals	N2 Terminal Features *Improved* over N1
Transmission plan	Wider channel gain-frequency response
Frequencies, powers, levels, and type of modulation	Smaller long- and short-term net-loss variations
Built-in compandors, and channel and group regulation	Closer to ideal compandor tracking
Frequency frogging of high and low groups	Greater channel overload capacity
Separate, single-frequency, in-band signaling (N1 also provides built-in out-of-band signaling)	Tighter channel and group regulation
Layout flexibility of repeatered line	Reduced intrasystem crosstalk
Repeater spacings up to 7 miles	Solid state components instead of electron tubes
Control of impulse noise	Built-in carrier failure alarm with automatic restoration
Power feed over line to remote repeaters	Built-in power supply for −48-volt operation with filtering and regulation for improved stability
Variety of plug-in units	In-service, hit-free switching of power unit (Not required for N1)
Message (compandored)	Simplified maintenance
Noncompandored	Packaged bay including signaling
Schedule C and D program	
Small, lightweight, portable test equipment	
Sets for in-service switching of group units	
Simple order wire and alarm system	

V. SYSTEM DESIGN ANALYSIS

5.1 *Over-All Transmission Requirements*

The terminals for the N2 carrier system have been designed to meet the performance objectives that apply for transmission of message, program, data and other special services on intertoll trunks. Table II summarizes the over-all transmission requirements on a four-wire voice-frequency channel which were established for this system.

TABLE II — N2 CARRIER SYSTEM: OVER-ALL TRANSMISSION REQUIREMENTS

Channel gain-frequency response (3-db points)	200–3400 cps
Net-loss stability	
Long-term stability vs time and temperature	σ = 0.5 db
Short-term variations from intrasystem sources ("beats")	0.1 db peak-to-peak
Compandor tracking and load capacity	±1.0 db for inputs +8 to −40 dbm0; ±2.0 db for inputs +10 to −52 dbm0
Compandor advantage (measured)*	28 db minimum 30 db average
Channel noise referred to 0 db SL	
Terminals	16 dbrnc
Over-all System:	
Measured	26 dbrnc
Effective	31 dbrnc
Channel crosstalk from intrasystem sources (equal level coupling loss)	70 db loss for all terminal sources
Channel intermodulation distortion	30 db average for one second- or third-order type product below either one of two simultaneous fundamentals of 0 dbm0

* The effective advantage is usually considered to be about 5 db less than the measured advantage due to the effect of increased noise present along with the syllabic speech spurts.

5.2 *Derived Transmission Requirements*

Based on over-all transmission objectives such as those given in Table II, analytical studies were made leading to the derivation of detailed technical requirements for the functional blocks of the N2 terminal. Since the overall performance of the N2 system is controlled by both the terminals and the repeatered line, it was necessary to develop requirements for the new N2 repeater as part of this study. The analysis involved the quantitative study and evaluation of all significant transmission and interference mechanisms, for the purpose of translating objectives into specific circuit design requirements (e.g., required filter suppression, modulator balance, repeater linearity or crosstalk). Table III lists some of the N2 circuit units for which detailed technical requirements were derived from system design studies.

5.3 *Example of System Analysis*

To illustrate a typical system analysis study, this section describes a problem concerned with the control of short-term net-loss variations, or beats, in short-haul carrier systems. The problem is important

because interference mechanisms causing beats of as much as several db characterized N1 and ON systems operating in the field. The following discussion describes the quantitative analysis of the important interference mechanisms producing beats, and shows how detailed technical requirements on the appropriate circuit blocks were derived. The analysis has been applied to the solution of the present beat problem in existing N1 repeatered lines, and to the design of circuits for the N2 carrier terminal and the new N2 repeater.

5.3.1 *Sources of Short-Term Net-Loss Variations (Beats)*

Nominal carrier and sideband power relationships of the transmitted signal may be modified by interferences that originate along the high-frequency line of an N2 system. Such interferences can cause beats, due to small frequency differences between the wanted and interfering signals. These frequency shifts occur because group frequencies in a given N system are not synchronized from one repeater section to the next, as they are frequency-frogged by a separate 304-kc oscillator in each repeater along an N line. It also follows that group frequencies are not synchronized among the different systems operating on a common cable.

Beat interference may originate from two sources: (1) intrasystem sources, which are subject to design control within the system itself, and (2) intersystem sources which are independent of internal system characteristics. To keep beats within desirable over-all limits, a total

TABLE III — DERIVED REQUIREMENTS AND INTERFERENCE CHARACTERISTICS

	Circuit Unit	Derived Requirement	Interference Characteristic
Terminals	receiving channel band filter	out-of-band suppression	system noise and interchannel crosstalk
	receiving channel low pass filter	out-of-band suppression	system noise and interchannel crosstalk
	group unit filters	out-of-band suppression	beats
Line repeaters	amplifier modulator output coupling networks	modulation performance and crosstalk between opposite directions of transmission	beats and intelligible crosstalk
	group bandpass and low-pass filters	out-of-band suppression	beats and system noise

TABLE IV — ALLOCATION OF 51-db S/I INTRASYSTEM LINE BEAT OBJECTIVE

Type of Interference	Allocation of S/I Objective
Crosstalk coupling between repeater halves	54 db
Low-high repeater image frequencies (344–432 kc)	58 db
Second- and third-order line repeater modulation products	58 db
Crosstalk coupling between transmit and receive terminal units	60 db
Total objective	51 db

high-frequency line objective of 0.1 db peak-to-peak allowable variation in transmitted carrier power was specified. Assuming that the total line objective was allocated equally between internal and external sources, the allowable peak-to-peak carrier variation for all intrasystem sources is 0.05 db. This can be translated to a signal-to-interference (S/I) ratio of 51 db for the high-frequency line. This also corresponds to a voice-frequency variation of 0.1 db peak-to-peak at the channel output because of the doubling action of the N2 expandor. The allocation of 51 db to external sources is in keeping with objectives for intelligible crosstalk from other short-haul systems in the same cable.

Within an N2 system, there are five independent mechanisms that can produce beats. Of these, four are allocated a significant share of the 51-db objective. The fifth mechanism, input leak through low-high repeaters, is so easily controlled through group filter suppression that it can be assigned a negligible portion of the 51-db requirement. The allocation of this requirement among the four significant sources is given in Table IV, assuming independent mechanisms which add in random phase. The allocation shown reflects the relative difficulty of controlling each type of interference.

5.3.2 *Derivation of Technical Requirements*

Based on the allocations given above, specific technical requirements were derived for individual parts of the system. In the allocation diagram of Fig. 5, for example, the repeater crosstalk objective of 54 db is equally divided among 40 repeaters of a long system to derive a per-repeater crosstalk objective of $54 + 10 \log 40 = 70$ db.

Similarly, the 58-db allocation to the low-high repeater image band leads to an allocation of $58 + 10 \log 20 = 71$ db per pair of repeaters. Meeting this objective assures that frequencies in the unwanted image band from 344 to 432 kc generated in the low-high repeater are ade-

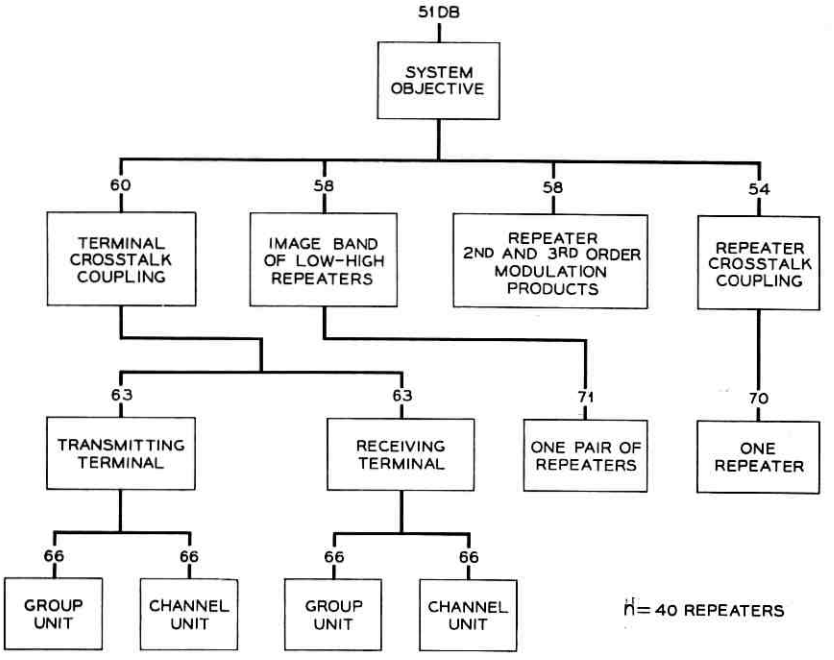

Fig. 5 — Allocation of beat objectives.

quately suppressed before reaching the next repeater, where modulation of the image band causes it to disturb the low-group band at the modulator output of the high-low repeater. From the stated 71-db objective, image band suppression requirements on the two filters which control this interference mechanism are readily derived. Fig. 6 shows minimum suppression requirements in the low-high repeater image frequency range which the low-high repeater output filter and the high-low repeater input filter taken together must provide.

This figure also shows the required combined suppression from the same pair of filters over the low-high repeater leak-through range of 36–140 kc. In this instance, an unwanted band of signals at low-group frequencies "leaks through" the low-high repeater modulator, is transmitted over the line and in turn leaks through the following high-low repeater modulator, where it disturbs the wanted low-group band of frequencies.

Similar image and leak-through bands of frequencies occur in the high-low repeater. Suppression requirements in the 468- to 572-kc

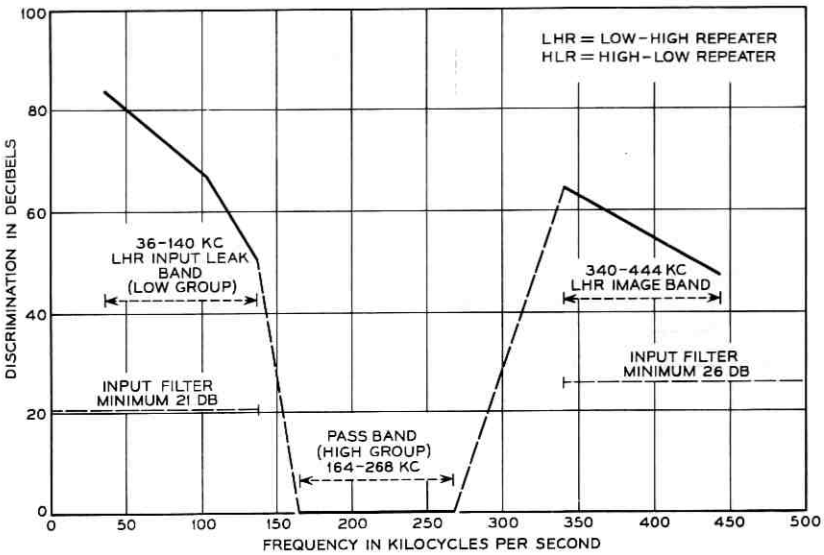

Fig. 6 — Minimum discrimination requirements of HL repeater input plus LH repeater output filters.

image band and the 164- to 268-kc leak-through bands necessary to reduce to acceptable limits the interferences from these sources are shown in Fig. 7.

The third item in the allocation table above provides for beats due to modulation products generated along the repeatered line. On the basis of a detailed analysis of all second- and third-order products generated in a 40-repeater system, it was concluded that significant improvement is needed in repeater linearity over that of the N1 repeaters to meet the beat objective of 58 db allocated to modulation sources. Table V summarizes these conclusions, indicates the required N2 repeater modulation coefficients, and compares them with the corresponding N1 performance.

Terminal crosstalk coupling, the final item in the allocation table, is allocated an objective of 60 db. This relatively severe objective reflects the fact that there are only two sets of terminal contributors per system and the expectation that terminal crosstalk should be more easily controlled than crosstalk along the repeatered line. As Fig. 5 shows, each terminal is allocated one half of the terminal objective, or 63 db. A further suballocation of 66 db is made within each terminal to each channel unit and group unit, respectively.

To summarize this illustrative example of the system analysis, in-

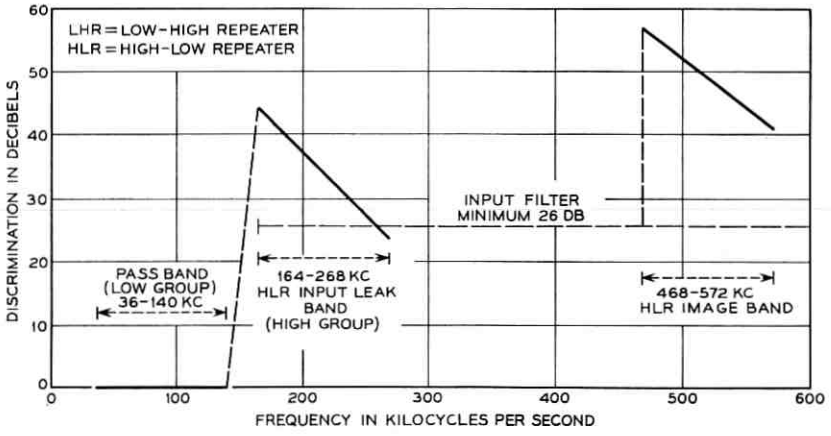

Fig. 7 — Minimum discrimination requirements of LH repeater input plus HL repeater output filters.

dividual circuits which meet the derived transmission objectives described above will assure that short-term net-loss variations from internal system sources will meet the over-all line beat objective of 51 db for systems of up to 40 repeaters. As a result, short-term net-loss stability in N2 systems will be governed by external sources of interference.

5.3.3 *Application of System Analysis of "Beats" to Type N Repeatered Lines*

The analysis described in the preceding sections has been applied to the problem of beats in existing N repeatered lines. On the basis of this analysis, it was possible to identify the dominant source of intrasystem beats in these systems as unwanted low-high repeater image band energy. The next step was to specify the design of a new high-low repeater input filter to control beats generated by this mechanism. Field tests of the new design demonstrated that the new filter virtually eliminated all intrasystem beats from this source generated on an N-repeatered line. The new filter is included in all current N1 and N1A repeater production, and is being installed widely on a replacement basis on existing N repeatered lines.

VI. N2 TERMINAL PERFORMANCE

This section summarizes the measured over-all performance of N2 terminals connected on a back-to-back basis. The performance is given

TABLE V — COMPARISON OF N1 AND N2 REPEATER MODULATION COEFFICIENTS

Product Magnitudes are Referred to 0 dbm Fundamental Power at Repeater Output.

Repeater	N1	Required N2
L-H	M2A = −75 dbm	M2A = −91 dbm
L-H	M2A-B = −99 dbm	M2A-B = −99.4 dbm
H-L	M2A = −71.5 dbm	M2A = −84.5 dbm
H-L	M2A-B = −86 dbm	M2A-B = −87.4 dbm

in terms of the characteristics of the over-all derived four-wire voice-frequency channel. Included are results of measurements of the channel gain frequency characteristic, compandor tracking, channel intermodulation, noise, crosstalk, and carrier failure arrangements. More detailed information on the performance of the functional circuit units is given in a companion paper.[1]

6.1 *Channel Frequency Characteristics*

Fig. 8 shows the gain frequency characteristics for typical N2 channels. The curve labeled "average" represents typical performance of N2 channels using the channel band filters incorporating quartz crystals. These filters have been in production since the fall of 1964. They replace an earlier design which was based on high-Q ferrite inductors. The performance of the channels with both types of filters is summarized in Table VI, which compares the frequencies at which the average channel gain is 3 db down from the gain at 1000 cps for N2 channels and for the A5 channel banks used to provide Bell System long-haul channels. It is expected that even the worst N2 channels with ferrite filters will provide a 3-db point at the high end of the channel characteristic whose frequency is greater than 3000 cycles.

A typical envelope delay distortion characteristic for N2 channels is shown in Fig. 9. The curve applies to channels with crystal filters. The bar values represent the standard deviation of the values measured at the indicated frequencies. The points represent average values for N2 channels with ferrite filters.

6.2 *Channel Net-Loss Stability*

Measurements have been made of the over-all net loss of N2 channels provided by two back-to-back N2 terminals installed at the Bell

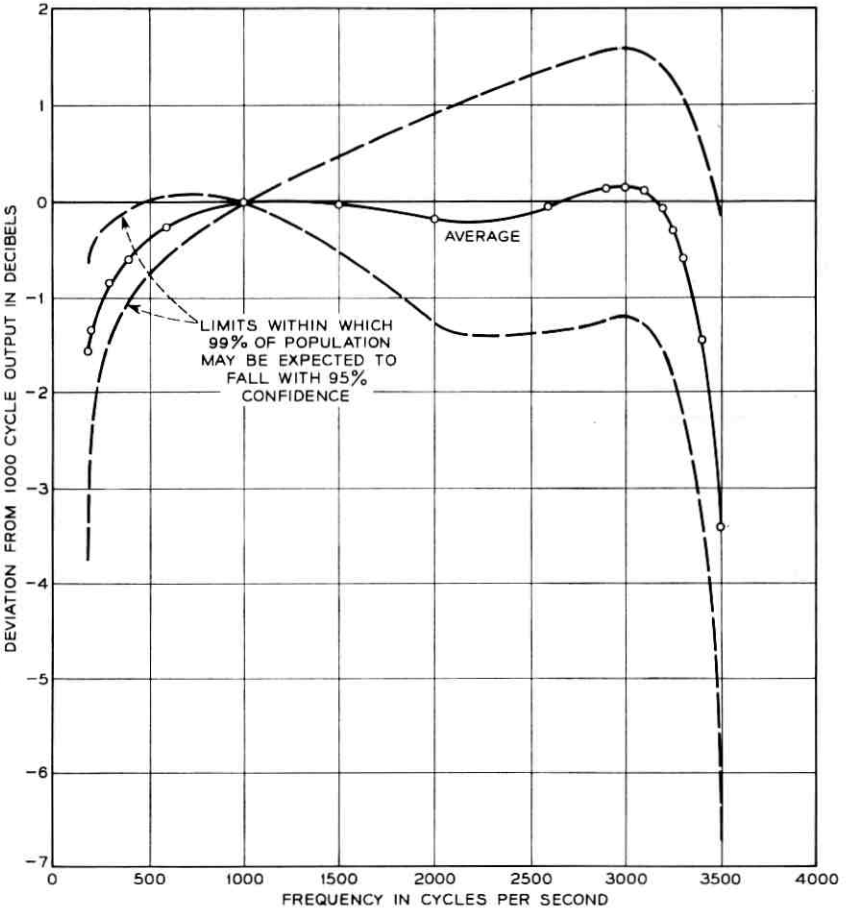

Fig. 8 — Typical over-all channel frequency characteristics provided by N2 terminals.

Laboratories field trial site in Virginia. (The over-all channel actually provides a net gain of 23 db between the −16-db input system level point and the +7-db output system level point. However, stability of the transmission of the channel is generally referred to in terms of the net loss of the derived trunk. Thus loss is used in this and following sections for convenience of association.) After six months without adjustment of over-all loss, the distribution of channel net losses had an average which deviated from nominal loss by 0.1 db and a standard devia-

TABLE VI — CHANNEL GAIN-FREQUENCY CHARACTERISTIC

Type of Channel	Frequencies in cps for Average Gain 3-db Down	
	Lower End	Upper End
N2 with crystal filters	100	3480
N2 with ferrite filters	100	3200
A5 channel bank	140	3370

tion of 0.3 db. Variation of net loss with changes in ambient temperature would increase those values by only small amounts.

6.3 *Compandor Tracking Performance*

The compressor and expandor which make up the compandor in an N2 channel must track each other, both statically and dynamically. The over-all input-output static characteristic of the compandor must minimize the changes in over-all channel net loss introduced for different powers of input signal within the design range. The input-output tracking of compressor and expandor separately and the dynamic tracking performance of the N2 compandor for step changes in input power are discussed in a companion paper.[1] Regarding the dynamic tracking, suffice it to say here that the N2 compandor has essentially no effect on the wave shape of signals for step changes in the input signal power.

The over-all compandor tracking performance for N2 channels versus input power is summarized in Fig. 10. The center curve represents the average deviations from the channel net gain measured for an input power of 0 dbm at 0 system level. A perfect tracking characteristic would have no deviation for any input power up to the point at which the channel overloads. The upper and lower curves show the limits within which 99 per cent of the compandor tracking characteristics may be expected to fall. For perspective, the shaded areas show the limits for tracking characteristics given in Section 5.1 and also those being proposed by the CCITT* for compandors to be used on long intercontinental circuits. The overload characteristic of the compandors shown in Fig. 10 is comparable to the performance of A5 channel banks.

The compandor tracking performance was measured at room temperature (26°C). If either the compressor or the expandor is operated at higher temperatures, there will be changes in the deviations from nominal. These changes are summarized in Table VII for a typical channel.

* International Telegraph and Telephone Consultative Committee.

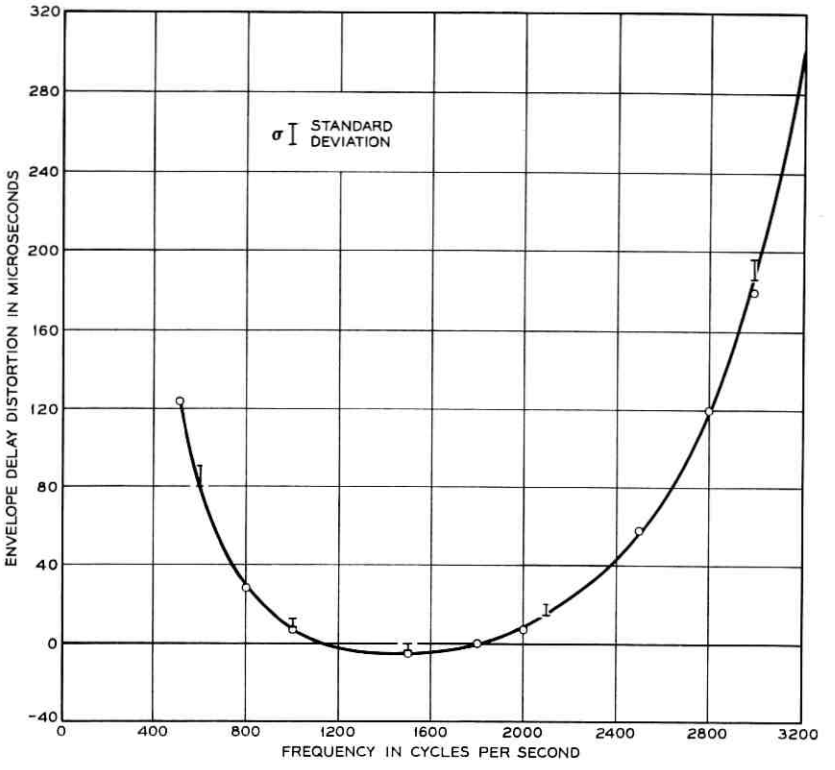

Fig. 9 — Typical envelope delay distortion characteristic provided by N2 channels.

These changes with temperature are not corrected by the channel regulator. However, they are small enough to allow the over-all channel to meet the requirements on tracking and net loss stability set forth in Section 5.1. The latter requirement may be interpreted as meaning that the distribution of net loss over a 6 month period for a given channel should have a standard deviation less than 0.5 db.

6.4 *Channel Distortion*

The distortion introduced by an N2 channel is an indication of the fidelity with which the channel can transmit voice signals, signaling tones such as multifrequency key pulsing, and voice-band data signals. The distortion can be characterized by measuring intermodulation products formed when two sine waves are transmitted over the

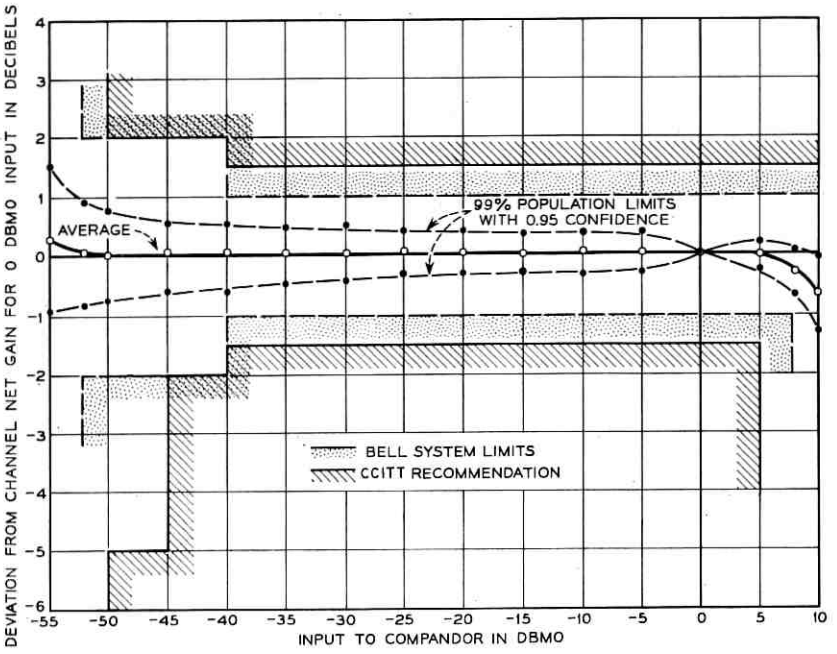

Fig. 10 — N2 carrier compandor tracking.

channel. The two frequencies used, 740 and 1250 cps, were chosen so that the important second-order products of the $f_1 \pm f_2$ type and third-order products of the $2f_1 \pm f_2$ type would fall within the transmitted voice band. Each of the two input signals was introduced at 0 dbm at 0 system level. The measured performance of recently manufactured compandors is summarized in Table VIII. Present indications are that the performance will be entirely satisfactory for transmission of one of the most critical signals, multifrequency key pulsing.

TABLE VII — COMPANDOR TRACKING VS TEMPERATURE

Deviations from Ideal Net Loss in db for Inputs from +8 to −52 dbm at 0 SL.

Temperature of		Max Deviation (Over Range of Input)	Max Change in Channel Net Loss (for 0 dbm at 0 SL Input)
Compressor	Expandor		
26°C	26°C	+0.2	0.0 (reference)
50°C	26°C	+0.4	0.0
26°C	50°C	+0.3	+0.2

TABLE VIII—DISTORTION FOR COMPANDORED N2 CHANNELS

Type of Product	Modulation Product — db Down on One Fundamental	
	Average	Worst
Second-order ($f_1 \pm f_2$)	44	35
Third-order ($2f_1 \pm f_2$)	36	31

6.5 *Channel Noise*

The back-to-back terminal noise performance of N2 channels is summarized in Table IX below. The performance of the channels was well within the objective of 16 dbrnC at 0 system level for the compandored channels. The noise measured on a noncompandored basis meets the objective of 48 dbrnC at 0 system level for voice-band special services provided on a private line basis.

An expandor advantage which averaged 30 db, with values ranging from 29.0 to 32.8 db, was provided by the N2 channels for noise from the carrier line and most of the carrier line terminal circuitry. The advantage is reduced by an average of 5 db when speech is present. The reduction results from the smaller expandor advantage during syllabic speech spurts and the consequently higher noise present during the speech spurt.

The background noise discussed thus far, however, is only part of the total noise which may be present in N2 channels. Babble noise will result from the multiplicity of small intrasystem crosstalk sources within the terminals, which will contribute both intelligible and unintelligible crosstalk. As a measure of such noise, several of the channels in an N2 terminal were loaded with white noise shaped to simulate speech, and the total noise was measured in the other channels. The total noise was separated into the noise power contributed by background noise and by crosstalk babble. Each of the contributions is summarized in Table X. The performance indicated in Table X con-

TABLE IX—BACKGROUND CHANNEL NOISE IN dbrnC AT 0 SL

Type of Channel	Channel Noise in dbrnC C at 0 SL	
	C Message Weighting	Flat Weighting
Compandored Channels	7–14	10–17
Noncompandored Channels (Private line data line-up)	25–27	26–37

TABLE X — MEDIAN CHANNEL NOISE IN dbrnC AT 0 SL DUE TO SYSTEM LOADING

(40 per cent of channels loaded with F1-shaped white noise at 88 dbrnC at 0 SL.)

Background noise	10
Babble noise	
in channels adjacent to loaded channels	13
in channels not adjacent to loaded channels	10

firmed the expectation that leak-over from adjacent channels through the receiving channel band filters would be the prime source of intra-system crosstalk. The value of 13 dbrn C at 0 system level implies an equal-level coupling loss of 75 db, which is 5 db better than the system design objective given in Section 5.1. It should be emphasized that this performance was achieved in the face of a stringent 40 per cent system loading with noise simulating loading of 0 vu at 0 system level in each active channel.

6.6 *Crosstalk*

The babble noise discussed in the preceding section is one form of crosstalk between channels within an N system which arises in the N2 terminal. Other forms of interchannel crosstalk were measured by using simulated speech as a disturbing signal in one channel and measuring total noise in the disturbed channel. The results are summarized in Table XI in terms of equal level coupling loss between channels. Even the worst coupling loss more than meets the 70-db system objective for crosstalk, with margin left for contributions from the repeatered line.

6.7 *Failure Alarm and Automatic Restoration*

Comprehensive tests of the N2 terminal in conjunction with its carrier group alarm (CGA) panel have demonstrated that the channels will provide satisfactory alarms and trunk processing in the event of transmission failure and satisfactory channel performance upon automatic restoration to service. The measured channel noise will come down

TABLE XI — FAR-END EQUAL LEVEL COUPLING LOSS IN db FOR WEIGHTED NOISE AS DISTURBER

Average	Range
77	74 to 84

from values which can exceed 90 dbnC at 0 system level during failure to values of about 30 dbrnC0 at 0 system level before the CGA circuitry will restore the channels to service.

VII. ACKNOWLEDGMENT

The authors wish to express their appreciation to the many people whose work and suggestions have made this paper possible. Particular thanks should go to C. W. Irby, T. W. Thatcher, Jr., and G. W. Bleisch for their contributions to Sections III, IV, and VI; to Miss F. C. Dunbar for her work, which is included in Section V; and to F. H. Blecher, A. J. Grossman, and D. D. Sagaser for their continued support.

REFERENCES

1. Lundry, W. R., and Willey, L. F., N2 Carrier Terminal — Circuit Design, B.S.T.J., this issue, p. 761.
2. Bell, D. T., Steiff, L. H., and Taylor, E. R., Equipment Aspects of Packaged N2 Carrier Terminals, B.S.T.J., this issue, p. 787.
3. Caruthers, R. S., The Type N1 Carrier Telephone System: Objectives and Terminal Characteristics, B.S.T.J., *30*, Jan., 1951, pp. 1–32.
4. Kahl, W. E., and Pedersen, L., Some Design Features of the N1 Carrier Telephone System, B.S.T.J., *30*, April, 1951, pp. 418–446.
5. Mahoney, J. Jr., and Osgood, D. T., New Uses for Short-Haul Carrier, Bell Laboratories Record, Feb., 1959, pp. 42–48.
6. Weaver, A., and Newell, N. A., In-Band Single-Frequency Signaling, B.S.T.J., *33*, Nov., 1954, p. 1381.

The N2 Carrier Terminal — Circuit Design

By W. R. LUNDRY and L. F. WILLEY

(Manuscript received December 7, 1964)

The N2 carrier circuit design takes advantage of solid-state technology to provide economies in power consumption and space requirements over earlier N-carrier equipment. At the same time, improvements in performance have been obtained in the areas of noise, signal distortion, and net loss stability. The last two are obtained through the extensive use of negative feedback in amplifiers and regulators and the use of variolosser diodes in the compandor which are stable over a wide operating range. Wider effective channel bandwidths have been attained. Where the terminal is used for nonvoice services, the resulting decrease in delay distortion is significant. Maintenance and testing on an in-service basis have been provided, including in-service switching of group units and power supplies.

I. INTRODUCTION

During the early 1950's there began to appear a series of carrier telephone systems intended for short-haul use. The first of these was N1,[1] a 12-channel double-sideband system for single-cable application. The rapid growth of the telephone plant resulted in large demands for these systems. Advances in the art, particularly those associated with the transistor and other solid-state devices, have made possible an improved design of the N system terminal. The circuits used in the new N2 terminal are described in this paper.

The N2 carrier system functions primarily as a connecting link over moderate distances for four-wire voice-frequency circuits. Basic considerations affecting the system layout are covered in a companion paper.[2] The principal features of the earlier N1 system are retained, including the use of compandors, double-sideband modulation and detection, and gain regulation on both a group and an individual channel basis, as well as provision for slope and flat gain adjustment of

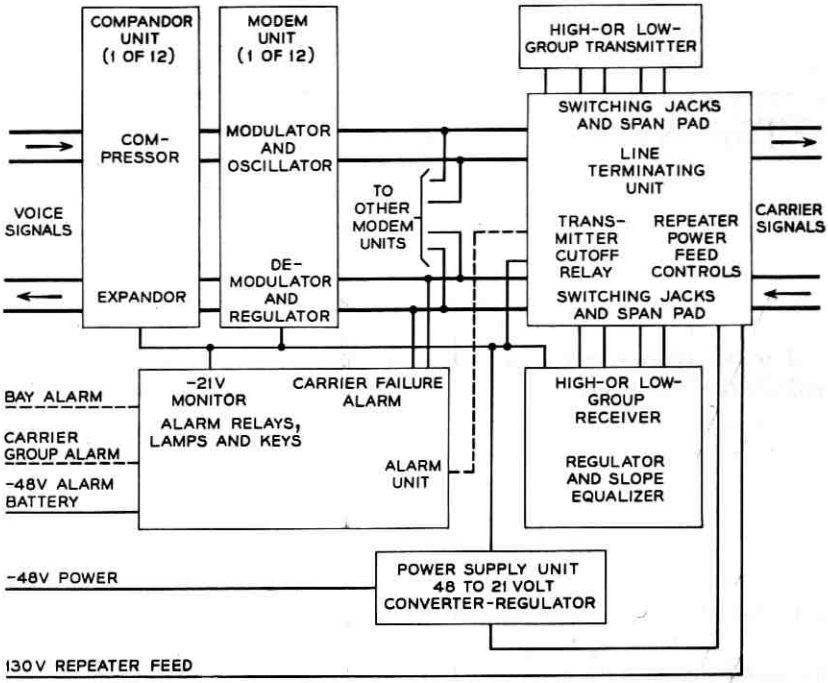

Fig. 1 — N2 carrier terminal.

levels at both ends of the line. Channel frequency assignments are unchanged.

Fig. 1 is a block diagram indicating the over-all circuit arrangement of the N2 terminal. The first step in preparing a signal for transmission over the N carrier system is to compress the volume range. Voice signals arriving at the carrier terminal may have a volume range as large as 60 db. For optimum system performance it is necessary to reduce this volume range to 30 db. This is accomplished by a compressor circuit which operates on a logarithmic basis. Each change of 2 db in the incoming signal is reduced to a variation of only 1 db in the outgoing signal. This increases the level of the lowest-volume signals by approximately 30 db, to a point where they are much less susceptible to noise interference which may be picked up in subsequent transmission over the line. The compressed signal is applied to a modulator where, after going through a band-limiting filter, it is modulated and added to a carrier frequency of accurately controlled amplitude. This fixes the carrier-to-sideband ratio of the signal

transmitted over the carrier line. The carrier frequency is generated by a crystal-controlled oscillator which is an integral part of the modulator unit. The modulator output is a double-sideband modulated signal. This is delivered to a summing circuit where it is combined with similar outputs from 11 other compandor and modulator circuits.

The carriers are spaced 8 kc apart, and both sidebands can be transmitted over the N line without interference between channels. The composite signal arriving at the summing point occupies the frequency band 172 to 268 kc. This is referred to as the high-group frequency band. Signals are transmitted over the line both in this frequency range and in another range commonly referred to as the low group. The low-group frequencies are from 36–132 kc.

The 12-channel composite signal is at a low level when it leaves the summing circuit. It also contains a large number of unwanted modulation products. These deficiencies are eliminated in the group transmitter. Two types of group transmitter are available. If transmission is to be in the high-group band, then a simple filter to select this band and a power amplifier to raise the levels to those desired on the line will be sufficient. On the other hand, if low-group transmission is desired, it will first be necessary to select the desired frequencies, modulate them with a carrier frequency of 304 kc, select the lower sideband resulting from this modulation step and amplify it to the desired line transmitting levels. In either case, the input and output of the group transmitter are wired through a line terminating unit which serves several functions in addition to the obvious one of connecting the carrier frequency signals to the line. First of all it provides access points for making in-service tests on the group transmitters and also for making in-service switches for maintenance purposes. In addition, the line terminating unit contains lightning protection devices and means for feeding simplex power over the lines to provide the energy required by remotely located repeaters.

Although the carrier line signals are transmitted and received over separate cable pairs, they are separated further by being placed in different frequency bands to simplify crosstalk suppression problems. Thus a particular terminal may be receiving high-group and transmitting low-group signals or vice versa. Except for this distinction and the obvious level differences, signals arriving at the carrier terminal from a distant repeater are treated in a manner inverse to that accorded to the outgoing signals. The 12-channel composite signal enters the line terminating unit with its protection and power supply

features. Provision is also made for adding loss to the circuit in cases where the distance from the nearest repeater is relatively short. This loss brings the signal levels within operating range of the regulation circuit in the group receiver. Two designs of group receiver are available. If the received signals are in the high-group band, then the unit known as the high-group receiver will select the desired frequency band and amplify the signals to the desired output level. If the incoming signals are in the low-group band, then the group receiver will select this band, perform a modulation process to convert these to the high-group frequency band, equalize for line slope, amplify the resultant signal and regulate its level. Therefore the signals coming out of a group receiver are always in the high-group frequency range. Regulation is on a total power basis, so that any errors in line equalization must be further compensated by regulation in the channel receiving units.

The channel demodulator contains a highly selective bandpass filter which selects one of the 12 channels available at this point. Since this is a carrier-transmitted system, a simple full-wave rectifier is adequate for demodulation purposes. Moreover, the carrier is transmitted at a known level relative to the sidebands, and this information can be used for regulating the voice-frequency output level. The resultant signal is passed to an expandor circuit which inverts the operation performed at the far end by a compressor. Each change of 1 db in the input signal to the expandor appears as a 2-db change in the level of the signal coming out of the expandor. Thus the received signal is restored to the full 60-db volume range it had when it arrived at the transmitting N carrier terminal.

Power for the N2 carrier terminal is obtained from the 48-volt office battery. However, substantial economies are obtained by interposing a dc-dc converter between this battery and the active circuits of the N2 terminal. By doing this, a tightly regulated power supply is obtained which is quite free of the usual types of noise encountered in battery supplies. The output of the converter is regulated to 21 volts $\pm$ 1 per cent for input voltage variations of 10 per cent or less. Both the lower voltage and the tight regulation make possible greater efficiency in the choice of bias conditions for the transistors. The power supply itself has an over-all efficiency of better than 85 per cent.

The state of the terminal is continually monitored by an alarm unit. One function of this unit is to maintain a check on the power supply and to provide a warning in case the 21-volt output goes out-

side working limits. Another function is to monitor the received carriers and to provide a warning whenever their level changes significantly from the normal value. In case there is a total failure of the received carriers such as would happen in the case of a line break, a processing circuit takes over and initiates certain alarm and checking conditions which take the system out of service until normal operation has been restored. The alarm unit also contains an access jack through which terminal power may be supplied from an external source, permitting removal and replacement of the normal power supply unit in the terminal without interrupting service.

II. COMPANDOR

Each terminal contains compressor and expandor circuits, the combination being designated a compandor. Although these units are closely associated in a given terminal, it should be kept in mind that it is the performance correlation (tracking) between a compressor at one terminal and an expandor at a distant terminal which affects channel net loss.

The compression and expansion functions are performed by controlled resistance pads known as variolossers. The heart of the variolosser is a pair of diffused silicon diodes whose ac impedance is an accurately calibrated inverse function of a small dc bias current of the order of 10 to 300 microamps. Using the diode pair as a shunt element in a high-impedance circuit and making the bias current proportional to the magnitude of the compressed signal gives the desired compressor action. Using the diodes as series elements in a low-impedance circuit and making the bias current proportional to the compressed signal gives the desired expandor action. Voice currents in the diodes must be kept small relative to the bias currents to avoid harmonic distortion.

The diode bias current is obtained by rectifying the compressed voice signal. This is done in a full-wave rectifier having relatively fast response time so as to follow the syllabic variations of speech. The voltage to be rectified is obtained from a transistor amplifier.

The compressor is "backward acting," since the compressed signal required to drive its variolosser appears at its output, while the low signal levels suitable for control by the variolosser appear at its input. This creates a potential feedback path which may produce low-frequency oscillations. The present design depends on the longitudinal balance of the variolosser to give high loss in this feedback path. This

requires a high degree of inherent balance in the diodes and in their individual bias currents. The amplifier must be capable of delivering 80 milliwatts to the rectifier to develop the required bias current. This is almost the total output, since only 1 mw is delivered to the channel modulator.

The expandor receives a compressed signal from the channel demodulator and hence must be "forward acting." The received voice frequency current is split, with the major part going to a control amplifier, while a small fraction is fed through the variolosser and an amplifier to the four-wire voice circuit. Both of these amplifiers are capable of delivering 80 mw to their respective loads. The output of the control amplifier is rectified and delivered as a dc bias to the variolosser diodes, with the result that for each 1 db increase in the voice signal the loss of the variolosser is reduced 1 db. This results in a 2-db increase in output to the four-wire circuit.

A schematic of the compressor circuit is shown in Fig. 2 and the expandor schematic is shown in Fig. 3.

In designing these circuits economy and compactness have been stressed. To this end semiconductor devices have been used throughout. The amplifiers use alloy junction pnp germanium transistors in the first stages and diffused silicon npn transistors in the output stage. The germanium units are Western Electric 12B transistors, which have been in production for several years. The silicon epitaxial 24B transistor was developed for this application. It operates as a linear amplifier over an extremely wide range of current and voltage swings. The higher temperature tolerable in silicon transistors permits class A operation at 340 milliwatts bias and ambient temperature of 140°F with a heat radiator of modest size.

The diffused silicon variolosser diodes also were developed for this application. They have a well-defined and stable ac impedance characteristic as a function of bias current. Variolosser action is obtained by varying this current over the range 7 to 300 microamps, corresponding to an impedance range of 7000 to 160 ohms for each diode. Diodes are used in matched pairs in order to maintain good circuit balance, low modulation distortion and, in the compressor, freedom from singing.

From the previous discussion it will be apparent that when there is no voice-frequency input signal to an N2 terminal, the compressor gain is at its maximum. Such a condition may occur during dialing and switching operations, and it is just at this time that supervisory relays associated with the trunk circuits may produce high-level tran-

sient voltages. Such voltages, if allowed to reach the compressor amplifier, would cause serious overload.

Protection against high-level transients is provided by a "click-reducer" varistor connected across the primary of the compressor input transformer. The varistor consists of two parallel silicon diodes which limit the transients to a maximum value of 0.5 volt. This limitation, coupled with the fact that the duration of the transients is less than 4 milliseconds, restricts amplifier overloading to a tolerable range. Recovery from such overloads is sufficiently rapid to permit proper transmission of signaling tones. Those same transient voltages can induce longitudinal impulse noise in voice-frequency inputs to other N2 channels. A grounded shield in the compressor input transformer blocks that source of impulse noise interference to N2 terminals.

The signal-carrying amplifiers in both compressor and expander consist of three grounded-emitter transistor stages. To bring the voice-frequency currents from the low levels necessary for satisfactory variolosser operation to the levels required at amplifier outputs requires about 60 db of voltage gain. This gain is made independent of transistor parameters and their variation by using approximately 35 db of negative feedback. The expandor control amplifier uses a grounded-collector followed by a grounded-emitter stage. Higher signal levels can be delivered to this amplifier, so that only about 20 db of voltage gain is required. Low sensitivity to parameter variations is obtained by using approximately 20 db of negative feedback.

A full-wave voltage-doubler rectifier fed by the compressed speech signal is used to derive the bias currents for the variolosser diodes. A voltage which may be anywhere between 2 and 105 volts, depending on the signal level, is coupled to the diodes through large resistances to produce bias currents proportional to the signal level. Strict proportionality is not obtained because of the finite forward voltages of both the rectifier and variolosser diodes. Because these voltages are subject to relatively wide manufacturing variations, it is necessary to provide low-level tracking adjustments.

Before discussing these adjustments it is necessary to consider the compandor characteristics shown by Fig. 4. As practical matter the 2:1 compressor characteristic can be maintained over only a limited range. This is partly due to the properties of the variolosser diodes and partly to the finite circuit impedance levels required. When the attenuation of the compressor variolosser has been reduced to its minimum value the input vs output curve has a 1:1 slope. The behavior of the N2 compressor is approximated by two solid-line

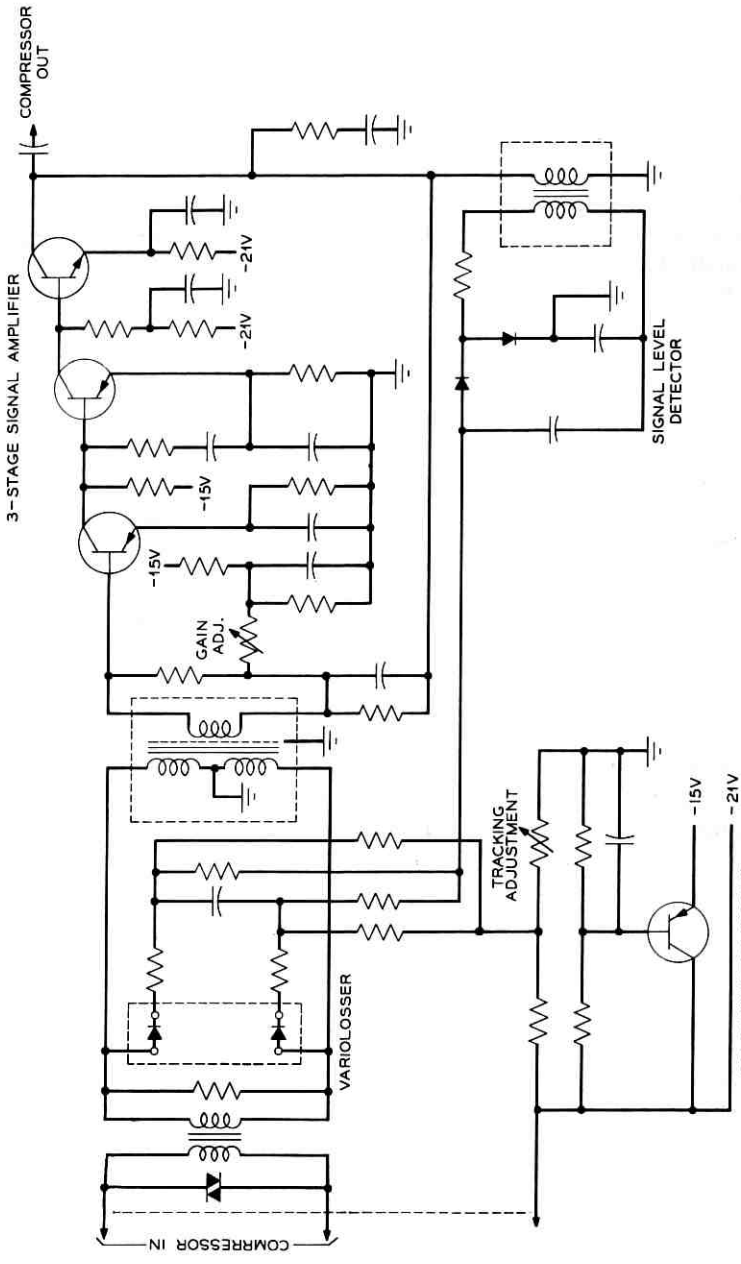

Fig. 2 — Compressor circuit schematic.

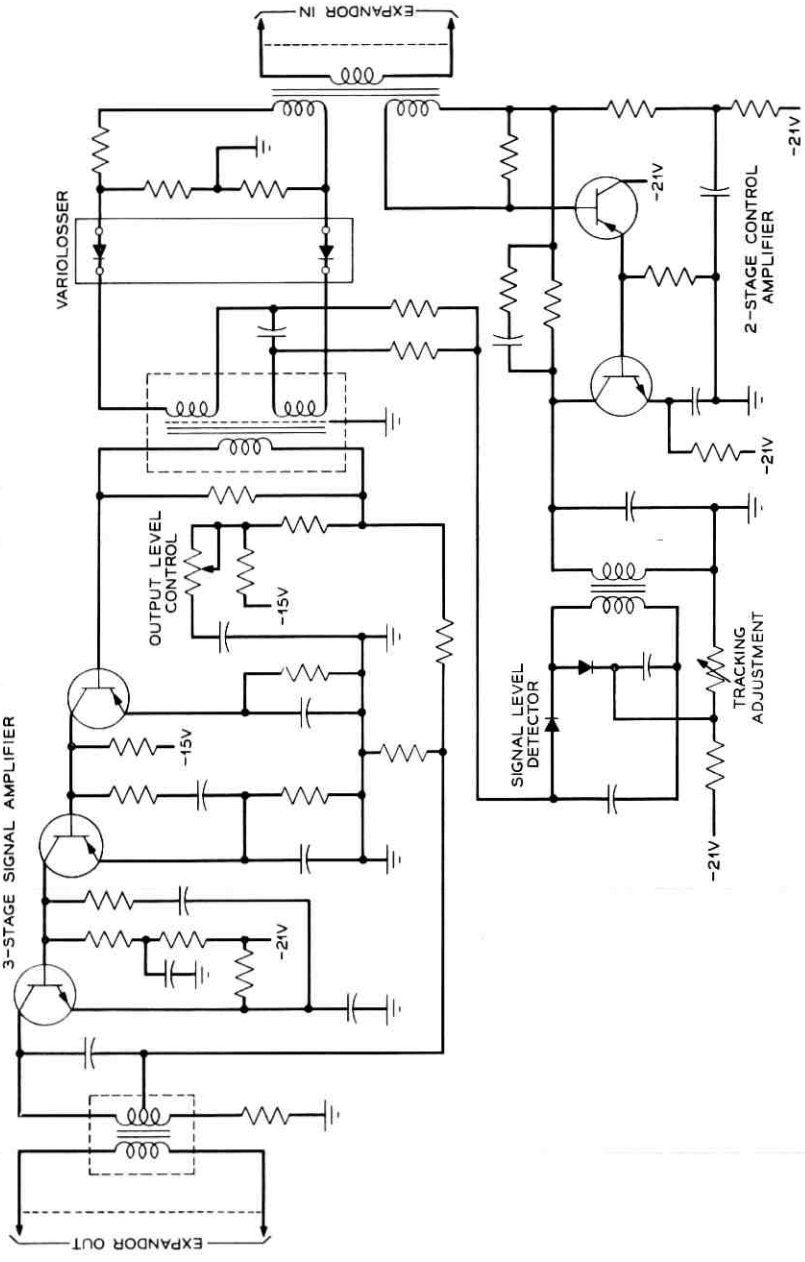

Fig. 3 — Expandor circuit schematic.

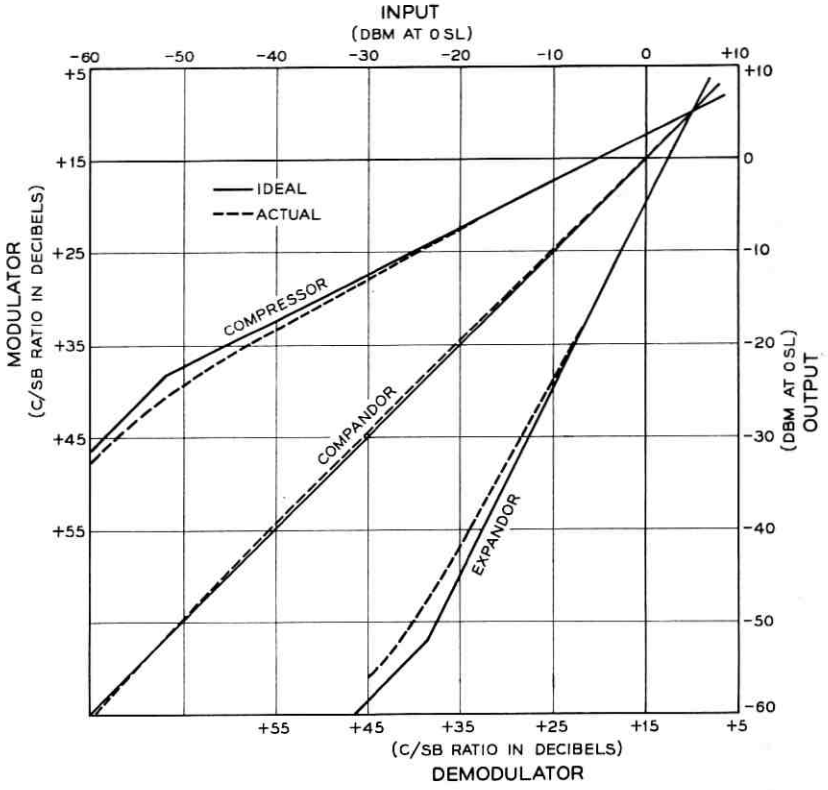

Fig. 4 — Compandor tracking characteristic.

segments, one having a 2:1 slope from +8 to −52 dbm0 and the other a 1:1 slope from −52 dmb0 through all lower values. For good compandor tracking the expandor must follow a complementary pattern, also shown by solid lines in Fig. 4. The actual performance is shown by the dashed curves.

To insure that these curves are matched exactly, so that every compressor will track every expandor, adjustments are provided. Three of these are factory set and one is available for service line-up adjustment. The gains of the amplifiers are set to give the desired outputs with 0 dbm0 input signal. Then the low-level adjustments are made with an input signal level of −52 dbm0. In the compressor the tracking adjustment furnishes a small forward bias to the variolosser diodes and hence determines the minimum loss at extremely low signal levels. There is also developed a small reverse bias on the rectifier

diodes which reduces the amount of low-level noise rectified. The tracking adjustment is set to force the compressor output to be −28.5 dbm0 rather than the "ideal" value of −26 dbm0. Similarly, in the expandor the variolosser diodes are forward biased to limit their maximum loss and the rectifier diodes are forward biased to improve efficiency at low signal levels. The tracking adjustment is set to force an expandor input of −28.5 dbm0 to give an output of −52 dbm0.

The end result of these adjustments is the over-all compandor tracking characteristic shown by the central dotted curve of Fig. 4. The maximum deviation from ideal is typically less than 0.2 db and in rare cases may be as much as 0.5 db.

Another important aspect of compandor performance is the response to suddenly applied signals. For example, a 12-db increase in the input signal will appear initially as a 12-db increase in compressor output but will immediately start dropping toward its ultimate value 6 db above the original output. The converse action holds for a decrease in input signal. The rates at which such changes take place have been defined as the "attack" and "recovery" times respectively of the compressor. These times are essentially determined by the rectifier filtering. The full-wave rectifier requires only simple filtering and hence permits fast attack and recovery times.

The CCITT-proposed recommendations furnish a precise definition of these times. The test signal is a 2-kc tone and at a zero system level point its magnitude is switched between −16 dbm and −4 dbm. The envelope of the compressor output wave resulting from such an input is indicated by the solid line of Fig. 5(a), where unit voltage represents the output for a steady input of −16 dbm0. The 12-db input level change temporarily changes the output level to 4 voltage units. However, the variolosser immediately starts reducing the output toward its ultimate value of 2. Attack time, t_a, is defined as the interval between switching and the point where the output envelope reaches the value 3.

After the input signal has been held at the −4-dbm0 level long enough for the output to reach its steady-state value of 2, the input level is reduced 12 db. The envelope of the output signal follows the pattern shown at the right of Fig. 5(a). There is an instantaneous drop to 0.5 followed by a gradual increase to the steady-state value 1. Recovery time, t_r, is defined as the time required to reach the value 0.75. In the first case (attack) the variolosser must increase its loss by the ratio $4/3 = 1.33$. In the second case the variolosser must re-

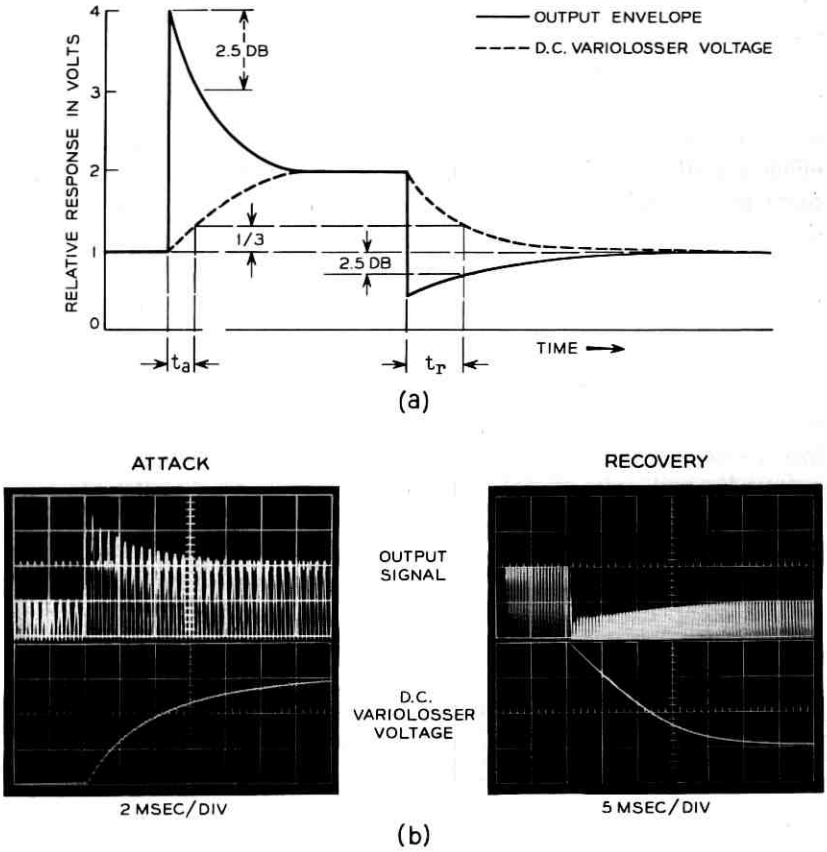

Fig. 5 — Compressor response times: (a) comparison of output voltage envelope to dc variolosser voltage; (b) oscillograms of response times with 2-kc input signal.

cover to the point where its loss is $1/0.75 = 1.33$ times its value with the low-level signal. Hence both times are measured at the instant when the loss is 2.5 db ($20 \log_{10} 1.33$) greater than its value with the steady −16-dbm0 tone.

This fact makes the measurement of the compressor response time relatively simple. The variolosser itself responds instantaneously to changes in the bias current of the diodes. To a reasonable approximation the ac impedance of the diodes is inversely proportional to the bias current and the loss is inversely proportional to the ac impedance. Hence there is a direct proportionality relation between the diode

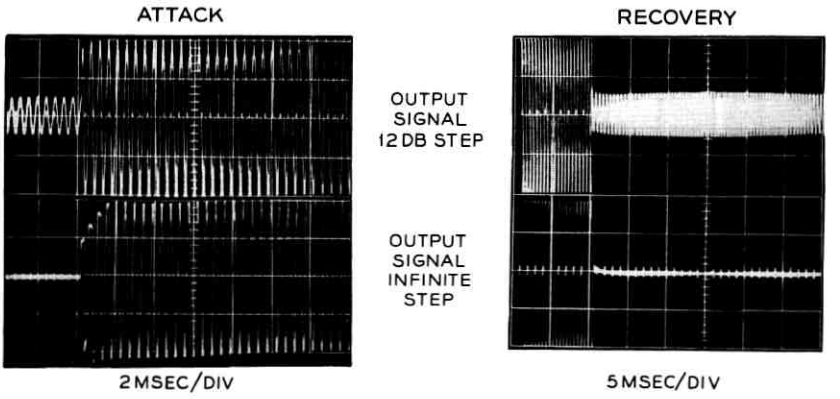

Fig. 6 — Compandor response times to step changes of 2-kc input signal amplitude.

bias current and the loss. Either may be represented by the dashed line of Fig. 5(a).

The behavior of a typical N2 compressor when tested in this way is shown in Fig. 5(b). The upper oscillogram of each pair shows the instantaneous magnitude of the 2-kc output wave and the lower shows a control voltage proportional to the diode bias current. From these it is apparent that the attack time is 1.7 and the recovery time 13 milliseconds. Tentative recommendations of CCITT are 3 ± 2 and 13.5 ± 9 milliseconds for attack and recovery time respectively. The N2 compressor meets these requirements.

In order to obtain good compandor response to stepped changes in signal level, the attack time of the expandor has been made faster than that of the compressor. The over-all result is a very rapid settling of the output signal, as shown by the oscillograms of Fig. 6.

This figure also shows the response to an "infinite" step where the input is suddenly changed from no signal to a tone level of -5 dbm0. At the start of such a tone the compressor is in its maximum gain condition and the output stage overloads. This situation lasts for about 3 cycles, or less than 2 milliseconds.

III. CHANNEL MODEM UNIT

The channel modem unit for the N2 carrier telephone terminal comprises both the transmitting modulator circuit and the receiving demodulator circuit for a particular carrier-frequency channel. Thir-

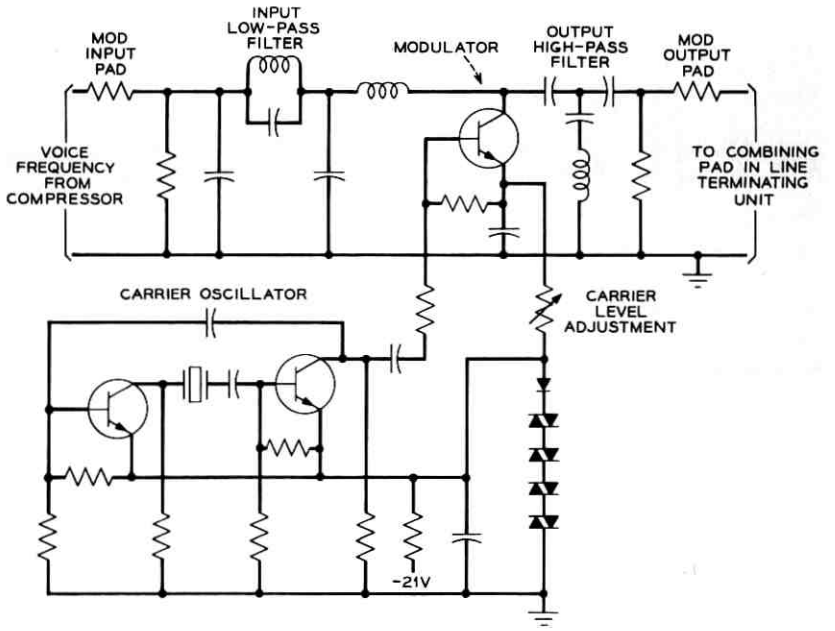

Fig. 7 — Channel modulator circuit schematic.

teen separate channel modem units are available to provide for transmission of double-sideband channels with carrier frequencies spaced at 8-kc intervals in the frequency range from 168 kc to 264 kc inclusive. A complete terminal employs 12 channels, usually channels 2 to 13 inclusive. The lowest-frequency channel (channel 1 at 168 kc) is provided as an option for use in place of any other channel which may be unavailable or unsatisfactory for use at a particular installation.

3.1 *Modulator Circuit*

The output voltage of the compressor is applied to the modulator circuit, which translates the voice-frequency input into an appropriate portion of the frequency spectrum between 164 kc and 268 kc. As shown in Fig. 7, the modulator circuit consists of input and output pads, low- and high-pass filters, a transistor switch modulator, and a crystal-controlled transistor carrier-frequency oscillator.

The input pad is designed to provide a high-impedance load (11,500 ohms) to the compressor and to terminate the transmitting low-pass filter in its design impedance of 3000 ohms. The low-pass filter band-

limits the voice frequencies from the compressor, passing frequencies up to 3250 cps with minimum distortion, and suppressing frequencies above 4000 cps by at least 20 db to reduce interchannel interference.

The modulator in the N2 modem is a simple shunt modulator consisting of a transistor switch driven by a square-wave carrier-frequency generator. The transistor switch interrupts the voice-frequency voltage periodically at a carrier-frequency rate, producing upper and lower sidebands on the carrier frequency. The carrier-frequency voltage component of the modulator output is directly proportional to the dc bias voltage on the transistor switch. Since the net-loss stability of the N2 system is directly related to the carrier-frequency output stability, a special temperature-compensated diode voltage regulator is used for deriving the dc bias voltage applied to the emitter of the transistor switch. A variable resistor in the emitter circuit permits factory adjustment of the carrier output to establish a precise carrier-to-sideband ratio for a known input signal.

The output voltage of the modulator is applied to a high-pass filter which eliminates the voice-frequency component of the signal. The output pad terminates the high-pass filter and provides isolation for paralleling the 12 modulator outputs at the combining pad in the line terminating unit. Unwanted components of the modulator output voltage, composed primarily of sidebands on the harmonics of the carrier frequency, are suppressed by the bandpass filter at the input of the group transmitting unit.

The carrier-frequency generator in the modulator circuit is a two-stage multivibrator circuit with a quartz crystal unit as the frequency controlling component. This circuit provides the square-wave output for driving the transistor switch, and for manufacturing convenience permits the 13 modulator circuits to be identical except for selection of the crystal unit.

3.2 *Demodulator Circuit*

The channel demodulator circuit shown in Fig. 8 includes the channel bandpass filter, a regulating amplifier, a demodulator, and a receiving low-pass filter. The bandpass filter selects the particular channel from the output of the group receiving unit. The bandpass filter and the carrier-frequency oscillator crystal unit are at the same frequency for a given channel modem plug-in unit, and these filters and crystal units are the only differences among the 13 channel modem units.

Initial production of N2 channel modem units employed a channel

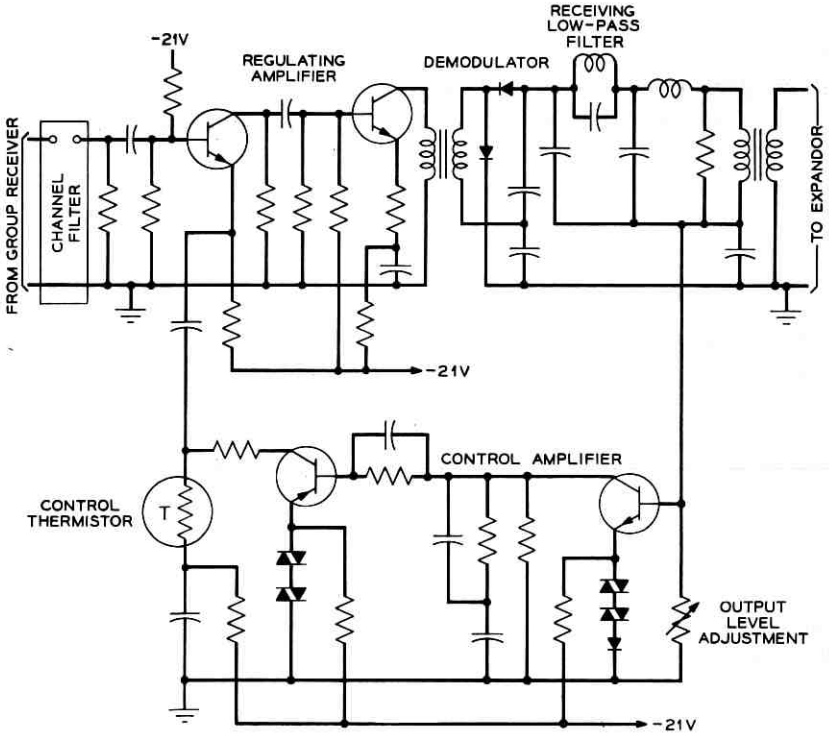

Fig. 8 — Channel demodulator circuit schematic.

filter design using ferrite inductors and capacitors. A new filter design using quartz crystal units has recently been introduced to give improved performance, particularly with regard to time and temperature stability. A typical characteristic of the crystal filter is shown in Figs. 9 and 10. As noted, the filter introduces less than 1 db distortion over the desired voice band and provides a minimum of 45 db suppression to adjacent channel carriers.

The output of the channel bandpass filter is amplified in a two-stage amplifier whose gain is regulated by a thermistor in the emitter circuit of the first stage. The output signal of the amplifier is detected by a full-wave rectifier and transmitted through a low-pass filter to the expandor. The dc component of the detected signal, obtained from carrier-frequency rectification, is compared to the voltage drop across a temperature-compensated diode voltage regulator. The difference voltage is then amplified by the dc control amplifier

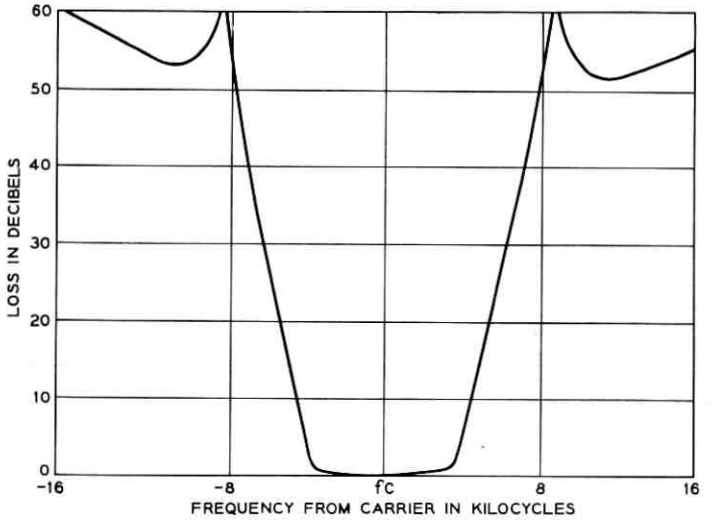

Fig. 9 — Channel bandpass filter characteristic.

and applied to the thermistor to control the gain of the regulating amplifier and maintain the carrier level input to the detector at a constant level. The regulation obtained by this arrangement is shown on Fig. 11. As noted, the variation of a 1000-cycle test tone at the voice-frequency output is less than ±0.25 db for carrier level changes of ±10 db at the input to the channel bandpass filter.

The receiving low-pass filter is designed to equalize at voice fre-

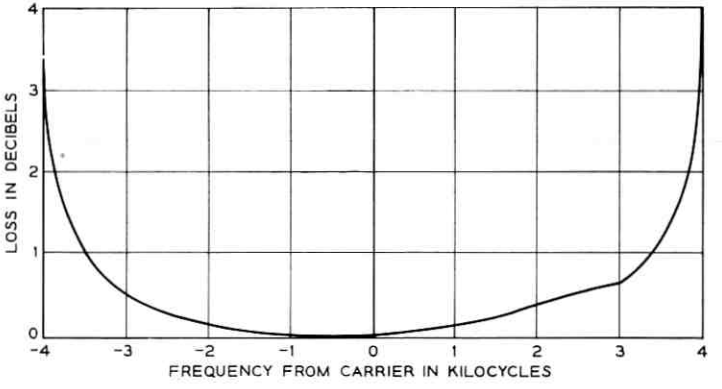

Fig. 10 — Passband distortion of typical channel filter.

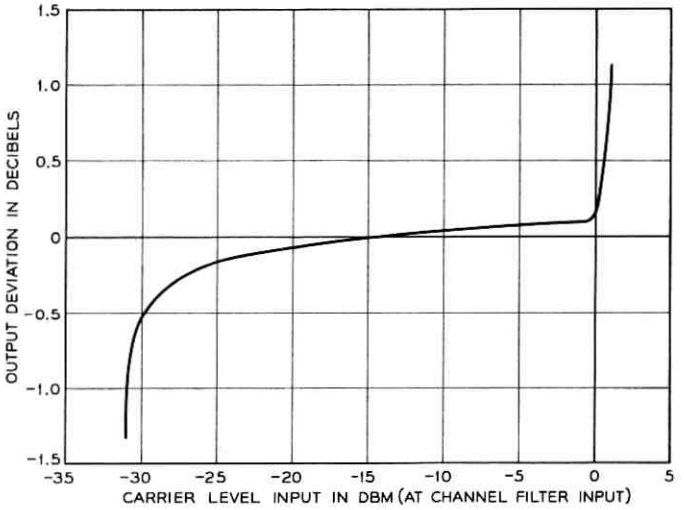

Fig. 11 — Channel modem regulator characteristic.

quencies the distortions introduced by the modulator circuit, the channel bandpass filter, and the compandor. The low distortion and reproducibility of the frequency characteristics of the above components have made possible a single fixed design of the receiving low-pass filter for all channels. The response characteristic of a typical channel for back-to-back terminal measurements is shown in Fig. 12. This figure also shows the expected 2σ limits for the manufactured product.

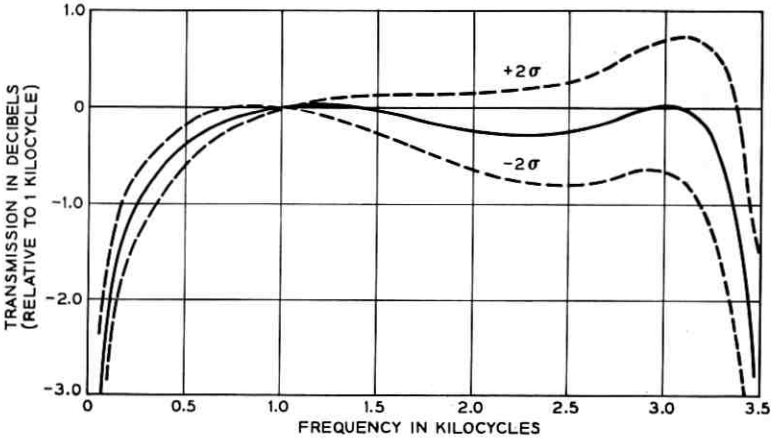

Fig. 12 — Gain-frequency characteristic of an N2 channel.

IV. GROUP UNITS

Four different group unit designs are required for complete flexibility in engineering N2 carrier systems. All of the functions performed by group units are found in the low-group receiver. Therefore a full description will be given of only this one unit, shown schematically in Fig. 13. Signals in this unit first go through a filter which selects only the low-group frequency band and rejects out-of-band noise which may have been picked up due to crosstalk or other sources in the incoming line. The wanted signals are delivered to a ring modulator consisting of a silicon diode bridge and appropriate coupling transformers. A modulating carrier frequency of 304 kc is introduced longitudinally through the transformers. This carrier frequency is obtained from a crystal-controlled oscillator and a driver stage which are integral parts of the low-group receiver. The output of the modulator consists of two sidebands and, because of transformer balance, very little carrier power. The lower of these sidebands is in the high-group frequency range and is selected by an appropriate bandpass filter.

A slope equalizer, selected to correct the attenuation distortion of the preceding line section, is inserted between the modulator and the filter. Therefore the wanted signals are essentially equal in level across the frequency band of interest at the filter output. However, the levels are quite low and amplification is required to obtain appropriate driving levels for the channel demodulators. This amplification is obtained in a three-stage transistor amplifier. To provide adequate power levels, the third stage of this amplifier uses two transistors operating in parallel. Degenerative feedback is also used to further enhance the modulation performance of the amplifier.

The feedback circuit is connected in shunt with the output transformer and to a hybrid tap on the input transformer. This provides the resistive input impedance required by the bandpass filter and a very low output impedance to drive the 12 paralleled channel bandpass filters. To make the impedance terminating the output stage substantially independent of the type and number of channel units actually in service, the output transformer is further terminated in a very low resistance.

Automatic gain control of the amplifier is obtained by including in the feedback circuit a thermistor whose impedance will depend upon the total power being delivered by the output stage. High output power will increase the energy absorbed by the thermistor, heating its internal element, thus reducing its resistance and hence reducing the

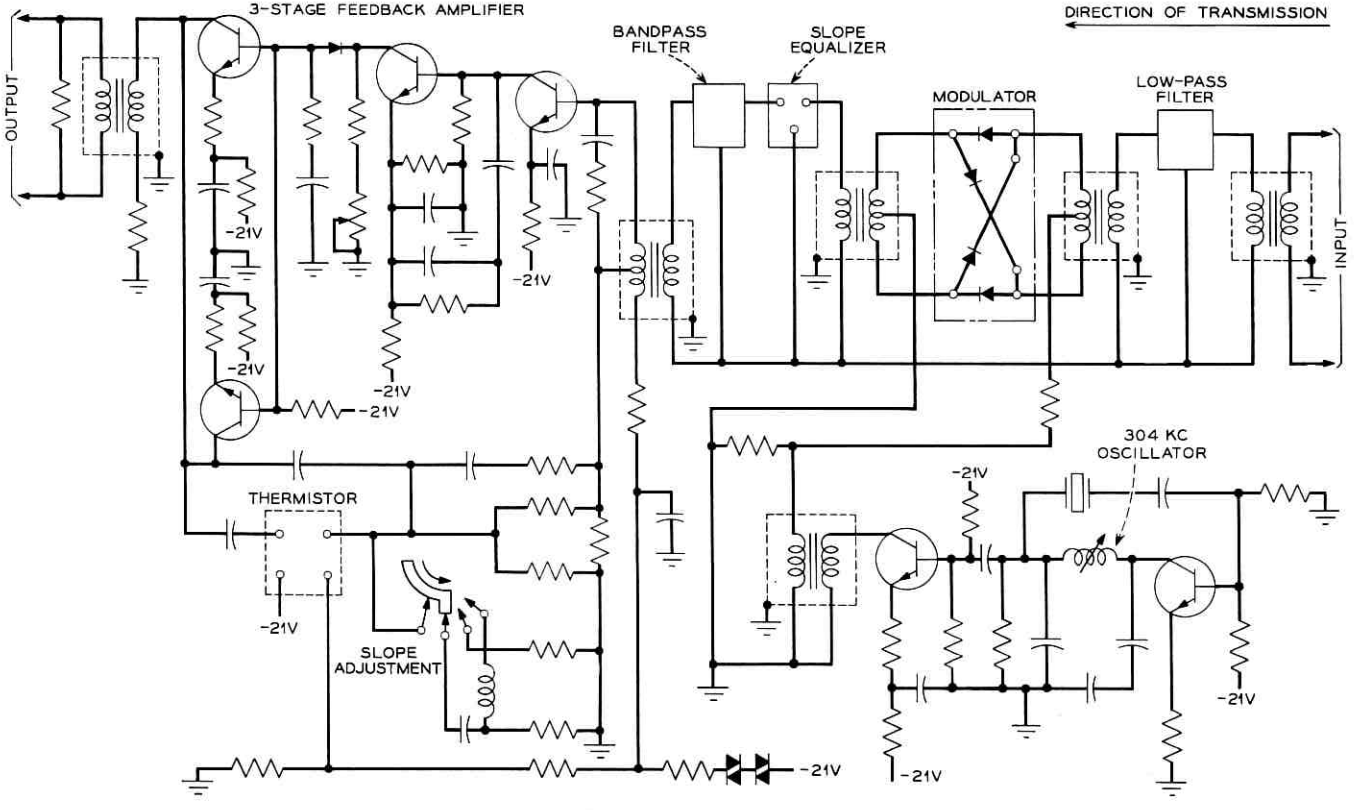

Fig. 13 — Low-group receiver schematic.

gain of the amplifier. In this manner, input level changes of ±8 db can be reduced to less than ±0.5 db. Also included in the feedback circuit are slope networks which can be used to provide fine adjustments of equalization in steps smaller than are obtainable in the slope equalizers. The desired slope network can be selected by operating a switch on the front of the unit.

The high-group receiver may be thought of as a stripped-down version of the low-group receiver. Since the signals received by it are already in the high-group range, the first selective filter, the modulator, and its associated oscillator are not required. Otherwise its performance is the same and it delivers the same signals at the same levels to the channel units. The group transmitters are simpler than the group receivers in two respects. First, since the signals delivered to them are at a constant and desired level, it is not necessary to provide gain regulation; moreover, since they are at the beginning end of a transmission circuit, the fine degree of slope equalization also is not needed. Provision is made, however, to provide some slope equalization which will act to pre-equalize the transmitted signal levels and facilitate line engineering. The low-group transmitter receives the high-group band, modulates it with the 304-kc group carrier, and selects the lower sideband for amplification and transmission to the line. The high-group transmitter does not require the modulation step and hence filters the received signal to eliminate unwanted modulation products developed in the channel modulators, amplifies the resultant signal after slope equalization and transmits this signal to the line.

V. DATA TRANSMISSION

Provision has been made for the transmission of a wideband 40.8-kilobit signal over the N2 system. This is done by removing the channel equipment associated with channels 5 through 11 and substituting equipment which has been designated as the N2WM1 wideband modem. These circuits are designed to accept signals in the band 10.2 to 51.0 kilocycles as generated by the 301B data set and modulate them into the band 203.5 to 244.3 kilocycles in the high-group N2 band. The resultant signal is summed with the six voice channels obtained by using channels 1, 2, 3, 4, 12 and 13 to give a composite signal which is delivered to the group unit and from there transmitted over the N carrier line.

The channel filters used in the wideband modem are delay equalized.

Provision is also made for additional delay equalization to compensate for high-frequency line distortion. The smallest delay equalization unit provides delay equalization for two high-frequency line repeater sections, including one high-low and one low-high repeater. Any number of delay equalization units between 1 and 18 can be obtained by selecting appropriate equalizer units and plugging them into sockets of the equalizer. Thus any number of line sections between 2 and 36 can be equalized. Since provision has not been made to equalize a single repeater section, any line involving an odd number of sections will have the delay errors associated with a single section.

VI. SPECIAL-SERVICE VOICE UNITS

Situations can arise where N2 circuits are connected together on a voice-frequency patching basis. In such cases, if standard N2 equipment were used, there would be several compandors operating in tandem on the same voice-frequency signal. This is undesirable. Therefore a special service unit has been made available which provides the necessary gains without the use of the compandor and its variolossers. By substituting the special unit, designated a VF amplifier, for the compandor at the intermediate points the circuit can be set up so as to operate with only one compressor at the transmitting end and one expandor at the receiving end, or the compandors may be eliminated from the circuit entirely. The latter situation is often used in transmitting certain kinds of voice-band data. In this case the signal levels used are somewhat higher than those encountered with voice signals, and therefore the VF amplifier units have been designed for such levels.

Another type of special service unit is the modem designed to handle the wider bands required to provide Schedule C & D program transmission service. These modem units are available only for channels 3 through 7. They differ from the standard units used in these channels only in their wider bandwidth. The wider band is obtained by using higher-quality transformers and improved filter designs.

VII. TESTING AND MAINTENANCE

The N2 terminal requires relatively few maintenance checks. The long life of the solid-state circuit components, adequate feedback in amplifiers and careful temperature compensation require only the simplest of maintenance measurements at intervals as infrequent as 6 or 12 months. Most of the measurements can be made on a working system by bridging conventional test gear at pin jacks located on

the front of the plug-in units. A special test set is provided, separate from the N2 terminal, for making out-of-service measurements on compandors, modems or alarm units. Another special test set makes possible the in-service substitution of a stand-by unit for either the power supply or any one of the group units. This permits removal and replacement of these units without a service interruption.

Only one operating adjustment is required. This is provided by the OUTPUT ADJUST potentiometer (Fig. 3) which controls the gain of the expandor amplifier. This sets the over-all net gain of the channel and compensates for any variations from channel to channel occurring in the compressor, modulator and resulting carrier-to-sideband ratio, receiving channel band filter losses relative to its center frequency, channel regulator and expandor variolosser. In addition, the adjustment also mops up variations in the common equipment and carrier line not completely eliminated by the channel regulator. The adjustment is made by introducing a 1000-cycle tone at a standard power at the distant terminal and setting the local output power to a standard value.

VIII. SUMMARY

Circuits for the N2 Carrier System have been designed to take advantage of the special properties of semiconductor devices. The low-current, low-voltage operating points of these devices coupled with the use of a closely regulated dc-dc power supply makes possible a highly efficient design with power requirements approximately one-fourth that of the earlier N1 system.

Significant performance improvements have been achieved in freedom from signal distortion, lower noise and better net loss stability. The latter is obtained by extensive use of negative feedback in amplifiers and regulators and partiuclarly by the uniform characteristics over a wide operating range of the diodes used in the compandor variolosser.

The performance improvements and the simple maintenance procedures have made N2 a very popular system. As of the end of 1964, more than 160,000 channels of N2 equipment have been sold.

APPENDIX

Levels

The concept of "system level" provides a convenient method for keeping track of signal levels at various points in a system. The

level diagram shown as Fig. 4 in the companion paper[2] is constructed by assuming 1 milliwatt of power to be injected into the system at a convenient point, which is arbitrarily designated as the zero-level point, and calculating the resultant levels at other points in the system. This is especially convenient while the signals are traveling through linear circuits where the ratio of the magnitude of the output signal to the magnitude of the input signal is totally independent of either magnitude.

However, when we consider the input-output characteristics of compressors and expandors a new dimension is added and further definition is required. Since the primary purpose of a compressor is to permit low-power message signals to produce greater carrier modulation and hence higher sideband levels than would be obtained without compression, it seems natural to define the power or "level" of the compressed signal relative to the power of the carrier signal at some arbitrary level point. The modulation ratio should be chosen as high as possible while still allowing some margin against exceeding 100 per cent modulation under extraordinary, high-power message conditions.

These considerations have led to the following definition for the N carrier systems:

After modulation, a zero system-level point (0SL) is any point where the energy in one sideband is +5 dbm when the signal producing this sideband is +5 dbm at a 0 voice level point. It is a requirement on the N2 system that the magnitude of the (unmodulated) carrier at such a point shall be +15 dbm.* The +5-dbm0 signal is thus the one which is not affected by the compressor and expandor characteristics; it is referred to as the "unaffected level" or "point of no compression" on the compandor characteristics.

The 2-db-for-1 characteristic of the compressor results in a power of +2.5 dbm for one sideband at this reference level point when the voice signal is 1 milliwatt at its 0-level point. Therefore, a test tone of 0 dbm 0 results in a carrier-to-(one)-sideband ratio of 12.5 db. Once the signal has been modulated, the broadband nature of the system insures that subsequent level changes will affect the carrier and sidebands equally. Therefore the carrier-to-sideband ratio is unchanged until the signals reach the channel demodulator.

The presentation of the compandor characteristics in Fig. 4 has

* This hypothetical point is not reached in the N2 system. In a normally operating system, the maximum level of a single carrier is nowhere greater than 4 milliwatts (+6 dbm).

been augmented by the addition of scales showing the carrier-sideband (C/SB) ratio over the appropriate range. The relation between voice frequency input and output of the compandor or either of its parts is shown by the top and right-hand scales respectively. Alternatively, the output of the compressor may be interpreted as that which will result in the carrier-to-sideband (C/SB) ratio at the modulator shown by the left-hand scale. Similarly, the input to the expandor may be interpreted as that resulting from an input signal to the modulator with the carrier-to-sideband ratio shown by the bottom scale.

REFERENCES

1. Caruthers, R. S., The Type N-1 Carrier Telephone System — Objectives and Transmission Features, B.S.T.J., *30*, Jan., 1951; Kahl, W. E., and Pedersen, L., Some Design Features of the N-1 Carrier Telephone System, B.S.T.J., *30*, April, 1951.
2. Boyd, R. C., and Herr, F. J., The N2 Carrier Terminal — Objectives and Analysis, B.S.T.J., this issue, p. 731.

Equipment Aspects of Packaged N2 Carrier System Terminals

By D. T. BELL, L. H. STEIFF and E. R. TAYLOR

(Manuscript received December 30, 1964)

Packaging the N2 carrier terminal has significantly reduced space requirements, the number of interbay wiring connections, and the number of wired options as compared to former N-type carrier equipment. At the same time noise and crosstalk performance has been improved, and ease of maintenance and operation has been provided. Isolation of "noisy" (high signal level) leads from "quiet" (low signal level) leads has been effected by routing the two sets of leads in separate ducts on newly developed double-bay duct-type frames. Wired options are simplified by furnishing optional equipment on a plug-in basis, with wiring confined to simple strapping. Automatic alarm, circuit condition and restoral equipment is furnished as part of the packaged terminal. Maintenance of multichannel equipment can be performed on an in-service basis

I. INTRODUCTION

The N2 carrier system provides twelve two-way message, program or narrow-band data channels over two pairs of a single toll or exchange plant cable. Occasionally, a two-way wideband channel for 40.8-kilobit data will be used in lieu of six narrow-band channels. Like the earlier N1 carrier system, double-sideband transmitted carrier signals in the 36- to 268-kc frequency band are used with repeaters at intervals up to eight miles, depending on the type of cable.

This paper describes the operational, functional, and mechanical aspects of N2 carrier terminals packaged to include not only the carrier circuits but also the signaling, circuit conditioning, patching and monitoring circuits required for the complete channels derived from the N2 system. Transmission performance, ease of maintenance and economy are optimized by the use of a packaged terminal framework containing mounting shelves, connectors and shop wiring such

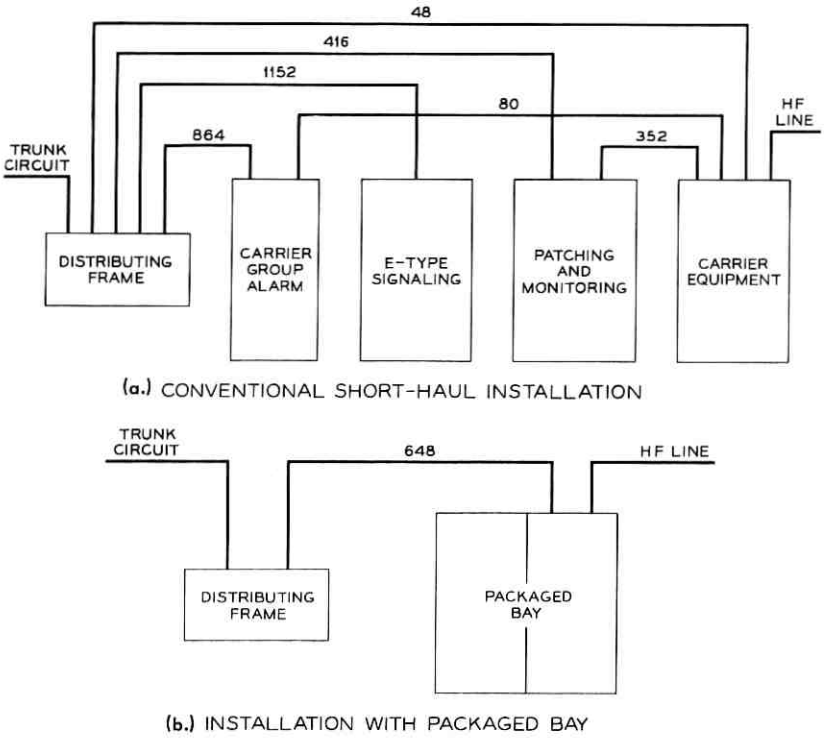

Fig. 1 — (a) Earlier N-type carrier installation (96 channels) with equipment mounted on separate installer-connected bays; (b) N2 packaged shop-wired frame installation (72 channels).

that any desired combination of circuits for message or program service may be obtained by the choice of suitable plug-in units and the placing of option strapping in accordance with standard templates.

Fig. 1 compares a 72-channel packaged frame installation to an early 96-channel installation wherein the carrier, patching and monitoring, E-type signaling and circuit conditioning equipment was located in four or more different frames. The 648 wires (9 per channel) from the distributing frame to the packaged terminal frame are about one-third of those required in the earlier installation. This is indicative of the savings in distributing frame space and installation effort because numerous long interconnections via the distributing frame are no longer required.

II. OBJECTIVES

The nature of type-N carrier systems and the numerous services using the channels derived therefrom result in wide ranges in frequencies, signal energy and susceptibility to interference, as well as numerous combinations of carrier, signaling and circuit conditioning options. Many of these options cannot be determined before a terminal is engineered and installed. Frequently, new options must be chosen to provide new services from time to time during the life of the terminal. In many cases, individual channels must be rearranged without affecting services using the remaining channels.

Consideration of these conditions and of economy in manufacture, installation and maintenance leads to the following objectives:

(a) transmission performance of terminals and office wiring suitable for all combinations of services with negligible contributions to impulse noise or crosstalk. This requires shielding, lead segregation, balancing and pairing, as well as the use of click-suppression networks.

(b) automatic circuit conditioning and alarms during failure of the carrier line between offices. This affords release of individual message channels, prevents billing for unusable time and seizure of disabled circuits.

(c) flexibility in message channel uses — same capability for all twelve channels.

(d) alternate use of wideband data channels in lieu of six message channels.

(e) universal terminal and office cabling suitable for all normal services.

(f) office engineering of economical numbers of system terminals before options are determined and with deferred expenditures for main optional (plug-in) units.

(g) functional compatibility with existing N2 carrier terminals. The inherent limitations of existing terminals still apply when a system includes both existing and packaged terminals.

(h) use of existing designs of E-type signaling units and tone supply panels.

(i) plug-in units for optional circuits and those requiring periodic maintenance adjustment. This not only aids trouble location and the substitution of spare units, but also permits expenditures for main circuit components to be deferred until needed.

(j) packaged terminal frames arranged for all of the plug-in units,

circuit conditioning, signaling tone supplies, alarms and fuses normally required for an economical number of systems.

(k) close association and simple identification of the components of each system or channel by physical proximity and numerous designations. This simplifies maintenance and restoration of service, minimizes exposure to electrical or maintenance interference, and facilitates meeting of strict resistance limitations between signaling and switching circuits.

(l) in-service monitoring of circuit performance and in-service switching of common units.

(m) portable test sets for in-service switching and testing of channel units and alarm and restoral unit.

(n) simplified circuit options through the use of option screws or wiring templates.

(o) seldom-used options obtainable with negligible effect on the cost or performance of the more usual options.

(p) optional high impedance monitoring and maintenance talking circuits.

(q) optional circuits for access to and remote control of centralized transmission or noise measuring equipment.

(r) separate power filters and fuses for odd and even numbered terminals so that a single failure will not affect more than half the terminals.

(s) application schematic covering all the options in the packaged terminals and all associated office wiring.

III. ELECTRICAL DESIGN FEATURES

In this section the electrical design of the packaged terminal and its external connections are discussed. Many of the requirements stem from the over-all system objectives discussed in a companion paper.[1] The electrical design of most of the plug-in units, both carrier and signaling, is discussed in other papers.[2,3]

Plug-in units for line terminating and for alarm and restoral functions are discussed with associated terminal circuits below.

3.1 *General*

A major factor in the electrical design is the control or avoidance of electrical interference or crosstalk. As indicated elsewhere,[1] nominal level differences between carrier frequencies may be as high as 65 db. This may be increased by as much as 30 db by differences in talker volume. In most installations this is alleviated by the use of separate

frequency bands for transmitting and receiving, but some systems in a terminal frame occasionally may use transmitting frequencies which are in the same band as receiving frequencies in other systems and vice versa. This occurs in rare junctions of carrier routes. In any case, however, received carrier frequency energies may be low enough for interference energies as low as −120 dbm to be significant.

Nominal voice-frequency level differences may be as high as 26 db and may be increased by as much as 60 db by talker volume differences. Interference energies as low as −100 dbm may be significant. The benefits to be obtained by lead segregation are limited because the same voice-frequency conductors are used for both transmitting and receiving in many types of message switching circuits to which the carrier-derived channels may be connected. Furthermore, many types of signaling and supervision circuits use the same conductors for speech (or data) signals and for low-frequency signaling and supervisory signals. This results in sizeable limitations on the degree of balance to ground obtainable in transmission conductors and correspondingly increases the susceptibility to crosstalk and interference.

It is evident that the electrical design of packaged terminals, including signaling and direct connection to switching circuits, is much more complicated than it was in the older terminals without signaling or direct connection. However, the same problems existed elsewhere in the telephone office and were aggravated by the long cables between the various locations of carrier, signaling and circuit conditioning equipments. The compactness and shop wiring of the packaged terminals afford much closer control of the end results.

Another factor in the electrical design is the control of conductor resistance and potential differences between signaling units, circuit conditioning equipment and switching circuts. Many telephone offices were designed and cabled to meet strict conductor resistance limits between signaling and switching circuits without allowance for circuit conditioning equipment inserted between them, because the latter had not been needed or invented. Accordingly, the packaged terminals, including both signaling and circuit conditioning equipment, are designed and cabled to the distributing frame to meet the same requirements as the older signaling equipment. This avoids expensive recabling of existing offices or the redesign of switching circuits.

3.2 *Packaged Terminals*

This section will make reference to the circuit units indicated on Fig. 2, which is a block schematic of a packaged N2 carrier terminal ar-

ranged for message use and includes common equipment for as many as six terminals (72 channels) within the same framework. The carrier group alarm (CGA) unit includes the circuit conditioning circuit for the twelve channels derived from one carrier system, and the 2600-cycle oscillator and transfer unit furnishes signaling tone for the entire frame. Other boxes marked UNIT rather than PANEL are plug-in units which may be replaced with other units for different services or frequencies or are removable for maintenance purposes.

High-level carrier- and voice-frequency pairs, and signaling tone supply, alarm, and power leads are run in so-called "noisy" ducts at the outside of the packaged terminal frame, as described in Section 4.3. Low-level pairs are run in a "quiet" duct in the center of the frame and are automatically shielded from "noisy" leads or circuits which may be in adjacent bays. Carrier-frequency pairs are normally balanced and individually shielded. Voice-frequency transmission and signaling tone supply leads are paired and balanced by twisting and by the use of similar resistors, relay contacts, etc., in both leads.

Exceptions are voice-frequency connections between compandor, modem and restoral oscillator units and carrier-frequency connections between modem and group units via the line terminating unit. These connections are two-wire, but are unbalanced in the interest of economy in the plug-in units, particularly in the filters. This is made possible by the physical arrangements shown in Fig. 3 and diagrammed in Fig. 4. The group and line terminating units are located together and close to the modem units. Each modem is located adjacent to its compandor and on a shelf immediately above or below the shelf containing the group units. Wiring and wiring terminals for the transmitting path are well separated from the receiving path. Wiring between shelves is shielded. The restoral switching relays for switching channel position A and B modem connections from compandors to the alarm and restoral unit and the restoral oscillator unit are located within inches of the compandor and modem connectors. Shielded pairs are used between these relays and the restoral units.

Common impedance between circuits in the two directions of transmission is minimized by the low impedance of the regulated 48- to 21-volt convertor and by the use of individually paired battery and ground leads from the power supply unit to each shelf.

Transmission leads traversing the CGA unit are subject to numerous options depending on the type of service, type or omission of signaling unit, etc. A terminal strip containing groups of terminals which are individual to each channel are provided for these options and for con-

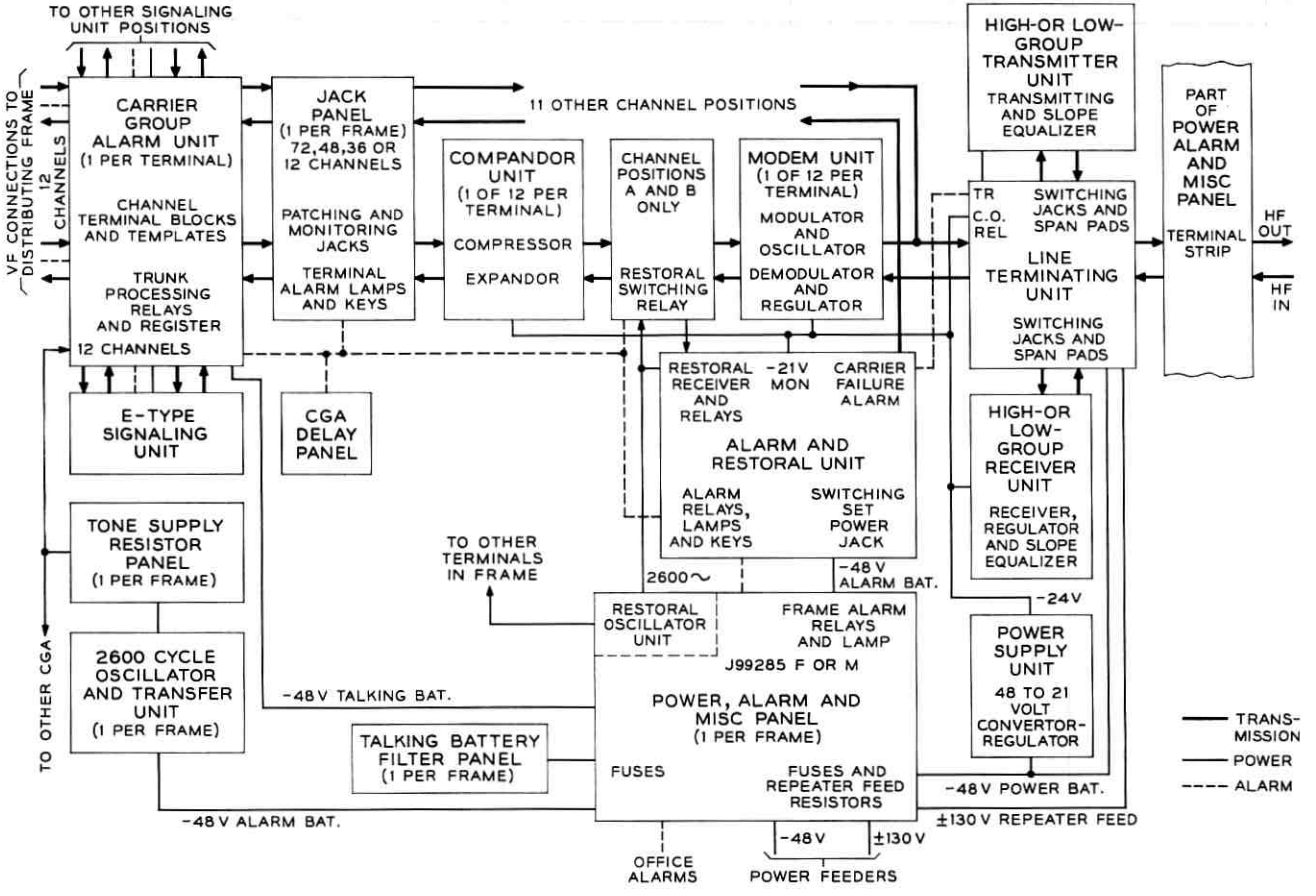

Fig. 2 — Block schematic of packaged N2 carrier terminal arranged for message use.

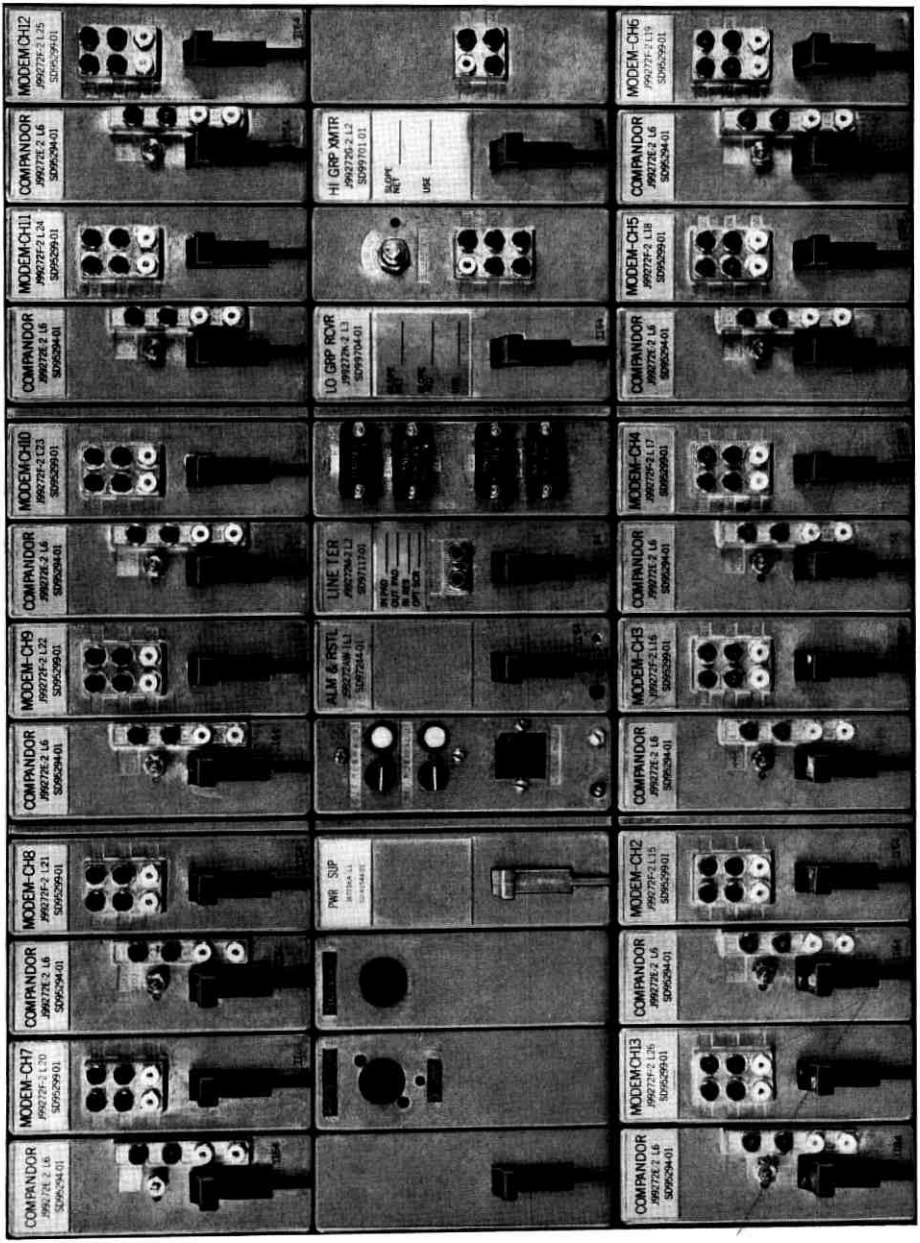

Fig. 3 — N2 carrier terminal, front view.

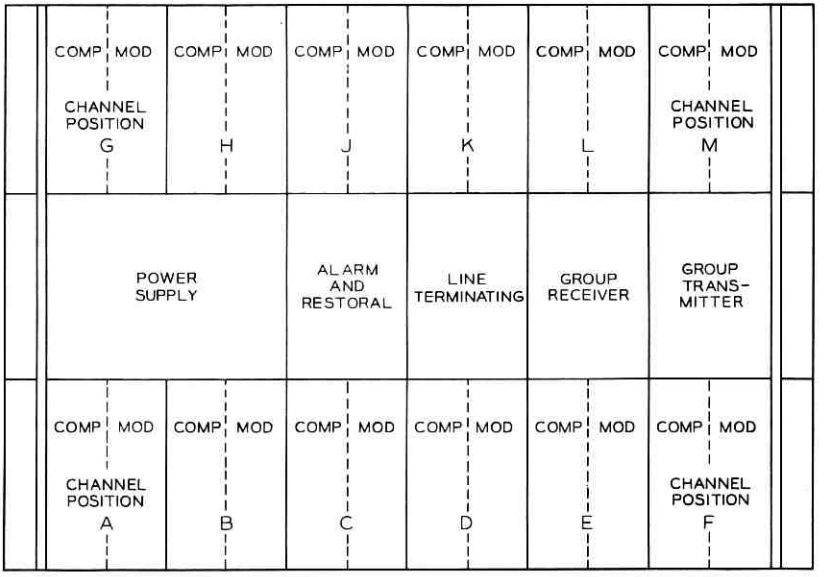

Fig. 4 — Diagram of N2 carrier shelf positions.

nections into and out of the CGA. These groups of terminals are located close to associated relays and signaling unit connectors to minimize noise and crosstalk exposures. Wiring terminals for executing the options in the transmission paths are separated from those for signaling or supervision options and are located close together for each direction of transmission. Normal option straps are less than one inch long. This is in contrast with optional cross connections for similar purposes in the older multibay arrangements, which often were many feet long and exposed to many interfering circuits.

Many types of signaling units connect battery to transmission pairs, making the transmission circuits vulnerable to noise on the power leads. Accordingly, a filter is provided to attenuate noise on the power feeders and to reduce the impedance common to signaling units. This and pairing of battery and ground feeders is helpful in meeting No. 1 ESS office objectives for preventing interframe interference. Separate power feeders, filters and fuses are used for odd- and even-numbered terminals to lessen the chance of a single failure affecting more than half the systems.

3.3 *Connections to and from Packaged Terminals*

Individually shielded pairs are used for carrier-frequency connections to and from the packaged terminals. The shields of these pairs are

connected to those of corresponding pairs in the terminal. In small offices, these pairs connect directly to the high-frequency interoffice cable protectors which may be located in one of the packaged frames. In somewhat larger offices, where the length of these pairs would exceed 200 feet, the individually shielded pairs terminate on a distributing terminal strip for connection to a multipair aluminum shielded cable which in turn is connected to the protectors or to a high-frequency cross-connecting cabinet which includes the protectors. In these cases, the shields of the individually shielded pairs connect to the shield of the ABAM cable which is connected to the shield of the interoffice cable. In any case, physical separation on the cable racks is specified between carrier pairs and noisy pairs or pairs carrying the same frequencies at substantially different levels. These provisions protect against carrier-frequency noise or interference from other circuits in the office.

Voice-frequency connections between the packaged terminals and the distributing frame are protected from noise and crosstalk by segregation of low-level pairs from high-level pairs or noisy conductors. This segregation may be obtained by the use of separate cables or by separate layers in the same cable.

3.4 *Wiring within the Packaged Bay*

Universal shop wiring is used in each frame for all of the options that may be required in the channels and terminals contained therein. This is made possible by liberal use of plug-in units — usually 41 per terminal or 247 per frame. The connectors for the plug-in units are multiconductor — 18 for signaling units, 20 for carrier units and 40 for line terminating or alarm and restoral units. With a few exceptions, the circuits assigned to the contacts of the connectors are chosen so that inadvertent insertion of a wrong unit into a position on a working terminal will not cause component failure or major service reaction. The exceptional positions where this electrical protection is not feasible are safeguarded by simple mechanical keying which is also used in the test stand.

The eighteen conductors for signaling unit connectors are required for the twelve different types of signaling units which may be used. The twenty conductors for carrier unit connectors are often more than are required for circuit connection, but the extras afford valuable lead segregation or shielding. In other cases, the extras permit close association of restoral switching relays or optional use of broadband data modems. The conductors of the carrier connectors are chosen so that

much of the critical wiring is run directly between adjacent positions avoiding the exposure to interference of more circuitous wiring.

3.5 *Alarm and Circuit Conditioning*

Relays common to all the terminals are used for both major and minor office alarms. They are multipled to fuses and individual terminals through isolating diodes. Separate relay contacts are provided for the numerous combinations of ground or battery and series contacts required for the variety of office alarm circuits which may be encountered in local and toll offices.

During complete system failures such as interruptions to the carrier line, alarms are brought in at both terminals of the system for two purposes — first, the customary alerting of maintenance personnel and second, circuit conditioning whereby the effects of the failure are minimized for the users or prospective users of the individual channels of the system. When the failure is corrected, the alarm indications are automatically restored and the channels are again offered for service. The circuit conditioning circuits have been designated "carrier group alarm" (CGA) because the alarms originate from loss of the group carrier energy. Subsequent testing and restoral to service processes are designated "restoral."

The circuit conditioning and restoral elements of Fig. 2 are CGA unit, alarm and restoral unit, restoral oscillator, CGA delay and restoral switching relays in channel positions A and B. These and similar elements at the distant terminal automatically proceed as follows:

(a) recognize carrier system failure,

(b) force transmission failure in other direction to start alarm sequence at distant terminal,

(c) alarm and start trunk conditioning at both terminals of the N2 system,

(d) transmit restoral tone (2600 cycles) over the carrier channel associated with channel position A and observe the restoral tone-to-noise ratio at the distant terminal,

(e) when this ratio has been satisfactory for an adequate time, transmit restoral tone in the opposite direction over the carrier channel associated with channel position B,

(f) restore alarms simultaneously at both terminals when restoral tone transmission is satisfactory over both channels in both directions of transmission, and

(g) return trunks to service at both terminals at the same time.

The alarm and restoral circuits include lamps which provide visual indication of the direction of transmission in which a carrier failure occurs and a counter which indicates the number of failures. These are maintenance aids obtained as byproducts of the restoral functions. The lamp is controlled by a CGA relay which is operated during the interval from failure to the receipt of satisfactory restoral tone at channel position A. Accordingly, the lamp lights on both carrier failure and carrier interruptions produced automatically to alarm the distant terminal. However, the latter lasts for less than a minute, so continued operation indicates a real receiving failure. The counter or register is controlled by a thermistor and CGA relay contacts to advance one step each time restoral tone is not received at channel position B immediately after satisfactory reception of restoral tone at channel position A. Accordingly, it ignores the forced interruptions and counts the real failures. This can be a real aid to detecting and locating intermittent troubles which otherwise would cause intangible service reactions.

3.6 *CGA Options*

Each message channel and associated supervision as well as each program channel passes through the CGA unit to a cable to the distributing frame. Any one of 33 different circuit configurations may be required and severe requirements on crosstalk and interference exist, as indicated in Section 3. The 54 leads required for the 33 configurations are connected to a terminal strip in a pattern which preserves pairing and level segregation and thus the optional straps which complete the circuit configurations are rarely more than a half-inch long. As described in Section 4.5, templates depicting the straps for each configuration are furnished for each channel.

IV. MECHANICAL DESIGN

This section discusses the general plans for mechanical design and describes the plug-in units designed specifically for N2 carrier terminals, the packaged terminal frames in which they are mounted, and the associated equipment common to several terminals.

4.1 *General Plans*

The flexibility and office engineering objectives discussed in Section II indicate liberal use of plug-in units. Detailed study of the circuit options determined the circuit subdivisions for the plug-in units. Approximate volume requirements for each unit were determined after

selection of the smallest reliable apparatus components economically obtainable and consideration of feasible packaging techniques, such as printed wiring and AMPLAS.

The requirement that existing E-type signaling units be used indicated the use of 23-inch bay frameworks. Later studies of frame sizes and capacities confirmed the desirability of using this standard along with the 12-inch standard for frame depth.

More detailed study of these factors and of manufacturing costs resulted in the use of die-cast aluminum mounting shelves, each divided into twelve modules $1\frac{5}{8} \times 4\frac{7}{8} \times 11$ inches, and the packaging of the plug-in units in 1-, 2- or 4-module sizes to afford the desired flexibility, interference-free interconnections and compactness. The twelve-module shelf is well adapted to a twelve-channel system, as the channel units are single module size — two per channel completely filling two shelves. The twelve single-module units on one shelf are replaced by four two-module and one four-module units when wideband data replaces six voice channels. A third shelf contains the common units of the terminal consisting of four two-module units and one four-module unit.

4.2 *Plug-in Units*

A full complement of terminal plug-in units consists of 12 compandors, 12 modems, a group receiving unit, a group transmitting unit, an alarm and restoral unit, a line terminating unit, and a power supply unit. In addition, a number of special-service units, such as schedule A and B program channel units, schedule C and D program channel units, voice-frequency amplifier unit and wideband data units either are or soon will be available to the field, to be plugged in to meet service requirements.

Most of the circuitry in the plug-in units is contained in AMPLAS assemblies. The word "AMPLAS" is an acronym of "apparatus mounted in plastic."

The AMPLAS process, developed by the Western Electric Company at the Merrimack Valley Works, begins with the casting of a cellulose acetate butyrate mold in the shape of a shallow tray, as shown in Fig. 5. Cellulose acetate butyrate is a thermoplastic material. In its solid state, the particular formulation used is waxlike, flexible and translucent. The mold is placed over a full-size assembly drawing which shows the locations and positions of all the components of that particular assembly. The translucency of the material permits the operator to view the layout through the bottom of the mold. Components are furnished to the operator with their leads cut to length and formed. The

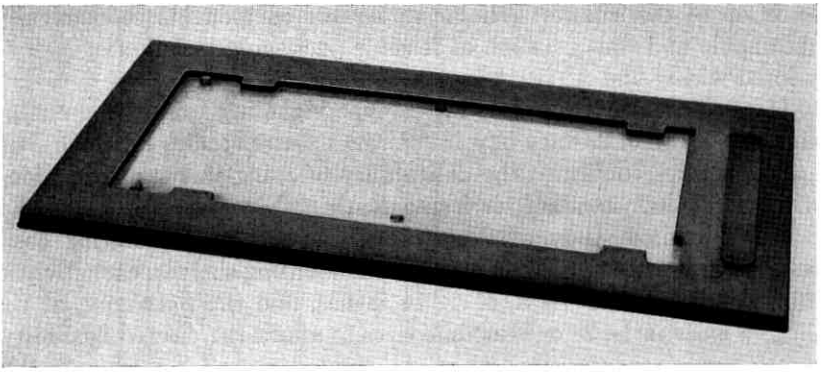

Fig. 5 — Cellulose acetate butyrate mold for Amplas process.

operator inserts the components by pressing the leads into the floor of the mold in the positions indicated by the assembly drawing. The insertion operation is shown in Fig. 6. The leads easily pierce the wax-like cellulose acetate butyrate and extend through to the bottom surface of the mold. After the last of the necessary components is inserted into the mold, a properly proportioned mixture of epoxy resin and hardener is metered into the mold. The mold, full of components and epoxy resin, is placed in an oven at 150°F for one-half hour, during which the epoxy cures into a hard, tough substrate. After cooling, the cellulose acetate butyrate is stripped away, leaving an epoxy substrate to which all of the component leads are bonded. The thickness of the substrate now corresponds to the depth of the tray-like mold, and the length of the protruding leads corresponds to the thickness of the floor of the mold. Pencil wiring is applied to the protruding leads to provide circuit continuity as shown in Fig. 7. Completed AMPLAS assemblies are shown in Fig. 8.

Special die-cast aluminum unit frames were designed to carry the circuit-laden substrates. The unit frame slides into tracks in the mounting shelf, which provides precise alignment of the plug-in unit with its mating connector. The front of the unit frame is a face plate which provides space for unit identification, test jacks, and a latch. When placed side-by-side in a terminal, the face plates provide an attractive facade for the terminal.

Each message channel requires two units — a compandor and a modem. The compandor and modem units occupy alternate spaces in and completely fill the top and bottom shelves of a three-shelf stack which constitutes a complete terminal housing. Each channel unit is

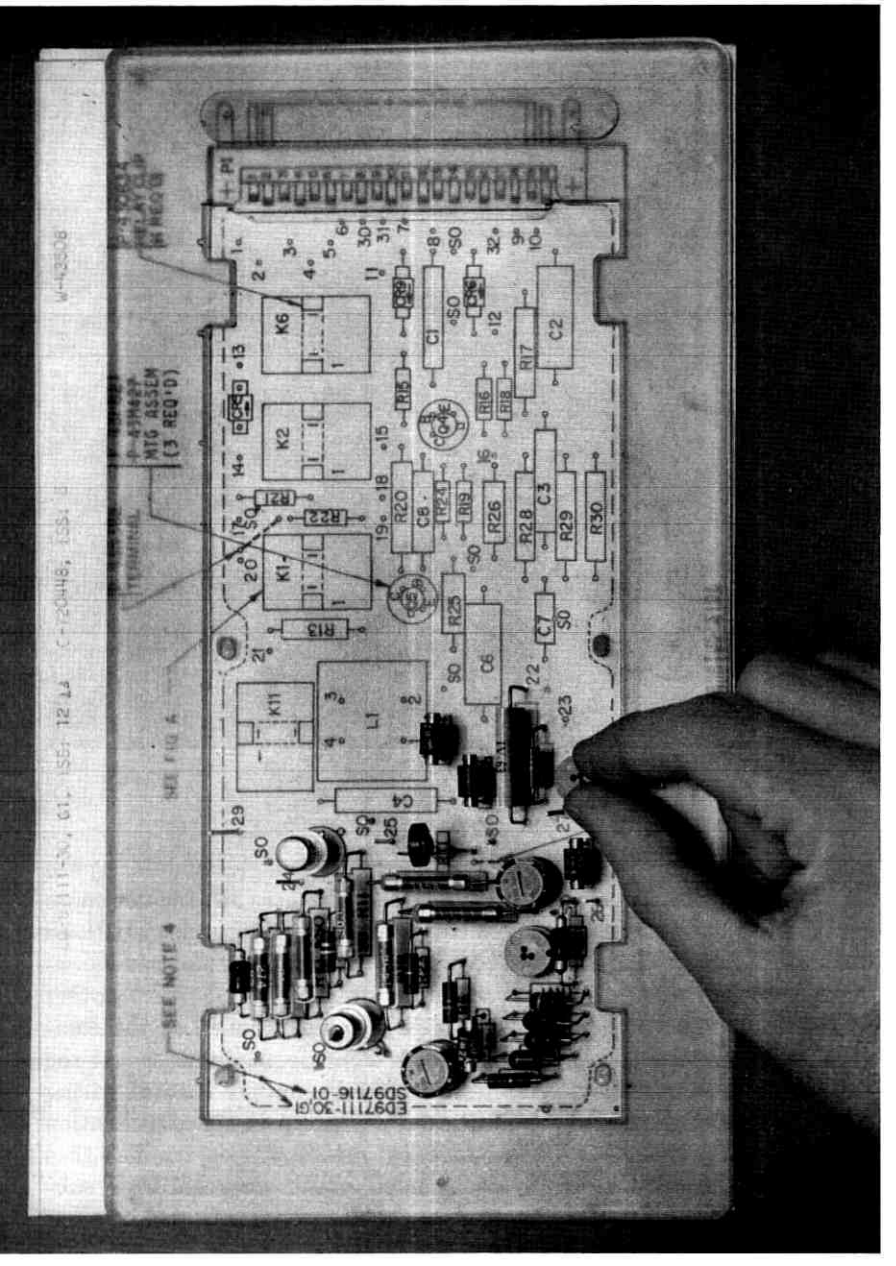

Fig. 6 — Insertion of components in mold.

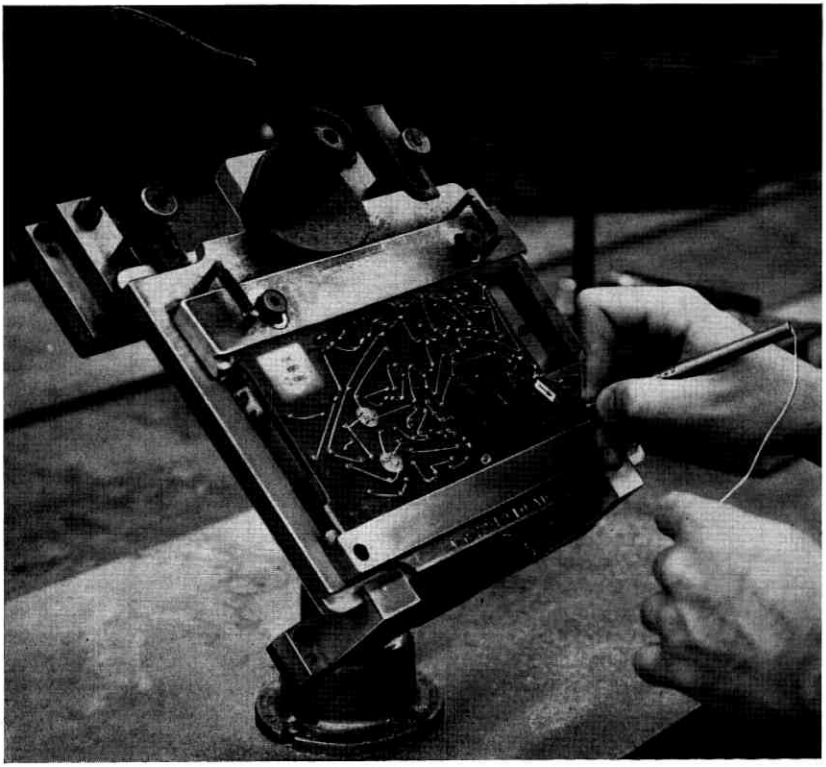

FIG. 7 — Application of pencil wiring to component leads in AMPLAS subassembly.

mounted on a single-unit frame and occupies a single modular space on the shelf. The compandor and the modem circuits each include more components than could be conveniently mounted in a single AMPLAS assembly. This problem was resolved with a rather unusual sandwich-type structure, wherein one subassembly was suspended over another with their component faces toward each other. Inasmuch as the complete structure was designed to fit into a single modular space, the distance between the subassemblies was severly limited. Printed wiring assemblies are lighter and somewhat more efficient in space utilization than are equivalent AMPLAS assemblies; and it was determined that a printed wiring subassembly, suspended above an AMPLAS subassembly, would fit into the available space. Hence was effected a rather neat marriage between AMPLAS and printed wiring. To assure interference-free positioning of the sandwich halves, care was taken to lay

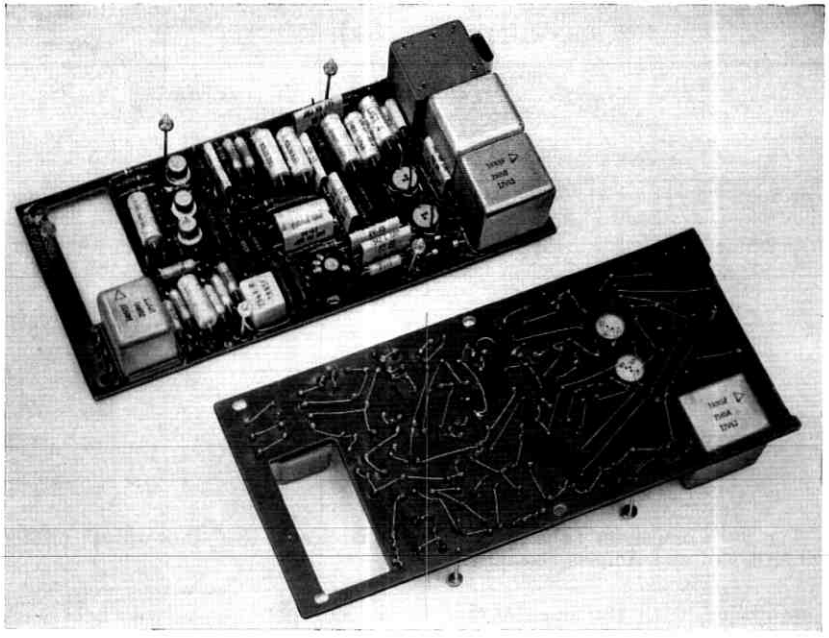

Fig. 8 — Completed AMPLAS subassemblies.

out the circuity in such a manner that tall components on either subassembly fall opposite short components on the other subassembly. This caused some difficulty in incorporating changes after the initial layouts were finished but was not an insurmountable obstacle.

Care was taken to keep intersubassembly wiring within the "sandwich" to a minimum. In the compandor, this was accomplished by putting the entire expandor circuit on the AMPLAS subassembly and almost all of the components of the compressor circuit on the printed wiring subassembly. All interboard wiring was dressed to a common edge of the structure so that the halves of the sandwich could be opened like a book for servicing.

The design of the modem is similar to that of the compandor. Here the wiring between the subassemblies was minimized by placing the entire demodulator circuit on the AMPLAS subassembly, and most of the components of the modulator circuit on the printed wiring subassembly. A compandor and a modem are shown in Fig. 9.

An electrostatic shield is placed between the two circuits of the modem to inhibit crosstalk. Because of the contours of the two sub-

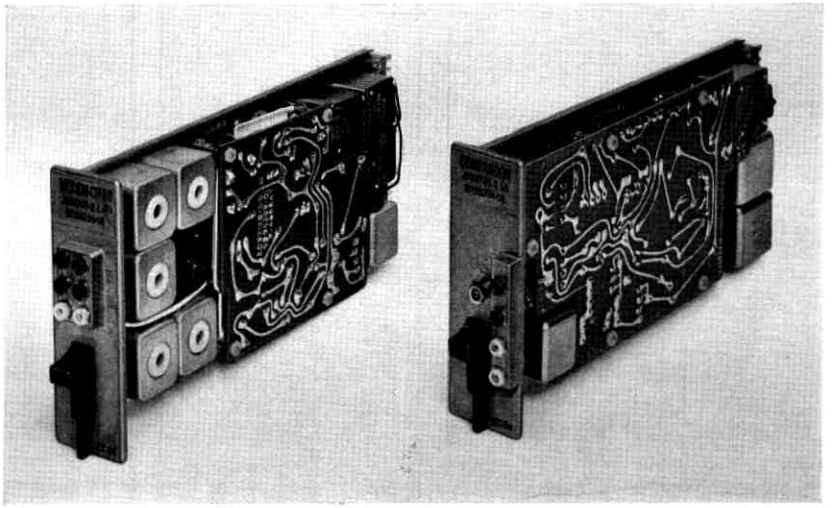

Fig. 9 — Compandor (a) and modem (b) fabricated by "sandwiching" printed wiring board and Amplas assembly.

assemblies, with the interleaving of the components, it was necessary to form the shield to the contours of one of the subassemblies. It was decided to form the shield to the contour of the printed wiring subassembly, the smaller of the two. A rigid vinyl sheet is vacuum formed to the contour of the printed wiring subassembly. Two-ounce copper foil is formed by hand over the formed rigid vinyl sheet, and a second rigid vinyl sheet is vacuum formed over the copper foil. The resulting product is a shield, which consists of a copper foil between two rigid vinyl sheets, and which conforms to the contours of the printed wiring subassembly. The shield is shown in Fig. 10.

The middle shelf of a terminal houses the common units. These are the power supply, alarm and restoral, line terminating, and two group units — either a high-group receiver and a low-group transmitter or a low-group receiver and a high-group transmitter. The group units are double units: that is, they occupy two modular spaces in the terminal housing. The group unit circuity is contained in two AMPLAS subassemblies, each supported on its own die-cast unit frame. The frames are attached to each other and accurately spaced by means of four specially designed brackets. Accurate spacing is essential to enable the two frames to slide in adjacent tracks in the housing without binding. Here again, all intersubstrate wiring is dressed to a common edge so that the units can be opened for servicing. For the sake of

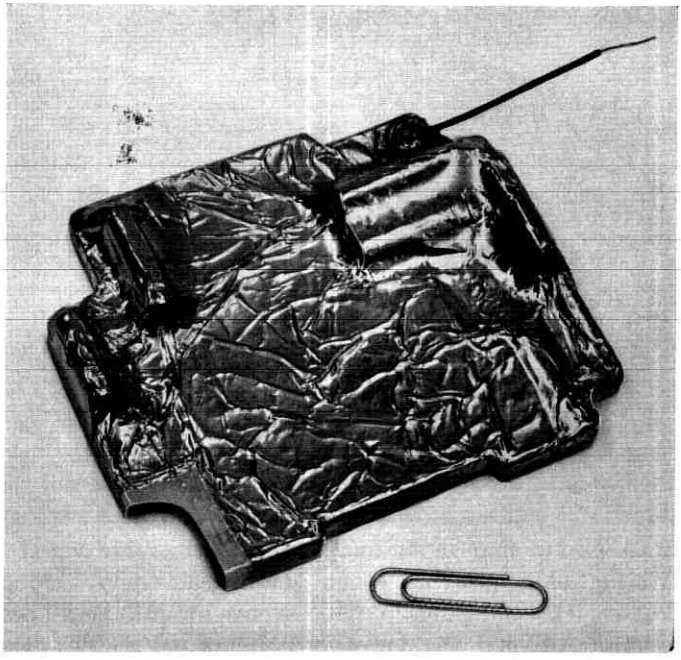

Fig. 10 — Electrostatic shield to isolate subassemblies of modem.

simplicity, the two substrates are referred to as "lower substrate" and "upper substrate," the lower being the one the components of which fall between the two substrates. Plug-in slope equalizers are carried on the lower subassembly, and can be inserted or extracted through the side of a completely assembled group unit. The group units are equipped with only one connector, and it is carried on the lower subassembly.

The line terminating unit, like the group units, is a double-module unit, but unlike the group units houses part of its circuitry in an AMPLAS subassembly and the remainder on an aluminum panel, and is equipped with two connectors. The two subassemblies are electrically independent of one another; there is no intersubstrate wiring. The aluminum panel carries a slide-wire resistor and screw-operated power options for feeding dc power to either one electron tube line repeater or one, two, or three transistorized line repeaters. The AMPLAS subassembly carries the line terminating circuitry. Suspended in the epoxy substrate are two tube sockets into which line build-out span pads can be plugged. The face plate of the unit frame on which the amplas substrate is mounted is equipped with two pairs of paralleled

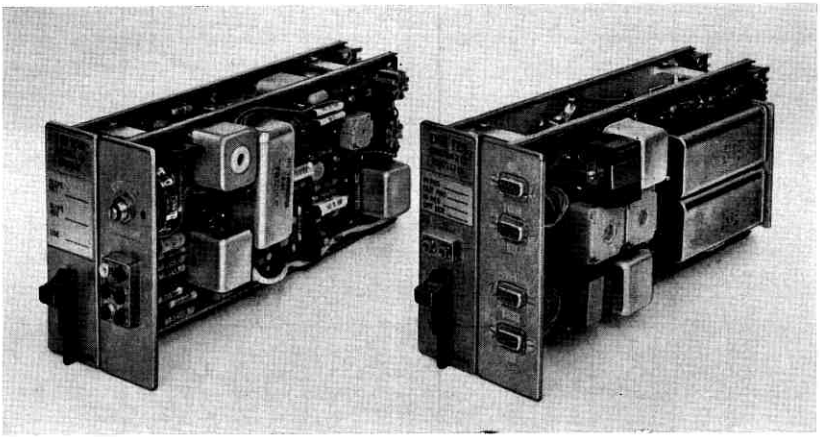

Fig. 11 — Low-group receiving unit (left) and line terminating unit (right).

9-pin jacks. Group unit input and output circuits are wired to the jacks. Mating plugs provide group circuit continuity through the jacks. These jacks and plugs provide access for in-service switching of group units. Although the power feed circuit and the line terminating circuit are carried on separate unit frames and are electrically independent of one another, they are tied together mechanically simply because the four switching jacks on the face plate on the line terminating unit frame leave no room for a latch. The power feed unit frame is equipped with a latch, and when tied mechanically with the line terminating unit frame, the pair is treated like an ordinary double unit. A low-group receiver unit and a line terminating unit are shown in Fig. 11.

The alarm and restoral unit, another double unit, contains its circuitry on printed wiring boards only. Because of the very high component density and the complexity of the printed wiring pattern, the upper subassembly consists of a "mother" board and four "slave" boards. Two of the slave boards contain no components — only wiring — and provide numerous crossovers. The alternative would have been either double-sided printed wiring boards, with the attendant through-connection problem, or many surface wires, always susceptible to operator error. The slave board approach to the crossover problem provides a compact, reliable, easy-to-manufacture solution which is virtually immune to operator error. The alarm and restoral unit is shown in Fig. 12.

Shown in Fig. 13 is the power supply unit. Occupying four modular spaces, it is the largest plug-in unit in the N2 terminal. The circuitry

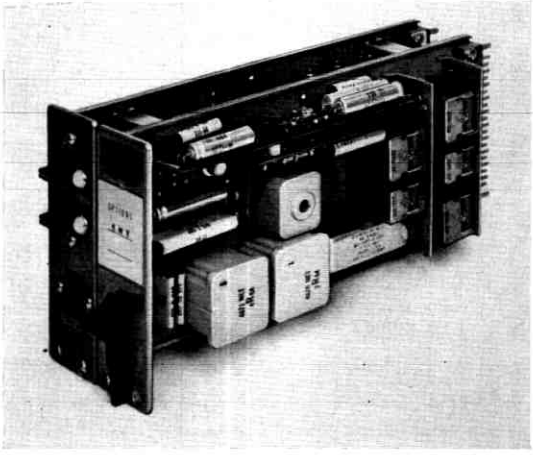

Fig. 12 — Alarm and restoral unit.

is mounted on a steel panel except for the power transistors. These are mounted on a massive heat sink at the connector end of the unit where they are free to generate convection currents.

Laboratory tests have indicated that the N2 terminal is capable of withstanding acceleration forces considerably in excess of the 3g

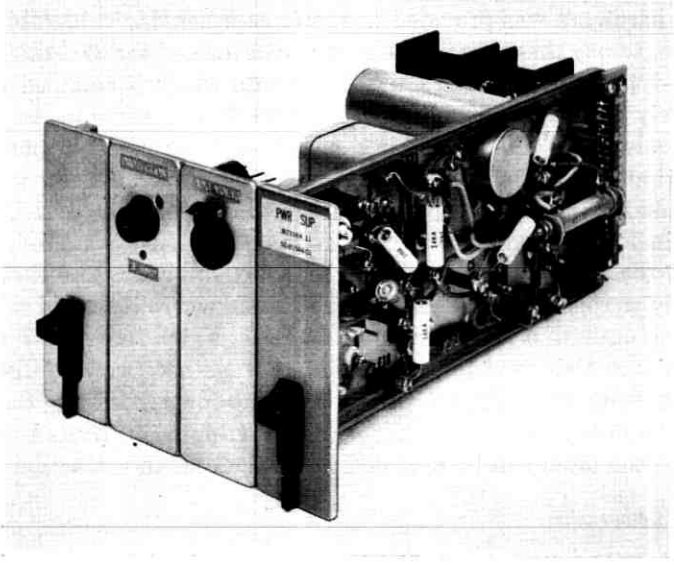

Fig. 13 — Power supply unit.

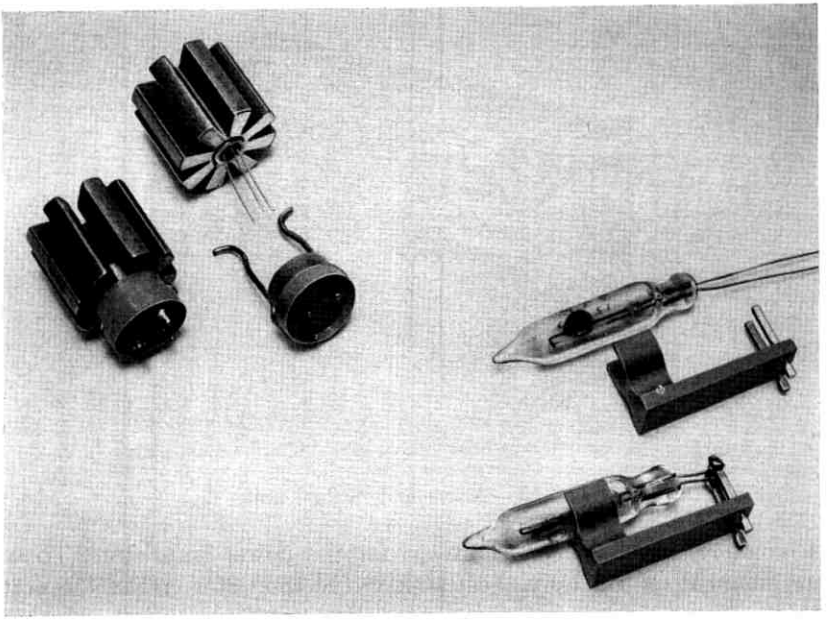

Fig. 14 — Socket and clip for 24-type transistor (left) and plastic clamp for 37C thermistor.

required of all new Bell System equipment. In hardening the equipment, special hardware was provided to reinforce otherwise vulnerable components. Among these are the 24-type transistor and the 37C thermistor. The 24-type transistor has a tiny body onto which is clamped a comparatively massive heat radiator. A number of transistors tore from their leads when the assemblies of which they were a part were vibrated at slightly more than 3g. To overcome this difficulty a special socket and reinforcing clip were designed. The socket and clip are shown in Fig. 14 together with a plastic clamp which was designed to provide the necessary support for the 37C thermistor. The thermistor is simply snapped into the jaws of the clamp, where it is held securely.

Each plug-in unit is equipped with a latch, which serves both to lock the unit into place and to provide the leverage necessary to disengage the unit from the terminal. In the pushed-down position, the latch cannot be operated and the unit is locked in position. In its pulled-up position, the latch can be operated and serves as an extraction tool.

4.3 *Frameworks*

The newly developed double-bay duct-type framework, shown in Fig. 15, as used for all packages containing more than one twelve-chan-

Fig. 15 — Double-bay duct-type framework: (a) front, (b) rear.

nel terminal. Use of a double-bay rather than a single-bay framework is more economical and permits much more efficient utilization of space. The frame uprights are drilled for mounting plates both front and rear. This affords more efficient utilization of bay space through location of some units in otherwise unused space in the rear of units mounted on the front. The frame will be shipped and installed without plug-in units, so the extra size and weight are not objectionable.

The five-inch deep uprights, with wide flanges in front and narrow flanges in the rear, form cable ducts at the sides and in the middle of the frame. The narrow flanges afford access for shop wiring or installer cabling. Low-level wiring is placed in the middle duct and is automatically shielded from the high-level or noisy wiring in adjacent bays or in the outside ducts of the frame. The wide front flanges not only

afford duct space, but also permit the use of assignment card holders adjacent to shelves for plug-in units or jack strips as shown in Fig. 17 (Section 4.4).

The single-bay 7-foot framework used for the twelve-channel packaged terminal is similar to the double-bay except that the cable ducts are between frames and automatic shielding is not feasible. The bases of the frames include guard rails and commercial power outlets both front and rear for convenience in powering maintenance and installation equipment. The bases and the upper sections of the frameworks are designed for ready attachment of tools to facilitate handling in the shop and in the telephone office during installation. These tools permit upending the frame and rolling it into position between other frames separated by only nominal tolerances.

4.4 *Packaged Terminal Frames*

Fig. 16 is a photograph of the six-terminal (72-channel) packaged terminal frame which is diagrammed in Fig. 17. The main components of a twelve-channel N2 terminal — the carrier units, the CGA and E-signaling units — are located as close to one another as economy of space and shelf capacities permit. For example, positions for channel units 1A to 1M, CGA No. 1, E-type signaling unit positions 1A to 1M, all in the lower part of the frame, and the lowest of six horizontal jack strips in the patching and monitoring jack field, are all components of terminal No. 1.

The patching and monitoring jack field includes a horizontal group of patching and monitoring jacks and associated terminal alarm lamps and keys for each twelve-channel terminal. Monitoring and talking jacks, keys and lamps; transmission and noise measuring control switch, keys and lamps; and trunk jacks and lamps, when required, are mounted at one end of the horizontal groups mentioned above.

4.5 *Carrier Group Alarm (CGA) and Associated Signaling Shelves*

Fig. 18 shows how the CGA is designed and located to take advantage of otherwise unused space and to permit short wiring runs between the terminal blocks, the CGA relays, and the majority of the connectors for the signaling units. This is not only economical, but also affords minimum crosstalk and noise exposure for the numerous critical leads which have been sources of trouble in other installations. The CGA relay and register strip is located immediately below the lower mounting of the shelf which mounts ten signaling units associated with the

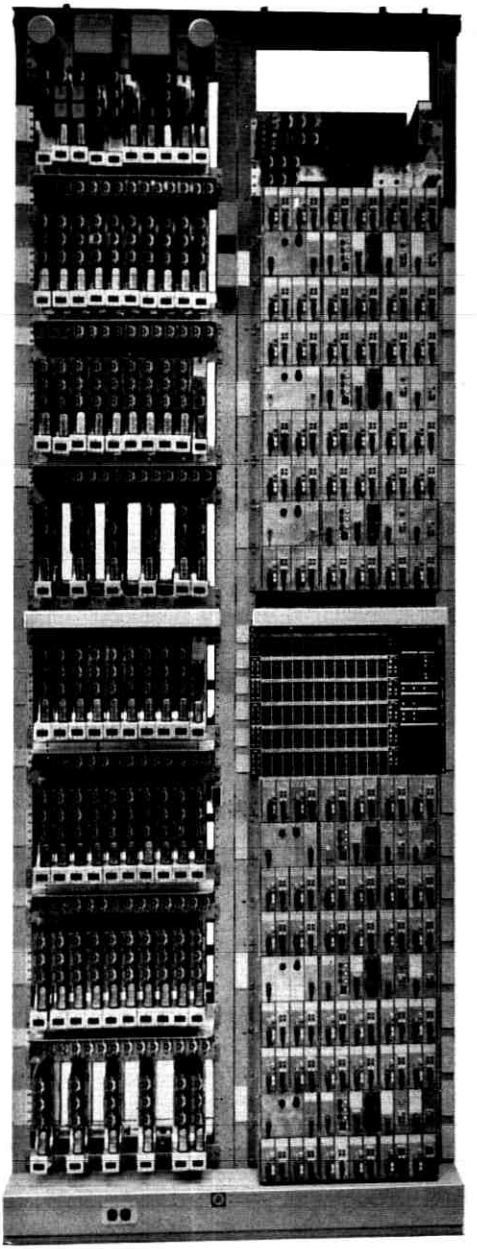

Fig. 16—Front view of six-terminal (72-channel) packaged carrier terminal frame.

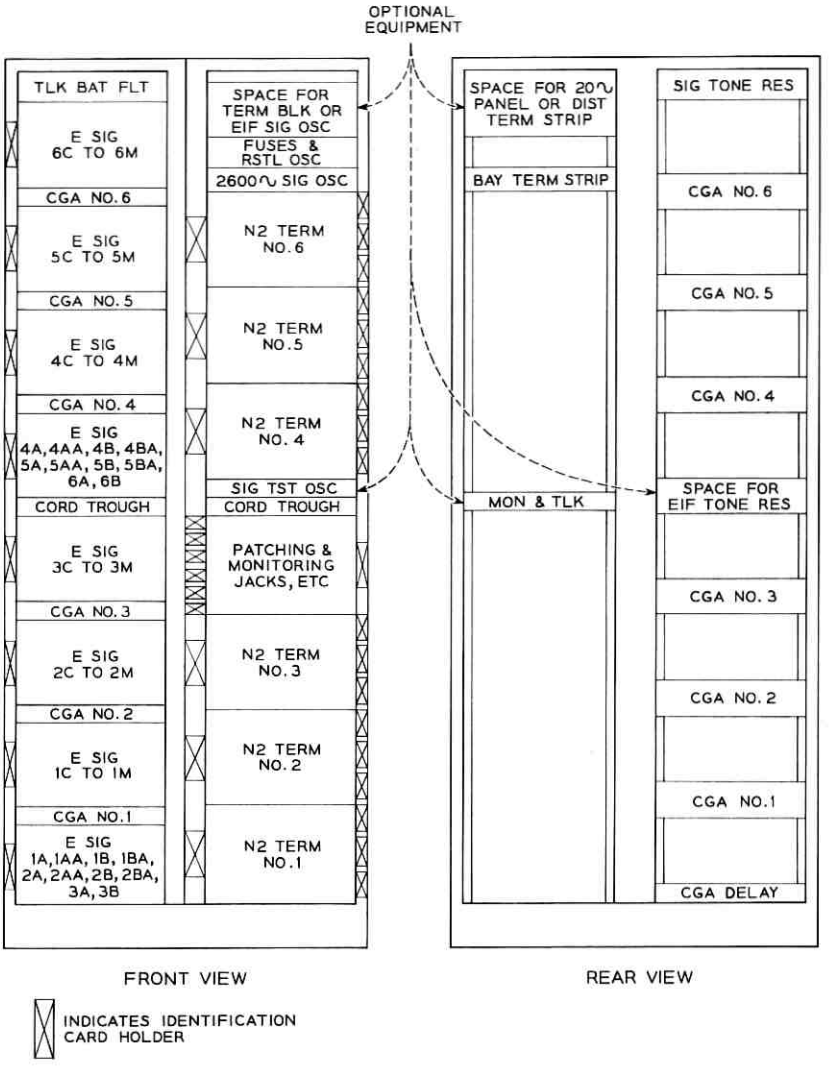

Fig. 17 — Equipment arrangement of 72-channel packaged six-terminal frame shown in Fig. 16.

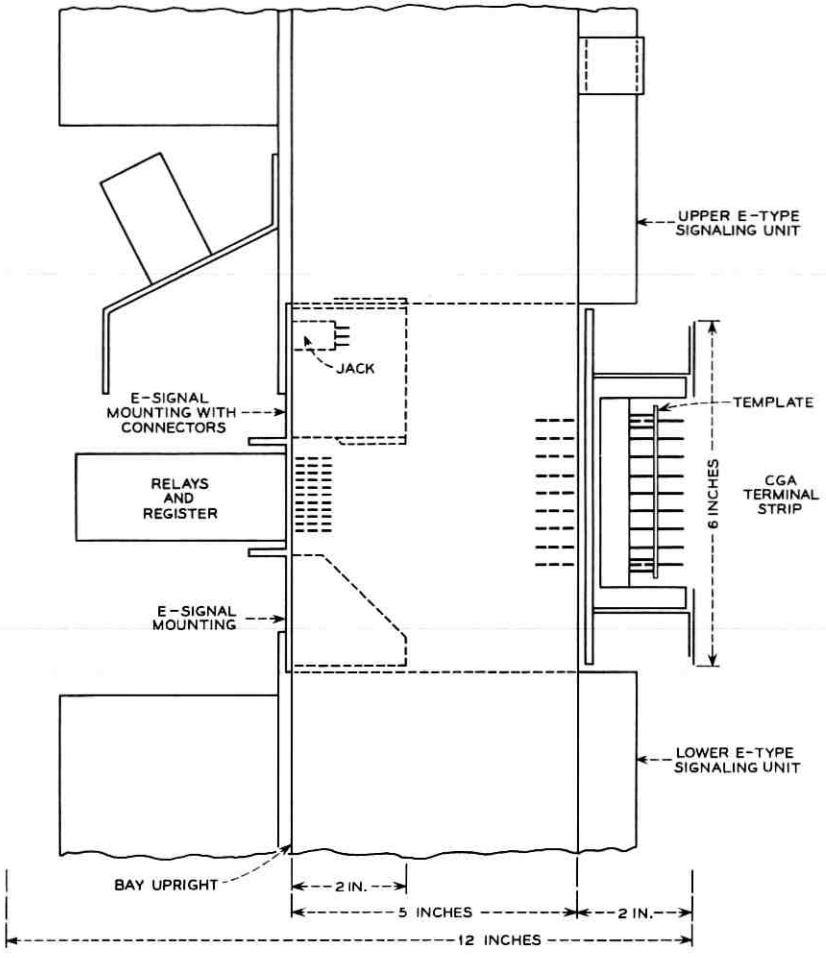

Fig. 18 — Carrier group alarm and E-type signaling shelves, showing arrangement of equipment.

CGA, and immediately above the upper mounting of another shelf. The terminal strip and cabling of the CGA occupy the space behind the two signaling shelf mountings and the relay and register strip. The terminal strip is hinged to give wiring and maintenance access to the relay and register terminals and the shop-wiring side of the terminal strip.

Fig. 19 shows how each CGA terminal strip is divided into twelve regular channel blocks, three spare channel blocks, and a miscellaneous

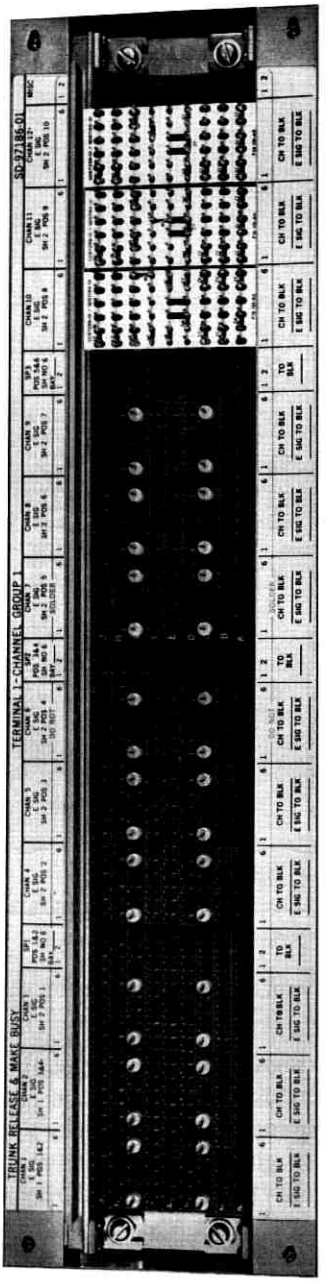

Fig. 19 — Arrangement of carrier group alarm terminal strips and wiring templates.

block, thereby providing for the twelve channels of an N2 terminal and for three spare signaling channels which may be used for circuits requiring auxiliary as well as regular signaling units, i.e., two signaling unit positions per channel. These auxiliary signaling units and those for the A and B channels cannot be accommodated on the nearby signaling shelf and are wired with shop cabling to connectors on other signaling shelves (refer to Fig. 17). The shelves for the A and B channels are located in the same frame as the CGA, but the spares may or may not be in that frame. Fortunately, the spare or auxiliary unit channels have less critical noise and crosstalk requirements.

The regular CGA channels may be used with any one of eleven different signal units, program or noncompandored channels and with a wide variety of office trunk circuits. Accordingly, the CGA's are universally wired so that any condition may be established by suitable straps between terminals on the terminal blocks. This includes optional straps to tone supply resistors for E1F signaling units and for 20-cycle ringing supply which may be required for E1S units.

Only wire-wrapped connections are used on the side of the CGA terminal blocks where the optional strapping is done. Solder connections are avoided because a terminal once soldered cannot be wire-wrapped. Option changes on a working CGA are often required because it is impracticable to "turn down" twelve channels when one is reassigned. Soldering of option straps on a working CGA unit would constitute an unnecessary service hazard. Changes in wire-wrapped connections may be made with negligible hazard because the tool is insulated and pliers are not required.

Plastic templates for use with the CGA unit afford a simple guide for applying optional strapping between terminal block terminals to provide the various trunk circuit and signaling features for the different channels. These templates are perforated so they may be slipped over the terminals of the CGA terminal block of the desired channel after installer wiring has been connected, but before the option straps are placed. Each template is designated as to the channel use, the associated E-type signaling unit, and the trunk condition to which it applies. As illustrated in Fig. 19, the desired straps are indicated by heavy lines and option letters between the terminals which must be strapped. When a new or different type signaling unit or trunk circuit is to be used, existing straps must be removed using an unwrapping tool, the existing template removed, the new template added, and then new straps following the option lines will be added using a hand-actuated or power-driven wrapping tool. Visual inspection at any time

can determine whether the template, strapping, signaling unit, and trunk circuit have been coordinated.

V. MAINTENANCE AND LINE-UP FEATURES

Maintenance and line-up are considered together because many of the provisions for one are used for the other and both are performed by operating company personnel. Line-ups may be a part of routine maintenance but they also occur when systems or channels are first put into service or are rearranged for changes in type of service. They include the choice of plug-in units and the administration of circuit options as well as the adjustment of controls.

5.1 *Portable Test Stands and Switching Sets*

All twelve channels of the N2 terminal depend on the proper operation of the group units and the power supply unit. Therefore it is essential to provide the means for routine maintenance of these units without interrupting service. Accordingly, portable switching sets capable of effecting in-service switching of group units, power supply units, and wideband 40.8-kilobit data channel units which operate in conjunction with the 301-type DATAPHONE data sets are available to the operating companies. The switching set is powered from the power receptacle on the face plate of the alarm and restoral unit.

The routine for in-service switching of group units begins with the removal of one of the plugs from either of the paralleled receiving 9-pin jacks or either of the paralleled transmitting 9-pin jacks (depending on which group unit is being switched) on the face of the line terminating unit. The group circuit is maintained intact by the remaining plug. The group-switching cord of the switching set is plugged into the vacated jack, and an alternate group unit is plugged into the switching set. At this point, the switching set and the remaining plug provide parallel paths for the group circuit. Now, upon removal of the plug, group circuit continuity is maintained by the switching set. Controls are provided so that the output of the alternate group unit in the switching set can be adjusted to the level of the regular group unit in the in-service terminal. When the levels are equal, a switch is operated, removing the regular group unit from service and simultaneously inserting the alternate unit. The regular group unit is now removed from its position in the terminal and a spare regular group unit inserted in its place. The output level of the alternate group unit is readjusted, if necessary, to match the output of the unit now in the

terminal shelf. The switch is returned to its original position, removing the alternate group unit from service and simultaneously inserting the spare regular unit. The plug is replaced in the line terminating unit, the group switching cord is removed, and the second plug is replaced, completing the process. In-service switching of group units produces a 1-db hit per channel on the system.

In-service switching of power supply units is completely hitless. The procedure begins by powering the switching set from the power receptacle on the face of the alarm unit. An alternate power supply is plugged into the switching set. Controls are operated, adjusting the output voltage of the alternate power supply to the level of the regular power supply. A switch is thrown, removing the regular supply from service and simultaneously inserting the alternate supply. The regular unit is removed from its position in the terminal, and a spare unit is inserted into the vacated position. The output voltage of the spare unit is adjusted, the switch returned to its normal position to transfer the load to the new regular unit and the switching set is disconnected from the terminal. The switching set is shown at the left in Fig. 20.

A portable test stand — capable of routine tests on channel units, the alarm and restoral unit and the restoral oscillator unit — is also available to the operating companies. The test stand serves as an extender, providing convenient access to each contact of any plug-in unit connector into which it is plugged. Tracks and connectors are provided into which one or more plug-in units can be inserted. This permits testing or trouble shooting the units in question in a convenient fixture while the units are electrically connected to the terminal circuit. Jacks, keys, lamps, switches and fuses provide the capability for a number of tests, including a compandor-modem loop-around test, permitting one end of a channel to be tested by itself.

Alarm and restoral units and restoral oscillator units can be tested and adjusted in the test stand without connection from the stand to the alarm or oscillator positions in the terminal. These tests require only a power connection to an N2 terminal. These plug-in units may be replaced with spares during the tests or the terminal may be operated without them, providing a failure requiring circuit conditioning does not occur. The test stand is shown at the right in Fig. 20.

The performance of signaling units is monitored by a portable set which may be plugged into a connector located on the face of each signaling unit. Adjustment and minor repairs are made with the unit removed to a more elaborate test set.

Signaling tone oscillators are furnished in pairs so that either may

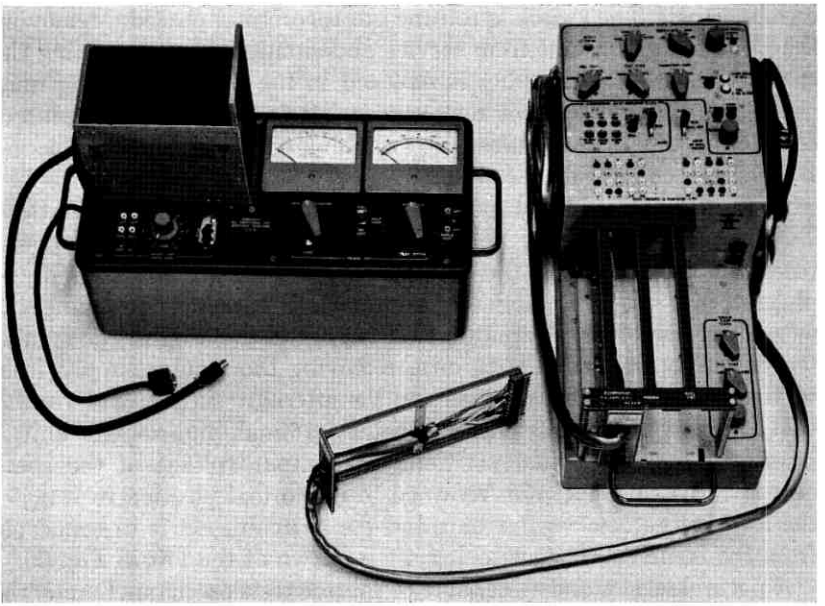

Fig. 20. — Portable switching set (left); test stand (right).

carry the entire load while the other is disabled or switched out for maintenance. The oscillator panels are equipped with transfer circuits, controls and jacks for maintenance adjustment.

Jacks are provided for headsets used in talking over N-carrier order circuits or idle channels to the distant terminal.

5.2 *Patching and Monitoring*

Patching and monitoring jacks which provide voice-frequency access to the various channels are furnished to permit circuits assigned to individual channels of one system to be routed over channels of another system during carrier line failures. These jacks may also be used for monitoring or for terminated measurements both toward and away from the carrier system, although all essential maintenance may be done without them, as indicated in Section 5.1. When centralized transmission and noise measuring equipment is available, it may be connected to these jacks and controlled by circuits and equipment associated with the patching jacks. This affords somewhat faster and more accurate circuit line-up than is possible with the portable equipment.

The monitoring jacks associated with the patching jacks may be connected to a high-impedance four-wire monitoring set which not only affords two directional monitoring without service impairment but also provides amplification for monitoring at a comfortable level and for the detection of low-energy interference.

This monitoring option also includes a telephone set and intraoffice communication circuit terminations for maintenance purposes.

VI. ENGINEERING AND INSTALLATION FEATURES

Many features can be provided to simplify engineering and expedite installation. They may increase or decrease manufacturing and merchandising costs. Accordingly, a prime objective of equipment design is an economic balance between manufacturing, engineering, and installation costs in light of the probable percentages of installations requiring various options. In general, shop wiring is much less expensive, more uniform and more readily tested than installer wiring.

Certain options, such as the use of E1F signaling units for revertive pulsing, are seldom required and involve considerable equipment and wiring which must be coordinated with the wiring for other options. In the N2 packaged frame, wiring terminals for administering this option are always provided, but the equipment and wiring is furnished only when specified. Preferably it is installed in the shop, but it may be added in the field.

As other options, such as monitoring and talking circuits, are required in an appreciable number of N2 terminals, the wiring is always furnished but the equipment may be added in the shop or field when required.

The broadband data option is seldom required when a terminal is initially installed, but is expected to be required in increasing numbers during the life of the equipment. The conversion to broadband data transmission involves simple but critical wiring and special terminal equipment. The wideband data terminal equipment consists of plug-in units which may be provided as needed, but the terminal wiring is always provided and terminated on the frame.

6.1 *Basic Plan*

Each packaged terminal frame includes plug-in unit mountings and common equipment suitable for any circuit option for a number of terminals. The complete frame is shop wired to terminal strips so that the frame wiring may be continually checked in the shop on a

machine routined basis. Installers' connections are made to terminal strips rather than apparatus terminals. This not only minimizes installer wiring and testing costs but also affords close control of critical wiring. The range of telephone office ceiling heights results in frame heights of 11 feet, 6 inches, 9 feet and 7 feet, with corresponding numbers of terminals per frame of 6, 4 and 3, respectively.

The N2 packaged terminal meets the A.T.&T. Co. objective of office engineering and installation of a sufficient number of circuits to permit economical anticipated growth over a reasonable period of time. Major equipment expenditure may be deferred until service needs require specific plug-in units. The initial installation suffices until growth exceeds the anticipated number of terminals. Initial cabling to the distributing frame is suitable for all circuit options. Distributing frame space requirements and the size of installer cables are minimized by the use of the same leads for various options. This is made feasible by provision of terminal strips and wiring options at the packaged terminal and optional cross connections to trunk circuits at the distributing frame. Both types of optional wiring are normally made by operating company personnel.

An interesting variation occurs in 7-foot packaged terminal frames for 3 terminals because the common equipment for the first frame is adequate for a second frame. Accordingly, alternate frames do not contain common equipment but are furnished with factory-connected cable for installer connection to an adjacent frame. Connections to the distributing frame are the same as those for other terminals.

Additional cabling to and from the packaged terminals is required when wideband data channels are desired but additional frame wiring is not required. The additional external cabling is engineered and installed as a part of the wideband data project.

In some of the larger offices where centralized patching jacks are furnished for other systems, the telephone company may elect to use jacks in the centralized location rather than in the packaged terminals. An 11-foot, 6-inch frame for six terminals with terminal strips instead of jacks is available for this choice. It obviously requires greater engineering and installation effort as well as more expense. These terminals include all the other features of packaged terminals.

A single-terminal 7-foot frame is available for installation where more than one terminal will not be required in the foreseeable future. Engineering and installation are exceedingly simple, but per-terminal costs are high because there are no terminals to share the expense of common equipment.

6.2 *Power and Alarm*

The packaged terminals include fuses and alarm circuits suitable for individual or multiple connection to the power distribution boards and office alarms of all types of offices, including No. 1 ESS. Busy hour power requirements and permissible feeder voltages are specified so that the engineering of power plants and power feeder cables to the frames is simplified. Optional connections permit operating company choice of alarm division between major and minor alarms, if the normal arrangement is not desired. Special engineering or additional apparatus is not required for the alarm circuits of different types of offices.

6.3 *Optional Features*

Most optional features may be included in the original specifications and furnished in the shop without the need for additional installation. Others, such as protector blocks for distributing interoffice cable pairs, may be mounted in the packaged terminal frame and connected by the installer. Any of the options may be engineered and installed at a later date if unforeseen need arises.

Options requiring engineering and installation consideration include:

(a) E1F revertive signaling,
(b) E1LA or E1SA auxiliary signaling,
(c) 20-cycle supply for E1S signaling,
(d) 2000- or 2400-cycle oscillator for testing certain signaling units,
(e) interbay transmission trunks,
(f) 4-wire monitoring and intraoffice talking trunks, and
(g) centralized transmission and noise measuring equipment.

In each case, engineering and installation are simplified by the provision of equipment codes and space assignments on packaged terminal frames. However, some auxiliary signaling units will require space outside the packaged terminals. Shelves for this purpose include connector-terminated cables, the free ends of which are installer-connected to terminal strips in the packaged terminals. No other connections are required.

6.4 *Telephone Office Considerations*

Telephone offices often differ from one another in many important aspects, including ceiling heights, power facilities, alarm circuits, cabling plans, grounding practices and noise environs. N2 carrier terminals are intended for installation in many types of offices—frequently

in some, and rarely in others. Accordingly, packaged N2 terminals are designed with the built-in capability of being used in any office without special engineering or equipment except for conditions that are rarely encountered.

The packaged N2 terminals are capable of satisfactory operation from near freezing to the maximum normal telephone office temperatures. Extremely low temperatures are sometimes encountered in community dial or other offices, and under these conditions some transmission impairment is permissible.

VII. CONCLUSIONS

Recent technological advances were used to advantage in the design of the packaged N2 carrier terminal. Solid-state devices, improved circuitry and modern packaging techniques were combined to produce a terminal which provides many advantages heretofore unavailable. It is felt that the features described in this paper are of considerable importance to the Bell System, and will earn for the packaged N2 carrier system a large share of the short-haul carrier market.

REFERENCES

1. Boyd, R. C., and Herr, F. J., The N2 Carrier Terminal — Objectives and Analysis, B.S.T.J., this issue, p. 731.
2. Lundry, W. R., and Willey, L. F., The N2 Carrier Terminal — Circuit Design, B.S.T.J., this issue, p. 761.
3. Weaver, A., and Newell, N. A., In-Band Single-Frequency Signaling, B.S.T.J., *33*, Nov., 1954, p. 1381.

Miniaturized *RC* Filters Using Phase-Locked Loop

By G. S. MOSCHYTZ

(Manuscript received September 15, 1964)

It is shown that an automatic phase-locked loop (APLL) can be used as a bandpass filter and FM discriminator while satisfying the circuit conditions imposed by microminiaturization techniques. It has the advantage over other methods of ease of adjustment and reduction in the number of circuit components. For this reason it is to be assumed that the APLL will be added to the few practical solutions of the frequency selection problem in microminiaturization available to date.

In contrast to most other applications, no assumptions can be made here about the ratio of the natural undamped system frequency to the lock range of the loop. This is taken account of in an analysis of the APLL which enables design equations and considerations to be derived. An approximation of the so-called capture ratio, defined as the ratio of capture-to-lock range, is presented. It produces results that correspond very well with the equivalent measurements.

The description of an APLL designed as RC channel filter and discriminator for a multichannel data receiver is presented. RC active circuits that are compatible with current microminiaturization techniques are employed exclusively. They are described in detail, and an analysis of the voltage-controlled oscillator used is included in the Appendix.

I. INTRODUCTION

One of the main problems in the microminiaturization of frequency-selective circuits is the considerable deterioration of Q of magnetic components when these are reduced in size to an extent comparable with other miniaturized components.[1,2,3] For the design of highly selective bandpass filters and frequency discriminators, completely new circuit design concepts that eliminate magnetic components and overcome other limitations imposed by current microminiaturization techniques must therefore be employed. At low frequencies they invariably

lead to the synthesis of narrow-band active *RC* filter circuits that are compatible with the state of the microminiaturization art. To this end a number of new circuit techniques have been described in the literature, but practical solutions to the problem are still few and not wholly satisfying.[4]

In this paper a new approach in this direction has been taken in the field of FM multiplex signal filtering by utilizing both the filtering properties of the APLL and the possibility of designing its circuits along the lines prescribed by present-day miniaturization technology. The resulting network, characterized by its simplicity, ease of adjustment and stability, shows that the APLL is a very useful filtering device in the field of circuit miniaturization. At the same time, if so desired, it also serves as a frequency discriminator.

For the purpose of introducing the APLL into the area of circuit miniaturization and to facilitate its implementation the characteristics and performance of the APLL covered in present literature are briefly reviewed. In designing the APLL for the specific function of signal filtering, the capture ratio, defined as the ratio of capture-to-lock range, has to be considered in detail. The exact capture range is given by the solution of a second-order nonlinear differential equation that is only soluble graphically by phase plane methods. To make it available in more tangible form to the circuit designer, a relatively simple yet sufficiently accurate approximation for the capture range is derived.

Design details pertinent to the realization of the APLL as an active *RC* network are given, and considerations involving the transition from discrete *RC* to thin film or integrated circuits presented. Finally, to demonstrate the effectiveness of the APLL in the field of *RC* frequency-selective circuits and to illustrate its implementation, the circuit design of an APLL employed as combined FM filter and discriminator in a multichannel data receiver operating in the audio-frequency range is described in detail. The circuits were designed to satisfy the circuit restrictions imposed by miniaturization techniques, in particular by that of tantalum thin film technology.

II. THE AUTOMATIC PHASE LOCKED LOOP

A block diagram of the automatic phase locked loop (APLL) is shown in Fig. 1. A signal corresponding to one of numerous incoming channels, which is frequency modulated about the channel center frequency ω_c, is assumed at the input. Here it is multiplied by the output of a voltage controlled oscillator (VCO) whose frequency is also centered at ω_c.

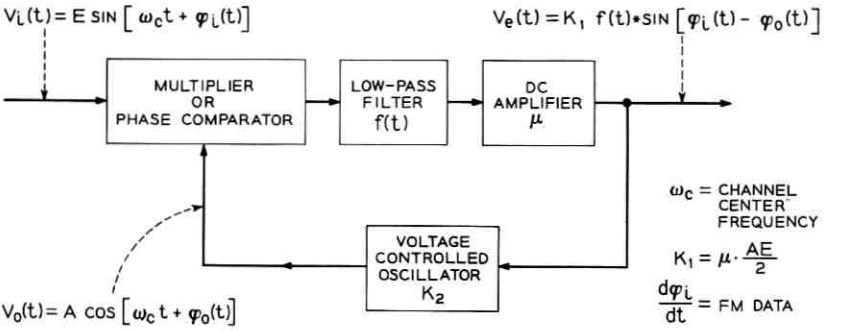

Fig. 1 — Functional block diagram of APLL.

The product is passed through a low-pass filter such that only a voltage proportional to the instantaneous phase difference, in other words to the integral of the frequency difference between the two multiplier inputs, remains. This voltage is amplified, controls the frequency of the VCO and can be considered as the output of a frequency discriminator since it is a measure of the input frequency. The multiplier serves here as a phase comparator. Furthermore it is assumed here that the frequency deviation at the input is limited to that characteristic range of steady-state frequencies within which tracking by the VCO is possible. This frequency range is called the "lock range" of the APLL, and will be discussed in more detail further on.

The frequency integration by the phase comparator results in zero error between the frequencies of the input signal and the VCO output signal in the steady state. It is maintained by a finite error voltage which is proportional to the phase difference between these two signals. This integration results in bandpass properties of the APLL with respect to noise and to interference from neighboring channels. They can be modified by the type of low-pass filter incorporated in a given loop, and are of particular significance for the application in question here.

The APLL can be designed to perform not only as a frequency demodulator but simultaneously as a selective device in the form of an FM signal separator. The extent to which either or both of these two properties is emphasized depends on the distribution and total value of the loop gain as well as on the parameters of the low-pass filter.

It will prove useful in the following discussion of APLL characteristics to apply conventional control theoretical terminology such as loop-gain, transfer function, phase margin, or cutoff frequency. The definitions

of these terms, which are somewhat unconventional, will be given as they appear in the following analysis.

2.1 *The Lock Range*

Assuming a phase comparator whose output is proportional to the sine of the error phase angle, the error voltage $v_e(t)$ after passing through the low-pass filter (see Fig. 1) is given by:

$$v_e(t) = K_1 f(t) * \sin [\varphi_i(t) - \varphi_0(t)] \tag{1}$$

where

$$K_1 = \mu(AE/2) \text{ volts}$$

and the $*$ denotes convolution. K_1 represents the sensitivity of the phase comparator multiplied by the gain μ of the amplifier, and $f(t)$ is the impulse response of the low-pass filter.

The VCO frequency ω_0 is assumed to be a linear function of the error voltage. This condition is not generally difficult to fulfill for a limited deviation ratio. With the same center frequency ω_c as that of the incoming signal, the VCO frequency is then given by:

$$\omega_0 = \omega_c + \frac{d\varphi_0}{dt}. \tag{2}$$

With K_2 in radians/(sec, volt) given as the voltage sensitivity of the VCO:

$$\omega_0 = \omega_c + K_2 v_e(t). \tag{3}$$

Substituting (1) for the error voltage and solving for the FM input signal:

$$\frac{d\varphi_i(t)}{dt} = \frac{d\varphi}{dt} + Kf(t) * \sin \varphi(t) \tag{4}$$

where

$$K = K_1 K_2 \text{ rad/sec.}$$

K is a system parameter corresponding to loop gain in conventional control theory. It equals the VCO frequency shift occurring at phase errors of $\pm \pi/2$.

For unity dc gain of the low-pass filter and an input frequency deviation from the center frequency of $\Delta\omega_i = d\varphi_i/dt$, the steady-state solution of (4) is given by:

$$\sin \varphi = \Delta\omega_i / K. \tag{5}$$

The sine of the static phase error of the APLL is therefore proportional to the input frequency deviation $\Delta\omega_i$ and inversely proportional to the gain constant K. The system locks as long as

$$| \Delta\omega_i | \leqq K. \tag{6}$$

Only within this range does (4) have a steady-state solution and a zero frequency difference $d\varphi/dt$; thus the oscillator lock range is equal to the gain constant K.

2.2 *The Capture Range*

The capture range defines the range of input frequencies to which the VCO can be synchronized when initially in the unlocked state. With an idealized low-pass filter network in the loop that rejects the high-frequency components of the phase comparator output and passes the low-frequency components unattenuated, the capture and lock ranges coincide. Modifications of this idealized case that narrow the system bandwidth with respect to the lock range can reduce the capture range appreciably.

For the application being considered here, the reduction in capture ratio (defined as the ratio of capture-to-lock range) due to narrowing of system bandwidth has to be taken into account. If, namely, the system is both to lock rapidly onto any initial frequency within the specified input frequency deviation as well as to relock rapidly after extraneous perturbations (e.g., impulse noise) might have thrown the system out of lock, the capture range has to cover the maximum frequency deviation. It is this last requirement that limits the attainable selectivity of a system all of whose parameters except those of the incorporated low-pass filter are given. However, substituting the impulse response of a generalized, or even of a simple, low-pass network in (4) and solving for the capture range is only possible by graphical phase plane methods.[5,6] The resulting second-order differential equation is nonlinear and becomes indeterminate at the values of input frequency for which the system just returns to lock. It is possible, however, to make an approximate analysis of the capture process that gives a good estimate of the capture range for a specified filter network and also some insight into the mechanism of the APLL at the time it recaptures the locked condition. The minimum attainable bandwidth compatible with a given frequency deviation thus results.

As long as the APLL is locked and in its steady state (e.g., constant

input frequency) the frequency response of the low-pass filter incorporated in the loop has no influence on the behavior of the loop. This is governed entirely by the loop gain K which defines the lock range. Out of lock, the VCO is modulated by a periodically varying signal. It therefore produces an ac error voltage at the input terminals of the low-pass filter which is subsequently attenuated according to the transfer function of the filter. This has an effect equivalent to attenuating the loop gain and with it the capture range, resulting in a capture-to-lock ratio smaller than one.

The transfer function of the low-pass filter can be expressed by

$$F(j\omega) = F_\omega e^{j\psi(\omega)} \tag{7}$$

where F_ω and $\psi(\omega)$ respectively denote the frequency and phase response. With the loop open at the VCO input terminals the error voltage for an input deviation $\Delta\omega_i$ is given by

$$v_e(t) = K_1 F_{\Delta\omega_i} \sin \Delta\omega_i t \tag{8}$$

whose peak value equals $K_1 F_{\Delta\omega_i}$. At the capture frequency $\Delta\omega_{ic}$, the corresponding peak error voltage therefore equals

$$\hat{v}_{ec} = K_1 F_{\Delta\omega_{ic}}. \tag{9}$$

By definition of the capture frequency this peak voltage is just large enough to produce the VCO frequency $\Delta\omega_{oc} = K_1 K_2 F_{\Delta\omega_{ic}}$ such that $\Delta\omega_{oc} = \Delta\omega_{ic}$ when the loop is closed. In other words, at this peak error voltage the difference frequency between input and VCO becomes zero, thus enabling the loop to go into its locked state. To do so the peak error voltage must take on its equivalent steady-state dc value

$$v_{ec} = K_1 \sin \varphi_c \tag{10}$$

which it does within the so-called "capture time" T_c. This time is in turn a function of loop parameters and initial conditions (see below). According to (5), (10) can also be written

$$v_{ec} = K_1 (\Delta\omega_{ic}/K). \tag{11}$$

Equations (9) and (11) give the peak ac and steady-state dc error voltages corresponding to the unlocked and locked states respectively. Both produce the same VCO frequency $\Delta\omega_{oc}$ such that $\Delta\omega_{oc} = \Delta\omega_{ic}$. Thus both voltages must be equal, and (9) and (11) can be combined to give an approximate expression for the capture frequency $\Delta\omega_{ic}$ in

terms of filter attenuation and loop gain, namely

$$\Delta\omega_{ic} \approx KF_{\Delta\omega_{ic}}.\dagger \tag{12}$$

As is briefly explained further on, this is an approximation inasfar as the frequency $\Delta\omega_{ic}$ in the term $F_{\Delta\omega_{ic}}$ of (9) is actually not identical but somewhat smaller than $\Delta\omega_{ic}$ in (11).

For an idealized low-pass filter for which $F_\omega = 1$, (12) gives the capture range $\Delta\omega_{ic} = K$, as is to be expected. For a single low-pass filter, as shown in Fig. 2(a), the attenuation as a function of frequency is given by

$$F_\omega = 1/[1 + \omega^2 T_{co}^2]^{\frac{1}{2}} \tag{13}$$

where $T_{co} = RC$. Substituted in (12), we get

$$\frac{\Delta\omega_{ic}}{K} \approx \left[\frac{\omega_{co}}{K}\left\{1 + \frac{1}{4}\left(\frac{\omega_{co}}{K}\right)^2\right\}^{\frac{1}{2}} - \frac{1}{2}\left(\frac{\omega_{co}}{K}\right)^2\right]^{\frac{1}{2}} \tag{14}$$

where ω_{co} is the 3-db cutoff frequency of the low-pass filter and equals $1/T_{co}$.

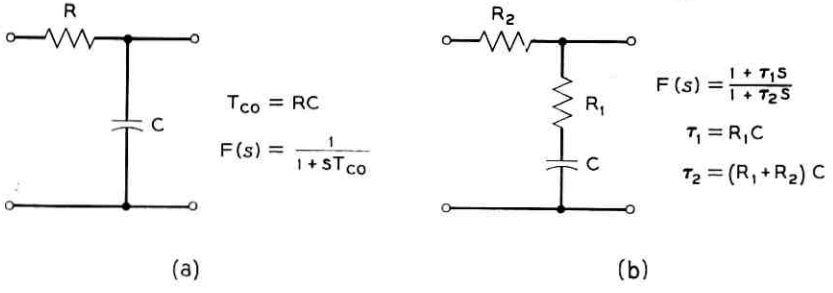

Fig. 2 — (a) Low-pass filter; (b) lag network low-pass filter.

Fig. 3 displays (14) graphically. Clearly, a simple low-pass filter in the APLL reduces the capture ratio if it attenuates frequencies within the lock range. The capture ratio decreases with the filter cutoff frequency ω_{co} and approximately equals the square root of the ratio of filter cutoff frequency to loop gain, when this ratio is small.

For a low-pass lag network as shown in Fig. 2(b), the attenuation as a function of frequency is given by

---

† It was brought to the author's attention that this expression was also derived by J. A. Narud in unpublished work.

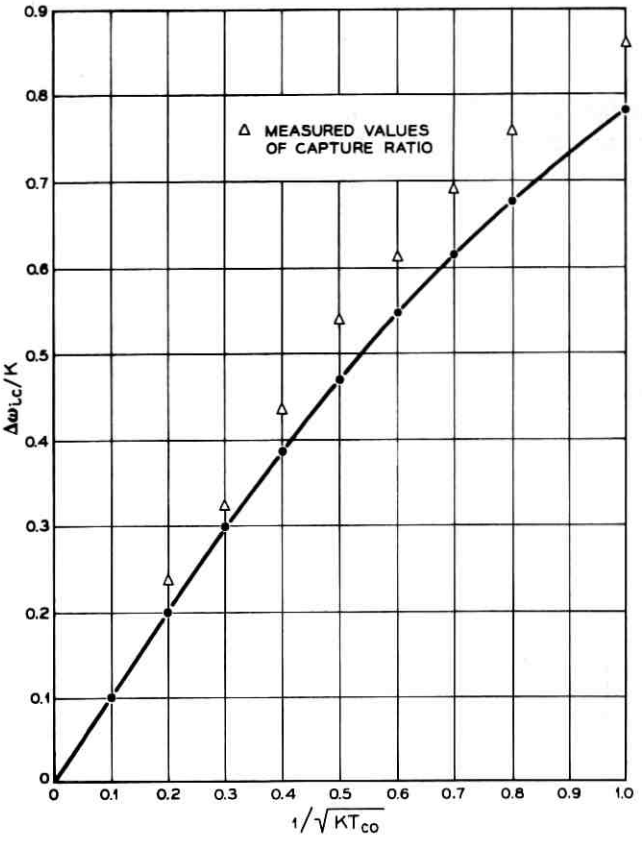

Fig. 3 — Capture range versus low-pass filter bandwidth.

$$F_\omega = \left[\frac{1 + \omega^2\tau_1^2}{1 + \omega^2\tau_2^2}\right]^{\frac{1}{2}} \tag{15}$$

where

$$\tau_1 = R_1C \tag{15a}$$

and

$$\tau_2 = (R_1 + R_2)C. \tag{15b}$$

The following substitutions may be made

$$2\zeta\frac{\omega_n}{K} = \frac{1 + K\tau_1}{K\tau_2} \tag{16}$$

$$\left(\frac{\omega_n}{K}\right)^2 = \frac{1}{K\tau_2} \tag{17}$$

where ω_n represents the undamped natural system frequency and ζ the ratio of actual to critical damping in the small-signal closed-loop transfer function of the APLL. Combining (12), (15), (16) and (17), the capture ratio as a function of filter parameters results as

$$\frac{\Delta\omega_{ic}}{K} \approx \frac{\omega_n}{K}\left[\left\{\left[2\zeta\left(\frac{\omega_n}{K} - \zeta\right)\right]^2 + 1\right\}^{\frac{1}{2}} - 2\zeta\left(\frac{\omega_n}{K} - \zeta\right)\right]^{\frac{1}{2}}. \tag{18}$$

This expression is plotted graphically as a function of ζ with the parameter ω_n/K in Fig. 4. As is to be expected, the capture ratio increases with

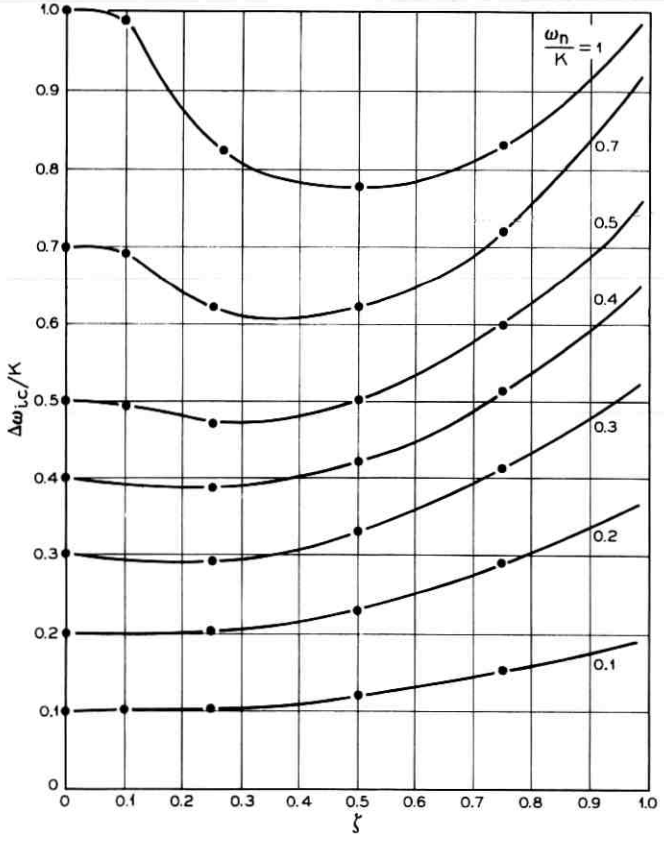

Fig. 4 — Capture range versus damping factor with parameter ω_n/K for APLL with lag network.

damping for a given system natural frequency. By making an appropriate choice of the additional parameter supplied by the lag filter, the capture ratio can now be kept closer to unity for a much smaller system bandwidth than is the case with a simple *RC* filter. In the next section the considerations that determine this choice will be discussed.

Measurements give capture ratio values that are larger than those derived from (12). The reason for this is that by equating the peak open-loop error voltage (9) with the steady state error voltage at capture (11) the assumption is made that when out-of-lock the oscillator mean frequency is equal to its rest frequency ω_c. Actually, when out-of-lock, the waveform at the phase detector output is not a pure sine wave with zero mean voltage.[7] As the VCO is frequency modulated by it, the beat frequency between input and VCO is varied, causing the rate at which the phase error characteristic (5) is traversed to vary accordingly. When the rate of traverse is slow (corresponding to a small beat frequency) the error voltage half cycle will be wider than when it is fast (large beat frequency). A net dc voltage, varying inversely with the instantaneous beat frequency, results at the phase comparator output. It causes the VCO mean frequency to leave its rest frequency by an amount $\overline{\Delta\omega_0}$ in the direction of the locked state. The exact equivalent to (12) is therefore given by

$$\Delta\omega_{ic} = KF_{\Delta\omega} \tag{19}$$

where

$$\Delta\omega = \Delta\omega_{ic} - \overline{\Delta\omega_0}\,. \tag{20}$$

As a result of (20)

$$\Delta\omega \leqq \Delta\omega_{ic} \tag{21}$$

so that

$$F_{\Delta\omega} \geqq F_{\Delta\omega_{ic}} \tag{22}$$

since F_ω is the amplitude response of a low-pass filter. Thus the actual capture range given by (19) is larger than the approximation given by (12).

Richman has shown that $\overline{\Delta\omega_0}$ can be derived for the idealized case in which $F_\omega \equiv 1$ by integrating (4) over a cycle and dividing by the period of the asynchronous loop beat frequency T_b. The results are shown qualitatively in Fig. 5. Unfortunately, as mentioned earlier (4) cannot be integrated in closed form when the impulse response of a general or even a lag network low-pass filter is substituted in the equa-

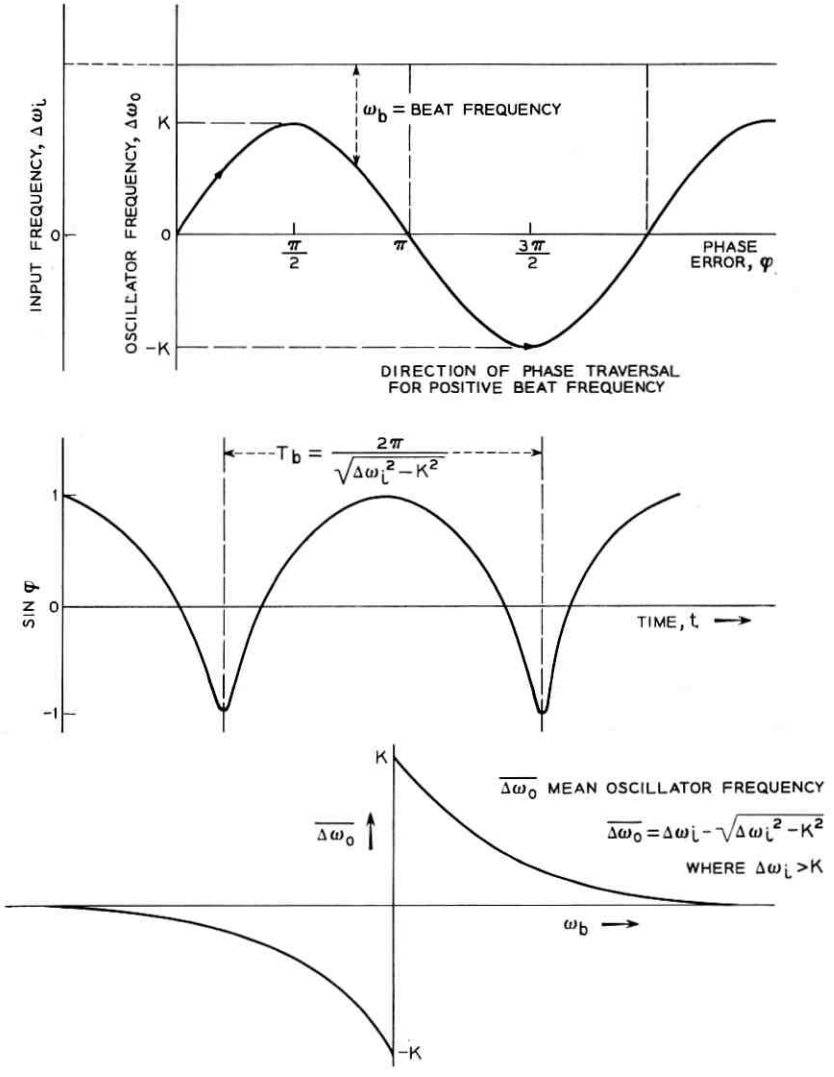

Fig. 5 — Schematic presentation of the capture mechanism for an idealized APLL without a low-pass filter incorporated in it.

tion. For this reason (19) cannot be derived analytically for anything but the idealized case.

In some circumstances, to eliminate the incompatibility of narrow bandwidth with wide capture range inherent in the APLL, the frequency-dependent control voltage mentioned above can be artificially

increased. One way of doing this is by adding a detector to the loop that is sensitive to the absolute value and polarity of the beat frequency.[8,9] Another[10] is to use a simple peak detector sensitive only to the beat frequency to disconnect the shunt elements of the low-pass filter when the loop is out-of-lock.

2.3 *Small-Signal Properties of the APLL*

Using the APLL as a frequency discriminator, the limits on phase error are set by the required system stability (lock and capture range) and by the desire to avoid distortion of the output error voltage (non-constant loop gain). To satisfy these conditions it is important to operate within the linear range of the phase comparator characteristic, e.g., within one radian phase error. Equation (4) can then be linearized and, respresenting the loop variables by their Laplace transforms (upper-case symbols), a linearized servo block diagram may be derived (see Fig. 6). The $1/s$ term takes the integration of the frequency difference by the phase comparator into account. The open-loop transfer function is given by

$$\frac{\Omega_0}{\Delta\Omega} = \frac{K_1K_2F(s)}{s} \tag{23}$$

and the closed-loop transfer function in terms of the output voltage V_e and input deviation Ω_i is given by

$$T(s) = \frac{V_e}{\Omega_i} = \frac{1}{K_2}\frac{K_1K_2F(s)}{s + K_1K_2F(s)}. \tag{24}$$

Thus the APLL has an inherent low-pass filter characteristic that is modified by the type of network $F(s)$ inserted in it. Obviously, the filter bandwidth is understood to be with respect to sinusoidal modulation of the input phase or frequency caused by noise or neighboring signal channels, and not to amplitude modulation as in conventional

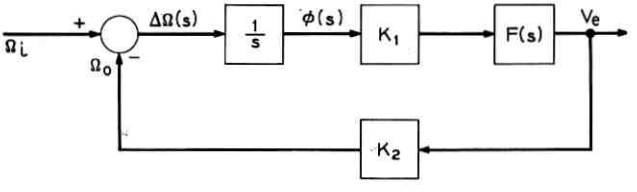

Fig. 6 — Linearized servo block diagram of APLL.

filter terminology. In this sense the APLL operates as a frequency noise or a jitter filter with respect to the input signal or signals.

The 3-db closed-loop bandwidth of the system characterized by (24) is equal to its open-loop crossover frequency or the frequency for which the open-loop gain given by (23) is equal to unity. Evaluation of this frequency results in precisely the same expression as was obtained in the capture frequency derivation given by (12). This means that an approximation for the capture range of an APLL is given by its 3-db closed-loop bandwidth.

It has been shown[11] that the mean square jitter of the VCO frequency and of the phase error caused by random interference at the input is proportional to the noise bandwidth of the system given by

$$B = \int_{-\infty}^{+\infty} | T(j\omega) |^2 \, d\omega. \tag{25}$$

This is constant for a simple *RC* filter (Fig. 2a), because the decrease in damping compensates for the area of the squared transfer function decreased by a reduced cutoff frequency. It can, however, be reduced by the lag network shown in Fig. 2(b), and previous papers[12,13,14] have demonstrated how the proper choice of its parameters can optimize system performance for various applications. The lag network also allows for the static phase error to be determined independently of the 3-db closed loop or of the noise bandwidth, which is important for the present application. Unfortunately, as was shown earlier, a reduction in bandwidth with respect to lock range always results in a reduced capture ratio. The ratio of capture range to noise bandwidth therefore represents a useful figure of merit that characterizes the APLL adequately for certain applications.

To calculate the noise bandwidth, (15), (16) and (17) are substituted into (25). Integrating gives

$$\frac{B}{K} = \frac{\pi}{2\zeta} \frac{\omega_n}{K} \left[1 + \left(2\zeta - \frac{\omega_n}{K} \right)^2 \right]. \tag{26}$$

Minimizing the noise bandwidth results in the following condition for the network parameters

$$\zeta_0 = \tfrac{1}{2}[1 + (\omega_n/K)^2]^{\frac{1}{2}}. \tag{27}$$

Substituting this in (26)

$$B_{\min}/K = 2\pi(\omega_n/K)^2[\{1 + (K/\omega_n)^2\}^{\frac{1}{2}} - 1] \tag{28}$$

which represents the two-sided minimum bandwidth as a function of

natural frequency and lock range, (see Fig. 7). Substituting (27) into (18) gives

$$(\Delta\omega_{ic}/K)\,|_{\text{cpt}} \approx x\big[\,\{[x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)]^2 + 1\}^{\frac{1}{2}} - \{x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)\}\,\big]^{\frac{1}{2}} \tag{29}$$

where

$$x = \omega_n/K.$$

Equation (29), plotted in Fig. 7, represents the approximate capture ratio given by (12) after the system has been optimized for minimum

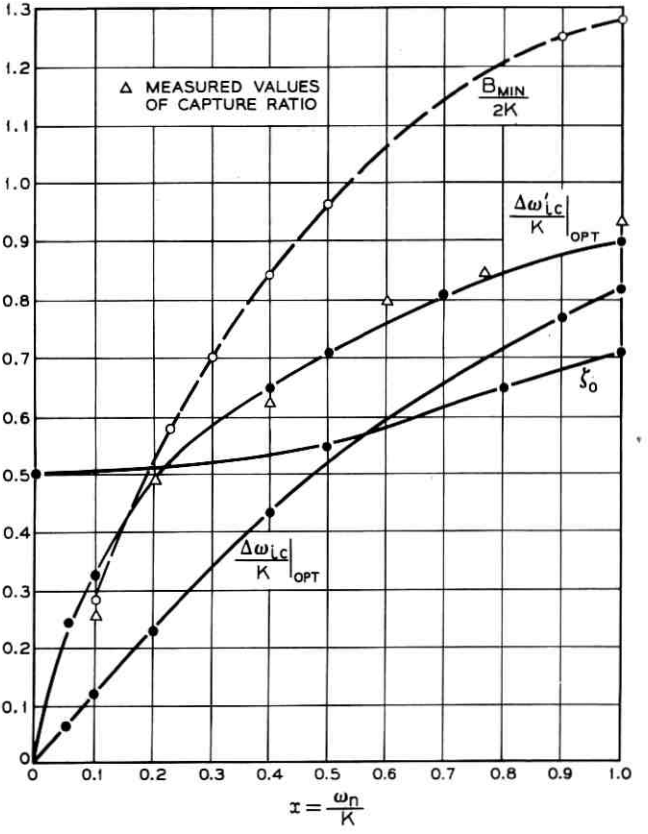

Fig. 7 — Relative noise bandwidth ($B_{\min}/2K$), capture range approximations ($\Delta\omega_{ic}/K$) and ($\Delta\omega_{ic}'/K$) and damping factor ζ_0 as a function of (ω_n/K) for an APLL optimized for minimum noise bandwidth.

noise bandwidth. Simultaneously, as was pointed out earlier, it exactly defines the open-loop crossover frequency or the 3-db closed-loop bandwidth of the APLL for an optimized lag filter.

Measurements for the capture ratio of an APLL incorporating a lag filter optimized with respect to noise bandwidth were made. The results suggested that for this case a closer approximation of the capture ratio is given by the geometrical mean of the crossover frequency derived from (12) and the lock range K. Denoting this improved approximation of capture range by $\Delta\omega_{ic}'$, we get from (29)

$$(\Delta\omega_{ic}'/K)\,|_{\text{opt}} \approx [x\{\{[x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)]^2 + 1\}^{\frac{1}{2}} - \{x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)\}\}^{\frac{1}{2}}]^{\frac{1}{2}}. \tag{30}$$

This expression has also been plotted in Fig. 7.

Combining (28) with (30) to give the figure of merit F_0 for a system optimized for minimum one-sided noise bandwidth gives

$$F_0 = \frac{\text{capture range}}{\text{minimum noise bandwidth}} \approx \frac{[x\{\{[x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)]^2 + 1\}^{\frac{1}{2}} - \{x(1+x^2)^{\frac{1}{2}} - \tfrac{1}{2}(1+x^2)\}\}^{\frac{1}{2}}]^{\frac{1}{2}}}{\pi x[(1+x^2)^{\frac{1}{2}} - x]}. \tag{31}$$

This expression is plotted in Fig. 8. It can be seen that the capture range is generally smaller than the minimum noise bandwidth. For narrow noise bandwidth systems, this may make unduly high demands on the stability of the system, in particular on the VCO and, if present, on the dc amplifier. It may then be worthwhile incorporating the already mentioned additional frequency-sensitive loop or the peak detector, sensitive to the out-of-lock beat frequency at the phase detector output, to increase the capture ratio.

For the application described here, the ratio of capture range to closed loop bandwidth $\Delta\omega_{ic}'/\Delta\omega_{ic}$ designated by F is a more practical figure of merit than the conventional F_0. It is therefore also plotted in Fig. 8. The resulting curve determines the degree of stability with respect to frequency drift that is required by a given system. This can best be illustrated by the channel frequency distribution of part of an FM multiplex system shown in Fig. 9. The assumption was made there that the bandwidth of each receiving channel filter equals the total frequency deviation in that channel. The total bandwidth BW of each channel extends halfway into the guardband on either side of a channel.

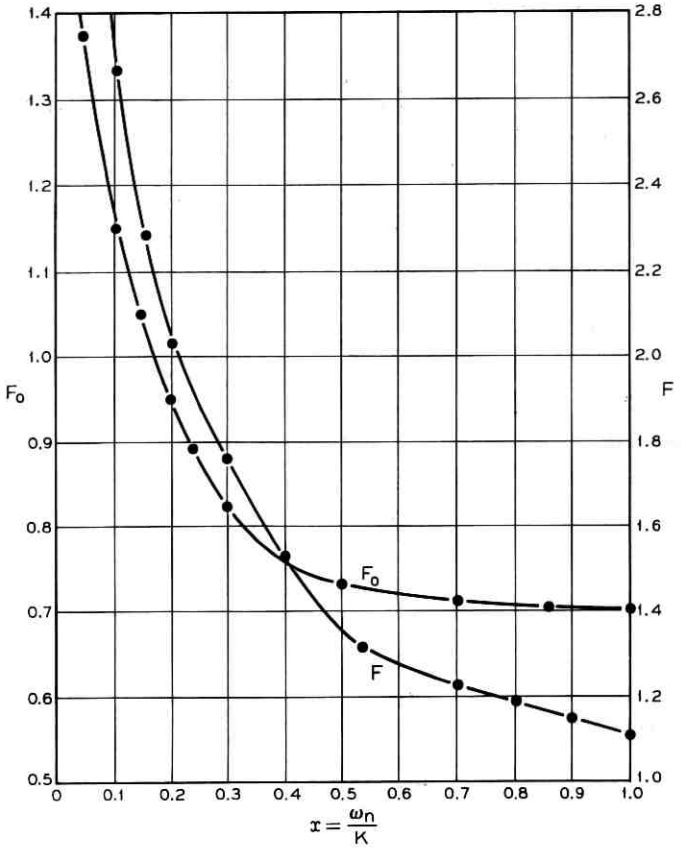

Fig. 8 — Figures of merit F_0 = capture range/noise bandwidth and F = capture range/closed-loop bandwidth for minimum noise-bandwidth APLL.

Limits on the capture frequency range can then be obtained by inspection, namely

$$\text{total frequency deviation} < \text{capture range} < \text{channel bandwidth } BW. \tag{32}$$

Expressing (32) in terms of F, we get

$$1 \leqq F \leqq (BW/2\Delta\omega_{ic}). \tag{33}$$

The closer F is to its lower limit the less frequency drift may be tolerated in the receiving APLL. The more it approaches its upper limit the smaller the ratio ω_n/K must be, as Fig. 8 shows. Referring to Fig. 7, a decreasing

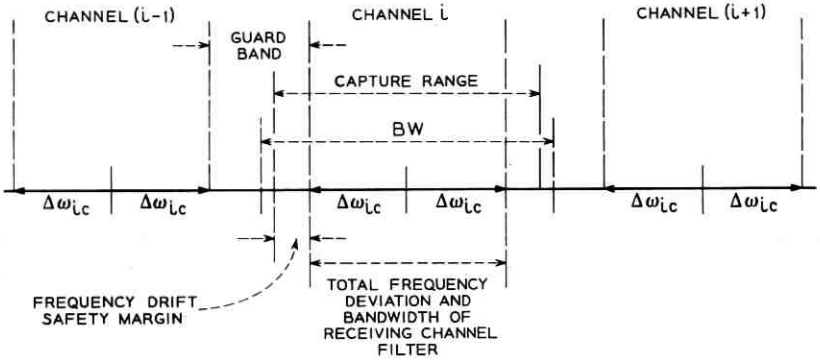

Fig. 9 — Channel frequency distribution of part of an FM multiplex system.

value of ω_n/K in turn results in a decreasing ratio $\Delta\omega_{ic}/K$. In other words, the extent by which the lock range K must exceed the closed-loop system bandwidth $\Delta\omega_{ic}$ increases as F is varied from its lower to its upper limits. Using conventional circuit techniques, the choice of F would thus be reduced to the question of whether it is more economical to attain a high degree of circuit stability or a high loop gain. Using circuit microminiaturization techniques, however, the choice depends on which of the two major approaches existing to date is taken. If integrated semiconductor circuits are to be used, low-precision components and therefore substantial frequency drift in the APLL must be expected. In this case F should be chosen as close to the upper limit given by (33) as possible. If high-stability thin film circuit components are available for design, the choice of F will be as close to the lower limit as the attainable over-all circuit stability and the lower limit on lock range allow.

The limits on lock range (or loop gain) K are determined by two different aspects of loop operation. The upper limit results directly from the upper limit on capture frequency range. With the same approximation for the capture range as that used to derive (30), namely

$$\Delta\omega_{ic}' \approx (K\Delta\omega_{ic})^{\frac{1}{2}} \tag{34}$$

and referring to Fig. 9, the upper limit on lock range in terms of channel and receiving filter bandwidth of a given system is $(BW)^2/\Delta\omega_{ic}$. It must be remembered, however, that (34) was empirically verified only with a lag network low-pass filter such as that shown in Fig. 2(b) connected in the APLL. By inspection of Fig. 9 the lower limit on lock range might be considered to equal the input frequency deviation. However,

this is true only for the case that a linear phase detector (e.g., sawtooth output) is used, since only then is loop operation linear throughout the frequency range covered at the input. With sinusoidal phase detectors which are simpler in implementation and therefore more widely used, the output is linear only to within a given percentage for phase error values that do not exceed a corresponding maximum value φ_{LIN}. Referring to Fig. 9 and (5) and (33), the limits on lock range then result as follows

$$\Delta\omega_{ic}/\sin\varphi_{\text{LIN}} \leqq K \leqq (BW)^2/\Delta\omega_{ic}. \tag{35}$$

Using a sinusoidal phase detector accordingly reduces the lower bound on F, thereby modifying (33) to give

$$(1/\sin\varphi_{\text{LIN}})^{\frac{1}{2}} \leqq F \leqq BW/2\Delta\omega_{ic}. \tag{36}$$

Besides decreasing the capture ratio, the inclusion of a low-pass filter network in the loop also increases the capture time T_c. Richman has shown[7] that for input frequencies within the capture range, this can be approximated by:

$$T_c \approx 32\Delta f_i^2/B^3 \tag{37}$$

where Δf_i is the maximum initial frequency offset and B the noise bandwidth. For many applications, the maximum frequency offset is equal to the system bandwidth. Approximating this by the noise bandwidth, the capture time is then given by

$$T_c \approx 8/\pi^2 B. \tag{38}$$

III. RC CIRCUIT DESIGN OF APLL

In this section general considerations for the design of an APLL filter and discriminator are discussed using *RC* active circuits that are compatible with microminiaturization techniques.

3.1 *The Phase Comparator*

Analog multipliers can be built[15] such that the output contains a dc term proportional to the input phase difference, but these are either complex or unsuitable for a composite input signal. Semiconductor switched modulators or phase-sensitive choppers can be used for the same purpose and are very much simpler and more economical to build. This is why they have been described quite extensively in the literature,[16-19] particularly with the advent of the epitaxial planar transistor,

which makes an excellent switch when used in the inverted mode.[20] However, since these circuits are most easily operated when driven from a transformer the methods usually described are not directly applicable here. To eliminate the transformers while retaining the isolation and performance they afford it was found that the transistors can be driven from an ac-coupled current source.

The equivalent diagram of a simple series switched modulator is shown in Fig. 10(a). A sinusoidal signal is periodically switched from the output of a voltage generator with source resistance R_S to the load R_L during the time T_0. The input signal period and the switching period both equal T, and φ is the phase angle between the two signals. Fig. 10(b) illustrates a schematic interpretation of this process, in that the input sinusoidal voltage is multiplied by a square wave of unity amplitude, period T and pulse width T_0. The dc component of the resulting output voltage is then given by

$$V_{OUT} = \frac{E}{\pi} \frac{R_L}{R_S + R_L} \sin (\pi T_0/T) \cos \varphi. \tag{39}$$

The switch S can be replaced by a transistor operating in the inverted mode to give a configuration such as that shown in Fig. 10(c). The transistor must be driven from a constant current source; therefore R_1 must be large, yet still allow the transistor to be saturated throughout the ON state. The ratio of reference to input signal level E_c/E, which should be as large as possible, determines the actual values of R_1 and R_2 in a straightforward way. To prevent emitter breakdown, a diode is connected in series with the transistor base.

With the transistor connected in the inverted mode, it is the base collector junction that is driven by the reference voltage. For this reason and to sustain saturation both the ratio R_S/R_L and R_S itself must be made as small as possible, the limits again being determined by the ratio of E_c/E. This suggests the use of an emitter follower from which to apply the input signal, as shown in Fig. 10(d). Here the variable resistor R_v is initially adjusted to cancel the offset voltage by making the voltage difference between points A and B equal to zero. The switched modulator is shown feeding into a simple dc amplifier stage T_3. If necessary, this amplifier can be increasingly stabilized by well known means, as was done in the application described in the next section. If — as is the case when operating as a phase detector — only the dc component of the switched modulator output voltage is required, the low-pass filter R,C shown in Fig. 10(d) can be added.

In the switched modulator shown in Fig. 10(d) the transistor offset

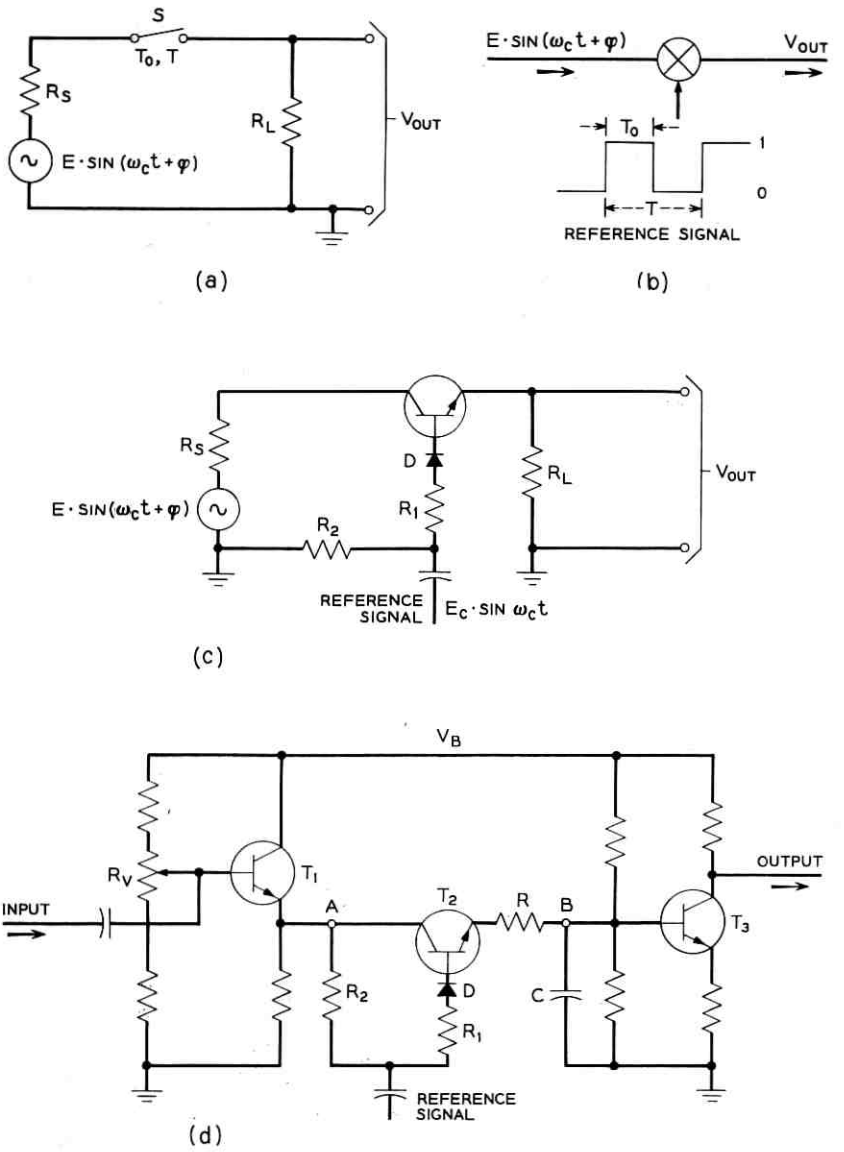

Fig. 10 — Transformerless single-transistor series-switched modulator: (a) equivalent diagram (b) functional interpretation (c) Switch S in (a) replaced by transistor in inverted mode (d) final circuit.

voltage can be entirely eliminated initially, but no compensation takes place for subsequent drift caused by temperature variations. However the circuit is operated essentially as a high-level switch, inasmuch as the input signal can be increased by more than half the reference voltage level, particularly if the latter signal is a square wave. In such cases offset voltage drift, being in the order of microvolts for silicon planar epitaxial transistors,[20] is usually negligible. For extreme stability with ambient temperature or for low-level switching applications a balanced modulator using two inverted transistors connected back-to-back gives excellent results.[16] Here again the transformer customarily driving the transistors can be replaced by a current source. Such a circuit is shown in Fig. 11 with the switching transistors driven by a high output impedance phase inverter. Offset voltages and currents can be canceled here over a wide temperature range without having to preselect or match the transistors.

Equation (39) shows that the output dc voltage of the switched modulator depends not only on the phase difference between the input and reference signals but also on the input signal amplitude. To eliminate this latter dependence an AGC circuit should precede the APLL when input level variations are anticipated.

3.2 *The Voltage-Controlled Oscillator (VCO)*

Stability of the VCO frequency is one of the main factors that determines the stability of the APLL. For a narrow filter characteristic,

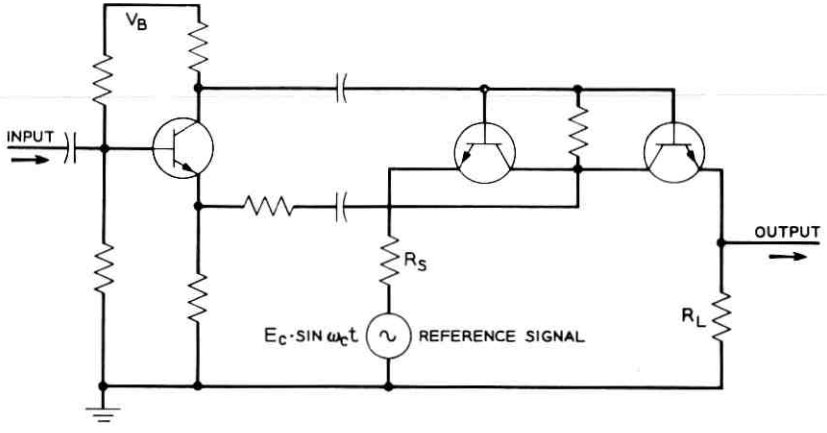

Fig. 11 — Balanced series-switched modulator fed from a phase inverter.

frequency stability with respect to temperature and power supply variations is therefore important. This can be achieved by designing the oscillator so that its frequency depends only on reliable low-tolerance passive circuit components. Highest stability can be attained with precision thin film resistors and capacitors that are matched with respect to temperature coefficients.

For circuit simplicity and economy it is desirable to minimize the number of voltage controllable elements needed to give a specified frequency deviation. Nevertheless the frequency versus control voltage must be linear for linear operation of the loop. As for the components presently available as voltage controlled elements, this depends on the frequency range of operation. Applications at high frequencies using varactors are not uncommon, whereas the mean capacitance value of these components is generally not large enough for low frequencies. In the latter case field effect transistors, varistors or diodes can be used as voltage controllable resistors over a certain range of their voltage-current characteristics.

RC oscillators suitable for frequency modulation by a control voltage can be grouped into three categories, i.e., relaxation oscillators, zero phase shift oscillators and 180° phase shift oscillators.

With relaxation oscillators, the period of oscillation is controlled by the input signal voltage. If so desired, the fundamental frequency component can be filtered out by a low-pass filter. For frequency stability, threshold compensation due to temperature variation and line voltage compensation are necessary. This entails additional regulating circuitry.

Zero phase shift voltage-controlled oscillators (i.e., Wien bridge, twin-T, bridged-T, or zero phase shift ladder network in the feedback loop) require more than one voltage-controlled element or the incorporation of some means of AGC. Changing any single element of the frequency determining network changes network loss and results in amplitude modulation.

Over a limited frequency range, amplitude variations can be eliminated in a 180° phase shift oscillator when varying only one element in the frequency determining ladder network. This is shown in Fig. 12, where attenuation and frequency curves for three and four equal-section ladder networks in which the value of a single element is varied are plotted. The ladder networks considered differ in the mode of operation (voltage or current driving source) and in the choice of variable element. all cases the attenuation curve goes through a broad minimum rather lling monotonically as is the case with networks used in zero oscillators. Experiments have verified this, in that negligible

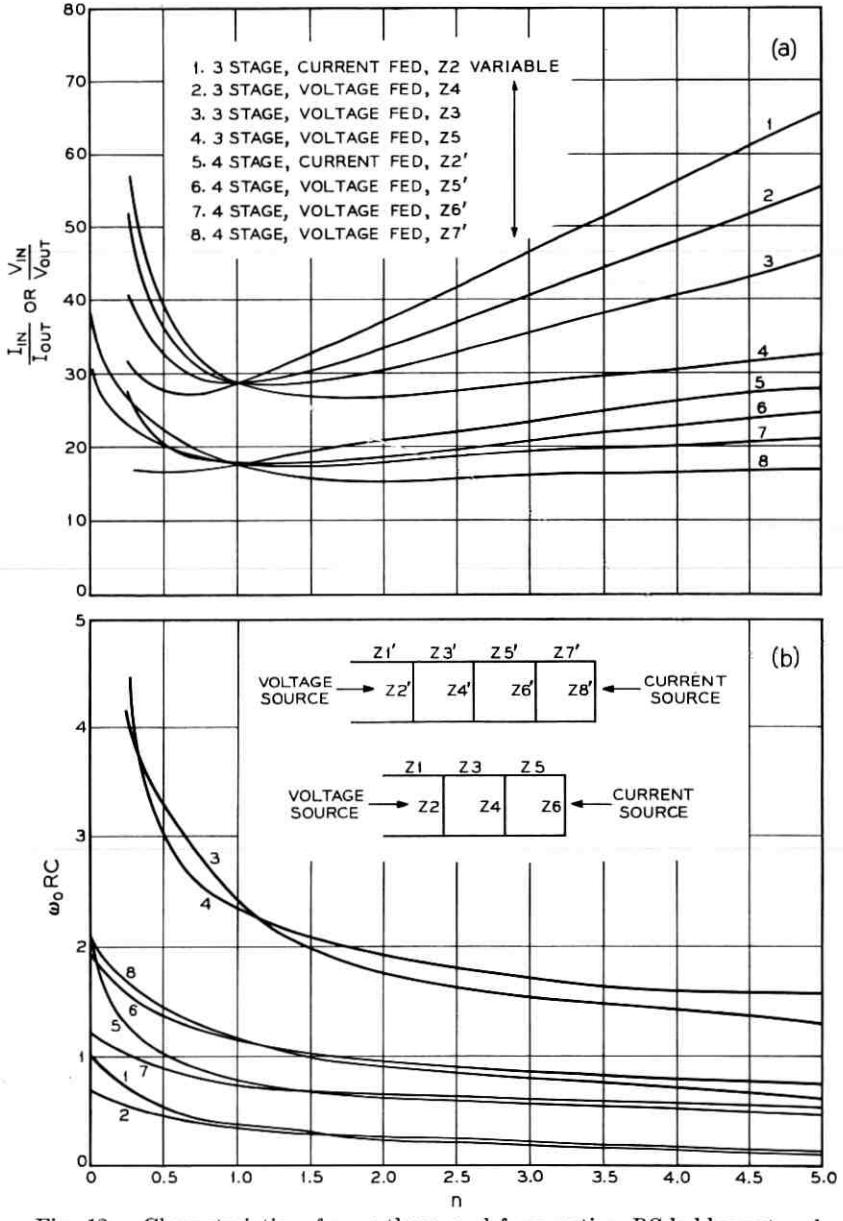

Fig. 12 — Characteristics of some three- and four-section *RC* ladder networks fed from a voltage or current source while varying a single ladder resistor. All *R*'s and *C*'s are equal except for the variable resistor $Z_i = nR$. All *R*'s are either in series or shunt, depending on the location of the respective Z_i: (a) attenuation versus *n;* (b) frequency versus *n*.

amplitude change was measured with up to 20 per cent frequency change. This range allows for modulation and initial tuning to be performed either by one and the same element or by two separate ones. If the amplifying section of the oscillator can be designed to be independent of transistor parameters and to have negligible phase shift at the frequencies in question, then the oscillator frequency is dependent only on the ladder network and control element. In the next section and the Appendix, the design of an *RC* 180° phase shift VCO utilizing these features is described in detail.

IV. RC FILTER AND FREQUENCY DISCRIMINATOR FOR THIN FILM DATA RECEIVER

In this section an application of the APLL as *RC* channel filter and frequency discriminator for a multichannel binary FM data receiver is described. The channel and frequency distribution of the transmitted data signal are shown in Fig. 13. Eight signal channels and one timing channel spaced evenly over a 1.43-kc frequency band are received simultaneously from the transmission line. Each channel is frequency modulated with a deviation of ±35 cps at a maximum rate of 75 bps. Adjacent channels are separated by a guard band of 100 cps.

The block diagram in Fig. 14(a) shows the circuits that are normally used in a receiver for this kind of binary FM data.[21] An AGC system is connected to the receiver input to compensate for losses suffered in the transmission line. Each channel has a bandpass filter and FM discriminator assigned to it. A post-detection filter follows to separate the actual data output from the higher-frequency detection products. It might also be used to give additional noise rejection and shaping in the baseband region. The filtered data signal drives a slicer whose output is

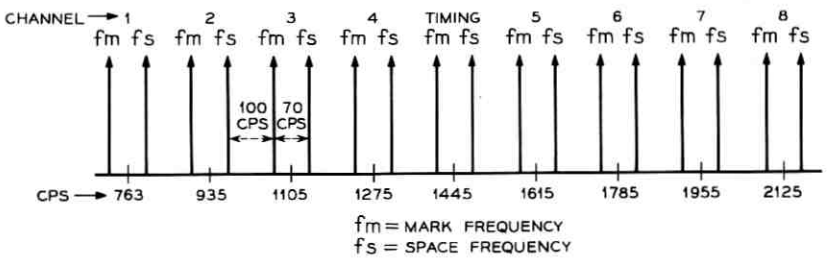

Fig. 13 — Frequency distribution of a multichannel data receiver.

sampled at appropriate instants by the timing channel to recover the binary data sequence. Fig. 14(b) shows the equivalent block diagram using the APLL. An APLL replaces the receiving bandpass filter and FM discriminator in every channel. Each APLL is followed by an *RC* active low-pass filter that serves the same purpose as the post-detection filter mentioned above. The transmitted binary data sequence can be obtained from a sampled slicer in the conventional way.

This particular transmission system was chosen because its very stringent filtering requirements are well suited to test the efficiency of *RC* active circuits that satisfy the restrictions imposed by microminiaturization techniques.

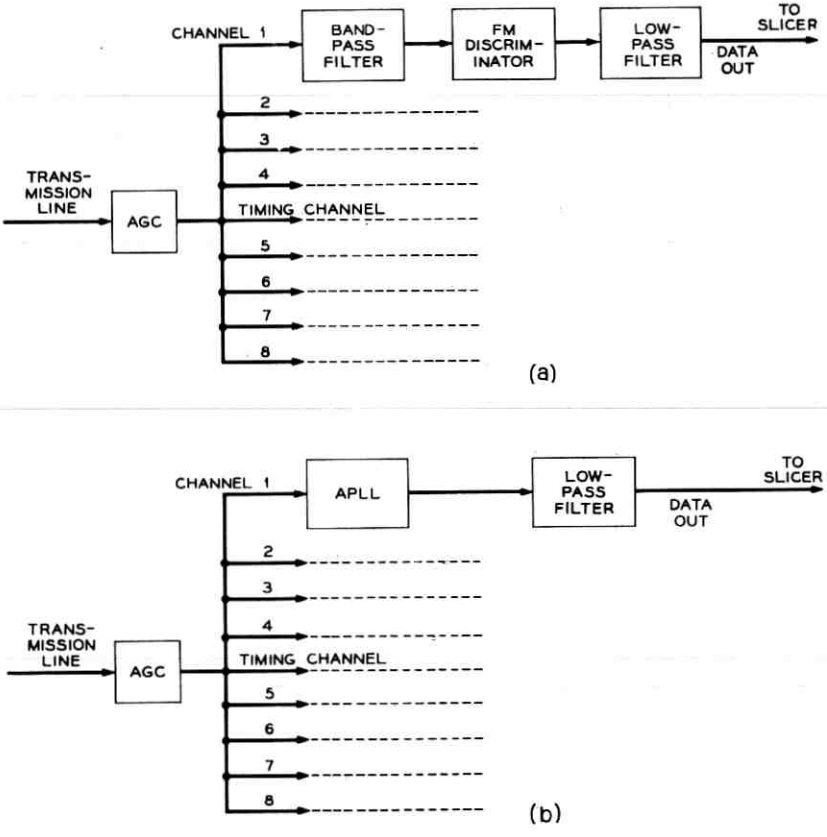

Fig. 14 — Block diagram of multichannel data receiver: (a) conventional circuits, (b) bandpass filters and FM discriminators replaced by APLL's.

4.1 *Design Requirements*

The miniaturization technique that has been chosen for these data sets combines tantalum thin film passive components with specially packaged active components.[22] The requirements that the circuits must then fulfill can be summarized as follows:

(a) for compatibility with thin film techniques:

no magnetic elements (inductors, transformers, etc.)

resistors not larger than 100 kilohms, not smaller than 100 ohms

high-quality capacitors not larger than 0.01μf; coupling capacitors not larger than 1μf

only one resistor adjustment at a time (by anodization); ease of tuning. This implies minimum interdependence between Q and frequency adjustments of a circuit.

(b) for stability, size reduction and economical compatibility:

circuit characteristics dependent only on passive components

as few components, in particular active ones, as possible.

4.2 *Loop Gain Requirements*

The over-all loop gain or lock range K is determined either by the linearity or capture frequency range, depending on the stability of the circuitry involved. High-stability tantalum thin film components are available for the application described here. The choice for high system stability and poor figure of merit F was therefore a natural one, so that the smallest loop gain compatible with loop linearity (35) was used.

Assuming a linearity error of approximately 10 per cent, the maximum phase error φ_{LIN} should not exceed $\pm 45°$. With an input frequency deviation of ± 35 cps the lock range is then given by:

$$K = 2\pi \times \sqrt{2} \times 35 \approx 2\pi \times 50 \text{ rad/sec.} \tag{40}$$

The distribution of gain within the loop is determined by the required discriminator output voltage for a specified frequency deviation at the input, in other words by the VCO sensitivity K_2. If the switched modulator cannot supply the remaining gain $K_1 = K/K_2$, a dc amplifier has to be incorporated in the system.

The discriminator output level was chosen at 4 volts peak-to-peak. Therefore

$$K_2 = \frac{2\pi \times 35}{2} = 110 \text{ rad/volts sec,} \tag{41}$$

and

$$K_1 = K/K_2 = 2.86 \text{ volts.} \tag{42}$$

The switched modulator output voltage (39) is maximum for equal ON/OFF switching times. With $R_S \ll R_L$ and E designating the amplitude of the input signal being filtered, (39) then simplifies to:

$$v_{out} = (E/\pi) \cos \varphi. \tag{43}$$

For n independent channels of equal average energy, the amplitude of a single-input signal in terms of the total rms input voltage $\overline{E_T}$ is given by:

$$E = (2/n)^{\frac{1}{2}}\overline{E_T} \tag{44}$$

which is approximately 0.35 v for each of the nine channels of the system here in question. The constant factor in (39) then equals 0.11 v. To satisfy (42), a dc amplifier with a voltage gain of 26 has to be added in the loop. An emitter-coupled differential dc amplifier was used for this purpose.

4.3 *Bandwidth Requirements*

For the data system considered here, optimum detection results when the receiving filter bandwidth is equal to the total frequency deviation in each channel. This corresponds to a bandwidth or crossover frequency of the open-loop gain (18) of ± 35 cycles. To minimize distortion caused by interference at the input, a system with minimum noise bandwidth is chosen. From (29) or Fig. 7, setting $\Delta\omega_{ic} = 35/50 = 0.7$, the relative system frequency

$$x = \omega_n/K = 0.77 \tag{45}$$

and with (27), the damping factor

$$\zeta_0 = 0.63. \tag{46}$$

τ_1 and τ_2 can now be calculated from (16) and (17). Referring to Fig. 2(b), with $R_2 = 5.1$ kilohms, R_1 and C follow from (15a) and (15b), giving 3 kilohms and 0.67 μf respectively.

The switched modulator is followed by a simple RC low-pass filter section that greatly attenuates the ac components of the error voltage before applying it to the dc amplifier. The ratio of channel frequency to system bandwidth is large enough to permit this without affecting the capture ratio (see Section 2.2). With the 3-db cutoff frequency

$\omega_{co} = 1.8K = 2\pi \times 90$ cps the fundamental frequency (i.e., twice the channel frequency) is attenuated by approximately 30 db in the lowest channel.

The baseband low-pass filter following the APLL is designed to reject the remaining high-frequency products, generated in the APLL by the switched modulator, while passing the data output with minimum attenuation. To obtain sufficient selectivity for this purpose an active low-pass filter with a cutoff frequency of 60 cps and a slope of 18 db/oct is used.

4.4 *Stability Requirements*

The maximum phase error $\varphi_{LIN} = \pm 45°$ that was used to derive the lock range given by (40) also defines the ratio of capture range to system bandwidth. From (36) we get:

$$F = \left(\frac{1}{\sin \varphi_{LIN}}\right)^{\frac{1}{2}} = (2)^{\frac{1}{2}} = 1.19 \tag{47}$$

which could have been obtained directly from Fig. 8. With the frequency deviation of ± 35 cps the capture range therefore extends over ± 42 cps. This allows for a drift in frequency of the APLL of ± 7 cps or 0.33 per cent at the highest channel (2125 cps). Designing the active circuitry of the APLL to depend solely on the high-precision thin film components, this stability can well be maintained. Initial tuning of the VCO (see Appendix) is accomplished by means of precision anodization of a resistor in the ladder network. This can be performed with up to 0.02 per cent accuracy.[23] Frequency stability is maintained by balancing the temperature coefficient of the capacitors by that of resistors. In this way the temperature coefficient of frequency can be held well under the permissible limit.

4.5 *Circuit Design*

The breadboard model of an APLL for one channel of the data receiver described above was built with conventional *RC* components. Channel center frequency was 1955 cps. A schematic of the circuit is shown in Fig. 15. High quality NPN Si planar epitaxial transistors and Si diffused-junction diodes were used throughout.

A 180° phase shift oscillator with one voltage-variable ladder element was used as VCO. This was quite adequate, as the maximum relative frequency deviation occurring at the lowest channel does not exceed 5 per cent. To insure linear operation of the voltage-variable element, it was inserted in a network section of low signal level. Since the attenua-

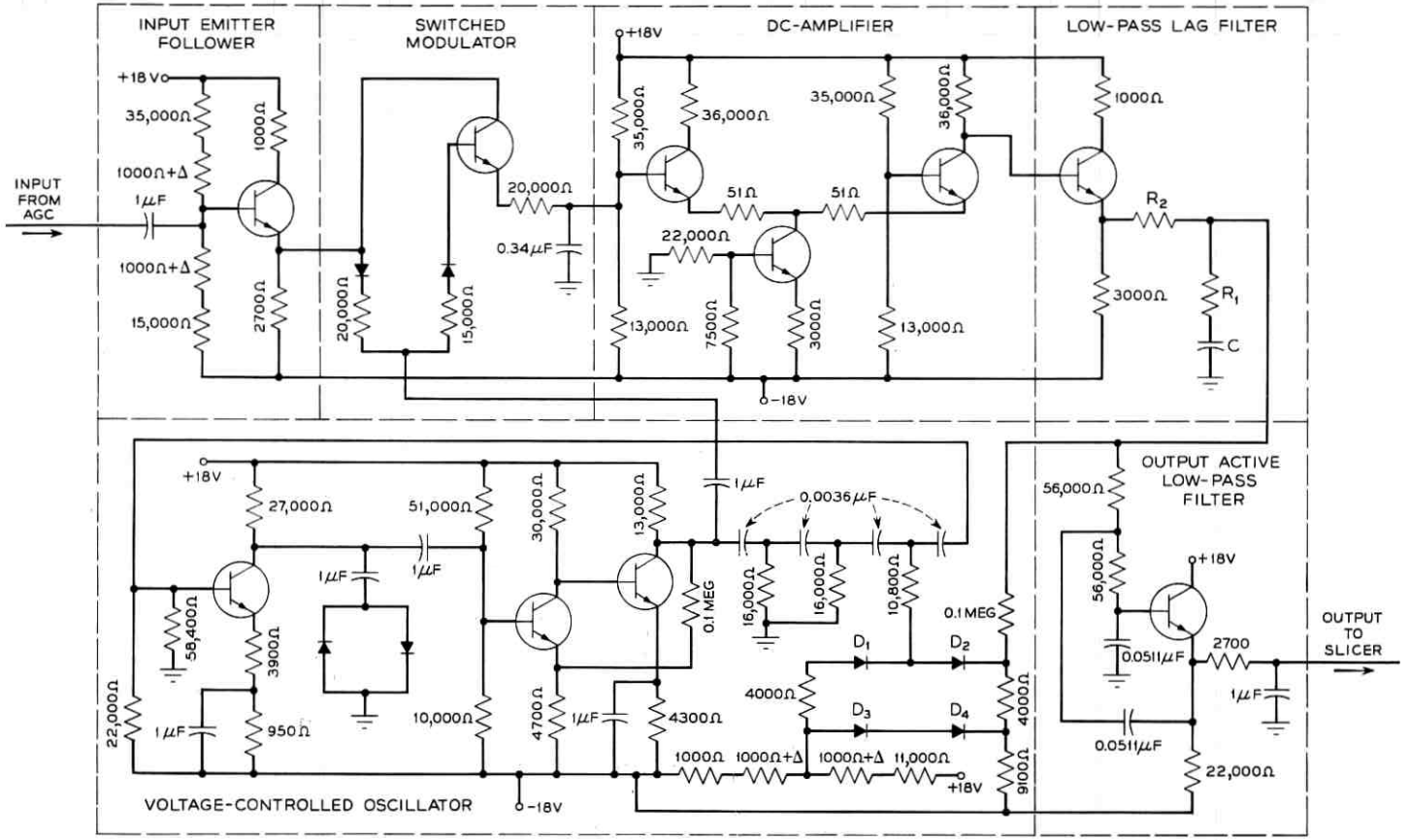

Fig. 15 — Circuit diagram of the APLL designed as filter and discriminator for one of nine parallel FM data channels, in a multichannel data receiver. All transistors are high-quality *NPN* Si planar-epitaxial, all diodes are Si diffused-junction types.

tion minimum is flatter for four- than for three-section ladder networks, while the difference in frequency sensitivity is not appreciable, the former configuration was used. The necessary gain is also lower for this configuration, allowing for sufficient current and voltage feedback to make the 3-stage voltage amplifier independent of transistor parameters. A varistor limits the oscillation amplitude, enabling the amplifier to be operated in its linear range. This improves temperature and voltage stability. It also eliminates signal distortion, so as to maintain constant switching intervals at the modulator.

The small-signal ac resistance of two forward-biased parallel silicon diodes was used as the voltage-variable shunt element in the ladder network. It is inversely proportional to the biasing current, if the semiconductor and lead resistance can be neglected. To reduce temperature variations, the diodes are biased by a constant current which is much larger than the leakage current. The frequency-voltage characteristic for small frequency deviations is derived in the Appendix.

A one-transistor switched modulator is used as multiplier. It is driven by the 15-vpp reference signal coming from the VCO. A preceding emitter follower feeds the composite signal into the APLL of each channel. The biasing voltage divider is adjusted for minimum voltage offset across the switched transistor. The simple *RC* low-pass filter mentioned in Section 4.3 connects the multiplier with an emitter-coupled dc amplifier.

The amplifier output is passed through a lag network and current controls diodes D_1 and D_2 in the oscillator. The initial steady-state biasing current in the diodes is set by a voltage divider. The diodes D_3 and D_4 serve to temperature stabilize the oscillator (see Appendix).

The discriminator output voltage is taken from the lag filter output terminal. To reject still remaining high-frequency components it is passed through an active low-pass filter before being further utilized.

4.6 *Experimental Results*

Measurements on the over-all performance of the APLL both as bandpass filter and discriminator and with respect to stability with temperature and voltage supply variations were made. The bandpass characteristic was measured while varying the bit rate of the incoming signal. One side of it is plotted in Fig. 16 in terms of the corresponding modulation frequency. It coincides well with the requirements listed in Section 4.3. The influence of the post-detection active low-pass filter

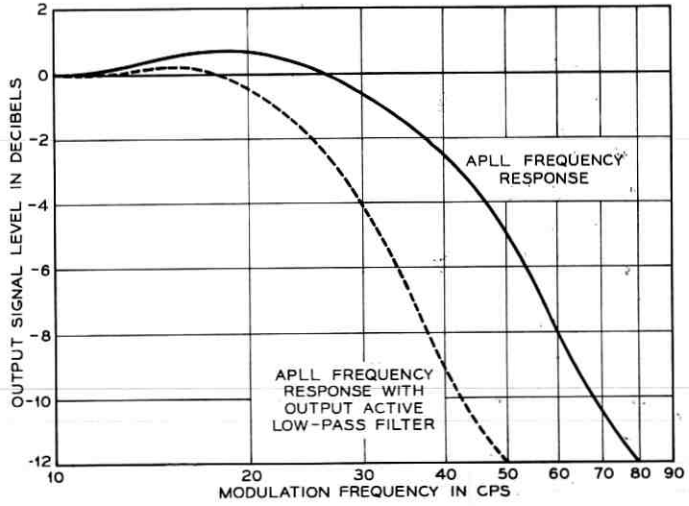

Fig. 16 — Measured frequency response of the APLL used as channel-filter and FM discriminator in a multichannel data receiver.

following the loop is illustrated in the same figure by the dashed curve. Fig. 17 shows waveforms for single- and multiple-channel inputs and for two different data speeds as they appear at the output of the post-detection filter. Fig. 18 shows the measured and calculated characteristics of the voltage-controlled oscillator used. It was built for a rest frequency of 1955 cps following the design procedure given in the Appendix. With the means of temperature compensation described there the relative frequency drift attained with conventional RC components was reduced to 0.25 per cent over a temperature range varying from 70°F to 140°F. Due to the amount of feedback possible in the three-stage amplifier section of the oscillator, the relative frequency change with ± 2-v supply voltage variations was in the range of 0.1 per cent. The transistors of the dc differential amplifier were preselected and housed in a ten-lead TO-5 package. In this way the voltage drift with temperature variations was negligible.

Error rate measurements on the APLL designed as a combined filter and discriminator for one channel of the data receiver with different loop parameters have been performed. Further experimental and analytical work must, however, be completed before an evaluation of the error performance of the APLL or a comparison with conventional discriminators can be made.

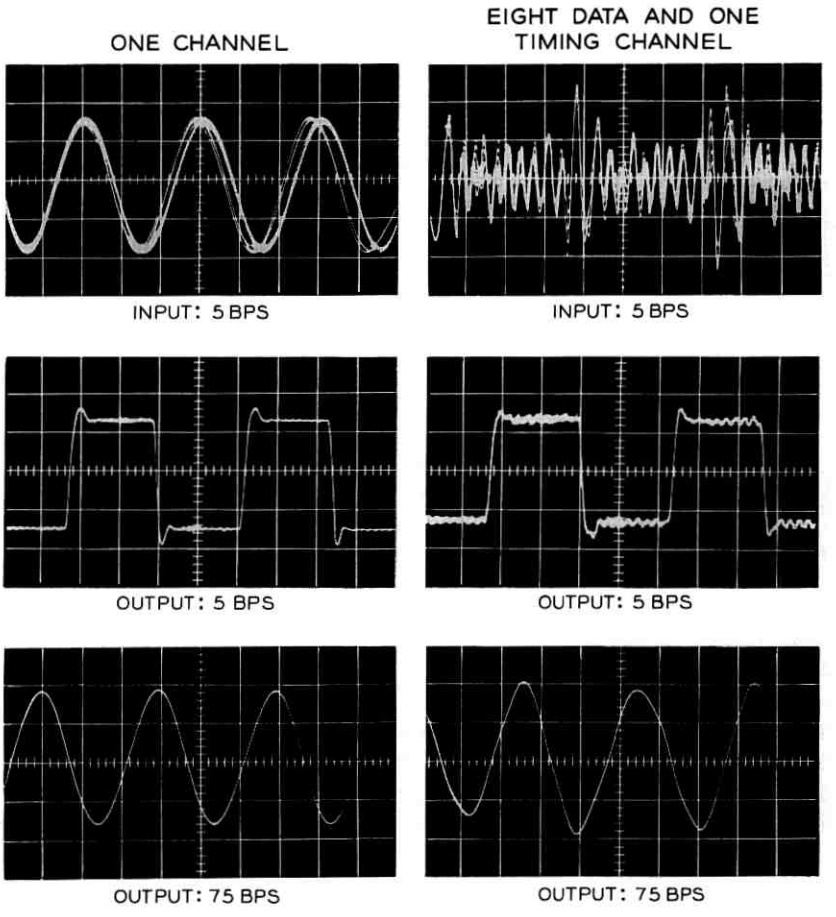

Fig. 17 — Output waveforms for single- and multiple-input signals and variable data speeds.

V. CONCLUSION

It has been shown that the APLL can be used as an FM signal separator and discriminator. Its signal separating properties allow for it to replace conventional bandpass filters in FM multiplex applications. The circuits entailed can be designed to satisfy the restrictions imposed by *RC* circuit miniaturization techniques. This makes it a useful device in the microminiaturization field, where practical solutions to the frequency selection problem are still few. Furthermore, it has the advantage

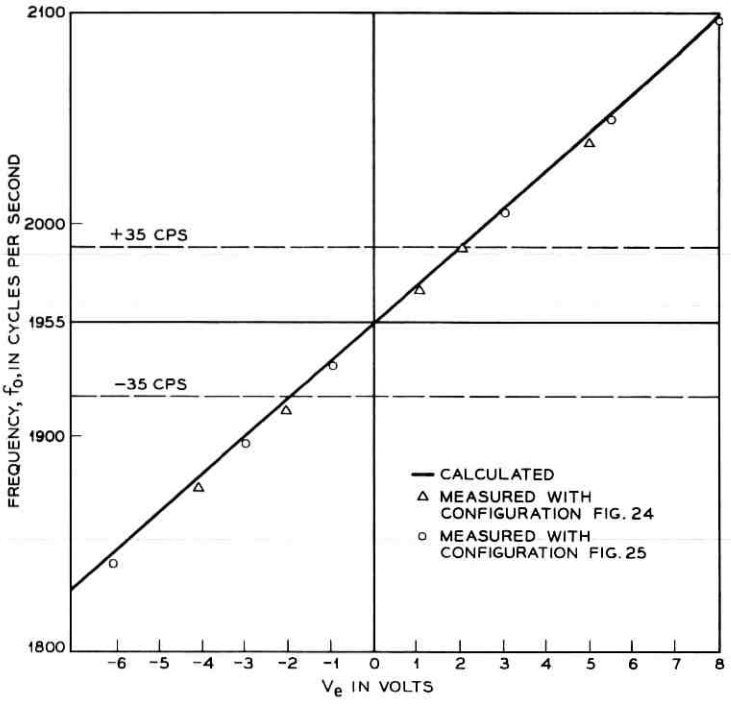

Fig. 18 — Oscillator frequency versus control voltage for the VCO shown in Fig. 12.

of ease of adjustment and reduction in circuit components over other comparable *RC* filtering methods.

In most applications the basic purpose of the APLL is to track narrow-band signals in a high-noise environment while maintaining minimum phase error. To do this the loop is designed so that the slowly varying frequency deviations to be tracked and the effective system bandwidth are very small with respect to the lock range. For this case the fact that the narrow bandwidth severely decreases the capture range is of no consequence.

For the type of application described in this paper, the system requirements of the APLL differ somewhat from those usually encountered. The limitations on phase error magnitude are given only by the requirement of linear loop action and are, therefore, not severe, i.e., the phase error may be as large as one radian. When the APLL is employed specifically as a digital data filter and demodulator the input frequency deviation is constant; for analog data applications the maximum devia-

tion must be given. The modulation frequency at the input may vary within a wide range, the only limit being given by the system bandwidth itself. The bandwidth of a given system cannot be arbitrarily reduced by low-pass filter modifications within the loop, because of the ensuing decrease in capture range. If, namely, the system is both to lock rapidly onto any initial frequency within the specified input frequency deviation as well as to relock rapidly after extraneous perturbations (e.g., impulse noise) might have thrown the system out of lock, the capture range has to cover the maximum frequency deviation. It is this last requirement that limits the attainable selectivity of a system all of whose parameters except those of the incorporated low-pass filter are given. An approximate expression relating system bandwidth and capture range has been derived for two types of low-pass filters. The minimum usable bandwidth compatible with a given frequency deviation thus results. It extends over a frequency band comparable to the capture range. Frequency deviation, minimum bandwidth, and capture range do not, therefore, differ greatly from one another.

As long as the lock range extends sufficiently beyond the maximum given frequency deviation, its value is not critical theoretically. Since it represents the total loop gain, however, it is uneconomical to make it larger than necessary. Its minimum value is determined by the requirement for linear loop operation. At the same time allowance must be made for a degree of circuit instability. This becomes increasingly important for narrow band systems operating at high channel frequencies. Using conventional circuit components and techniques the choice of lock range is therefore decided by the economics of designing for high circuit stability or for high loop gain. On the other hand with the major current microminiaturization techniques involving semiconductor integrated or thin film circuits the decision depends mainly on the inherent stability characteristics of the particular technique being used.

VI. ACKNOWLEDGMENT

The author would like to thank G. Malek for having carried out much of the laboratory work connected with the experimental results described in this paper.

APPENDIX

Analysis of 4-Section 180° Phase Shift Voltage-Controlled Oscillator with Single Variable Resistance

An equivalent diagram of the type of phase shift oscillator used is shown in Fig. 19. For an ideal voltage amplifier, $q = 0$, $m = 1$. The

variable element is nR. For no control voltage, the oscillator is at its rest frequency ω_c and $n = n_0$. The generalized oscillator frequency derived from Fig. 19 is given by:

$$\omega_0 = \frac{1}{RC}\left(\frac{2n + m + q + 4}{mn(q + 3) + n(5q + 4) + 3m(q + 1) + q}\right)^{\frac{1}{2}} \tag{47}$$

and the voltage gain necessary to compensate for network losses by:

$$-\frac{V_2}{V_1} = \frac{1}{mn\omega^4C^4R^4} - \frac{1}{\omega^2C^2R^2} \cdot\left[\frac{3(1 + q)}{mn} + \frac{7 + 2q}{m} + \frac{4 + q}{n} + 1\right] + q\left[\frac{1}{m} + \frac{1}{n} + 2\right] + 1. \tag{48}$$

Usually at least one of the two conditions for a perfect voltage amplifier can be fulfilled, sometimes at the cost of the other. Assuming for instance that the amplifier input impedance is much higher than the ladder shunt resistance ($m = 1$), (47) simplifies to:

$$\omega_0 = \frac{1}{RC}\left(\frac{2n + q + 5}{n(7 + 6q) + 4q + 3}\right)^{\frac{1}{2}}. \tag{49}$$

Substituting (49) in (48):

$$-\frac{V_2}{V_1} = \frac{1}{n}\left[\left(\frac{n(7 + 6q) + 4q + 3}{2n + q + 5}\right)^2 - \{n(8 + 2q) + 4q + 7\} \cdot\left\{\frac{n(7 + 6q) + 4q + 3}{2n + q + 5}\right\} + n(3q + 1) + q\right]. \tag{50}$$

Equations (49) and (50) have been plotted for different values of q in Figs. 20 and 21, respectively.

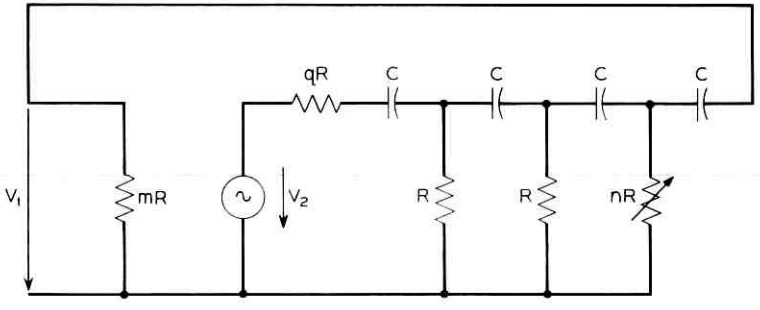

Fig. 19 — Equivalent diagram of a four-section 180° phase shift audio oscillator with a single variable resistance, when fed from a nonideal voltage source.

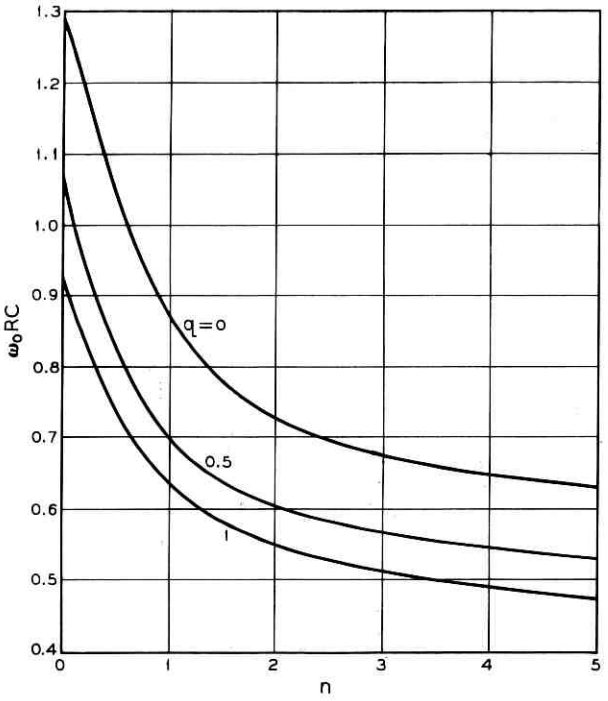

Fig. 20 — Frequency curves of a four-section ladder network with one variable resistor and the parameter q.

Fig. 21 shows that the point of minimum voltage loss does not occur when the ladder configuration is uniform. In the vicinity of the attenuation minimum, the curves are flat and it is possible over a given frequency range to choose a point n_0 for negligible voltage variation.

Expanding the frequency in a Taylor series around the rest frequency $\omega_c = \omega_0(n_0)$, (49) becomes:

$$\omega_0(n) = f(n)$$

$$= f(n_0) + \frac{df(n_0)}{dn}\frac{(n - n_0)}{1!} + \frac{d^2f(n_0)}{dn^2}\frac{(n - n_0)^2}{2!} + \cdots. \tag{51}$$

Retaining only the first two terms of the series, whereby

$$\omega_c = \frac{1}{RC}\left(\frac{2n_0 + q + 5}{n_0(7 + 6q) + 4q + 3}\right)^{\frac{1}{2}} \tag{52}$$

and

$$\nu = -\frac{\omega_c}{2}\frac{6q^2 + 29q + 29}{(2n_0 + q + 5)[n_0(7 + 6q) + 4q + 3]} \text{ rad/sec} \quad (53)$$

(51) becomes:

$$\omega_0(n) = \omega_c + \nu(n - n_0). \quad (54)$$

A voltage-dependent shunt element consisting of a combination of current-driven silicon diodes, a resistor and a capacitor is shown in Fig. 22. The diodes are dc-biased in series, but due to the large capacitor shunting them are ac-connected in parallel. The resulting symmetrical ac resistance seen from the ladder network reduces distortion that might arise from nonlinear operation of the diodes if the signal level is not sufficiently small. To keep the level as small as possible, this configura-

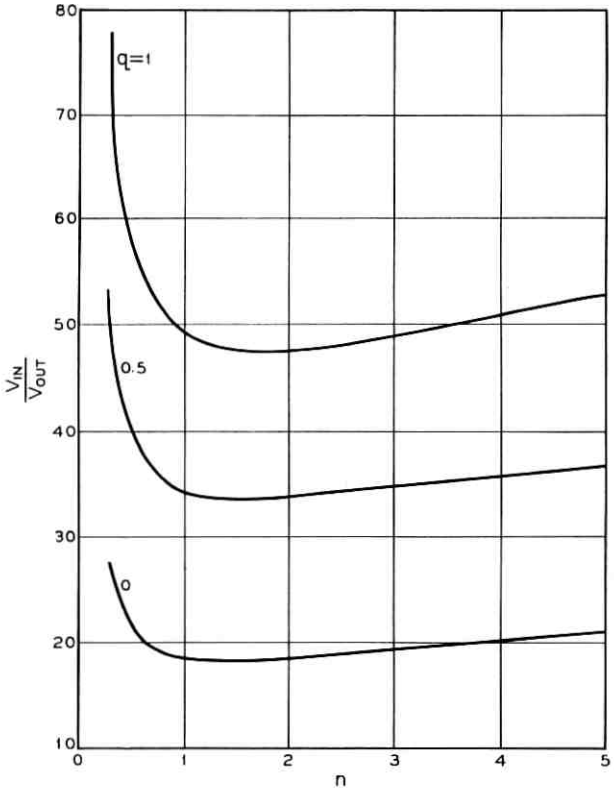

Fig. 21 — Attenuation curves of a four-section ladder network with one variable resistor and the parameter q (see text).

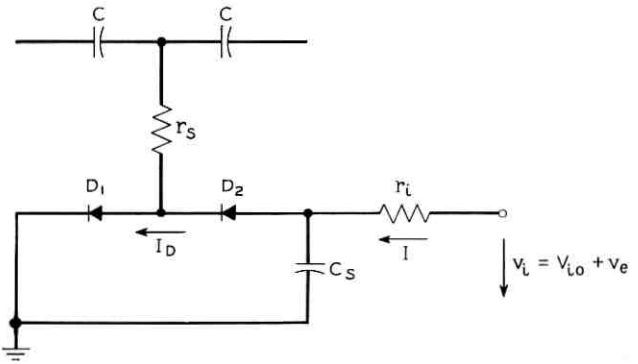

Fig. 22 — Voltage-variable, ladder shunt resistance using diodes.

tion substitutes the third shunt element of the ladder network. The fourth is reserved as a voltage divider to bias the first amplifier stage of the oscillator.

The incremental ac resistance of an ideal forward-biased diode is given by:

$$r_D = \frac{kT}{e} \frac{1}{I_D + I_0} \tag{55}$$

where the familiar term kT/e equals 25 millivolts at room temperature. I_D and I_0 denote the forward biasing and leakage currents, respectively. Silicon diodes are used, which have very small leakage currents, so that $I_D \gg I_0$ and

$$r_D \approx \frac{kT}{e} \frac{1}{I_D} = \frac{V_0}{I_D}. \tag{56}$$

The small-signal resistance is therefore inversely proportional to the diode current.

Fig. 23 shows the measured small-signal conductance of two silicon diodes of the type used in a configuration as shown in Fig. 22. The curve extends linearly over a wide range as theoretically predicted. Furthermore, measurements with numerous silicon diodes were very consistent. Because no account is taken of the lead and semiconductor resistance of the diode, nor of surface leakage effects, a steeper slope would, however, be expected. Nevertheless, since the modified diode configurations actually used reduce the influence of the diode charac-

teristics, the resulting VCO performance can be predicted sufficiently accurately.

It can be shown[24] that r_D is considerably less dependent on temperature change when fed from a current rather than from a voltage source. If necessary, additional temperature compensation is easily effected for current-fed diodes, as will be shown later.

Referring to Fig. 22, it is now of interest to calculate the change in oscillator frequency caused by the control voltage deviation v_e. To do so, the Taylor series (51) can be expanded around the steady-state diode current I_{D0} corresponding to the steady-state control voltage V_{i0} and to the oscillator rest frequency ω_c. With the results already obtained in (54) and referring to Fig. 22, the first two terms of (51) then become:

$$\omega(v_i) = \omega_c - \nu \frac{dn}{dI_D}(I_{D0}) \frac{dI_D}{dv_i}(I_{D0}) v_e \tag{57}$$

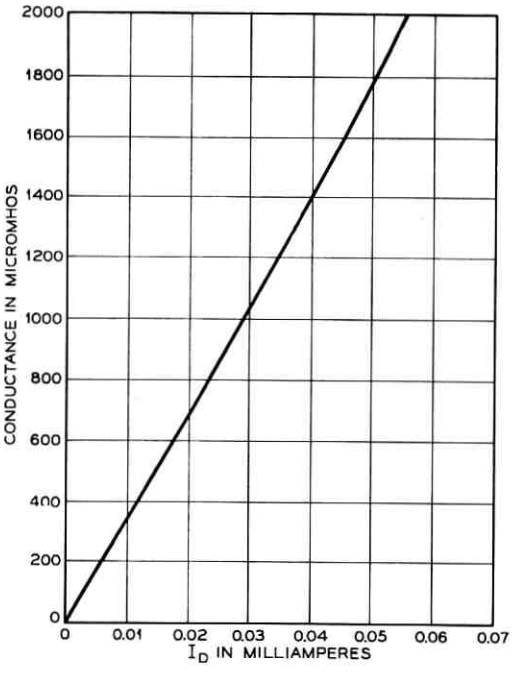

Fig. 23 — Small-signal conductance of two silicon diodes versus biasing current in the configuration shown in Fig. 22.

where

$$v_i = V_{i0} + v_e$$

and V_{i0} is the steady-state component of the control voltage and v_e its voltage increment. With the substitutions

$$\epsilon(I_D) = dn/dI_D \tag{58}$$

$$\kappa(I_D) = dI_D/dv_i \tag{59}$$

(57) becomes:

$$\omega(v_i) = \omega_c - \nu\epsilon\kappa v_e . \tag{60}$$

Comparing Figs. 19 and 22,

$$nR = r_s + \frac{1}{2}\frac{V_0}{I_D} \tag{61}$$

so that:

$$\epsilon(I_{L0}) = -\frac{1}{2R}\frac{V_0}{I_{D0}^2} . \tag{62}$$

From inspection of Fig. 22, we may write:

$$v_i = r_i I + 2V_D \tag{63}$$

and

$$I = I_D \tag{64}$$

where V_D is the voltage across one of the diodes. Assuming idealized diodes with negligible lead resistance:

$$V_D = V_0 \ln\left(\frac{I_D}{I_0} + 1\right). \tag{65}$$

Combining (63), (64) and (65) and differentiating with respect to the input voltage v_i,

$$\kappa(I_{L0}) = \frac{I_{L0}}{r_i I_{D0} + 2V_0} . \tag{66}$$

In (60) ν is determined by the oscillator frequency ω_c, the optimum choice of n_0, and the ratio of amplifier output to ladder resistance q. To insure current biasing of the diodes, r_i is limited in size either by the values attainable with thin film technology or by the output capa-

bilities of the preceding dc amplifier. This leaves the diode current I_{D0} to be chosen appropriately for a specified VCO sensitivity K_2. Since

$$K_2 = \nu\epsilon\kappa \tag{67}$$

it follows from (62) and (66) that:

$$I_{D0} = \left(\frac{V_0^2}{r_i^2} + \frac{\nu}{2K_2}\frac{V_0}{Rr_i}\right)^{\frac{1}{2}} - V_0/r_i. \tag{68}$$

As would be expected from the forward-biased diode characteristic or from Fig. 23, the current operating point decreases with an increase in the specified VCO sensitivity K_2.

The remaining network parameter r_s can now be determined by satisfying the condition that $n = n_0$ when $I_D = I_{D0}$. From (61)

$$r_s = n_0 R - \frac{1}{2}\frac{V_0}{I_{D0}}. \tag{69}$$

The frequency characteristic of the VCO incorporating the network shown in Fig. 22 and given by (53), (60), (62) and (67) is linear only over very small voltage increments. This is due to the nonlinear diode characteristics. Linearity can be greatly improved by modifying the voltage variable network according to Fig. 24. The additional resistor r_p shunting the diodes reduces κ and with it the influence of the nonlinear diode characteristics. Going through the same derivations as above

$$\kappa(I_{D0}) = \frac{r_p}{r_i + r_p}\frac{I_{D0}}{\frac{r_i r_p}{r_i + r_p} I_{D0} + 2V_0} \tag{70}$$

while ν and ϵ remain unchanged. This simplifies to

$$\kappa(I_{D0}) \approx \frac{r_p}{r_i}\frac{I_{D0}}{r_p I_{D0} + 2V_0} \tag{71}$$

if $r_i \gg r_p$.

A value of r_p can now be chosen such that r_i is at least an order of magnitude larger. The diode current necessary to establish a specified VCO sensitivity K_2 then follows from (62) and (70), namely:

$$I_{D0} = \left[\left(\frac{V_0}{r_p'}\right)^2 + \frac{\nu V_0}{2K_2 R r_i}\right]^{\frac{1}{2}} - \frac{V_0}{r_p'} \tag{72}$$

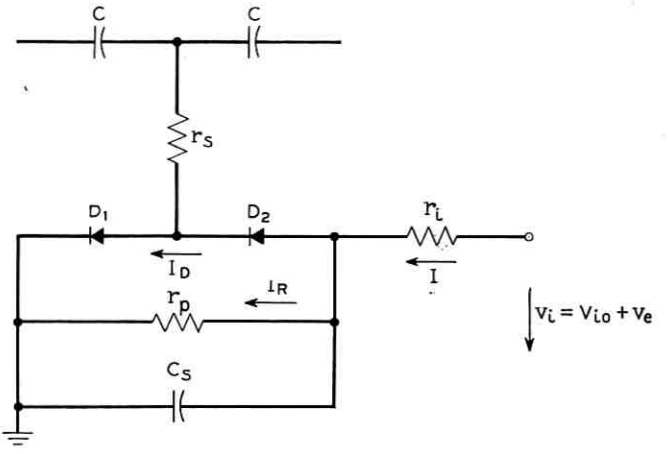

Fig. 24 — Modification of Fig. 22.

where

$$r_p' = \frac{r_p r_i}{r_p + r_i}. \tag{73}$$

Since $r_i \gg r_p$

$$I_{D0} \approx \left[\left(\frac{V_0}{r_p}\right)^2 + \frac{\nu V_0}{2K_2 R r_i}\right]^{\frac{1}{2}} - \frac{V_0}{r_p}. \tag{74}$$

It is apparent again here that I_{D0} increases with reduced VCO sensitivity K_2.

As in the preceding case, the remaining resistor r_s can now be determined by comparison of Figs. 19 and 24. The same value as before, given by (69), is found.

The characteristics of the low-pass filter consisting of r_i and C_s in the diode configuration of Fig. 24 will interfere with the operation of the APLL if its 3-db cutoff frequency cannot be maintained higher than the lock range of the loop (see Section II of main text). If this is the case the configuration shown in Fig. 25 must be used. If r_i is sufficiently larger than r_{p1} and r_{p2}, the latter two resistors can be made equal. Derived in the same manner as before,

$$\kappa(I_{D0}) = \frac{r_p}{r_i + r_p} \frac{I_{D0}}{\frac{r_p}{r_i} + r_p(2r_i + r_p)I_{D0} + 2V_0} \tag{75}$$

where again ν and ϵ remain unchanged. If $r_i \gg r_p$

$$\kappa(I_{L0}) \approx \frac{1}{2}\frac{r_p}{r_i}\frac{I_{D0}}{r_p I_{D0} + V_0}. \tag{76}$$

With the choice of r_p at least an order of magnitude smaller than r_i and substituting (62) into (75)

$$I_{L0} = \left\{V_0^2\left[\frac{r_i + r_p}{r_p(2r_i + r_p)}\right]^2 + \frac{\nu V_0}{2K_2R(2r_i + r_p)}\right\}^{\frac{1}{2}} - V_0\left[\frac{r_i + r_p}{r_p(2r_i + r_p)}\right]. \tag{77}$$

With $r_i \gg r_p$

$$I_{D0} \approx \left[\left(\frac{V_0}{2r_p}\right)^2 + \frac{\nu V_0}{4K_2Rr_i}\right]^{\frac{1}{2}} - \frac{V_0}{2r_p}. \tag{78}$$

From a comparison of Figs. 19 and 24

$$r_s = n_0R - \frac{r_p}{2} - \frac{V_0}{2}\frac{1}{I_{D0}}. \tag{79}$$

Eliminating the diode current from (78) and (79), a simplified expression relating r_p with r_s results, namely:

$$r_s \approx R\left(n_0 - \frac{r_p}{2R}\right) \tag{80}$$

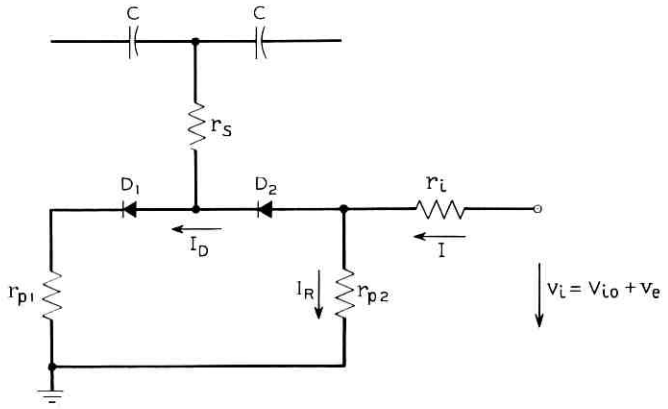

Fig. 25 — Another modification of Fig. 22.

when the specified VCO sensitivity

$$K_2 \ll \frac{r_p}{r_i} \frac{\nu}{V_0}.$$

The oscillator design can now proceed as follows: R is chosen such that with the largest available high-precision thin film capacitors (0.01 μf) the lowest necessary oscillator frequency can be obtained (in the data receiver described in the text, this corresponds to the lowest channel frequency). For all higher frequencies only the ladder capacitors are changed. If the capacitor value is not critical, R is chosen to conform with amplifier biasing considerations and to minimize q. q is given by the ratio of amplifier output impedance to R. n_0 is chosen for minimum voltage loss over a given frequency range and is obtained from Fig. 21. C results from (52) for a given frequency ω_c. ν is obtained from (53). The largest value of r_i compatible with thin film techniques and the output capabilities of the preceding stage is chosen, and r_p is made at least an order of magnitude smaller. The steady-state diode current is then given by one of the appropriate equations above, using the desired VCO frequency-voltage slope K_2. To obtain oscillation at the desired rest frequency, r_s is then determined by the appropriate expression derived above.

At this point a word must be said about the accuracy of (47) or (52) in connection with thin film tuning techniques. (47 might just as well have been used for the above computations, but would have been somewhat cumbersome.) In spite of the input and output impedances of the oscillator being taken into account, the frequency of the actual circuit may be lower than the predicted value owing to coupling, blocking and stray capacitance in the amplifier section not taken into account by the calculations. Initial oscillator frequency adjustment is therefore imperative. Thin film resistor adjustments can be made either by additional anodizing of the resistor films or by scratch-pad techniques in which the sides of thin film lattice configurations are progressively opened mechanically. Both methods are irreversible and only increase resistor values. To accommodate this circuit peculiarity, a frequency approximately 10 per cent higher than required is substituted in the equations defining the ladder capacitor C. In doing so n_0 is assumed equal to unity. Tuning the oscillator now consists of increasing the ladder resistance r_s (see Figs. 22, 24 and 25) and with it n_0, until the required frequency is reached. The necessary increase of n_0 can be computed from (54) and is approximately 25 per cent. Fortunately, this coincides with the region of minimum loss shown in Fig. 21. However, since the

attenuation curve is relatively flat for values of n larger than unity the exact value of initial frequency offset is not critical.

When biasing diodes from a constant-current source such that the leakage current is negligibly small, only the increase with temperature of $V_0 = kT/e$ has to be compensated for. The increase of this voltage has the over-all effect of increasing the oscillator frequency. By inserting two forward-biased silicon diodes in series with r_{p2} (see Fig. 25), the relative frequency variation is reduced to 0.25 per cent when the ambient temperature is varied from 70°F to 140°F. Due to the amount of feedback possible in the three-stage amplifier section of the oscillator, (see Fig. 15) the relative frequency change with ±2-v supply voltage variations is in the range of 0.1 per cent.

GLOSSARY

A: amplitude of VCO output
B: noise bandwidth of APLL
C: capacitance in uniform RC ladder network of 180° phase shift VCO
C_s: capacitance shunting diodes in VCO voltage variable network defined in Appendix
E: amplitude of single-channel input signal
E_c: amplitude of voltage controlling the switched modulator
$\overline{E_T}$: total rms voltage at APLL input
F: APLL figure of merit: capture range/3-db closed-loop bandwidth
F_0: APLL figure of merit: capture range/minimum-noise bandwidth
$F(j\omega)$: transfer function of low-pass filter
F_ω: amplitude-vs-frequency response of low-pass filter
$F_{\Delta\omega_i}$: filter frequency response at input frequency $\Delta\omega_i$
$F_{\Delta\omega_{ic}}$: filter frequency response at capture frequency $\Delta\omega_{ic}$
$f(t)$: impulse response of low-pass filter
Δf_i: maximum initial frequency offset before ultimate locking by APLL
I: input control current of VCO
I_D: diode bias current
I_{D0}: steady-state diode bias current
I_0: diode leakage current
I_R: current through resistance parallel with diodes in VCO voltage variable network defined in Appendix

K_1 : gain of phase detector and dc amplifier in volts

K_2 : gain (voltage sensitivity) of VCO in radians/(sec, volts)

$\pm K = K_1K_2$: steady-state loop gain and lock range of APLL in rad/sec

m: ratio of amplifier input impedance to ladder resistance R in 180° phase shift VCO

n: ratio of variable resistance to ladder resistance R in 180° phase shift VCO

n_0 : ratio of variable resistance to ladder resistance R at VCO rest frequency

q: ratio of amplifier output impedance to ladder resistance R in 180° phase shift VCO

R: resistance in uniform RC ladder network of 180° phase shift VCO

r_D : incremental ac resistance of ideal forward-biased diode

r_{p1}, r_{p2}, r_p, r_s, r_i : resistors in diode configurations dealt with and defined in Appendix

T: period of VCO output frequency

T_b : period of beat frequency ω_b

T_c : capture time, i.e., time it takes APLL to lock onto input frequency

T_{co} : time constant of simple RC low-pass filter

T_0 : switching time of switched modulator

$T(j\omega)$: discriminator transfer function of APLL

V_{io} : dc component of VCO control voltage v_i

$V_0 = kT/e$: where k is Boltzmann's constant, e the charge of an electron and T the temperature in degrees Kelvin

v_e : increment of VCO control voltage (error voltage) and discriminator output

v_{ec} : error voltage at capture instant

$\hat{v}_{ec}$: amplitude of error voltage at capture instant

v_i : control voltage of VCO

v_{out} : dc component of switched modulator output voltage

x: ratio of APLL undamped natural frequency ω_n to lock range K

$\epsilon = dn/dI_D$: sensitivity of resistance ratio n to incremental change in diode current for diode configuration dealt with in Appendix

ζ: ratio of actual to critical damping of APLL

ζ_0 : damping ratio for system with minimized noise bandwidth

$\kappa = dI_D/dv_i$: sensitivity of diode current to incremental change in control voltage for diode configurations dealt with in Appendix

μ: dc amplifier gain
$\nu = d\omega/dn$: sensitivity of VCO frequency to incremental change in n
τ_1, τ_2: time constants of low-pass lag network defined in Fig. 2(b)
$\varphi(t)$: instantaneous difference (error) between input and VCO phase
φ_e: phase error at capture instant
$\varphi_i(t)$: instantaneous input phase
φ_{LIN}: maximum phase error corresponding to given percentage linearity error in phase detector output characteristic
$\varphi_0(t)$: instantaneous VCO phase
$\psi(\omega)$: phase versus frequency response of low-pass filter
ω_b: out-of-lock beat between input and VCO frequency
ω_c: channel center frequency; VCO rest frequency
ω_{co}: 3-db cutoff frequency of simple RC low-pass filter
$\omega_i(t)$: instantaneous frequency of APLL input signal
ω_n: undamped natural frequency of APLL
$\omega_0(t)$: instantaneous VCO frequency
$\Delta\omega(t)$: difference between instantaneous input and VCO frequency
$\Delta\omega_i(t)$: instantaneous input frequency deviation from ω_c
$\Delta\omega_{ic}$: first approximation of capture frequency deviation at input and 3-db closed-loop bandwidth of APLL
$\Delta\omega_{ic}'$: closer approximation of capture frequency than $\Delta\omega_{ic}$
$\Delta\omega_0(t)$: instantaneous VCO frequency deviation from ω_c
$\overline{\Delta\omega_0}$: mean VCO frequency in the unlocked state
$\Delta\omega_{oc}$: first approximation of VCO capture frequency deviation; by definition $\Delta\omega_{ic} \equiv \Delta\omega_{oc}$
$\Delta\omega_{ic}/K$: first approximation of capture ratio, defined as ratio of capture-to-lock range
$\Delta\omega_{ic}'/K$: closer approximation of capture ratio than $\Delta\omega_{ic}/K$.

REFERENCES

1. Rand, A., Inductor Size vs Q: A Dimensional Analysis, IEEE Trans. Comp. Parts, *CP-10*, March, 1963, pp. 31–35.
2. Newell, W. E., The Frustrating Problem of Inductors in Integrated Circuits Electronics, *37*, March, 1964, pp. 50–52.
3. Liebsch, H., Miniaturisierung von Induktivitaetsbauelementen, Nachrichtentechnik, *12*, Nov., 1962, pp. 409–413.
4. Newell, W. E., Tuned Integrated Circuits — A State-of-the-Art Survey, Proc. IEEE, *52*, Dec., 1964, pp. 1603–1608.
5. Preston, G. W., and Tellier, J. C., The Lock-In Performance of an AFC Circuit, Proc. IRE, *41*, Feb., 1953, pp. 249–251.
6. Viterbi, A. J., Acquisition and Tracking Behavior of Phase-Locked Loops,

Proc. Symposium on Active Networks and Feedback Systems, Polytechnic Institute of Brooklyn, *10*, April, 1960, pp. 583–619.
7. Richman, D., Color-Carrier Reference Phase Synchronization Accuracy in NTSC Color Television, Proc. IRE, *42*, Jan., 1954, pp. 106–133.
8. Richman, D., The DC Quadricorrelator, A Two-Mode Synchronization System, Proc. IRE, *42*, Jan., 1954, pp. 288–299.
9. Gassmann, G. G., Neue Phasen-und Frequenzvergleichschaltungen, Archiv der Elektrischen Übertragung, *15*, Aug., 1961, pp. 359–376.
10. White, E. L. C., The Pull-In Range of an APC Loop, Report No. KR/94, EMI Research Laboratories, Nov., 1955.
11. George, T. S., Analysis of Synchronizing Systems for Dot Interlaced Color TV, Proc. IRE, *39*, Feb., 1951, pp. 124–131.
12. Jaffe, R., and Rechtin, E., Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels, IRE Trans. Inf. Theory, *IT-1*, March, 1955, pp. 66–76.
13. Gruen, W. J., Theory of AFC Synchronization, Proc. IRE, *41*, Aug., 1953, pp. 1043–1048.
14. McAleer, H. T., A New Look at the Phase-Locked Oscillator, Proc. IRE, *47*, June, 1959, pp. 1137–1143.
15. Highleyman, W. H., and Jacob, E. S., An Analog Multiplier Using Two Field Effect Transistors, IRE Trans. Comm. Systems, *CS-10*, Sept., 1962, pp. 311–317.
16. Holcomb, S. W., Opp, F., and Walston, J. A., Choppers: Their Characteristics and Applications, T. I. Application Report, July 15, 1963.
17. Wright, M. J. A., A Transistor Phase Sensitive Demodulator of High Performance, Electronic Engng., *34*, Oct., 1962, pp. 698–700.
18. Bright, R. L., Junction Transistors Used as Switches, Trans. AIEE, *74*, pt. 1, March, 1955, pp. 111–121.
19. Vogelsberg, D., Transistoren als Phasenselektive Gleichrichter, Frequenz, *17*, 4, 1963.
20. Verster, T. C., Silicon Planar Epitaxial Transistors as Fast and Reliable Low-Level Switches, IEEE Trans. on Electron Devices, *ED-11*, May, 1964, pp. 228–237.
21. Bennett, W. R., and Salz, J., Binary Data Transmission by FM over a Real Channel, B.S.T.J., *42*, Sept., 1963, pp. 2387–2426.
22. McLean, D. A., Schwarz, N., and Tidd, E. D., Tantalum-Film Technology, Proc. IEEE, *52*, Dec., 1964, pp. 1450–1462.
23. Orr, W. H., Precision Tuning of a Thin-Film Notch Filter, ISSCC Digest of Technical Papers, Feb., 1964, pp. 56–57.
24. Greiner, R. A., *Semiconductor Devices and Applications*, McGraw-Hill Book Co., Inc., New York, N. Y., 1961; Voltage Controlled Wide-Range Oscillator, Electronics, *34*, Dec. 22, 1961, pp. 31–35.

Some Results on the Theory of Physical Systems Governed by Nonlinear Functional Equations

By I. W. SANDBERG

(Manuscript received February 2, 1965)

The main purpose of this paper is to present a reasonably complete picture of the results of the first phase of some recent research on the properties of solutions of nonlinear functional equations that frequently arise in the study of physical systems. We consider in detail the properties of a vector nonlinear Volterra integral equation of the second kind, and some conditions are presented for the norm boundedness of solutions of a functional equation of similar type defined on an abstract space.

More specifically, concerning the Volterra equation, conditions are presented under which the solutions (a) approach zero as $t \to \infty$, (b) approach zero exponentially as $t \to \infty$, (c) are uniformly bounded on $t \geqq 0$, (d) are square integrable on $[0, \infty)$, or (e) are ultimately periodic. On the basis of these results, it appears that an input-output stability theory of a large class of time-varying nonlinear systems of engineering interest is well within sight.

I. PRELIMINARY NOTATION AND DEFINITIONS

The set of real measurable N-vector-valued functions of the real variable t defined on $[0, \infty)$ is denoted by $\mathcal{H}_N(0, \infty)$ and the jth component of $f \varepsilon \mathcal{H}_N(0, \infty)$ is denoted by f_j.

The sets $\mathcal{L}_{\infty N}(0, \infty)$ and $\mathcal{L}_{2N}(0, \infty)$ are defined by

$$\mathcal{L}_{\infty N}(0, \infty) = \{f \mid f \varepsilon \mathcal{H}_N(0, \infty), \sup_{t \geqq 0} [f'(t)f(t)] < \infty\}$$

$$\mathcal{L}_{2N}(0, \infty) = \left\{f \mid f \varepsilon \mathcal{H}_N(0, \infty), \int_0^\infty f'(t)f(t)\, dt < \infty\right\},$$

in which $f'(t)$ denotes the transpose of $f(t)$. In order to be consistent with standard notation, we let $\mathcal{L}_2(0, \infty) = \mathcal{L}_{2N}(0, \infty)$ when $N = 1$.

Let $y \varepsilon (0, \infty)$ and define f_y by

$$f_y(t) = f(t) \quad \text{for} \quad t \in [0, y]$$
$$= 0 \quad \text{for} \quad t > y$$

for any $f \in \mathcal{K}_N(0,\infty)$, and let

$$\mathcal{E}_N = \{f \mid f \in \mathcal{K}_N(0,\infty),\ f_y \in \mathcal{L}_{2N}(0,\infty) \quad \text{for} \quad 0 < y < \infty\}.$$

With A an arbitrary real measurable $N \times N$ matrix-valued function of t with elements $\{a_{nm}\}$ defined on $[0,\infty)$, let $\mathcal{K}_{pN}$ $(p = 1,2)$ denote

$$\left\{A \mid \int_0^\infty |a_{nm}(t)|^p\, dt < \infty \qquad (n, m = 1,2, \cdots, N)\right\}.$$

For an arbitrary $f \in \mathcal{K}_N(0,\infty)$, let $\psi[f(t),t]$ denote

$$(\psi_1[f_1(t),t],\psi_2[f_2(t),t], \cdots, \psi_N[f_N(t),t])'$$

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 $w \in (-\infty,\infty)$ and $t \in [0,\infty)$ such that

(*i*) $\psi_n(0,t) = 0$ for $t \in [0,\infty)$ and $n = 1,2, \cdots, N$

(*ii*) $\psi_n[w(t),t]$ $(n = 1,2, \cdots, N)$ is a measurable function of t whenever $w(t)$ is measurable.

Let α and β denote real numbers such that $\alpha \leqq \beta$. We shall say that $\psi[\cdot\ ,\cdot] \in \Psi_0(\alpha,\beta)$ if and only if

$$\alpha \leqq \frac{\psi_n(w,t)}{w} \leqq \beta \qquad (n = 1,2, \cdots, N)$$

for $t \in [0,\infty)$ and all real $w \neq 0$; and we shall say that $\psi[\cdot\ ,\cdot] \in \Psi(\alpha,\beta)$ if and only if

$$\alpha \leqq \frac{\psi_n(w_1,t) - \psi_n(w_2,t)}{w_1 - w_2} \leqq \beta \qquad (n = 1,2, \cdots, N)$$

for $t \in [0,\infty)$ and all real w_1, w_2 such that $w_1 \neq w_2$.

II. INTRODUCTION

In the study of physical systems containing time-varying nonlinear elements, attention is frequently focused on the properties of the equation

$$g(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau),\tau]\, d\tau, \qquad t \geqq 0 \tag{1}$$

in which $g \in \mathcal{E}_N$, $f \in \mathcal{E}_N$, $k \in \mathcal{K}_{1N}$ and $\psi[\cdot\ ,\cdot] \in \Psi_0(\alpha,\beta)$ for some α and β. For example, consider the multi-input multi-output nonlinear feedback

system of Fig. 1 in which u, v, f, w, x, and y are assumed to denote elements of $\mathcal{E}_N$ with the input $u \in \mathcal{L}_{2N}(0,\infty)$. Let the block labeled ψ represent N memoryless time-varying nonlinear elements which introduce the constraint $w(t) = \psi[f(t),t]$ for $t \geqq 0$ with $\psi[\cdot\,,\cdot] \in \Psi_0(\alpha,\beta)$, and let the blocks labeled $\mathbf{K}_1$, $\mathbf{K}_2$, and $\mathbf{K}_3$, which represent linear time-invariant portions of the system, introduce the constraints

$$f(t) = \int_0^t k_1(t-\tau)v(\tau)\,d\tau + g_1(t)$$

$$x(t) = \int_0^t k_3(t-\tau)w(\tau)\,d\tau + g_3(t)$$

$$y(t) = \int_0^t k_2(t-\tau)x(\tau)\,d\tau + g_2(t)$$

for $t \geqq 0$, in which the impulse response matrices k_1, k_2, and k_3 are elements of $\mathcal{K}_{1N}$ and the initial condition functions g_1, g_2, and g_3 are elements of $\mathcal{L}_{2N}(0,\infty)$. Then

$$g(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

where k, the inverse Fourier transform of

$$\int_0^\infty k_1(t)e^{-i\omega t}\,dt \int_0^\infty k_2(t)e^{-i\omega t}\,dt \int_0^\infty k_3(t)e^{-i\omega t}\,dt,$$

is an element of $\mathcal{K}_{1N}$, and g defined for $t \geqq 0$ by

$$g(t) = g_1(t) + \int_0^t k_1(t-\tau)u(\tau)\,d\tau - \int_0^t k_1(t-\tau)g_2(\tau)\,d\tau - \int_0^t k_1(t-\tau)\int_0^\tau k_2(t-q)g_3(q)\,dq\,d\tau$$

is an element of $\mathcal{L}_{2N}(0,\infty)$.

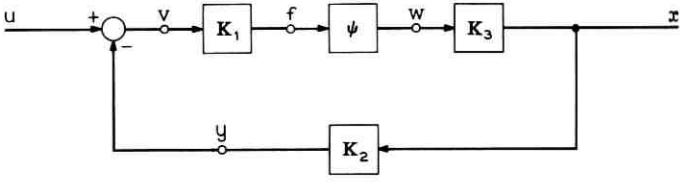

Fig. 1 — Nonlinear feedback system.

Equations of the form

$$\frac{d^n f}{dt^n} = g(t) + \int_0^t k(t-\tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0 \tag{2}$$

also arise in a natural way in the study of physical systems. For example, this type of equation with $n = 1$ is encountered in the theory of feedback control systems containing a motor in the forward path.

This paper is addressed primarily to the engineer interested in the mathematical aspects of nonlinear systems. Its main purpose is to present a reasonably complete picture of the results of the first phase of some recent research on the properties of solutions of (1) and of equations of similar type defined on an abstract space. The essentials of the material to be presented are drawn largely from Refs. 1 and 2.

With the exception of some observations in connection with the problem of determining lower bounds on the decay rate or upper bounds on the growth rate of solutions of (1), each of the results of Section III constitutes a set of sufficient conditions, in which a certain frequency-domain condition†, †† plays a central role, under which the solutions of (1) are stable in one of several significant senses. More specifically, conditions are presented under which: $f(t) \to 0$ (i.e., the zero vector) as $t \to \infty$, $g \in \mathfrak{L}_{2N}(0, \infty)$ implies $f \in \mathfrak{L}_{2N}(0, \infty)$, $g \in \mathfrak{L}_{\infty N}(0, \infty)$ implies $f \in \mathfrak{L}_{\infty N}(0, \infty)$, and g ultimately periodic with period T implies that f is ultimately periodic with period T. Conditions are also presented under which f depends continuously on g. On the basis of these results, it appears that an input-output stability theory of a large class of time-varying nonlinear systems is well within sight. In particular, the class is not restricted to lumped-parameter systems. An example is presented concerning the necessity of some of the conditions.

Section IV is devoted to the proof of a previously unpublished result concerning (2) with n an arbitrary nonnegative integer, and, for simplicity, f and g scalar functions (i.e., $N = 1$). A set of conditions is established under which $g \in \mathfrak{L}_2(0, \infty)$ implies that $f \in \mathfrak{L}_2(0, \infty)$.

In Section V, we consider some properties of equations defined on an

---

† For some other results concerned with frequency-domain conditions for the stability of nonlinear or time-varying systems, see Refs. 3–6. In particular, Ref. 3 describes in detail the work of V. M. Popov.

†† After this paper had been submitted for publication, the following related papers came to the writer's attention: B. N. Naumov and Ya. Z. Tsypkin, A Frequency Criterion for Absolute Process Stability in Nonlinear Automatic Control Systems, Automation and Remote Control, *25*, Jan. 1965; and V. A. Yakubovich, The Matrix-Inequality Method in the Theory of Stability of Nonlinear Control Systems: I. The Absolute Stability of Forced Vibrations, Automation and Remote Control, *25*, Feb. 1965.

abstract space. More specifically, the basic problem considered is a direct generalization of the problem of establishing the $\mathfrak{L}_{2N}(0,\infty)$-boundedness of solutions of nonlinear functional equations (i.e., a generalization of the central problem of Ref. 1). In particular, Theorem 8 is of immediate utility in obtaining results similar to those of Section III for other types of equations [e.g., the discrete analog of (1) which is of interest in the theory of sampled-data systems]. Section 5.2 is essentially a restatement of the key argument of Ref. 1 in a more abstract setting. In Section 5.3 some new results are proved.

In the Appendix we state some results concerning an integral equation similar to (1) that arises in the study of ordinary linear differential equations, and conditions are presented under which all solutions of a system of second-order equations with varying coefficients approach zero exponentially as $t \to \infty$.

III. RESULTS CONCERNING THE PROPERTIES OF (1)

3.1 *Further 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 $\Lambda\{M\}$, and 1_N denotes the identity matrix of order N.

The norm of $f \in \mathfrak{L}_{2N}(0,\infty)$ is denoted by $\| f \|$ and is defined by

$$\| f \| = \left(\int_0^\infty f'(t) f(t)\, dt \right)^{\frac{1}{2}}.$$

The symbol s denotes a scalar complex variable with $\sigma = \text{Re}\,[s]$ and $\omega = \text{Im}\,[s]$.

We shall say that k is an element of the set $\Phi(\alpha,\beta)$ if and only if $k \in \mathcal{K}_{1N}$ and, with

$$K(s) = \int_0^\infty k(t) e^{-st}\, dt \quad \text{for} \quad \sigma \geqq 0,$$

(*i*) $\det [1_N + \frac{1}{2}(\alpha + \beta) K(s)] \neq 0 \quad \text{for} \quad \sigma \geqq 0$

(*ii*) $\frac{1}{2}(\beta - \alpha) \sup_{-\infty < \omega < \infty} \Lambda\{[1_N + \frac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} K(i\omega)\} < 1.$

Comments: It can be shown that conditions (*i*) and (*ii*) above are satisfied if $\alpha \geqq 0$ and $[K(i\omega) + K(i\omega)^*]$ is nonnegative definite for all ω.

For $N = 1$, conditions (i) and (ii) above are met if $\beta > 0$ and 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).

Concerning the condition for $\alpha > 0$, if $\beta = \alpha$, then the circle C_1 degenerates to a point, and the criterion becomes the well-known Nyquist stability criterion.

3.2 *Results*

Our first theorem, which is proved in Ref. 1 as an application of an abstract result similar to Theorem 8 of Section V, is the key result of this section. It is of direct interest in the theory of stability of dynamical systems, and it plays an important role in the proof of each of the other theorems of this section.

Theorem 1: Let $k \in \Phi(\alpha,\beta)$, *let* $\psi[\cdot\ ,\ \cdot] \in \Psi_0(\alpha,\beta)$, *and let*

$$g(t) = f(t) + \int_0^t k(t - \tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0$$

where $g \in \mathcal{L}_{2N}(0,\infty)$ *and* $f \in \mathcal{E}_N$. *Then* $f \in \mathcal{L}_{2N}(0,\infty)$, *and there exists a positive constant* ρ *which depends only on* k, α, *and* β *such that*

$$\| f \| \leqq \rho \| g \|.$$

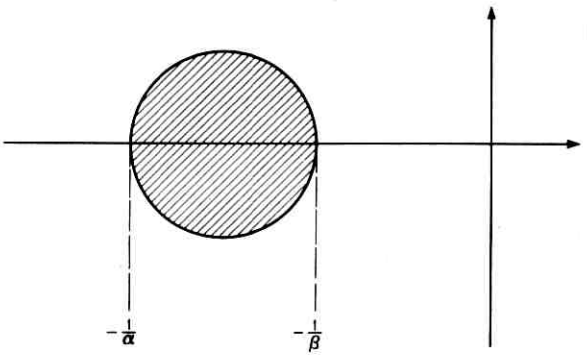

Fig. 2 — Location of the "critical circle" C_1 in the complex plane ($\alpha > 0$).

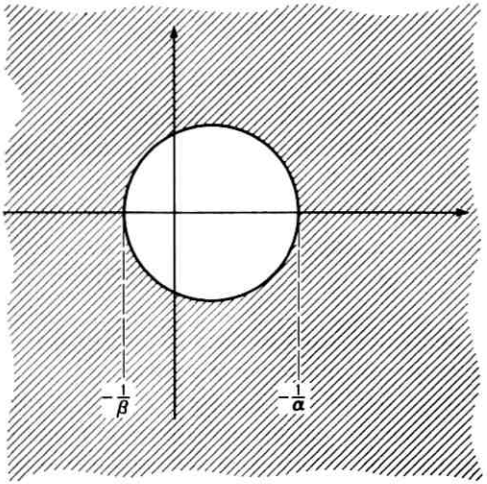

Fig. 3 — Location of the "critical circle" C_2 in the complex plane ($\alpha < 0$).

Corollary 1 (a): Let $k \in \Phi(\alpha,\beta)$, *let* $\psi[\cdot\ ,\cdot] \in \Psi(\alpha,\beta)$, *and let*

$$g_1(t) = f_1(t) + \int_0^t k(t-\tau)\psi[f_1(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

$$g_2(t) = f_2(t) + \int_0^t k(t-\tau)\psi[f_2(t),\tau]\,d\tau, \qquad t \geqq 0$$

where g_1, g_2, f_1, $f_2 \in \mathcal{E}_N$ *and* $(g_1 - g_2) \in \mathcal{L}_{2N}(0,\infty)$. *Then* $(f_1 - f_2) \in \mathcal{L}_{2N}(0,\infty)$, *and there exists a positive constant* ρ *which depends only on* k, α, *and* β *such that*

$$\| f_1 - f_2 \| \leqq \rho \| g_1 - g_2 \|.$$

Proof of Corollary 1 (a): Let $q_j(t)$ be defined on $[0,\infty)$ by

$$q_j(t) = \frac{\psi_j[f_{1j}(t),t] - \psi_j[f_{2j}(t),t]}{f_{1j}(t) - f_{2j}(t)}, \qquad t \in \{t \mid t \geqq 0, f_{1j}(t) \neq f_{2j}(t)\}$$

$$= \tfrac{1}{2}(\alpha + \beta), \qquad t \in \{t \mid t \geqq 0, f_{1j}(t) = f_{2j}(t)\}$$

for $j = 1,2,\cdots,N$; and let $q(t)$ denote the diagonal matrix diag $[q_1(t),q_2(t),\cdots,q_N(t)]$. Then $\alpha \leqq q_j(t) \leqq \beta$ for $j = 1,2,\cdots,N$ and

$$g_1(t) - g_2(t) = f_1(t) - f_2(t) + \int_0^t k(t-\tau)q(\tau)[f_1(\tau) - f_2(\tau)]\,d\tau, \qquad t \geqq 0.$$

The conclusion of the corollary follows from this equation and the theorem.

Remarks: A direct application of the Schwarz inequality and the Riemann-Lebesgue Lemma shows† that if the hypotheses of Theorem 1 are satisfied and $g(t) \to 0$ as $t \to \infty$ [i.e., $g_j(t) \to 0$ as $t \to \infty$ for $j = 1,2, \cdots, N$], then $f(t) \to 0$ as $t \to \infty$ provided that $k \in \mathcal{K}_{2N}$ [observe that $k \in \mathcal{K}_{2N}$ if $k \in \mathcal{K}_{1N}$ and the elements of k are uniformly bounded on $[0, \infty)$]. Similarly, if the hypotheses of Corollary 1(a) are met and $[g_1(t) - g_2(t)] \to 0$ as $t \to \infty$, then $[f_1(t) - f_2(t)] \to 0$ as $t \to \infty$ provided that $k \in \mathcal{K}_{2N}$.

Theorem 8 of Section V leads to a result for the integral equation (1) that is actually somewhat stronger than that stated as Theorem 1. If k, $\psi[\cdot\,, \cdot]$ and f are as defined in Theorem 1, and if $g \in \mathcal{E}_N$ satisfies the integral equation, then it can be shown that Theorem 8 implies the existence of a positive constant ρ which depends only on k, α, and β such that $\| f_\nu \| \leqq \rho \| g_\nu \|$ for all $\nu > 0$.

Theorem 2: Let

$$g(t) = f(t) + \int_0^t k(t - \tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0 \tag{3}$$

in which $\psi[\cdot\,, \cdot] \in \Psi_0(\alpha,\beta)$, $f \in \mathcal{E}_N$, *and there exists a real constant* c_1 *such that*

(*i*) $ge^{c_1 t} \in \mathcal{L}_{2N}(0, \infty)$
(*ii*) $ke^{c_1 t} \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$
(*iii*) $ke^{c_1 t} \in \Phi(\alpha,\beta)$.

Then there exists a positive constant c_2 such that

$$|f_j(t)| \leqq |g_j(t)| + c_2 e^{-c_1 t}, \qquad t \geqq 0$$

for $j = 1, 2, \cdots, N$.

Corollary 2(a): Let

$$g(t) = f(t) + \int_0^t k(t - \tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0$$

in which $\psi[\cdot\,, \cdot] \in \Psi_0(\alpha,\beta)$, $k \in \Phi(\alpha,\beta)$, $f \in \mathcal{E}_N$, *and there exists a positive constant* c_1 *such that*

(*i*) $ge^{c_1 t} \in \mathcal{L}_{2N}(0, \infty)$
(*ii*) $ke^{c_1 t} \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$.

† See the proof of Theorem 6 of Ref. 1.

Then there exist positive constants c_2 and c_3 such that

$$|f_j(t)| \leqq |g_j(t)| + c_2 e^{-c_3 t}, \qquad t \geqq 0.$$

Proof of Theorem 2: From the fact that f and g satisfy (3), we have

$$e^{c_1 t} g(t) = \hat{f}(t) + \int_0^t e^{c_1(t-\tau)} k(t-\tau) e^{c_1 \tau} \psi[e^{-c_1 \tau} \hat{f}(\tau), \tau]\, d\tau, \qquad t \geqq 0 \quad (4)$$

in which $\hat{f}(t) = f(t)e^{c_1 t}$. Since

$$\alpha \leqq \frac{e^{c_1 t} \psi_n[e^{-c_1 t} x, t]}{x} \leqq \beta \qquad (n = 1, 2, \cdots, N)$$

for all real $x \neq 0$ and $t \geqq 0$, it follows from Theorem 1 that $\hat{f} \in \mathfrak{L}_{2N}(0, \infty)$. Thus $e^{c_1 t}\psi[f(\cdot), \cdot] \in \mathfrak{L}_{2N}(0, \infty)$, and by the Schwarz inequality, there exists a positive constant c_2 such that the modulus of the jth component of

$$\int_0^t e^{c_1(t-\tau)} k(t-\tau) e^{c_1 \tau} \psi[f(t), \tau]\, d\tau$$

does not exceed c_2 for $t \geqq 0$ and $j = 1, 2, \cdots, N$. Thus, using (4),

$$|f_j(t)| \leqq |g_j(t)| + c_2 e^{-c_1 t}, \qquad t \geqq 0$$

for $j = 1, 2, \cdots, N$.

Proof of Corollary 2(a): Let

$$K(i\omega - \rho) = \int_0^\infty k(t) e^{-(i\omega - \rho)t}\, dt$$

for $\rho \leqq c_1$ and $-\infty < \omega < \infty$. It clearly suffices to prove that there exists a positive constant $c_4 \leqq c_1$ such that $ke^{\rho t} \in \Phi(\alpha,\beta)$ for $0 \leqq \rho \leqq c_4$. The existence of such a constant follows easily from the fact that each element of $[K(s) - K(s - \rho)]$ approaches zero uniformly in $\sigma \geqq 0$ as $\rho \to 0+$. The details are omitted.

For some results related to Theorem 2 and Corollary 2(a), see the Appendix.

The following theorem is proved in Ref. 2 with the aid of Theorem 1.

Theorem 3: Let $k \in \Phi(\alpha,\beta)$ with $t^p k \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$ for $p = 0,1,2$. Let $\psi[\cdot\,, \cdot] \in \Psi_0(\alpha,\beta)$, and let

$$g(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0$$

where $g \in \mathfrak{L}_{\infty N}(0,\infty)$ *and* $f \in \mathcal{E}_N$. *Then* $f \in \mathfrak{L}_{\infty N}(0,\infty)$, *there exists a positive constant* ρ *which depends only on* k, α, *and* β *such that*

$$\max_j \sup_{t \geqq 0} |f_j(t)| \leqq \rho \max_j \sup_{t \geqq 0} |g_j(t)|,$$

and $f_j(t) \to 0$ *as* $t \to \infty$ *for* $j = 1,2, \cdots, N$ *whenever* $g_j(t) \to 0$ *as* $t \to \infty$ *for* $j = 1,2, \cdots, N$.

Corollary 3(a): Let $k \in \Phi(\alpha,\beta)$ *with* $t^p k \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$ *for* $p = 0,1,2$. *Let* $\psi[\cdot\,,\cdot] \in \Psi(\alpha,\beta)$, *and let*

$$g_1(t) = f_1(t) + \int_0^t k(t-\tau)\psi[f_1(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

$$g_2(t) = f_2(t) + \int_0^t k(t-\tau)\psi[f_2(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

where $g_1, g_2, f_1, f_2 \in \mathcal{E}_N$ *and* $(g_1 - g_2) \in \mathfrak{L}_{\infty N}(0,\infty)$. *Then* $(f_1 - f_2) \in \mathfrak{L}_{\infty N}(0,\infty)$, *there exists a positive constant* ρ *which depends only on* k, α, *and* β *such that*

$$\max_j \sup_{t \geqq 0} |f_{1j}(t) - f_{2j}(t)| \leqq \rho \max_j \sup_{t \geqq 0} |g_{1j}(t) - g_{2j}(t)|,$$

and $[f_{1j}(t) - f_{2j}(t)] \to 0$ *as* $t \to \infty$ *for* $j = 1,2, \cdots, N$ *whenever* $[g_{1j}(t) - g_{2j}(t)] \to 0$ *as* $t \to \infty$ *for* $j = 1,2, \cdots, N$.

Comments: If the hypotheses of Theorem 3 are altered to the extent that the integrability condition on $t^p k$ is replaced with the assumption that there exists a positive constant c_1 such that $e^{c_1 t}k \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$, then it is possible to give a considerably simpler proof (than that of Ref. 2) of the fact that $f \in \mathfrak{L}_{\infty N}(0,\infty)$. Specifically, under the new assumptions, it can be easily verified that for any positive constant $c_2 < c_1$, there exists a positive constant c_3 such that the modulus of the jth component of

$$\int_0^y k(y-\tau)\psi[f(\tau),\tau]d\tau$$

does not exceed $c_3 e^{-c_2 y} \| e^{c_2 t} f_y \|$ for all $y > 0$ and $j = 1, 2, \cdots, N$. By arguments very similar to those of the proofs of Theorem 2 and its corollary, it can be shown that there exist positive constants ρ and c_4 such that $c_4 < c_1$ and

$$\| e^{c_4 t} f_y \| \leqq \rho \| e^{c_4 t} g_y \|$$

for all $y > 0$. Since $g \in \mathfrak{L}_{\infty N}(0,\infty)$, and for $y > 0$

$$f(y) = g(y) - \int_0^y k(y-\tau)\psi[f(\tau),\tau]d\tau$$

and

$$\| e^{c_4 t} g_y \| \leqq \left(\frac{N}{2c_4}\right)^{\frac{1}{2}} e^{c_4 y} \max_j \sup_{t \geqq 0} | g_j(t) | ,$$

it follows that $f \in \mathfrak{L}_{\infty N}(0,\infty)$. This type of approach, when coupled with the techniques of Section V, can be used to establish the $\mathfrak{L}_{\infty N}(0,\infty)$-boundedness of solutions of more general functional equations.

For results similar to Theorems 1 and 3 concerning the discrete analog of (1), see Ref. 2.

Definition: Let T be a real positive constant, and let

$$\mathfrak{D} = \{f \mid f \in \mathfrak{L}_{\infty N}(-\infty,\infty), \quad f(t) = f(t+T) \quad \text{for} \quad -\infty < t < \infty\}$$

where $\mathfrak{L}_{\infty N}(-\infty,\infty)$ is the natural extension of the space $\mathfrak{L}_{\infty N}(0,\infty)$ to N-vector-valued functions defined on the entire real line.

Theorem 4: Let $k \in \Phi(\alpha,\beta)$ with $t^p k \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}$ for $p = 0,1,2$. Let $g_1 \in \mathfrak{D}$, $g_2 \in \mathfrak{L}_{\infty N}(0,\infty)$, $g_2(t) \to 0$ as $t \to \infty$, and $\psi[\cdot\,,\cdot] \in \Psi(\alpha,\beta)$ with $\psi_n(w,t) = \psi_n(w,t+T)$ for all real w and $t \geqq 0$. Let $f \in \mathcal{E}_N$ satisfy

$$g_1(t) + g_2(t) = f(t) + \int_0^t k(t-\tau)\psi[f(\tau),\tau]\,d\tau, \qquad t \geqq 0.$$

Then $\mathfrak{D}$ contains an element $\hat{f}$, which does not depend on g_2, such that $[f(t) - \hat{f}(t)] \to 0$ as $t \to \infty$.

If, in addition to the hypotheses stated above, there exist positive constants c_1, c_2, and c_3 such that

$$e^{c_1 t} k \in \mathcal{K}_{1N} \cap \mathcal{K}_{2N}; \qquad \int_t^\infty | k_{mn}(x) | \, dx \leqq c_2 e^{-c_3 t}, \qquad t \geqq 0$$

for $m,n = 1,2,\cdots,N$; and

$$| g_{2j}(t) | \leqq c_2 e^{-c_3 t}, \qquad t \geqq 0$$

for $j = 1,2,\cdots,N$; then there exist positive constants c_4 and c_5 such that

$$| f_j(t) - \hat{f}_j(t) | \leqq c_4 e^{-c_5 t}, \qquad t \geqq 0$$

for $j = 1,2,\cdots,N$.

Proof of Theorem 4: Assume that the hypotheses of the first part of Theorem 4 are satisfied. Let $\psi_n(w,t)$ be defined for $t < 0$ by the condi-

tion that $\psi_n(w,t) = \psi_n(w,t+T)$ for *all* real t, all real w, and $n = 1,2,\cdots,N$. We need the following result, which is easily provable (see the proofs of Lemmas 4 and 5 of Ref. 8) with the aid of Theorem 4 of Ref. 7 and the remarks relating to its proof.

Lemma: The set $\mathfrak{D}$ contains a unique element $\hat{f}$ such that

$$g_1(t) = \hat{f}(t) + \int_{-\infty}^{t} k(t-\tau)\psi[\hat{f}(\tau),\tau]\,d\tau, \qquad -\infty < t < \infty.$$

Thus we have

$$g_1(t) - \int_{-\infty}^{0} k(t-\tau)\psi[\hat{f}(\tau),\tau]\,d\tau = \hat{f}(t) + \int_{0}^{t} k(t-\tau)\psi[\hat{f}(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

$$g_1(t) + g_2(t) = f(t) + \int_{0}^{t} k(t-\tau)\psi[f(\tau),\tau]\,d\tau, \qquad t \geqq 0.$$

Since

$$g_2(t) + \int_{-\infty}^{0} k(t-\tau)\psi[\hat{f}(\tau),\tau]\,d\tau \to 0 \quad \text{as} \quad t \to \infty,$$

by Corollary 3(a), $[f(t) - \hat{f}(t)] \to 0$ as $t \to \infty$.

The second part of the theorem follows at once from Corollary 2(a) and the fact that here

$$g_2(t) + \int_{-\infty}^{0} k(t-\tau)\psi[\hat{f}(\tau),\tau]\,d\tau = [f(t) - \hat{f}(t)] + \int_{0}^{t} k(t-\tau)\{\psi[f(\tau),\tau] - \psi[\hat{f}(\tau),\tau]\}\,d\tau, \qquad t \geqq 0$$

with $\psi[\cdot\,,\cdot] \in \Psi(\alpha,\beta)$.

Comments: A result similar to the first part of Theorem 4 is proved in Ref. 8. There it is assumed that $g_2 \in \mathfrak{L}_{2N}(0,\infty)$.

Under the additional assumptions that $g_1(t)$ is a *constant* N-vector, and that $\psi_n(w,t)$ is independent of t for $n = 1,2,\cdots,N$, it can be shown that $\hat{f}(t)$ of the lemma is a *constant* N-vector, and hence that $f(t)$ of Theorem 4 approaches a limit as $t \to \infty$.

It is a simple matter to construct examples involving f's not contained

in $\mathfrak{L}_{\infty N}(0,\infty)$ which illustrate that the conclusion of the first part of Theorem 4 can be false if k does not belong to $\Phi(\alpha,\beta)$ (with the understanding that the remaining hypotheses are satisfied). The following example shows that the conclusion can be false in some relatively simple situations in which $f \in \mathfrak{L}_{\infty N}(0,\infty)$, if k does not belong to $\Phi(\alpha,\beta)$.

Let $N = 1$, and for $t \geqq 0$ let

$$\begin{aligned} \psi(w,t) &= w, && -\infty < w \leqq 1 \\ &= (w)^{\frac{1}{3}}, && 1 \leqq w \leqq 9 \\ &= \tfrac{1}{6}w + \tfrac{3}{2}, && w \geqq 9. \end{aligned}$$

Let $t^p k \in \mathfrak{K}_{11} \cap \mathfrak{K}_{21}$ for $p = 0, 1, 2$; and let $K(0) = 0$ and $K(i) = -4$. Here $\psi[\cdot,\cdot] \in \Psi(\frac{1}{6}, 1)$ and, since $K(i)$ is a point on the real-axis diameter of the disk of Fig. 2 when $\alpha = \frac{1}{6}$ and $\beta = 1$, it is clear that k does not belong to $\Phi(\frac{1}{6}, 1)$.

For $t \geq 0$, let

$$g_1(t) = \tfrac{9}{2} - \tfrac{1}{2}\cos 2t$$

$$\begin{aligned} g_{2a}(t) = e^{-t} &+ \int_0^t k(t-\tau)\{\psi[\tfrac{9}{2} + 4\sin\tau - \tfrac{1}{2}\cos 2\tau + e^{-\tau},0] \\ &- \psi[\tfrac{9}{2} + 4\sin\tau - \tfrac{1}{2}\cos 2\tau,0]\}d\tau \\ &- \int_{-\infty}^0 k(t-\tau)\{\psi[\tfrac{9}{2} + 4\sin\tau - \tfrac{1}{2}\cos 2\tau,0]\}d\tau. \end{aligned}$$

(Observe that $g_{2a}(t)$ is uniformly bounded on $[0,\infty)$ and that $g_{2a}(t) \to 0$ as $t \to \infty$.) Then, using the identity $(2 + \sin t) = (\frac{9}{2} + 4\sin t - \frac{1}{2}\cos 2t)^{\frac{1}{3}}$ which is valid for all real t, it can be verified that $f_1(t) = \frac{9}{2} + 4\sin t - \frac{1}{2}\cos 2t + e^{-t}$ satisfies

$$g_1(t) + g_{2a}(t) = f_1(t) + \int_0^t k(t-\tau)\psi[f_1(\tau),0]d\tau, \qquad t \geqq 0.$$

Note that although f_1 is ultimately periodic, it contains a component of one half the frequency of g_1.

At this point it is convenient to comment on the necessity of the hypotheses of Corollaries 1(a) and 3(a). Let g_1, f_1, k, and ψ be as defined in the preceding two paragraphs, and assume that

$$\int_t^\infty |k(\tau)|\,d\tau \in \mathfrak{L}_2(0,\infty).$$

For $t \geqq 0$, let

$$g_{2b}(t) = e^{-t} + \int_0^t k(t-\tau)\{\psi[\tfrac{9}{2} - 4\sin\tau - \tfrac{1}{2}\cos 2\tau + e^{-\tau},0] - \psi[\tfrac{9}{2} - 4\sin\tau - \tfrac{1}{2}\cos 2\tau,0]\}d\tau - \int_{-\infty}^0 k(t-\tau)\psi[\tfrac{9}{2} - 4\sin\tau - \tfrac{1}{2}\cos 2\tau,0]d\tau.$$

Then, using the identity mentioned above, it can be verified that $f_2(t) = \frac{9}{2} - 4\sin t - \frac{1}{2}\cos 2t + e^{-t}$ satisfies

$$g_1(t) + g_{2b}(t) = f_2(t) + \int_0^t k(t-\tau)\psi[f_2(\tau),0]d\tau, \qquad t \geqq 0.$$

Although $[g_{2a}(t) - g_{2b}(t)]$ approaches zero at infinity and belongs to both $\mathfrak{L}_{\infty 1}(0,\infty)$ and $\mathfrak{L}_2(0,\infty)$, it is obvious that $[f_1(t) - f_2(t)]$ (i.e., $8\sin t$) does not approach zero at infinity and does not belong to $\mathfrak{L}_2(0,\infty)$.

IV. SUFFICIENT CONDITIONS FOR THE $\mathfrak{L}_2$-BOUNDEDNESS OF THE SOLUTIONS OF (2)

Let f^n denote $d^n f/dt^n$ for $n = 1,2,\cdots$, and let $f^0 = f$.

Theorem 5: Let $\psi[\cdot\,,\cdot] \in \Psi_0(\alpha,\beta)$ *with* $N = 1$, $g \in \mathfrak{L}_2(0,\infty)$, $k \in \mathfrak{K}_{11}$, *and, with* n *a nonnegative integer,*

$$f^n(t) = g(t) + a\psi[f(t),t] + \int_0^t k(t-\tau)\psi[f(\tau),\tau]\,d\tau, \qquad t \geqq 0$$

where $f \in \mathfrak{K}_1(0,\infty)$, $f^n \in \mathfrak{E}_1$, *and* a *is a real constant. Suppose that, with*

$$K(s) = a + \int_0^\infty k(t)e^{-st}\,dt \quad \textit{for} \quad \sigma \geqq 0,$$

(*i*) $s^n - \frac{1}{2}(\alpha+\beta)K(s) \neq 0$ for $\sigma \geqq 0$ $(s^0 = 1)$, *and*

$$1 - \tfrac{1}{2}(\alpha+\beta)a \neq 0 \quad \text{if} \quad n = 0$$

(*ii*) $\frac{1}{2}(\beta-\alpha)\sup_\omega |[(i\omega)^n - \frac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega)| < 1.$

Then

$f \in \mathfrak{L}_2(0,\infty)$; *and* $\|f\| \leqq (1-\rho_0)^{-1}\rho_1\|g\| + (1-\rho_0)^{-1}\rho_2$, *where*

$$\rho_0 = \tfrac{1}{2}(\beta-\alpha)\sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega)|$$

$$\rho_1 = \sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}| \qquad [(i\omega)^0 = 1]$$

and

$$\rho_2 = \left[\frac{1}{2\pi}\int_{-\infty}^{\infty}\left|[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1} \cdot \sum_{j=0}^{(n-1)}(i\omega)^j f^{(n-1-j)}(0)\right|^2 d\omega\right]^{\frac{1}{2}}$$

for $n > 0$; *for* $n = 0$, $\rho_2 = 0$.

Proof: Assume that the hypotheses of the theorem are satisfied, and let y be an arbitrary positive number. From the fact that

$$f^n(t) = g(t) + a\psi[f(t), t] + \int_0^t k(t-\tau)\psi[f(\tau), \tau]\, d\tau, \qquad t \geqq 0$$

we have for $-\infty < \omega < \infty$

$$\int_0^y e^{-i\omega t} f^n(t)\, dt = \int_0^y e^{-i\omega t} g(t)\, dt + a\int_0^y e^{-i\omega t}\psi[f_y(t), t]\, dt + \int_0^y e^{-i\omega t}\int_0^t k(t-\tau)\psi[f_y(\tau), \tau]\, d\tau$$

in which

$$f_y(t) = f(t) \quad \text{for} \quad t \in [0, y]$$
$$= 0 \quad \text{for} \quad t > y.$$

For $-\infty < \omega < \infty$, let

$$F_y = \int_0^y e^{-i\omega t} f_y(t)\, dt$$

$$G_y = \int_0^y e^{-i\omega t} g_y(t)\, dt$$

$$H_y = \int_0^y e^{-i\omega t}\psi[f_y(t), t]\, dt$$

$$X = a\int_y^\infty e^{-i\omega t}\psi[f_y(t), t]\, dt + \int_y^\infty e^{-i\omega t}\int_0^t k(t-\tau)\psi[f_y(\tau), \tau]\, d\tau$$

$$U = \sum_{j=0}^{(n-1)}(i\omega)^j f^{(n-1-j)}(y), \qquad n > 0$$

$$V = \sum_{j=0}^{(n-1)}(i\omega)^j f^{(n-1-j)}(0), \qquad n > 0$$

$$U = V = 0, n = 0.$$

Then, using the fact that

$$\int_0^y e^{-i\omega t} f^n(t)\, dt = (i\omega)^n F_y + e^{-i\omega y} U - V$$

for $n \geqq 0$ and $-\infty < \omega < \infty$, we have

$$(i\omega)^n F_y + e^{-i\omega y} U - V = G_y + K(i\omega) H_y - X$$

for $n \geqq 0$ and $-\infty < \omega < \infty$. It follows that

$$\begin{aligned} F_y + [(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} [X + e^{-i\omega y} U] \\ = [(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} K(i\omega) [H_y - \tfrac{1}{2}(\alpha + \beta) F_y] \\ + [(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} [G_y + V] \end{aligned} \tag{5}$$

for $n \geqq 0$ and $-\infty < \omega < \infty$.

Observe that $[(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1}$ is uniformly bounded for $\omega \in (-\infty, \infty)$, and that $[(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} U$ is square integrable on the ω-set $(-\infty, \infty)$. Thus, since $\psi[f_y(t), t] \in \mathfrak{L}_2(0, \infty)$,

$$[(i\omega)^n - (\alpha + \beta) K(i\omega)]^{-1} [X + e^{-i\omega y} U]$$

is the Fourier transform of a square-integrable function $\hat{f}_y$. Further, since both

$$[s^n - \tfrac{1}{2}(\alpha + \beta) K(s)]^{-1} \quad \text{and} \quad [s^n - \tfrac{1}{2}(\alpha + \beta) K(s)]^{-1} \sum_{j=0}^{(n-1)} s^j f^{(n-1-j)}(y)$$

are analytic and uniformly bounded for $\sigma > 0$, it follows[9,10] that $\hat{f}_y(t) = 0$ almost everywhere on $(-\infty, y)$.

Using (5), Parseval's identity, and Minkowski's inequality,

$$\begin{aligned} &\left(\int_0^y | f_y(t) |^2 \, dt \right)^{\frac{1}{2}} \\ &\leqq \left(\int_0^\infty | f_y(t) + \hat{f}_y(t) |^2 \, dt \right)^{\frac{1}{2}} \\ &\leqq \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} | [(i\omega)^n - \tfrac{1}{2}(\alpha + \beta) K(i\omega)]^{-1} K(i\omega) \right. \\ &\qquad \left. \cdot [H_y - \tfrac{1}{2}(\alpha + \beta) F_y] |^2 \, d\omega \right)^{\frac{1}{2}} \\ &+ \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} | [(i\omega)^n - (\alpha + \beta) K(i\omega)]^{-1} [G_y + V] |^2 \, d\omega \right)^{\frac{1}{2}} \end{aligned}$$

$$\leqq \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} | [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega) \cdot [H_y - \tfrac{1}{2}(\alpha+\beta)F_y] |^2 \, d\omega\right)^{\frac{1}{2}}$$

$$+ \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} | [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}G_y |^2 \, d\omega\right)^{\frac{1}{2}}$$

$$+ \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \left| [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1} \cdot \sum_{j=0}^{(n-1)} (i\omega)^j f^{(n-1-j)}(0) \right|^2 d\omega\right)^{\frac{1}{2}}$$

with the understanding that the last integral vanishes if $n = 0$.

Thus

$$\left(\int_0^y | f_y(t) |^2 \, dt\right)^{\frac{1}{2}} \leqq \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} | [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega) \cdot [H_y - \tfrac{1}{2}(\alpha+\beta)F_y] |^2 \, d\omega\right)^{\frac{1}{2}} + \sup_{\omega} | [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1} | \left(\int_0^{\infty} | g(t) |^2 \, dt\right)^{\frac{1}{2}} + \rho_2$$

where

$$\rho_2 = \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} | [(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1} \cdot \sum_{j=0}^{(n-1)} (i\omega)^j f^{(n-1-j)}(0) |^2 \, d\omega\right)^{\frac{1}{2}}$$

if $n > 0$ and $\rho_2 = 0$ if $n = 0$.

Since $| x^{-1}\psi(x,t) - \tfrac{1}{2}(\alpha+\beta)| \leqq \tfrac{1}{2}(\beta-\alpha)$ for all real $x \neq 0$ and $t \geqq 0$, we have

$$\left(\int_{-\infty}^{\infty} | H_y(i\omega) - \tfrac{1}{2}(\alpha-\beta)F_y(i\omega) |^2 \, d\omega\right)^{\frac{1}{2}} = \left(2\pi\int_0^y | \psi[f_y(t), t] - \tfrac{1}{2}(\alpha+\beta)f_y(t) |^2 \, dt\right)^{\frac{1}{2}} \leqq \tfrac{1}{2}(\beta-\alpha)\left(2\pi\int_0^y | f_y(t) |^2 \, dt\right)^{\frac{1}{2}}.$$

It follows that

$$\left(\int_0^y |f_y(t)|^2\,dt\right)^{\frac{1}{2}} \leqq (1-\rho_0)^{-1}\sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}| \times \left(\int_0^\infty |g(t)|^2\,dt\right)^{\frac{1}{2}} + (1-\rho_0)^{-1}\rho_2 \tag{6}$$

in which

$$\rho_0 = \tfrac{1}{2}(\beta-\alpha)\sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}K(i\omega)|$$

(recall that $\rho_0 < 1$ by assumption). Thus, since ρ_2 and

$$\sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}|$$

are finite, and (6) is valid for all $y > 0$, it follows that $f \in \mathfrak{L}_2(0,\infty)$, and that

$$\|f\| \leqq (1-\rho_0)^{-1}\sup_\omega |[(i\omega)^n - \tfrac{1}{2}(\alpha+\beta)K(i\omega)]^{-1}|\cdot\|g\| + (1-\rho_0)^{-1}\rho_2 .$$

Comment: With regard to the hypothesis concerning f^n of Theorem 5, observe that if $f \in \mathcal{E}_1$, and if f^n exists and satisfies the integral equation, with g, a, k, and $\psi[\cdot\,,\cdot]$ as defined, then $f^n \in \mathcal{E}_1$.

V. SOME RESULTS ON THE PROPERTIES OF EQUATIONS DEFINED ON AN ABSTRACT SPACE

5.1 *Definitions and Notation*

Let $\mathfrak{K}$ denote an abstract linear space that contains a normed linear space $\mathfrak{L}$ with norm $\|\cdot\|$.

Let Ω denote a set of real numbers, and let P_y denote a linear mapping of $\mathfrak{K}$ into $\mathfrak{L}$ for each $y \in \Omega$.

Let $g \in \mathfrak{L}$ if and only if $g \in \mathfrak{K}$ and $\|P_y g\|$ is uniformly bounded on Ω.

The norm of a linear transformation A defined on $\mathfrak{L}$ is denoted by $\|A\|$, and I denotes the identity operator on $\mathfrak{K}$.

We shall say that a (not necessarily linear) operator T is an element of the set Θ if and only if T maps $\mathfrak{K}$ into itself and $P_yT = P_yTP_y$ on $\mathfrak{K}$ for $y \in \Omega$.†

† As a concrete example, we note that the development of Ref. 1 is concerned with the complex-valued-function version of the case in which $\mathfrak{L} = \mathfrak{L}_{2N}(0,\infty)$, $\mathfrak{K} = \mathcal{E}_N$, $\Omega = [0,\infty)$, and P_y is defined by: $(P_yf)(t) = f(t)$ for $t \in [0,y]$ and $(P_yf)(t) = 0$ for $t > y$ where f is an arbitrary element of $\mathcal{E}_N$. With that definition of P_y, an operator belongs to Θ only if it is "causal."

As a matter of convenience, we shall let g_y denote $P_y g$ for $g \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$. Thus, for example, if $f \varepsilon \mathcal{K}$ and $T \varepsilon \Theta$, then $(Tf_y)_y$ denotes $P_y T P_y f$ for $y \varepsilon \Omega$.

5.2 *Results of the Type Presented in Ref. 1*

Our first two observations, which are stated as Theorems 6 and 7, are instructive.

Theorem 6: Let $T \varepsilon \Theta$; $f_1, f_2 \varepsilon \mathcal{K}$; *and*

$$g_1 = f_1 + Tf_1$$

$$g_2 = f_2 + Tf_2 .$$

Suppose that there exists a positive constant $k < 1$ *such that*

$$\| (Tf_{1y})_y - (Tf_{2y})_y \| \leqq k \| f_{1y} - f_{2y} \|, \qquad y \varepsilon \Omega.$$

Then

(*i*) $\| f_{1y} - f_{2y} \| \leqq (1 - k)^{-1} \| g_{1y} - g_{2y} \|, \qquad y \varepsilon \Omega$

(*ii*) $(f_1 - f_2) \varepsilon \mathfrak{L}$ *provided that* $(g_1 - g_2) \varepsilon \mathfrak{L}$.

Proof of Theorem 6: For $y \varepsilon \Omega$,

$$\begin{aligned} f_{1y} - f_{2y} &= g_{1y} - g_{2y} - P_y[Tf_1 - Tf_2] \\ &= g_{1y} - g_{2y} - P_y[Tf_{1y} - Tf_{2y}]. \end{aligned}$$

Thus

$$\begin{aligned} \| f_{1y} - f_{2y} \| &\leqq \| g_{1y} - g_{2y} \| + \| (Tf_{1y})_y - (Tf_{2y})_y \| \\ &\leqq \| g_{1y} - g_{2y} \| + k \| f_{1y} - f_{2y} \| \end{aligned}$$

for $y \varepsilon \Omega$. Hence

$$\| f_{1y} - f_{2y} \| \leqq (1 - k)^{-1} \| g_{1y} - g_{2y} \|, \qquad y \varepsilon \Omega.$$

If $(g_1 - g_2) \varepsilon \mathfrak{L}$, then

$$\sup_{\Omega} \| f_{1y} - f_{2y} \| \leqq (1 - k)^{-1} \sup_{\Omega} \| g_{1y} - g_{2y} \| < \infty,$$

and hence $(f_1 - f_2) \varepsilon \mathfrak{L}$. This proves the theorem. A similar argument establishes

Theorem 7: Let $T \varepsilon \Theta$; $f \varepsilon \mathcal{K}$; *and*

$$g = f + Tf.$$

Suppose that there exists a positive constant $k < 1$ *such that*

$$\| (Tf_y)_y \| \leqq k \| f_y \|, \qquad y \in \Omega.$$

Then

(*i*) $\| f_y \| \leqq (1 - k)^{-1} \| g_y \|, \qquad y \in \Omega$

(*ii*) $f \in \mathfrak{L}$ *provided that* $g \in \mathfrak{L}$.

Our next result is quite useful. For example, it leads to Theorem 1.

Theorem 8: Let P_y *be such that* $\| P_y h \| \leqq \| h \|$ *and* $P_y P_y h = P_y h$ *for* $h \in \mathfrak{L}$ *and* $y \in \Omega$. *Let* $L,N \in \Theta$, *with* L *linear and* L *mapping* $\mathfrak{L}$ *into itself. Let* $f \in \mathfrak{K}$, *and*

$$g = f + LNf.$$

Suppose that there exists a scalar λ (λ *real if* $\mathfrak{L}$ *is a real space*) *such that*

(*i*) $(I + \lambda L)^{-1}$ *exists on* $\mathfrak{L}$, *and*

$$P_y(I + \lambda L)^{-1} = P_y(I + \lambda L)^{-1} P_y$$

on $\mathfrak{L}$ *for* $y \in \Omega$

(*ii*) $\| (I + \lambda L)^{-1} L \| < \infty$

(*iii*) *there exists a positive constant* k_λ *with the property that*

$$\| (Nf_y)_y - \lambda f_y \| \leqq k_\lambda \| f_y \| \quad \text{for} \quad y \in \Omega$$

(*iv*) $\| (I + \lambda L)^{-1} L \| k_\lambda < 1.$

Then

$$\| f_y \| \leqq (1 - r)^{-1} \| P_y(I + \lambda L)^{-1} g_y \|, \qquad y \in \Omega.$$

in which

$$r = \| (I + \lambda L)^{-1} L \| k_\lambda .$$

Corollary 8(a): Suppose that the hypotheses of Theorem 8 are satisfied and that $g \in \mathfrak{L}$. *Then* $f \in \mathfrak{L}$.

Proof of Theorem 8: Let $y \in \Omega$. From the fact that $g = f + LNf$, we clearly have

$$g_y = f_y + P_y LNf.$$

Since L and N belong to Θ,

$$\begin{aligned} g_y &= f_y + P_y L P_y N f \\ &= f_y + P_y L P_y N f_y . \end{aligned}$$

Thus

$$g_y = P_y(I + \lambda L)f_y + P_yLP_y(Nf_y - \lambda f_y).$$

Since $(I + \lambda L)^{-1}$ exists on $\mathfrak{L}$ and $P_y(I + \lambda L)^{-1} = P_y(I + \lambda L)^{-1}P_y$, we have

$$P_y(I + \lambda L)^{-1}P_y(I + \lambda L)f_y = f_y,$$

and hence

$$f_y = -P_y(I + \lambda L)^{-1}P_yLP_y(Nf_y - \lambda f_y) + P_y(I + \lambda L)^{-1}g_y.$$

Using the fact that

$$\| P_y(I + \lambda L)^{-1}P_yL \| = \| P_y(I + \lambda L)^{-1}L \|$$
$$\leqq \| P_y \| \cdot \| (I + \lambda L)^{-1}L \| \leqq \| (I + \lambda L)^{-1}L \|,$$

it follows that

$$\| f_y \| \leqq \| P_y(I + \lambda L)^{-1}P_yL \| \cdot \| (Nf_y)_y - \lambda f_y \| + \| P_y(I + \lambda L)^{-1}g_y \|$$
$$\leqq \| (I + \lambda L)^{-1}L \| k_\lambda \| f_y \| + \| P_y(I + \lambda L)^{-1}g_y \|.$$

Therefore

$$\| f_y \| \leqq (1 - r)^{-1} \| P_y(I + \lambda L)^{-1}g_y \|, \qquad y \varepsilon \Omega$$

in which

$$r = \| (I + \lambda L)^{-1}L \| k_\lambda.$$

This proves the theorem.

Proof of Corollary 8(a): Assume that the hypotheses of the corollary are satisfied. Then

$$\sup_{y\varepsilon\Omega} \| f_y \| \leqq (1 - r)^{-1} \sup_{y\varepsilon\Omega} \| P_y(I + \lambda L)^{-1}g_y \|$$
$$\leqq (1 - r)^{-1} \sup_{y\varepsilon\Omega} \| P_y(I + \lambda L)^{-1}g \| < \infty$$

and hence $f \varepsilon \mathfrak{L}$.

Arguments similar to those of the proof of Theorem 8 and its corollary establish the following theorem and corollary.

Theorem 9: Let P_y be such that $\| P_yh \| \leqq \| h \|$ and $P_yP_yh = P_yh$ for $h \varepsilon \mathfrak{L}$ and $y \varepsilon \Omega$. Let $L,N \varepsilon \Theta$; with L linear and L mapping $\mathfrak{L}$ into itself. Let $f_1, f_2 \varepsilon \mathfrak{K}$; and

$$g_1 = f_1 + LNf_1$$
$$g_2 = f_2 + LNf_2.$$

Suppose that there exists a scalar λ (λ *real if* $\mathfrak{L}$ *is a real space*) *such that*

(*i*) $(I + \lambda L)^{-1}$ *exists on* $\mathfrak{L}$, *and* $P_y(I + \lambda L)^{-1} = P_y(I + \lambda L)^{-1}P_y$ *on* $\mathfrak{L}$ *for* $y \in \Omega$

(*ii*) $\| (I + \lambda L)^{-1}L \| < \infty$

(*iii*) *there exists a positive constant* k_λ *with the property that*

$$\| (Nf_{1y})_y - (Nf_{2y})_y - \lambda(f_{1y} - f_{2y})\| \leqq k_\lambda \| f_{1y} - f_{2y} \|$$

for $y \in \Omega$

(*iv*) $\| (I + \lambda L)^{-1}L \| k_\lambda < 1.$

Then

$$\| f_{1y} - f_{2y} \| \leqq (1 - r)^{-1} \| P_y(I + \lambda L)^{-1}(g_{1y} - g_{2y})\|, \qquad y \in \Omega$$

in which

$$r = \| (I + \lambda L)^{-1}L \| k_\lambda .$$

Corollary 9(a): Suppose that the hypotheses of Theorem 9 are satisfied and that $(g_1 - g_2) \in \mathfrak{L}$. *Then* $(f_1 - f_2) \in \mathfrak{L}$.

5.3 *Results for the Case in Which* $\mathfrak{L}$ *is an Inner-Product Space*

In this section we employ the definitions and notation of Section 5.1. It is further assumed here that $\mathfrak{L}$ is an inner-product space, with inner product $\langle \cdot , \cdot \rangle$. The norm of $f \in \mathfrak{L}$ is $\| f \| = \langle f,f \rangle^{\frac{1}{2}}$.

It is also assumed throughout this section that A and B are elements of Θ.

Lemma 1: Let $f \in \mathfrak{K}$, $h = Bf$, *and* $g = f + Ah$. *Then*

$$| Re \langle (Bf_y)_y , f_y \rangle + Re \langle (Ah_y)_y , h_y \rangle | \leqq \| g_y \| \cdot \| h_y \|$$

for $y \in \Omega$.

Proof: Since A and B belong to Θ, we have

$$h_y = (Bf_y)_y \quad \text{and} \quad g_y = f_y + (Ah_y)_y \quad \text{for} \quad y \in \Omega.$$

Thus, for $y \in \Omega$,

$$\begin{aligned} \text{Re} \langle (Bf_y)_y , f_y \rangle + \text{Re} \langle (Ah_y)_y , h_y \rangle &= \text{Re} \langle f_y , (Bf_y)_y \rangle + \text{Re} \langle (Ah_y)_y , h_y \rangle \\ &= \text{Re} \langle f_y + (Ah_y)_y , h_y \rangle = \text{Re} \langle g_y , h_y \rangle \end{aligned}$$

and so, by the Schwarz inequality,

$$| \text{Re} \langle (Bf_y)_y , f_y \rangle + \text{Re} \langle (Ah_y)_y , h_y \rangle | \leqq \| g_y \| \cdot \| h_y \|$$

for $y \in \Omega$.

Remarks: An application of the Schwarz inequality similar to that of the proof of Lemma 1 shows that if $g = f + Af$ with $f \varepsilon \mathcal{K}$ and if there exists a constant $\delta > 0$ such that

$$| \| f_y \|^2 + \langle (Af_y)_y , f_y \rangle | \geqq \delta \| f_y \|^2, \qquad y \varepsilon \Omega$$

then $\| f_y \| \leqq \delta^{-1} \| g_y \|$ for $y \varepsilon \Omega$ and thus then $f \varepsilon \mathfrak{L}$ provided that $g \varepsilon \mathfrak{L}$. Observe that for $\mathfrak{L}$ an inner-product space, this is a stronger result than that of Theorem 7.

Theorem 10: Let $g \varepsilon \mathfrak{L}$, $f \varepsilon \mathcal{K}$, and $g = f + ABf$. Let

$$Re \langle (Aq_y)_y , q_y \rangle \geqq 0 \quad and \quad Re \langle (Bq_y)_y , q_y \rangle \geqq 0$$

for $q \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$. Then $f \varepsilon \mathfrak{L}$ provided that at least one of the following conditions is satisfied:

(i) A maps $\mathfrak{L}$ into $\mathfrak{L}$, and there exist constants $\alpha > 0$ and $p > 1$ such that $Re \langle (Aq_y)_y , q_y \rangle \geqq \alpha \| q_y \|^p$ for $q \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$

(ii) there exist constants $k > 0$, $\alpha > 0$, and $p > 1$ such that $\| (Bq_y)_y \| \leqq k \| q_y \|$ and $Re \langle (Bq_y)_y , q_y \rangle \geqq \alpha \| q_y \|^p$ for $q \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$

(iii) there exist constants $k_1 > 0$, $k_2 > 0$, and $p > 1$ such that $Re \langle (Aq_y)_y , q_y \rangle \geqq k_1 \| (Aq_y)_y \|^p$ and $\| (Bq_y)_y \| \leqq k_2 \| q_y \|$ for $q \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$

(iv) A maps $\mathfrak{L}$ into $\mathfrak{L}$, and there exist constants $k > 0$ and $p > 1$ such that $Re \langle (Bq_y)_y , q_y \rangle \geqq k \| (Bq_y)_y \|^p$ for $q \varepsilon \mathcal{K}$ and $y \varepsilon \Omega$

Proof of Theorem 10: Let $h = Bf$. Suppose that condition (i) is met Then, using Lemma 1,

$$\alpha \| h_y \|^p \leqq \operatorname{Re} \langle (Ah_y)_y , h_y \rangle \leqq \| g_y \| \cdot \| h_y \|, \qquad y \varepsilon \Omega$$

from which

$$\| h_y \| \leqq (\alpha^{-1} \| g_y \|)^{(p-1)^{-1}}, \qquad y \varepsilon \Omega.$$

Thus

$$\sup_{\Omega} \| h_y \| \leqq (\alpha^{-1})^{(p-1)^{-1}} (\sup_{\Omega} \| g_y \|)^{(p-1)^{-1}} < \infty$$

and hence $h \varepsilon \mathfrak{L}$. Since A is assumed to map $\mathfrak{L}$ into itself, $Ah \varepsilon \mathfrak{L}$ and since $\mathfrak{L}$ is a linear space, $f = (g - Ah) \varepsilon \mathfrak{L}$.

Suppose that condition (ii) is satisfied. Then, using Lemma 1,

$$\alpha \| f_y \|^p \leqq \operatorname{Re} \langle (Bf_y)_y , f_y \rangle \leqq \| g_y \| \cdot \| h_y \| \leqq k \| g_y \| \cdot \| f_y \|$$

for $y \varepsilon \Omega$, and so

$$\| f_y \| \leqq (k/\alpha)^{(p-1)^{-1}} (\| g_y \|)^{(p-1)^{-1}}, \qquad y \varepsilon \Omega.$$

Since $\sup_{\Omega} \| g_y \| < \infty, f \in \mathfrak{L}$.

Assume now that condition (*iii*) is met. Then, using the lemma,

$$k_1 \| (Ah_y)_y \|^p \leqq \operatorname{Re} \langle (Ah_y)_y , h_y \rangle \leqq \| g_y \| \cdot \| h_y \|$$

for $y \in \Omega$. Thus

$$k_1 \| g_y - f_y \|^p \leqq k_2 \| g_y \| \cdot \| f_y \|, \qquad y \in \Omega.$$

It follows that $\sup_{\Omega} \| f_y \| < \infty$.

Finally, assume that condition (*iv*) is satisfied. Then, using the lemma,

$$k \| h_y \|^p \leqq \operatorname{Re} \langle (Bf_y)_y , f_y \rangle \leqq \| g_y \| \cdot \| h_y \|$$

for $y \in \Omega$. Therefore

$$\| h_y \| \leqq (k^{-1})^{(p-1)^{-1}} (\| g_y \|)^{(p-1)^{-1}},$$

from which it is clear that $h \in \mathfrak{L}$. Since A is assumed to map $\mathfrak{L}$ into $\mathfrak{L}$, and $\mathfrak{L}$ is a linear space, $f \in \mathfrak{L}$.

Remarks: The requirement of conditions (*i*) and (*iv*) that A map $\mathfrak{L}$ into itself can be replaced with the condition that

$$\sup_{\Omega} \operatorname{Re} \langle (Bq_y)_y , q_y \rangle = +\infty \quad \text{for} \quad q \in (\mathfrak{K} - \mathfrak{L})$$

which (through Lemma 1) implies that $f \in \mathfrak{L}$ whenever $h \in \mathfrak{L}$ and

$$\operatorname{Re} \langle (Aq_y)_y , q_y \rangle \geqq 0 \quad \text{and} \quad \operatorname{Re} \langle (Bq_y)_y , q_y \rangle \geqq 0$$

for $q \in \mathfrak{K}$ and $y \in \Omega$.

It is possible to extend Theorem 10 in many other directions. For example, an argument essentially the same as that used to establish the sufficiency of condition (*iii*) of Theorem 10 shows that if $g \in \mathfrak{L}, f \in \mathfrak{K}$, and $g = f + ABf$, then $f \in \mathfrak{L}$ provided that (a) $\operatorname{Re} \langle (Bq_y)_y , q_y \rangle \geqq 0$ for $q \in \mathfrak{K}$ and $y \in \Omega$ and (b) there exist constants $k_1 > 0$, $k_2 > 0$, $k_3 > 0$, and $p > 1$ such that

$$\operatorname{Re} \langle (Aq_y)_y , q_y \rangle \geqq k_1 \| (Aq_y)_y \|^p - k_3 \| (Aq_y)_y \| - k_3$$

and $\| (Bq_y)_y \| \leqq k_2 \| q_y \|$ for $q \in \mathfrak{K}$ and $y \in \Omega$.

For some earlier material related to Theorem 10, see Ref. 1, Theorem 3, and the observations of Ref. 7 concerning the relation between passivity and conditions for certain nonlinear operators to be contraction mappings.

The following lemma and theorem can be proved with essentially the same arguments used in the proof of Lemma 1 and Theorem 10.

Lemma 2: Let $f_1, f_2 \varepsilon \mathcal{K}$; $h_1 = Bf_1$, $h_2 = Bf_2$; *and*

$$g_1 = f_1 + Ah_1$$
$$g_2 = f_2 + Ah_2$$

Then

$$| Re \langle (Bf_{1y} - Bf_{2y})_y, f_{1y} - f_{2y} \rangle + Re \langle (Ah_{1y} - Ah_{2y})_y, h_{1y} - h_{2y} \rangle | \leqq \| g_{1y} - g_{2y} \| \cdot \| h_{1y} - h_{2y} \|$$

for $y \varepsilon \Omega$.

Theorem 11: Let $f_1, f_2 \varepsilon \mathcal{K}$; *and let*

$$g_1 = f_1 + ABf_1$$
$$g_2 = f_2 + ABf_2$$

with $(g_1 - g_2) \varepsilon \mathfrak{L}$. *Let*

$$Re \langle (Aq_{1y} - Aq_{2y})_y, q_{1y} - q_{2y} \rangle \geqq 0$$

and

$$Re \langle (Bq_{1y} - Bq_{2y})_y, q_{1y} - q_{2y} \rangle \geqq 0$$

for $q_1, q_2 \varepsilon \mathcal{K}$ *and* $y \varepsilon \Omega$. *Then* $(f_1 - f_2) \varepsilon \mathcal{K}$ *provided that at least one of the following conditions is satisfied.*

(*i*) $(Aq_1 - Aq_2) \varepsilon \mathfrak{L}$ *whenever* $q_1, q_2 \varepsilon \mathcal{K}$ *and* $(q_1 - q_2) \varepsilon \mathfrak{L}$; *and there exist constants* $\alpha > 0$ *and* $p > 1$ *such that* $Re \langle (Aq_{1y} - Aq_{2y})_y, q_{1y} - q_{2y} \rangle \geqq \alpha \| q_{1y} - q_{2y} \|^p$ *for* $q_1, q_2 \varepsilon \mathcal{K}$ *and* $y \varepsilon \Omega$.

(*ii*) *there exist constants* $k > 0$, $\alpha > 0$, *and* $p > 1$ *such that*

$$\| (Bq_{1y} - Bq_{2y})_y \| \leqq k \| q_{1y} - q_{2y} \|$$

and $Re \langle (Bq_{1y} - Bq_{2y})_y, q_{1y} - q_{2y} \rangle \geqq \alpha \| q_{1y} - q_{2y} \|^p$ *for* $q_1, q_2 \varepsilon \mathcal{K}$ *and* $y \varepsilon \Omega$.

(*iii*) *there exist constants* $k_1 > 0$, $k_2 > 0$, *and* $p > 1$ *such that* $Re \langle (Aq_{1y} - Aq_{2y})_y, q_{1y} - q_{2y} \rangle \geqq k_1 \| (Aq_{1y} - Aq_{2y})_y \|^p$ *and*

$$\| (Bq_{1y} - Bq_{2y})_y \| \leqq k_2 \| q_{1y} - q_{2y} \|$$

for $q \varepsilon \mathcal{K}$ *and* $y \varepsilon \Omega$.

(*iv*) $(Aq_1 - Aq_2) \varepsilon \mathfrak{L}$ *whenever* $q_1, q_2 \varepsilon \mathcal{K}$ *and* $(q_1 - q_2) \varepsilon \mathfrak{L}$; *and there exist constants* $k > 0$ *and* $p > 1$ *such that* $Re \langle (Bq_{1y} - Bq_{2y}), q_{1y} - q_{2y} \rangle \geqq k \| (Bq_{1y} - Bq_{2y})_y \|^p$ *for* $q_1, q_2 \varepsilon \mathcal{K}$ *and* $y \varepsilon \Omega$.

Results similar to Theorems 10 and 11 can be established for the equation

$$g = Af + Bf.$$

In particular, we can very easily prove

Lemma 3: Let $f \varepsilon \mathcal{K}$, and $g = Af + Bf$. Then

$$| Re \langle (Bf_y)_y , f_y \rangle + Re \langle (Af_y)_y , f_y \rangle | \leqq \| g_y \| \cdot \| f_y \|$$

for $y \varepsilon \Omega$.

VI. ACKNOWLEDGMENT

The writer is indebted to H. O. Pollak for a valuable suggestion concerning the material of Section 5.3.

APPENDIX

Some Results Related to Theorem 2 and Corollary 2(a)

Let $\underline{\mathcal{K}}_{1N}$, $\underline{\mathcal{K}}_{2N}$, $\underline{\mathcal{E}}_N$ and $\underline{\mathcal{L}}_{2N}(0,\infty)$ denote the natural complex extensions of the real sets $\mathcal{K}_{1N}$, $\mathcal{K}_{2N}$, $\mathcal{E}_N$ and $\mathcal{L}_{2N}(0,\infty)$, respectively.

Using arguments very similar to those of the proofs of Theorem 2, Corollary 2(a), and the lemma of Ref. 11, it is not difficult to prove the following theorem and corollary.

Theorem 12: Let $Q(\cdot)$ denote a complex measurable $N \times N$ matrix-valued function of t defined on $[0,\infty)$, and let the elements of $Q(t)$ be uniformly bounded on $[0,\infty)$. Let

$$g(t) = f(t) + \int_0^t k(t-\tau)Q(\tau)f(\tau)\, d\tau, \qquad t \geqq 0 \tag{1}$$

in which $f \varepsilon \underline{\mathcal{E}}_N$, and there exists a real constant c_1 such that

(*i*) $ge^{c_1 t} \varepsilon \underline{\mathcal{L}}_{2N}(0,\infty)$
(*ii*) $ke^{c_1 t} \varepsilon \underline{\mathcal{K}}_{1N} \cap \underline{\mathcal{K}}_{2N}$
(*iii*) *with*

$$K(i\omega - c_1) = \int_0^\infty k(t)e^{(c_1 - i\omega)t}\, dt \quad \text{for} \quad -\infty < \omega < \infty,$$

$$\sup_{t \geqq 0} \Lambda\{Q(t)\} \quad \sup_{-\infty < \omega < \infty} \Lambda\{K(i\omega - c_1)\} < 1.$$

Then there exists a positive constant c_2 such that

$$|f_j(t)| \leqq |g_j(t)| + c_2 e^{-c_1 t}, \qquad t \geqq 0$$

for $j = 1,2, \cdots, N$.

Corollary 12(a): Let $Q(\cdot)$ denote a complex measurable $N \times N$ matrix-

valued function of t defined on $[0,\infty)$. and let the elements of $Q(t)$ be uniformly bounded on $[0,\infty)$. Let

$$g(t) = f(t) + \int_0^t k(t-\tau)Q(\tau)f(\tau)\,d\tau, \qquad t \geqq 0$$

in which $f \varepsilon \underline{\mathcal{E}}_N$, and there exists a positive constant c_1 such that

(*i*) $ge^{c_1 t} \varepsilon \underline{\mathcal{L}}_{2N}(0,\infty)$
(*ii*) $ke^{c_1 t} \varepsilon \underline{\mathcal{K}}_{1N} \cap \underline{\mathcal{K}}_{2N}$.

With

$$K(i\omega) = \int_0^\infty k(t)e^{-i\omega t}\,dt \quad \text{for} \quad -\infty < \omega < \infty,$$

let

$$\sup_{t\geqq 0} \Lambda\{Q(t)\} \sup_{-\infty<\omega<\infty} \Lambda\{K(i\omega)\} < 1.$$

Then there exist positive constants c_2 and c_3 such that

$$|f_j(t)| \leqq |g_j(t)| + c_2 e^{-c_3 t}, \qquad t \geqq 0$$

for $j = 1,2,\cdots,N$.

With the aid of Corollary 12(a) and arguments very similar to those used to establish the corollary of Ref. 11, it is a simple matter to prove the following result concerned with conditions under which all solutions of a well known type of linear differential equation approach zero *exponentially* at infinity.†

Theorem 13: Let A denote a constant positive-definite $N \times N$ Hermitian matrix. Let $B(t)$ denote an $N \times N$ positive-definite Hermitian-matrix-valued function of t for $t \geqq 0$, and let the elements of $B(t)$ be measurable and uniformly bounded for $t \geqq 0$. Let f be a complex N-vector valued function of t defined and twice differentiable on $[0,\infty)$ such that

$$\frac{d^2f(t)}{dt^2} + A\frac{df(t)}{dt} + B(t)f(t) = g(t), \qquad t \geqq 0$$

with

$$ge^{c_1 t} \varepsilon \underline{\mathcal{L}}_{2N}(0,\infty)$$

for some positive constant c_1.

† Theorem 13 can be obtained also from the corollary of Ref. 11 by using it to show that there exists a positive constant δ such that, under the conditions of Theorem 13, $f(t)$ can be written as $h(t)e^{-\delta t}$ for $t \geqq 0$ with $h(t) \to 0$ as $t \to \infty$. That is, the corollary can be applied to the differential equation in h for δ sufficiently small.

Let $\underline{\lambda}\{A\}, \underline{\lambda}\{B(t)\}$, and $\bar{\lambda}\{B(t)\}$ denote, respectively, the smallest eigenvalue of A, the smallest eigenvalue of $B(t)$, and the largest eigenvalue of $B(t)$. Suppose that

$$\inf_{t \geqq 0} \underline{\lambda}\{B(t)\} > 0$$

and that

$$\underline{\lambda}\{A\} > (\sup_{t \geqq 0} \bar{\lambda}\{B(t)\})^{\frac{1}{2}} - (\inf_{t \geqq 0} \underline{\lambda}\{B(t)\})^{\frac{1}{2}}.$$

Then there exist positive constants c_2 and c_3 such that

$$|f_j(t)| \leqq c_2 e^{-c_3 t}, \qquad t \geqq 0$$

for $j = 1,2, \cdots, N$.

REFERENCES

1. Sandberg, I. W., On the $\mathfrak{L}_2$-Boundedness of Solutions of Nonlinear Functional Equations, B.S.T.J., *43*, July, 1964, p. 1581.
2. Sandberg, I. W., On the Boundedness of Solutions of Nonlinear Integral Equations, B.S.T.J., *44*, March, 1965, p. 439.
3. Aizerman, M. A., and Gantmacher, F. R., *Absolute Stability of Regulator Systems*, Holden-Day, San Francisco, 1964.
4. Kalman, R. E., Lyapunov Functions for the Problem of Lur'e in Automatic Control, Proc. Natl. Acad. Sci., *49*, Feb., 1963, pp. 201–205.
5. Bongiorno, J. J., Jr., Real-Frequency Stability Criteria for Linear Time-Varying Systems, Proc. IEEE, *52*, July, 1964, pp. 832–841.
6. Narendra, K. S., and Goldwyn, R. M., A Geometrical Criterion for the Stability of Certain Nonlinear Nonautonomous Systems, IEEE Trans. Ckt. Theory, *CT-11*, Sept., 1964, pp. 406–408.
7. Sandberg, I. W., On Truncation Techniques in the Approximate Analysis of Periodically Time-Varying Nonlinear Networks, IEEE Trans. Ckt. Theory, *CT-11*, June, 1964, p. 195.
8. Sandberg, I. W., and Beneš, V. E., On the Properties of Nonlinear Integral Equations that Arise in the Theory of Dynamical Systems, B.S.T.J., *43*, Nov., 1964, p. 2839.
9. Paley, R. E., and Wiener, N., *Fourier Transforms in the Complex Domain*, publ. 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 Solutions of Systems of Second Order Differential Equations with Variable Coefficients, to appear in the SIAM J. Control, *2*, No. 2.

The Attenuation of the Holmdel Helix Waveguide in the 100–125-Kmc Band*

By WILLIAM H. STEIER

(Manuscript received January 4, 1965)

The attenuation of the Holmdel, N. J., Bell Telephone Laboratories two-inch diameter helix waveguide operating in the circular electric mode has been measured in the 100–125-kmc band. With these measurements the helix waveguide attenuation is now known from 33 kmc to 125 kmc. Above 100 kmc the loss increases with frequency, as contrasted to all previous lower-frequency measurements, which show the loss decreasing with increasing frequency. The minimum loss of 2 db/mile occurs around 90 kmc. The loss is below 3 db/mile from 40 kmc to 116 kmc. A loss peak observed at 122 kmc is believed due to the method of supporting the sections of the waveguide measured. An analysis of the loss peak is presented which indicates that this loss peak can be avoided by suitably supporting a helix installation.

I. INTRODUCTION

The advantages of overmoded circular waveguide using the TE_{01} mode for long-distance transmission in the millimeter-wave region have long been recognized.[1] The added advantage of the circular helix waveguide in giving spurious mode suppression has also been considered.[2] The propagation characteristics of the Holmdel, N. J., Bell Laboratories two-inch diameter helix waveguide, which is made to very precise tolerances, have been investigated extensively, and its attenuation has previously been measured from 33 kmc to 90 kmc.[3] The work reported on here extends these measurements to 125 kmc.

Over the previously measured band the attenuation decreases with increasing frequency and falls below 2 db/mile at 90 kmc. It was previously not known how high in frequency this decreasing trend would continue nor to how low the attenuation would fall. The measurements

* This paper was presented at the Conference on Precision Electromagnetic Measurements, June 1964, Boulder, Colorado.

in the 100–125-kmc band show that the minimum has been passed and that the losses in this band increase with frequency.

As far as is known, these are the only relatively precise measurements which have been reported on low-loss millimeter waveguide in the range from 100 to 125 kmc. They extend the known communications bandwidth of a practical waveguide by about 30 kmc.

II. MEASURING METHOD

The measuring technique used was the single-oscillator shuttle-pulse method described by others.[3]

The rectangular waveguide components were of the RG 138/U size, and the circular waveguide components were designed for use in the 120-kmc band. The millimeter-wave mixer diodes used were developed by Sharpless.[4] The mode transducer for exciting the circular electric mode is a two-stage transducer. The rectangular TE_{01} is converted to rectangular TE_{02} and then to circular TE_{01}.[5,6]

The waveguide being measured was filled with dry nitrogen at a positive pressure to avoid losses due to gas absorption.

III. ATTENUATION MEASUREMENTS

The measured attenuation of the 2-inch helix waveguide as a function of frequency is shown in Fig. 1. The attenuation from 33 to 90 kmc was

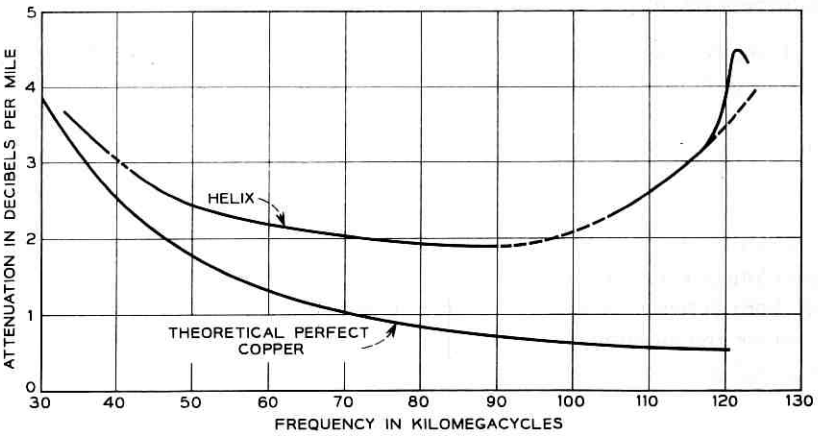

Fig. 1 — Measured attenuation of circular electric mode in 2-inch diameter helix waveguide.

measured by King and Mandeville.[3] The dashed curve from 90 to 105 kmc and around 40 kmc indicates that no measured data are available in these ranges. The dashed curve under the loss peak at 122 kmc indicates the expected curve in the absence of the loss peak. Also shown is the theoretical loss for perfect copper waveguide.

The measurements in the 120-kmc band were taken on several different sections of helix to give an indication of the performance of typical helix. The accuracy is ± 0.1 db/mile.

The new data show the losses increasing with frequency, with a minimum loss of 2 db/mile around 90 kmc. The loss of the 2-inch helix waveguide is less than 3 db/mile over a band from 40 kmc to 116 kmc.

The theoretical loss for the TE_{01} mode in perfect copper pipe continues to decrease with increasing frequency. The fact that the helix loss increases above 100 kmc can probably be attributed to imperfections in the helix. These imperfections, such as changes in diameter, ellipticity, and random deviations of the axis from a straight line, cause mode conversion and hence increased loss. Calculations indicate that the loss of the TE_{01} mode itself in the Holmdel helix is the same as in solid copper pipe even at the highest frequencies used here.

The additional loss peak at 122 kmc is believed not a property of the helix itself and is considered below.

IV. ANALYSIS OF LOSS PEAK

A peak in the loss curve is clearly shown in Fig. 1 at 122 kmc. The analysis presented in this section indicates that the loss peak is probably due to the way in which the helix was supported. This means that in a helix waveguide installation this loss peak can be avoided by proper support methods. The loss is believed caused by power being converted to the TE_{11} mode or possibly the TE_{12} mode and hence lost in the lossy lining of the guide. The TE_{11} and TE_{12} conversion is caused by the flexing of the waveguide between the support points due to its own weight. This is the serpentine bend effect described by Unger.[7] It is believed that this is the first time this effect has been observed in helix waveguide.

The measured waveguide was supported at 15-foot intervals and flexed ≈ 0.040 inch between support points. At a frequency where the beat wavelength between the TE_{01} and some spurious mode reaches 15 feet or some submultiple of 15 feet, the coupling to the spurious mode will be greatly increased.

From the calculations of Unger[8] and Miller[9] for helix waveguide, Fig. 2 shows a plot of beat wavelength vs frequency for several spurious

modes. At 122 kmc the beat wavelength for the TE_{11} mode is 7.5 feet, which is one-half the nominal 15-foot support interval. Hence a loss peak due to TE_{11} mode conversion is to be expected at 122 kmc.

At 122 kmc the beat wavelength for the TE_{12} mode is 4.7 feet, which would correspond to one-third of a support spacing slightly less than 15 feet. Since the actual support spacing varied somewhat around 15 feet, it may be possible that the TE_{12} also contributes to the loss peak.

To see if this explanation of the loss peak is reasonable, a calculation of the expected magnitude of the TE_{11} loss peak in helix can be made from the analysis for copper waveguide by Rowe and Warters[10] and by Unger.[7] Although the following calculation is for the TE_{11} mode, a similar calculation can be made for the TE_{12} mode.

Rowe and Warters show that the shapes of these mode conversion loss peaks vary with frequency, f, as

$$\left[\frac{\sin \pi \frac{L}{B_0 f_0}(f_0 - f)}{\pi \frac{L}{B_0 f_0}(f_0 - f)}\right]^2$$

for the case of no differential attenuation between the TE_{01} and the spurious mode. Here L is the distance over which the energy is fed in phase into the spurious mode, f_0 is the center frequency of the loss peak, and B_0 is the beat wavelength at f_0. The longer the interaction distance, the narrower and higher is the loss peak.

When differential loss is introduced between the TE_{01} and the spurious mode, as in the helix case, the first-order effect is to reduce the interaction length. This causes a lower and wider loss peak, but to first order the area under the curve remains the same as for the case of no differential loss. The expected area under the loss curve for helix can therefore be calculated from the theory for copper waveguide.

From equations (171) and (178) of Ref. 10, assuming a $[(\sin X)/X]^2$ shape, the area of the loss peak is expected to be:

$$\text{area} = 2.06 \times 10^4 B_0 f_0 \mid c_n \mid^2 (\text{db/mile}) \text{ kmc} \qquad (1)$$

where

f_0 = frequency at the peak of the loss curve in kmc
B_0 = beat wavelength in feet of the spurious mode at frequency f_0
c_n = Fourier coefficient in feet^{-1} of the curvature mode coupling coefficient corresponding to a spatial wavelength equal to B_0.

For axial bends the curvature mode coupling coefficient, $C(z)$, is inversely proportional to the radius of curvature, $R(z)$, of the bend.

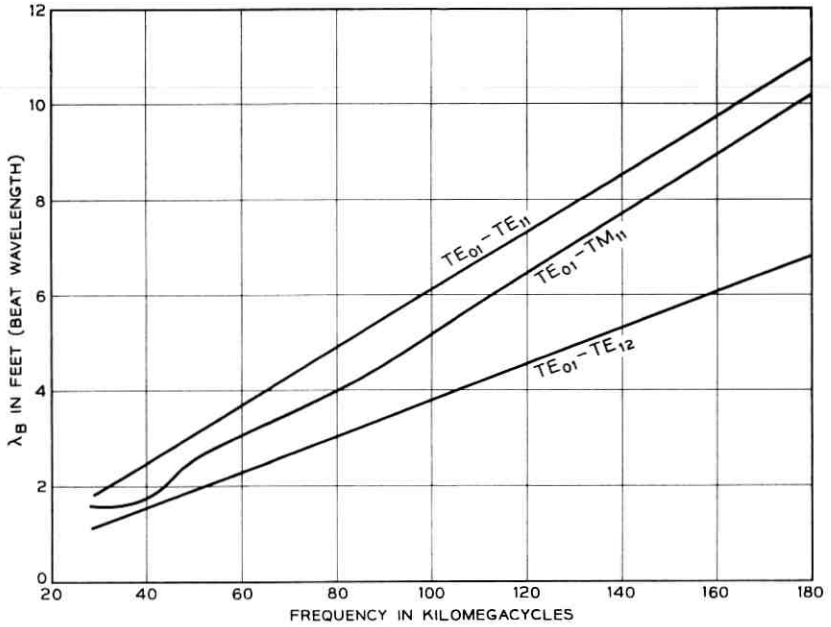

Fig. 2 — Beat wavelength vs frequency for 2-inch diameter helix waveguide.

$$C(z) = C_t/R(z)$$

where C_t is the tilt coupling coefficient for TE_{01} to TE_{11} conversion as computed by Morgan.[11] If the radius of curvature is computed assuming the guide bends from its own weight between the support points, the nth Fourier coefficient of the coupling coefficient is

$$c_n = C_t w L^2/4EI(\pi n)^2 \qquad (2)$$

where

w = weight per unit length of the pipe
L = distance between support points
E = Young's modulus for the pipe
I = moment of inertia of the pipe
$n = L/B$.

Using the appropriate value for 2-inch diameter helix waveguide and $C_t = 12$, $n = 2$, one finds:

$$\text{area} = 3.53 \text{ (db/mile) kmc.}$$

An estimation of the width at the half-height points on the loss hump

can be obtained from Young.[12] This assumes that the differential attenuation is relatively large, as is the case in helix:

$$\Delta f = (2\Delta\alpha/\Delta\beta)f_0 \tag{3}$$

where

f_0 = frequency at the peak of the loss curve
$\Delta\alpha$ = differential attenuation = $\alpha(TE_{01}) - \alpha(\text{spurious})$
$\Delta\beta$ = differential propagation constant = $\beta(TE_{01}) - \beta(\text{spurious}) = 2\pi/B_0$.

From the work of Unger[8] and Miller[9] the differential loss between the TE_{01} and TE_{11} in helix at 122 kmc is 8.3×10^{-3} nepers per foot. This gives an expected half-height width at 122 kmc of

$$\Delta f = 2.4 \text{ kmc.}$$

From Fig. 1, if the curve is assumed symmetrically, the area and half-height width of the measured hump can be computed. The area is measured as that above the extrapolated average-loss curve shown in Fig. 1 as a dashed line under the loss peak.

A comparison of the theoretical and measured properties of the loss peak is made in Table I.

The comparison between the measured and the theoretical values is reasonable and indicates that the observed hump is most probably due to mode conversion as discussed. Because of the underground installation it was not physically possible to check this conclusion by varying the support spacing.

It is expected that a similar phenomenon should occur when the beat wavelength is 5 feet, or $\frac{1}{3}$ of the support spacing. For the $TE_{01} - TM_{11}$ beat this occurs at 97 kmc, where good experimental data are not available. For the $TE_{01} - TE_{12}$ beat this occurs at 132 kmc and may contribute to the observed loss peak as mentioned earlier. The $TE_{01} -$

TABLE I — COMPARISON OF OBSERVED LOSS PEAK WITH THEORETICAL LOSS PEAK ASSUMING $TE_{01} \rightarrow TE_{11}$ MODE CONVERSION

	Position in Frequency	Area under Loss Peak	Half-Height Width of Loss Peak
Theoretical	122 kmc	3.53 $\frac{\text{db-kmc}}{\text{mile}}$	2.4 kmc
Observed	122 kmc	3.0 $\frac{\text{db-kmc}}{\text{mile}}$	3.0 kmc

TE_{11} beat wavelength is 5 feet at 82 kmc, which falls in a measured range. It is of interest to compute the expected magnitude of the loss peak at 82 kmc to see if the effect is observable.

Using the same appropriate constants as earlier, with $C_t = 8.1$, $n = 3$, the area from (1) is

$$\text{area at 82 kmc} = 0.149 \text{ (db/mile) kmc.}$$

The expected half-height width, using $\Delta\alpha = 3.6 \times 10^{-2}$ nepers per foot, is

$$\Delta f \text{ at 82 kmc} = 4.7 \text{ kmc.}$$

The height of the loss peak above the average curve is then approximately 3×10^{-2} db/mile. This small, wide peak would not be observable in the measurements.

V. SUMMARY

The attenuation of the Holmdel two-inch diameter helix has now been measured from 33 kmc to 125 kmc. Above 100 kmc the losses increase with frequency. This increase is attributed to mode conversion due to random imperfections in the waveguide. The loss is less than 3 db/mile from 40 kmc to 116 kmc.

The loss peak observed at 122 kmc is believed to be not a property of the waveguide itself but of the method of support. It is important therefore in a helix installation to take care in supporting the waveguide so that these loss peaks do not fall in a frequency range where the waveguide is to be used.

VI. ACKNOWLEDGMENTS

The author would like to acknowledge the assistance of H. M. James, who made some of the measurements reported here. The author would also like to thank S. E. Miller for the use of some unpublished calculations, W. M. Sharpless for supplying the millimeter-wave diode detectors, and D. H. Ring for helpful discussions.

REFERENCES

1. Miller, S. E., and Beck, A. C., Low-Loss Waveguide Transmission, Proc. IRE, *41*, March, 1953, pp. 348–358.
2. Morgan, S. P., and Young, J. A., Helix Waveguide, B.S.T.J., *35*, Nov., 1956, pp. 1347–1384.
3. King, A. P., and Mandeville, G. D., The Observed 33–90-kmc Attenuation of 2-inch improved Waveguide, B.S.T.J., *40*, Sept., 1961, pp. 1323–1330.
4. Sharpless, W. M., Point-contact Wafer Diodes for Use in the 90–140 kmc Frequency Range, B.S.T.J., *42*, Sept., 1963, pp. 2496–2499.

5. King, A. P., Status of Low-Loss Waveguide and Components at Millimeter Wavelengths, Microwave J., *7*, March, 1964, pp. 102–106.
6. Marie, P., transducer described in a French patent.
7. Unger, H. E., Circular Electric Wave Transmission through Serpentine Bends, B.S.T.J., *36*, Sept., 1957, pp. 1279–1291.
8. Unger, H. G., Helix Waveguide Theory and Application, B.S.T.J., *37*, Nov., 1958, pp. 1599–1648.
9. Miller, S. E., private communication.
10. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., *41*, May, 1962, pp. 1031–1170.
11. Morgan, S. P., Theory of Curved Circular Waveguide Containing an Inhomogeneous Dielectric, B.S.T.J., *36*, Sept., 1957, pp. 1209–1251.
12. Young, D. T., Effect of Differential Loss on Approximate Solutions to the Coupled Line Equations, B.S.T.J., *42*, Nov., 1963, pp. 2787–2794.

The Optical Ring Resonator

By W. W. RIGROD

(Manuscript received February 1, 1965)

Expressions are derived for the stability parameter, spot size, and wavefront curvature of a Gaussian beam in a ring resonator containing up to four spherical mirrors unequally spaced. Higher-order transverse modes and aperture effects are not considered. Two methods of analysis are used: (1) replacement of the mirrors by an infinite sequence of equally-spaced identical thick lenses, and (2) transformation of the beam into itself after one circuit of the ring, by means of a ray matrix representation of the equivalent thin lenses. The procedure can readily be extended to ring resonators with any number of spherical mirrors.

I. INTRODUCTION

In the ring laser, a light beam is directed about a closed loop, typically by three or four mirrors, and regeneratively amplified at frequencies for which the circuit path equals an integral number of wavelengths.[1-3] The only available analysis of a ring resonator with more than one spherical mirror appears to be that of Clark,[4] who used a ray-optical approach to derive the stability conditions for a ring with mirrors of two different curvatures and spacings. However, the means for a complete analysis of any arbitrary ring resonator are contained implicitly in the optical-mode theory developed in recent years in connection with two-mirror resonators.[5-8] The purpose of this note is to trace the connection between this theory and that of ring resonators in two different ways, and to derive the formulae defining the Gaussian (fundamental mode) beam in an arbitrary four-mirror resonator. A third method has been proposed recently by Collins in general form,[9,10] but will not be employed here because of its greater complexity.

In a two-mirror resonator, the wavefront curvatures of the light beam coincide with those of the mirrors, since the beams are reflected back on themselves. This is not the case in ring resonators, in which the beam is reflected obliquely. The boundary condition of the latter is merely that

the beam reproduce itself after each circuit, following its passage through a series of focusing elements or equivalent lenses.

As noted by Boyd and Kogelnik,[7] the stability conditions and beam size in spherical mirror resonators are the same as in an equivalent sequence of lenses. Thus the problem consists of applying the traveling-wave boundary condition to the appropriate equivalent-lens system. Owing to the astigmatism of concave mirrors in oblique reflection, they must be replaced by two different sequences of lenses, and each analyzed separately. For example, in the quadrilateral resonator of Fig. 1(a), the equivalent lens sequence for the clockwise traveling wave is shown in Fig. 1(b), where the focal length of the ith lens is given by[11]

$$f_{xi} = \tfrac{1}{2}\, b_i \cos\,(\varphi_i/2) \tag{1a}$$

in the plane of the ring, and

$$f_{yi} = b_i/2 \cos\,(\varphi_i/2) \tag{1b}$$

in the plane normal to the ring, for a mirror with radius of curvature b_i.

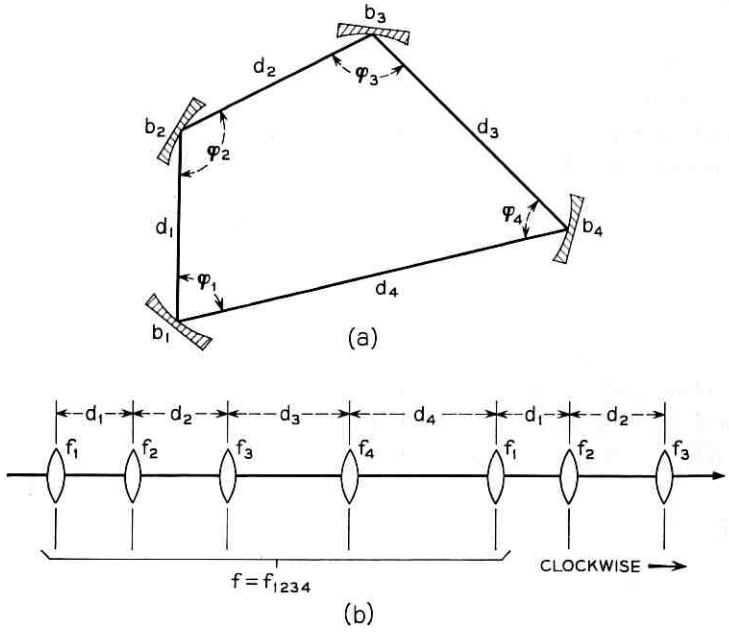

Fig. 1 — (*a*) Ring resonator with four spherical mirrors, unequally spaced. (b) Equivalent sequence of thin lenses for clockwise wave, with focal lengths given by equations (1a) and (1b) for tangential and sagittal planes, respectively.

subtending an included angle φ_i. The beam, consequently, is usually elliptical in cross section.

The circulating beam in a ring laser can be made circular in cross section, if need be, by the use of astigmatic mirrors or lenses, such that the effective curvatures of the ith mirror, b_{xi} and b_{yi}, in and normal to the plane of the ring, respectively, are related through

$$b_{yi} = b_{xi} \cos^2 (\varphi_i/2). \tag{2}$$

When the laser beam is plane-polarized, it is possible to design such correcting lenses for insertion at the Brewster angle to the optic axis, to minimize transmission losses. Alternatively, the elliptic output beam of a ring laser can be transformed into one of circular cross section by means of a single astigmatic element outside of the ring, placed where the spot is circular.

II. EQUIVALENT SEQUENCE OF THICK LENSES

The first method of analyzing the iterated sequence of four thin lenses shown in Fig. 1(b) is to replace them by a sequence of identical thick lenses f, whose principal planes are separated by a constant distance L. The stability condition for this system,[12]

$$0 < \frac{L}{f} < 4 \tag{3}$$

is then expressed in terms of the focal lengths and spacings of the equivalent thin lenses, and the beam description (radius, wavefront curvature, etc.) obtained from the known relations for an equivalent two-mirror resonator.[5-8]

The beam waist w_0 is located at a distance $L/2$ from the principal planes of each thick lens, and is given by

$$\frac{2\pi w_0^2}{\lambda} = L\left(\frac{4f}{L} - 1\right)^{\frac{1}{2}} \tag{4}$$

where λ is the wavelength. At a distance z from the waist in real space (when there are no intervening lenses), the beam radius is given by

$$w = w_0\left[1 + \left(\frac{z\lambda}{\pi w_0^2}\right)^2\right]^{\frac{1}{2}} \tag{5}$$

and the wavefront curvature R by

$$R = z\left[1 + \left(\frac{\pi w_0^2}{\lambda z}\right)^2\right] \tag{6}$$

where R has the same sign as z, positive when the center of curvature is to the left of the surface, i.e., when the surface is to the right of the beam waist.

The detailed procedure is illustrated in Fig. 2, in which the four thin lenses are taken in the order f_1, d_1 $\cdots$ f_4, d_4. First f_1 and f_2 are combined to form a thick lens f_{12} with principal planes located at distances h_1 and h_2 from the two lenses; then f_{12} and f_3 are combined to form the thick lens f_{123} with its principal planes located at distances h_1' and h_2' from the principal planes of the component lenses; and similarly for the combination of f_{123} and f_4 to form f_{1234}. The principal planes of f_{1234} are located at distances H_1 and H_2, respectively, from each of the thin lenses f_1 and f_4 as shown.

The value of $f = f_{1234}$, as well as of L_4, depends on the way in which the thin lenses are grouped, i.e., $f_{1234} \neq f_{2341}$. Thus, whereas L/f is invariant for the group of iterated thin lenses, there can be as many different beam waists as there are lenses. For the group shown in Fig. 2, the waist defined by L_4 in (4) above is located at a distance S_4 to the right of lens f_4 (i.e., measured clockwise from mirror b_4 in the ring resonator). Given the location and radius of the beam waist, the beam size and wavefront curvature at any distance z from the waist can be computed from the foregoing relations.

The expressions for the significant parameters of the equivalent thick-

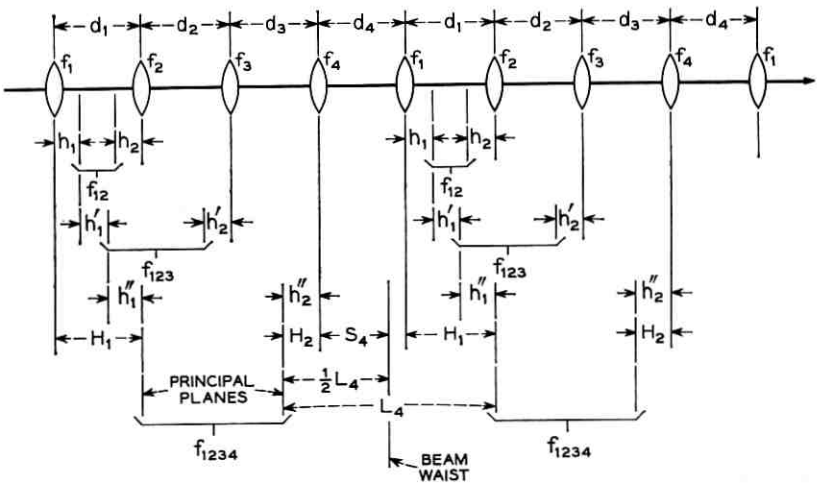

Fig. 2 — Reduction of iterated four-lens system to sequence of identical thick lenses ($f = f_{1234}$).

lens sequence of Fig. 2 are listed below. The remaining three beam waists in this system of four lenses can be found either by imaging the known waist about its nearest lens,[8,13] or simply by permuting the indices for the thin lenses and their spacings.

$$\begin{aligned}\frac{1}{f_{1234}} &= \left(\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}+\frac{1}{f_4}\right) - \frac{d_1}{f_1}\left(\frac{1}{f_2}+\frac{1}{f_3}+\frac{1}{f_4}\right) - \frac{d_3}{f_4}\left(\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}\right) \\ &\quad - d_2\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\left(\frac{1}{f_3}+\frac{1}{f_4}\right) + \frac{d_1 d_2}{f_1 f_2}\left(\frac{1}{f_3}+\frac{1}{f_4}\right) + \frac{d_2 d_3}{f_3 f_4}\left(\frac{1}{f_1}+\frac{1}{f_2}\right) \\ &\quad + \frac{d_1 d_3}{f_1 f_4}\left(\frac{1}{f_2}+\frac{1}{f_3}\right) - \frac{d_1 d_2 d_3}{f_1 f_2 f_3 f_4}\end{aligned} \tag{7}$$

$$\begin{aligned}\frac{L}{f} &= (d_1+d_2+d_3+d_4)\left(\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}+\frac{1}{f_4}\right) - \frac{d_1 d_2}{f_2}\left(\frac{1}{f_3}+\frac{1}{f_4}+\frac{1}{f_1}\right) \\ &\quad - \frac{d_2 d_3}{f_3}\left(\frac{1}{f_4}+\frac{1}{f_1}+\frac{1}{f_2}\right) - \frac{d_3 d_4}{f_4}\left(\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}\right) - \frac{d_4 d_1}{f_1}\left(\frac{1}{f_2}+\frac{1}{f_3}+\frac{1}{f_4}\right) \\ &\quad - d_1 d_3\left(\frac{1}{f_1}+\frac{1}{f_4}\right)\left(\frac{1}{f_2}+\frac{1}{f_3}\right) - d_2 d_4\left(\frac{1}{f_1}+\frac{1}{f_2}\right)\left(\frac{1}{f_3}+\frac{1}{f_4}\right) \\ &\quad + \frac{d_1 d_2 d_3}{f_2 f_3}\left(\frac{1}{f_4}+\frac{1}{f_1}\right) + \frac{d_2 d_3 d_4}{f_3 f_4}\left(\frac{1}{f_1}+\frac{1}{f_2}\right) + \frac{d_3 d_4 d_1}{f_4 f_1}\left(\frac{1}{f_2}+\frac{1}{f_3}\right) \\ &\quad + \frac{d_4 d_1 d_2}{f_1 f_2}\left(\frac{1}{f_3}+\frac{1}{f_4}\right) - \frac{d_1 d_2 d_3 d_4}{f_1 f_2 f_3 f_4}\end{aligned} \tag{8}$$

$$L_4 = f_{1234}(L/F) = H_1 + H_2 + d_4 \tag{9}$$

$$\begin{aligned}H_1 &= f_{1234}\left[\frac{d_3}{f_4} + d_2\left(\frac{1}{f_3}+\frac{1}{f_4}\right) + d_1\left(\frac{1}{f_2}+\frac{1}{f_3}+\frac{1}{f_4}\right) - \frac{d_2 d_3}{f_3 f_4}\right. \\ &\quad \left. - \frac{d_3 d_1}{f_4}\left(\frac{1}{f_2}+\frac{1}{f_3}\right) - \frac{d_1 d_2}{f_2}\left(\frac{1}{f_3}+\frac{1}{f_4}\right) + \frac{d_1 d_2 d_3}{f_2 f_3 f_4}\right]\end{aligned} \tag{10}$$

$$\begin{aligned}H_2 &= f_{1234}\left[\frac{d_1}{f_1} + d_2\left(\frac{1}{f_1}+\frac{1}{f_2}\right) + d_3\left(\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}\right) - \frac{d_1 d_2}{f_1 f_2}\right. \\ &\quad \left. - \frac{d_1 d_3}{f_1}\left(\frac{1}{f_2}+\frac{1}{f_3}\right) - \frac{d_2 d_3}{f_3}\left(\frac{1}{f_1}+\frac{1}{f_2}\right) + \frac{d_1 d_2 d_3}{f_1 f_2 f_3}\right]\end{aligned} \tag{11}$$

$$S_4 = \tfrac{1}{2}L_4 - H_2\,. \tag{12}$$

Less general ring resonators can be reduced to an iterated sequence of three, two, or one lens, respectively. The corresponding expressions for an iterated sequence of three thin lenses can be obtained from those for

four lenses by putting $d_4 = 1/f_4 = 0$ with appropriate redefinition of H_2 to locate the right-hand principal plane relative to lens f_3 :

$$\frac{1}{f_{123}} = \left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}\right) - \frac{d_1}{f_1}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) - \frac{d_2}{f_3}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) + \frac{d_1\, d_2}{f_1 f_2 f_3} \tag{13}$$

$$\begin{aligned}\frac{L}{f} = (d_1 + d_2 + d_3)\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}\right) - \frac{d_1\, d_2}{f_2}\left(\frac{1}{f_3} + \frac{1}{f_1}\right) \\ - \frac{d_2\, d_3}{f_3}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) - \frac{d_3\, d_1}{f_1}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) + \frac{d_1\, d_2\, d_3}{f_1 f_2 f_3}\end{aligned} \tag{14}$$

$$L_3 = f_{123}(L/f) = d_3 + H_1 + H_2 \tag{15}$$

$$H_1 = f_{123}\left[\frac{d_2}{f_3} + d_1\left(\frac{1}{f_2} + \frac{1}{f_3}\right) - \frac{d_1\, d_2}{f_2 f_3}\right] \tag{16}$$

$$H_2 = f_{123}\left[\frac{d_1}{f_1} + d_2\left(\frac{1}{f_1} + \frac{1}{f_2}\right) - \frac{d_1\, d_2}{f_1 f_2}\right] \tag{17}$$

$$S_3 = \tfrac{1}{2}L_3 - H_2 = f_{123}\left[\frac{1}{2}\frac{L}{f} - \frac{d_1}{f_1} - d_2\left(\frac{1}{f_1} + \frac{1}{f_2}\right) + \frac{d_1\, d_2}{f_1 f_2}\right]. \tag{18}$$

Similarly, for an iterated sequence of two thin lenses, f_1 , d_1 , f_2 , d_2 we obtain:

$$\frac{1}{f_{12}} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d_1}{f_1 f_2} \tag{19}$$

$$\frac{L}{f} = (d_1 + d_2)\left(\frac{1}{f_1} + \frac{1}{f_2}\right) - \frac{d_1\, d_2}{f_1 f_2} \tag{20}$$

$$L_2 = f_{12}(L/f) = d_2 + h_1 + h_2 \tag{21}$$

$$h_1 = \frac{f_{12}\, d_1}{f_2} \tag{22}$$

$$h_2 = \frac{f_{12}\, d_1}{f_1} \tag{23}$$

$$S_2 = \tfrac{1}{2}L_2 - h_2 = f_{12}\left[\frac{1}{2}\frac{L}{f} - \frac{d_1}{f_1}\right]. \tag{24}$$

III. MATRIX REPRESENTATION OF LENS GROUP

A second method of analyzing an infinite sequence of thin-lens groups has been derived recently by Kogelnik,[13] based on the representation of a lens system by a 2-by-2 matrix of transmission-line parameters A, B,

C, D. He has shown that the same $ABCD$ matrix which describes the transformation of position and slope of a ray, between input and output planes of the system, also serves to transform the radius w and wavefront curvature R of a Gaussian beam. The transformation is expressed by

$$q_j = (Aq_i + B)/(Cq_i + D) \tag{25}$$

where i refers to the input plane and j to the exit plane of the system, and

$$\frac{1}{q} = \frac{1}{R} - j\frac{\lambda}{\pi w^2}. \tag{26}$$

The ray matrix for any lens system is given by

$$[a] = \begin{vmatrix} A & B \\ C & D \end{vmatrix} = \begin{vmatrix} 1 - \dfrac{h_2}{f} & h_1 + h_2 - \dfrac{h_1 h_2}{f} \\ -\dfrac{1}{f} & 1 - \dfrac{h_1}{f} \end{vmatrix} \tag{27}$$

where h_1 and h_2 locate the principal planes relative to the input and output planes, respectively (Fig. 3), and f is the focal length of the system. Because of reciprocity,

$$AD - BC = 1. \tag{28}$$

For an element consisting of a thin lens f_1 followed by a distance d_1, we have

$$f = f_1, \qquad h_1 = 0, \qquad h_2 = d_1. \tag{29}$$

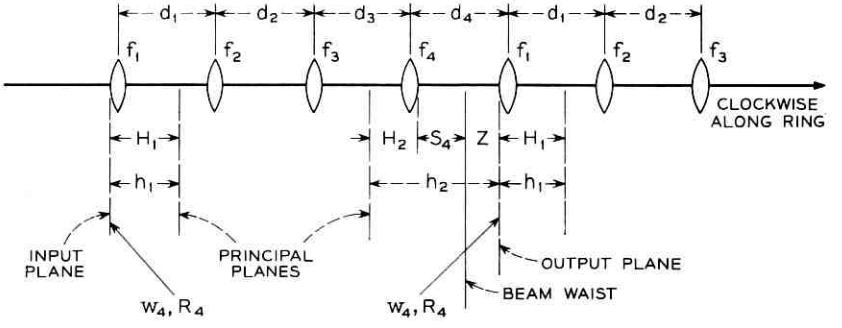

Fig. 3 — Optical system of four thin lenses for transformation of a beam into itself at the reference planes.

The ray matrix of this element is thus given by

$$[a_1] = \begin{vmatrix} 1 - \dfrac{d_1}{f_1} & d_1 \\ -\dfrac{1}{f_1} & 1 \end{vmatrix}. \tag{30}$$

The transformation of a beam into itself, after traversing a group of (say) four thin lenses, each followed by a spacing as indicated in Fig. 3, is then expressed by evaluating the $ABCD$ parameters of the product matrix:

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = [a_4] \times [a_3] \times [a_2] \times [a_1] \tag{31}$$

wherein each element matrix has the same form as (30); and then setting $q_j = q_i$ in (25), to obtain (at the reference plane):

$$R_4 = \frac{2B}{D - A} \tag{32}$$

$$\frac{\pi w_4^2}{\lambda} = \frac{2B}{\sqrt{4 - (A + D)^2}} \tag{33}$$

Combining these expressions with relations (3)–(6), we find the spot size w_0 of the beam waist to be given by

$$\pi w_0^2/\lambda = -\sqrt{4 - (A + D)^2}/2C \tag{34}$$

and the location of that waist given by

$$z = (A - D)/2C \tag{35}$$

where z is measured from the waist to the reference plane (just in front of f_1 in Fig. 3).

The expressions for the equivalent thick lens of Fig. 2, evaluated in the previous section, can be related to the $ABCD$ parameters of the lens system of Fig. 3 with the help of (27), as follows:

$$H_1 = h_1 = (D - 1)/C \tag{36}$$

$$H_2 + d_4 = h_2 = (A - 1)/C \tag{37}$$

$$L_4 = h_2 + h_1 = (A + D - 2)/C \tag{38}$$

$$\frac{1}{f_{1234}} = -C \tag{39}$$

$$\frac{L}{f} = 2 - (A + D) \tag{40}$$

$$z = d_4 - S_4 = \tfrac{1}{2}L_4 - H_1 \tag{41}$$

where z is also given by (35).

The $ABCD$ parameters for the four-lens system of Fig. 3 have been evaluated as indicated in (30) and (31) and listed in the Appendix. The parameters for a similar three-lens group f_1, d_1, $\cdots$ f_3, d_3 can be found by setting $d_4 = f_4^{-1} = 0$; and for a two-lens group by setting in addition $d_3 = f_3^{-1} = 0$.

Although the ray matrix formalism of Kogelnik offers no economy in computational labor over the straightforward derivation of the equivalent thick lens parameters, it has greater analytical flexibility and permits almost automatic extension to any number of lenses.

IV. ACKNOWLEDGMENT

The writer wishes to thank H. Kogelnik for helpful advice and stimulating discussions.

APPENDIX

Ray Matrix Parameters of Four-Lens System (Fig. 3)

$$\begin{aligned}
A = 1 &- \frac{d_1}{f_1} - d_2\left(\frac{1}{f_1} + \frac{1}{f_2}\right) - d_3\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}\right) \\
&- d_4\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4}\right) \\
&+ \frac{d_1\, d_2}{f_1 f_2} + \frac{d_2\, d_3}{f_3}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) + \frac{d_3\, d_1}{f_1}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) \\
&+ \frac{d_4\, d_1}{f_1}\left(\frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4}\right) + \frac{d_3\, d_4}{f_4}\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}\right) \\
&+ d_2\, d_4\left(\frac{1}{f_1} + \frac{1}{f_2}\right)\left(\frac{1}{f_3} + \frac{1}{f_4}\right) \\
&- \frac{d_1\, d_2\, d_3}{f_1 f_2 f_3} - \frac{d_2\, d_3\, d_4}{f_3 f_4}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) - \frac{d_3\, d_4\, d_1}{f_4 f_1}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) \\
&- \frac{d_4\, d_1\, d_2}{f_1 f_2}\left(\frac{1}{f_3} + \frac{1}{f_4}\right) + \frac{d_1\, d_2\, d_3\, d_4}{f_1 f_2 f_3 f_4}
\end{aligned} \tag{42}$$

$$B = (d_1 + d_2 + d_3 + d_4) - \frac{d_1}{f_2}(d_2 + d_3 + d_4) - \frac{d_4}{f_4}(d_1 + d_2 + d_3)$$
$$- \frac{(d_1 + d_2)(d_3 + d_4)}{f_3} + \frac{d_1 d_2}{f_2 f_3}(d_3 + d_4) + \frac{d_3 d_4}{f_3 f_4}(d_1 + d_2) \quad (43)$$
$$+ \frac{d_4 d_1}{f_4 f_2}(d_2 + d_3) - \frac{d_1 d_2 d_3 d_4}{f_2 f_3 f_4}$$

$$C = -\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4}\right) + \frac{d_1}{f_1}\left(\frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4}\right)$$
$$+ d_2\left(\frac{1}{f_1} + \frac{1}{f_2}\right)\left(\frac{1}{f_3} + \frac{1}{f_4}\right)$$
$$+ \frac{d_3}{f_4}\left(\frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3}\right) - \frac{d_1 d_2}{f_1 f_2}\left(\frac{1}{f_3} + \frac{1}{f_4}\right) - \frac{d_2 d_3}{f_3 f_4}\left(\frac{1}{f_1} + \frac{1}{f_2}\right) \quad (44)$$
$$- \frac{d_1 d_3}{f_4 f_1}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) + \frac{d_1 d_2 d_3}{f_1 f_2 f_3 f_4}$$

$$D = 1 - d_1\left(\frac{1}{f_2} + \frac{1}{f_3} + \frac{1}{f_4}\right) - d_2\left(\frac{1}{f_3} + \frac{1}{f_4}\right) - \frac{d_3}{f_4}$$
$$+ \frac{d_1 d_2}{f_2}\left(\frac{1}{f_3} + \frac{1}{f_4}\right) + \frac{d_1 d_3}{f_4}\left(\frac{1}{f_2} + \frac{1}{f_3}\right) + \frac{d_2 d_3}{f_3 f_4} - \frac{d_1 d_2 d_3}{f_2 f_3 f_4}. \quad (45)$$

REFERENCES

1. Rosenthal, A. H., J. Opt. Soc. Am. *52*, Oct., 1962, p. 1143.
2. Macek, W. M., and Davis, D. T. M., Appl. Phys. Lett. *2*, Feb., 1963, p. 67.
3. Macek, W. M., Schneider, J. R., and Salamon, R. M., J. Appl. Phys. *35*, Aug., 1964, p. 2556.
4. Clark, P. O., Proc. IEEE *51*, June, 1963, p. 949.
5. Boyd, G. D., and Gordon, J. P., B.S.T.J. *40*, March, 1961, p. 489.
6. Fox, A. G., and Li, T., B.S.T.J. *40*, March, 1961, p. 453.
7. Boyd, G. D., and Kogelnik, H., B.S.T.J. *41*, July, 1962, p. 1347.
8. Kogelnik, H., review article in *Advances in Lasers*, Dekker Pub., New York, 1964.
9. Collins, S. A., Appl. Opt., *3*, 1964, p. 1263.
10. Collins, S. A., and Davis, D. T. M., Appl. Opt. *3*, 1964, p. 1314.
11. Jenkins, F. A., and White, H. E., *Fundamentals of Optics*, 3rd ed., McGraw-Hill Book Co., New York, 1957, p. 94.
12. Pierce, J. R., *Theory and Design of Electron Beams*, D. Van Nostrand and Co., New York, 1954.
13. Kogelnik, H., Imaging of Optical Modes—Resonators with Internal Lenses, B. S. T. J., *44*, March, 1965, p. 455.

Diffraction Loss and Selection of Modes in Maser Resonators with Circular Mirrors

By TINGYE LI

(Manuscript received February 4, 1965)

The losses, phase shifts and field distribution functions for the two lowest-order modes of interferometer-type maser resonators consisting of spherically curved mirrors with circular apertures are computed by solving a pair of integral equations numerically on a digital computer. Solutions are obtained for the symmetric geometry of identically curved mirrors and for the half-symmetric geometry consisting of one plane and one curved mirror, with the radius of curvature of the mirrors as a variable parameter. The confocal or near-confocal configuration is shown to have good mode-selective properties in that the ratio of the loss of the second lowest-order (TEM_{10}) mode to that of the lowest-order (TEM_{00}) mode is the largest of the configurations considered. The numerical results should be of interest to those concerned with the problem of mode selection in optical masers and with the design of single-mode masers with relatively low gain.

I. INTRODUCTION

Interferometer-type resonators used for optical masers usually have a number of modal resonances falling under the gain profile of the active medium. Therefore optical maser oscillators generally can and often do oscillate in many modes, each mode having its own characteristic frequency and field pattern. Such a multimode, multifrequency output is undesirable for applications in communications and metrology. Many mode-selection schemes have been devised and tried, but most of them involve added complications and are beset by problems of stability. A simple solution was obtained by Gordon and White,[1] who built a stable single-frequency gas optical maser using a short, thin discharge tube. The resonator length was made short to reduce the number of longitudinal resonances and the mirror curvature was chosen so that only the lowest-order transverse mode had enough gain for oscillation. (The dif-

fraction losses of the higher-order modes were all greater than the gain of the tube.) Choosing the appropriate mirror curvatures for mode suppression is preferable to the commonly used method of aperturing the mirrors, because the former method does not restrict the amount of active material that can participate in maser action and therefore is capable of producing greater output power.

In order to select a pair of mirrors with the appropriate radii of curvature for single-mode operation, it is important to know accurately the diffraction losses of the modes as functions of the mirror size, spacing and curvature. Much work has been done on resonators with rectangular mirrors: Boyd and Gordon[2] and Boyd and Kogelnik[3] have derived approximate formulas for estimating the losses; Fox and Li[4] have made numerical calculations on a digital computer by solving the appropriate integral equations by the method of successive approximations; Streifer[5] recently made similar calculations using Schmidt expansion theory; Gloge[6] used perturbation techniques for his calculations. Unfortunately, these results are not very practical because the discharge tubes usually have circular cross sections and the diffraction losses for square mirrors and for circular mirrors are not simply related. Recently, Heurtley[7] has made some calculations on the modes and the eigenvalues of resonators with circular mirrors using Schmidt expansion theory. But published data on the diffraction losses suitable for practical design purposes are still lacking. It was with this purpose in mind that we computed the diffraction losses of the two lowest-order modes of the maser resonator with circular mirrors for various mirror curvatures. The results, which complement our earlier work,[4,8,9] should be of interest to those concerned with the problem of mode selection in optical masers.

II. MATHEMATICAL FORMULATION

The geometry of a maser resonator with circular mirrors is shown in Fig. 1. It is convenient to use the cylindrical coordinate system, with its axis coinciding with the axis of the resonator. The mirrors are spherically curved and have radii of curvature equal to R_1 and R_2. The mirror apertures are circular, their radii being a_1 and a_2. The separation between the mirrors along the axis is d.

The integral equations for the modes of the resonator with mirrors of arbitrary curvature and shape are given in Ref. 4. For the present case, they are in the form

$$\gamma^{(1)}\psi^{(1)}(r_1, \varphi_1) = \int_0^{a_2}\int_0^{2\pi} K^{(2)}(r_1, \varphi_1; r_2, \varphi_2)\psi^{(2)}(r_2, \varphi_2) r_2 d\varphi_2 dr_2 \quad (1)$$

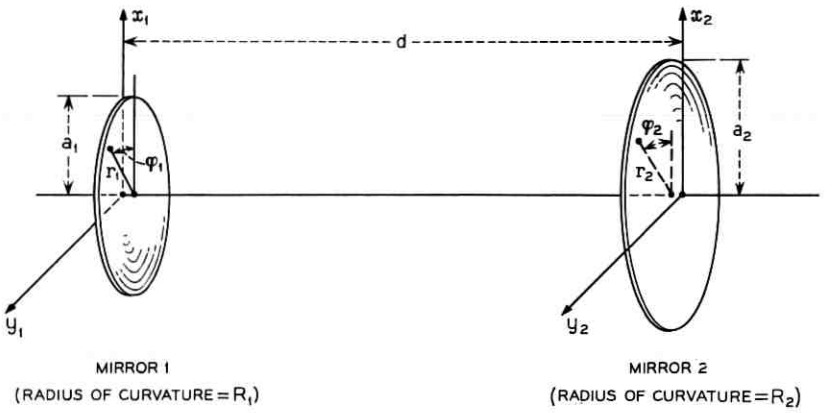

Fig. 1 — Geometry of the maser resonator consisting of spherically curved mirrors with circular mirror apertures.

$$\gamma^{(2)}\psi^{(2)}(r_2, \varphi_2) = \int_0^{a_1}\int_0^{2\pi} K^{(1)}(r_2, \varphi_2; r_1, \varphi_1)\psi^{(1)}(r_1, \varphi_1) r_1 d\varphi_1 dr_1 \quad (2)$$

where $K^{(1)}$ and $K^{(2)}$ are equal and are given by

$$K^{(2)}(r_1, \varphi_1; r_2, \varphi_2) = K^{(1)}(r_2, \varphi_2; r_1, \varphi_1)$$
$$= \frac{j}{\lambda d} \exp\left\{-\frac{jk}{2d}[g_1 r_1^2 + g_2 r_2^2 - 2r_1 r_2 \cos(\varphi_1 - \varphi_2)]\right\}. \quad (3)$$

In the above equations, $\gamma^{(1)}$ and $\gamma^{(2)}$ are the eigenvalues associated with the eigenfunctions $\psi^{(1)}(r_1, \varphi_1)$ and $\psi^{(2)}(r_2, \varphi_2)$, which are distribution functions of the reflected field at each mirror surface; g_1 is equal to $1 - (d/R_1)$ and g_2 is equal to $1 - (d/R_2)$; k is $2\pi/\lambda$ and λ is the wavelength in the medium between the mirrors. Making use of the relation[10]

$$\exp[jn(\pi/2 - \beta)]J_n(xy)$$
$$= \frac{1}{2\pi}\int_0^{2\pi} \exp\{j[xy\cos(\alpha - \beta) - n\alpha]\}\, d\alpha \quad (4)$$

and integrating (1) and (2) with respect to φ_1 and φ_2, respectively, it is seen that

$$\psi^{(1)}(r_1, \varphi_1) = R_n^{(1)}(r_1)e^{-jn\varphi_1}, \qquad (n = \text{integer}) \quad (5)$$

and

$$\psi^{(2)}(r_2, \varphi_2) = R_n^{(2)}(r_2)e^{-jn\varphi_2}, \qquad (n = \text{integer}) \quad (6)$$

satisfy (1) and (2), respectively. Thus the azimuthal variation is sinusoidal in form. The radial functions $R_n^{(1)}(r_1)$ and $R_n^{(2)}(r_2)$ satisfy the reduced integral equations

$$\gamma_n^{(1)} R_n^{(1)}(r_1) \sqrt{r_1} = \int_0^{a_2} K_n(r_1, r_2) R_n^{(2)}(r_2) \sqrt{r_2}\, dr_2 \tag{7}$$

$$\gamma_n^{(2)} R_n^{(2)}(r_2) \sqrt{r_2} = \int_0^{a_1} K_n(r_1, r_2) R_n^{(1)}(r_1) \sqrt{r_1}\, dr_1 \tag{8}$$

where

$$K_n(r_1, r_2) = \frac{j^{n+1}k}{d} J_n\left(k\frac{r_1 r_2}{d}\right) \sqrt{r_1 r_2} \exp\left\{-\frac{jk}{2d}(g_1 r_1^2 + g_2 r_2^2)\right\} \tag{9}$$

and J_n is a Bessel function of the first kind and nth order. It is unnecessary to show that the modes are orthogonal over their respective mirror surfaces, since this has been demonstrated for the general case of arbitrary mirrors.[4]

III. COMPUTED RESULTS AND DISCUSSION

An IBM 7094 computer was programmed to solve (7) and (8) for the two lowest-order modes (TEM_{00} and TEM_{10}) and their eigenvalues using the method of successive approximations.[4,8] Solutions were obtained for the symmetric and half-symmetric geometries; the symmetric geometry consists of identical mirrors ($a_1 = a_2 = a$, $g_1 = g_2 = g$) and the half-symmetric geometry consists of one plane and one curved mirror ($a_1 = a_2 = a$, $g_1 = 1.0$). In all cases, fifty or more intervals were used for the numerical integration of (7) and (8). When the losses were low and the convergence was slow, as many as one hundred intervals were used.

Figs. 2 and 3 show the relative field distributions of the lowest-order (TEM_{00}) and the next lowest-order (TEM_{10}) modes for the symmetric case. Except for geometries close to plane-parallel or concentric ($g \approx \pm 1$), the relative amplitude distributions can be closely approximated by Gaussian-Laguerre functions, with the spot sizes of the TEM_{00} mode equal to $(\lambda d/\pi)^{\frac{1}{2}}(1 - g^2)^{-\frac{1}{4}}$, as given by Boyd and Gordon.[2] Also, for geometries other than plane-parallel or concentric, the equiphase surfaces almost coincide with the mirror surfaces. The agreement between the computed field distributions and those of the generalized confocal theory[2,3] becomes closer with larger Fresnel numbers ($N > 1$).

The power losses per transit of the two lowest-order modes are given in Figs. 4 and 5 for various values of g lying in the low-loss region

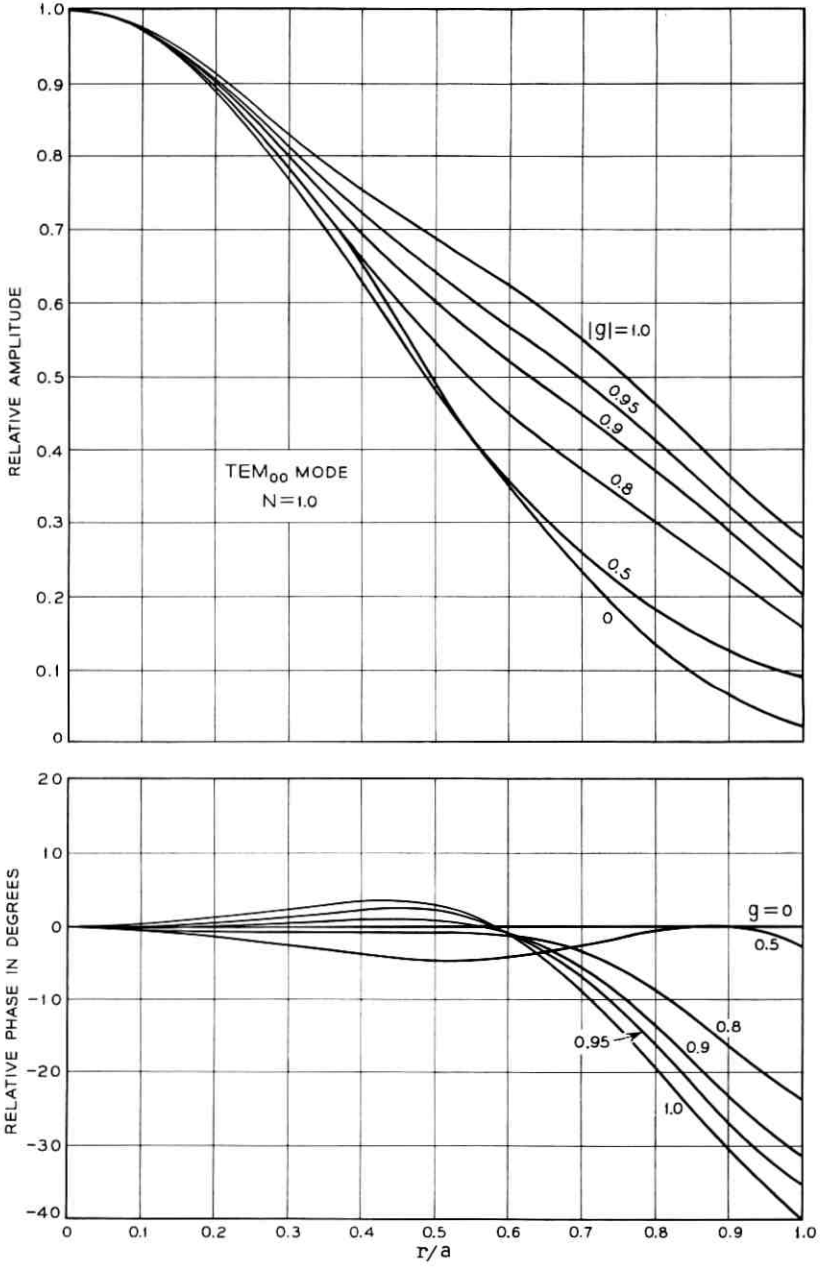

Fig. 2 — Relative field distributions of the fundamental (TEM_{00}) mode for the symmetric geometry ($N = a^2/\lambda d = 1.0$).

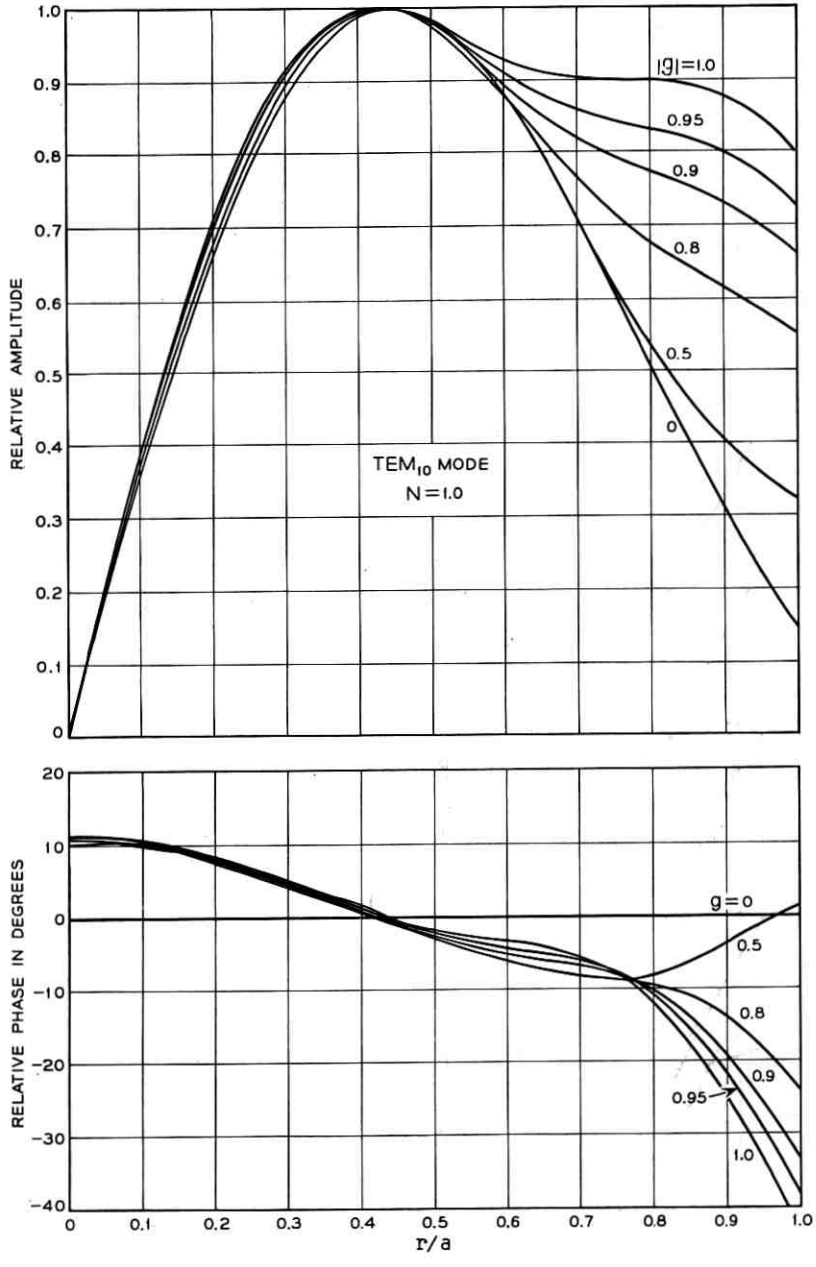

Fig. 3—Relative field distributions of the TEM_{10} mode for the symmetric geometry ($N = a^2/\lambda d = 1.0$).

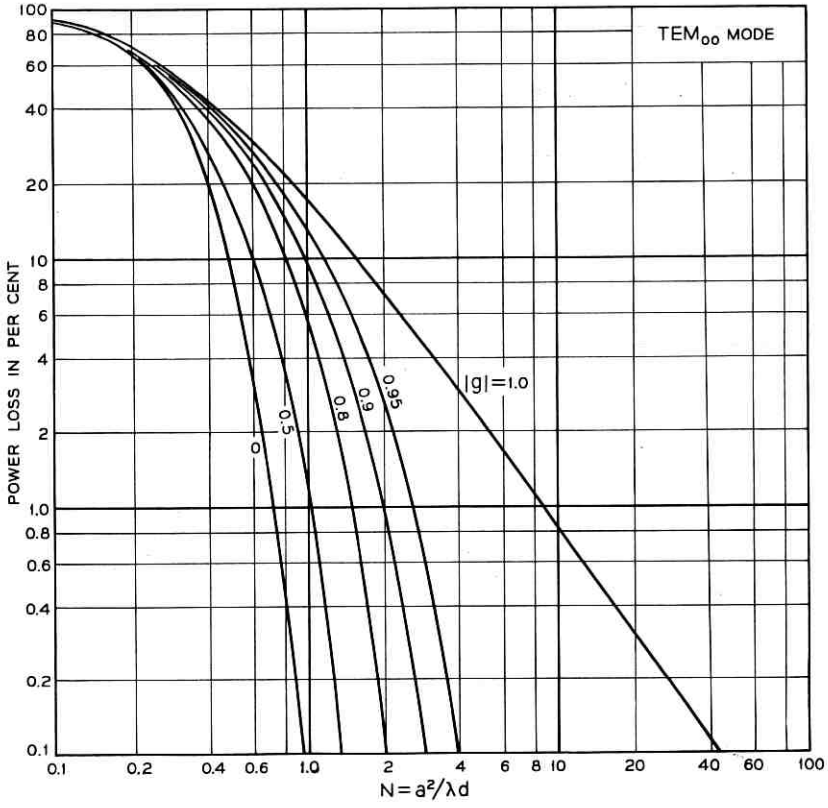

Fig. 4 — Power loss per transit of the fundamental (TEM_{00}) mode for the symmetric geometry.

($0 \leqq |g| \leqq 1$).[3,4] The curves for $|g| = 1$ (plane-parallel or concentric) and $g = 0$ (confocal) are the same as those given previously by Fox and Li.[8] Curves for other values of g lie between these two. The computed loss values for the confocal configuration are in perfect agreement with those of Slepian,[11] who has obtained various analytical expressions to approximate the losses. For our interest, the most useful form of the expression for the loss of the TEM_{nm} mode obtained by Slepian is

$$\text{loss} = \frac{2\pi (8\pi N)^{2m+n+1} e^{-4\pi N}}{\Gamma(m+1)\Gamma(m+n+1)} \cdot \left[1 + \frac{\text{constant}}{2\pi N} + \text{higher-order terms in } \frac{1}{2\pi N}\right] \tag{10}$$

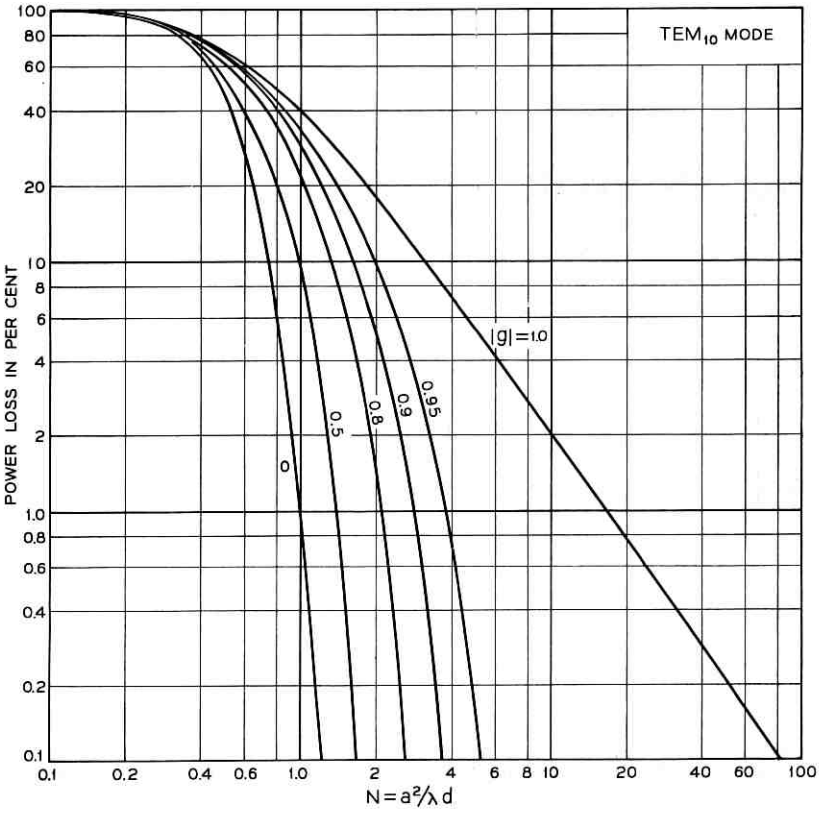

Fig. 5 — Power loss per transit of the TEM_{10} mode for the symmetry geometry.

where n and m are the azimuthal and the radial mode numbers respectively, and $\Gamma(x)$ is the gamma function. The computed curves for the plane-parallel or the concentric configuration are in excellent agreement with those of Vainshtein,[12] who has obtained an approximate formula for the loss of the TEM_{nm} mode. It is of the form

$$\text{loss} = 8\nu_{nm}^2 \frac{\beta(M + \beta)}{[(M + \beta)^2 + \beta^2]^2} \tag{11}$$

where ν_{nm} in the mth zero of the Bessel function $J_n(x)$, $\beta = 0.824$ and $M = (8\pi N)^{\frac{1}{2}}$. Both formulas (10) and (11) fail for small Fresnel numbers where losses are appreciable.

It is seen from Figs. 4 and 5 that the diffraction losses are very sensitive to changes in mirror curvature; also, provided $|g|$ is not very close

to unity, the losses decrease very rapidly as N increases. By choosing g and N suitably, it is possible to have a strongly oscillating fundamental mode in an optical maser with all higher-order modes suppressed. For example, if a maser tube has a gain of five per cent and the Fresnel number of the resonator is two, higher-order modes will not oscillate with $g = 0.90$. (Loss of TEM_{00} mode $\approx$1 per cent and loss of TEM_{10} mode $\approx$5.2 per cent).

While the loss of each mode is given by the magnitude of its eigenvalue, its resonant frequency is determined by the phase of its eigenvalue, which is equal to the phase shift (relative to the geometrical phase shift) per transit for the mode. The phase shifts for the two lowest-order modes are plotted in Figs. 6 and 7. The curves shown are for positive g only;

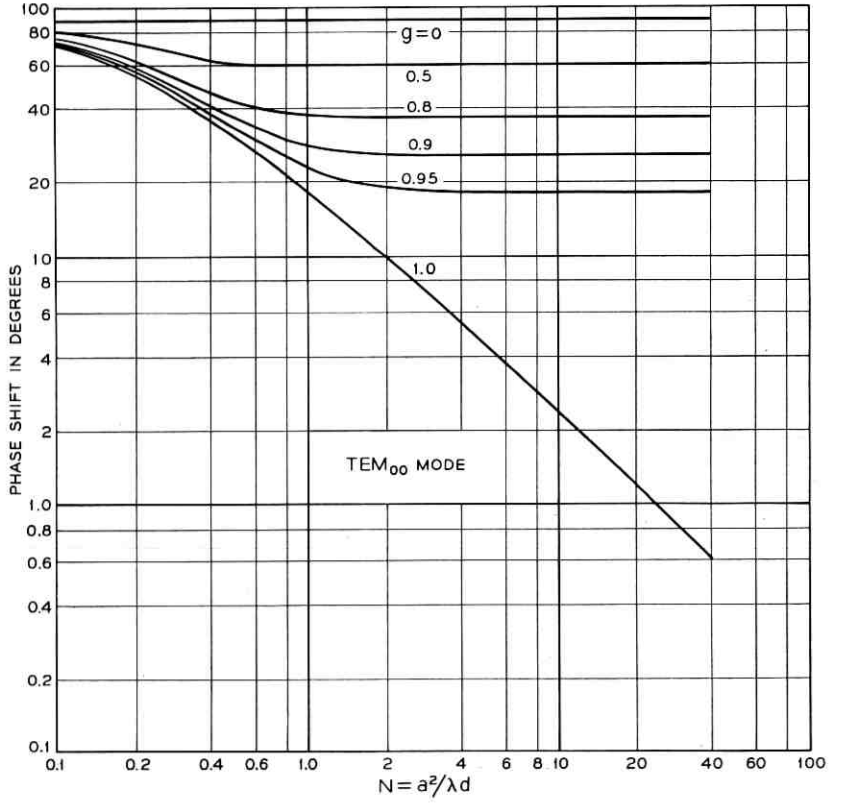

Fig. 6 — Phase shift per transit (leading relative to the geometrical phase shift) of the fundamental (TEM_{00}) mode for the symmetric geometry.

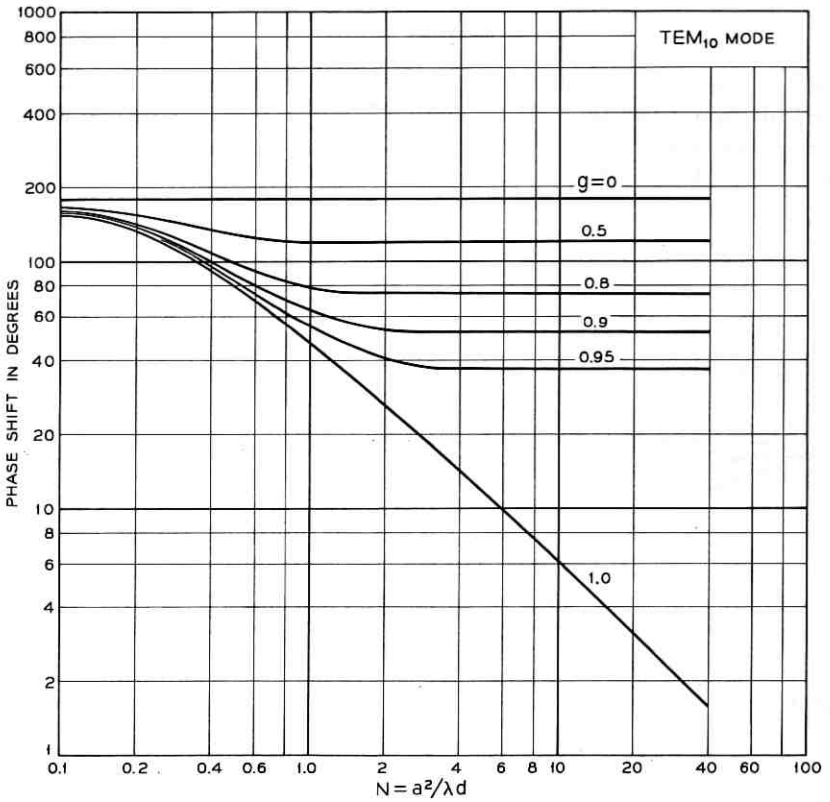

Fig. 7 — Phase shift per transit (leading relative to the geometrical phase shift) of the TEM_{10} mode for the symmetric geometry.

the phase shift for negative g is equal[4] to 180 degrees minus that for positive g. The horizontal portions of the curves can be calculated from the theory of Boyd and Kogelnik[3] as

$$\begin{aligned} \text{phase shift} &= (2m + n + 1) \arccos \sqrt{g_1 g_2} \\ &= (2m + n + 1) \arccos g, \quad \text{for} \quad g_1 = g_2 . \end{aligned} \tag{12}$$

The problem of mode discrimination in an optical maser can be approached in two different ways which depend on the available gain of the active medium. If the gain is large, the relative magnitudes of the eigenvalues of the different modes are of importance.[13,14,15] On the other hand, if the gain is small, such as in 6328Å He-Ne masers, the relative losses are important.[9] The ratio of the loss of the TEM_{10} mode to the loss

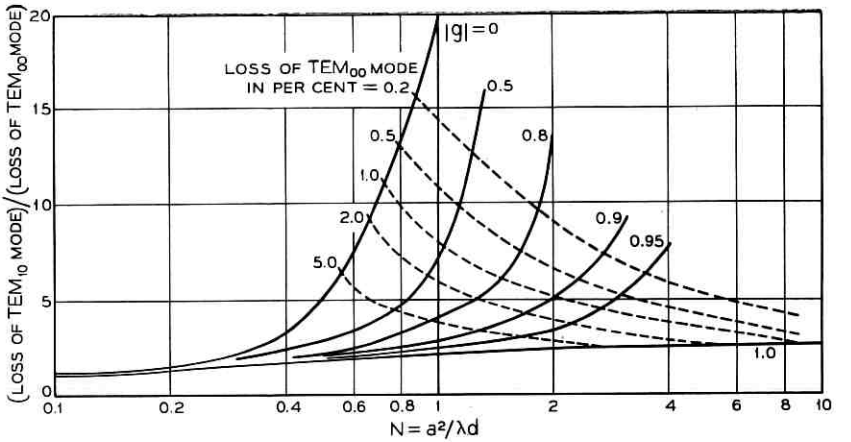

Fig. 8 — Ratio of the losses per transit of the two lowest-order modes for the symmetric geometry. The dotted curves are contours of constant loss for the TEM_{00} mode.

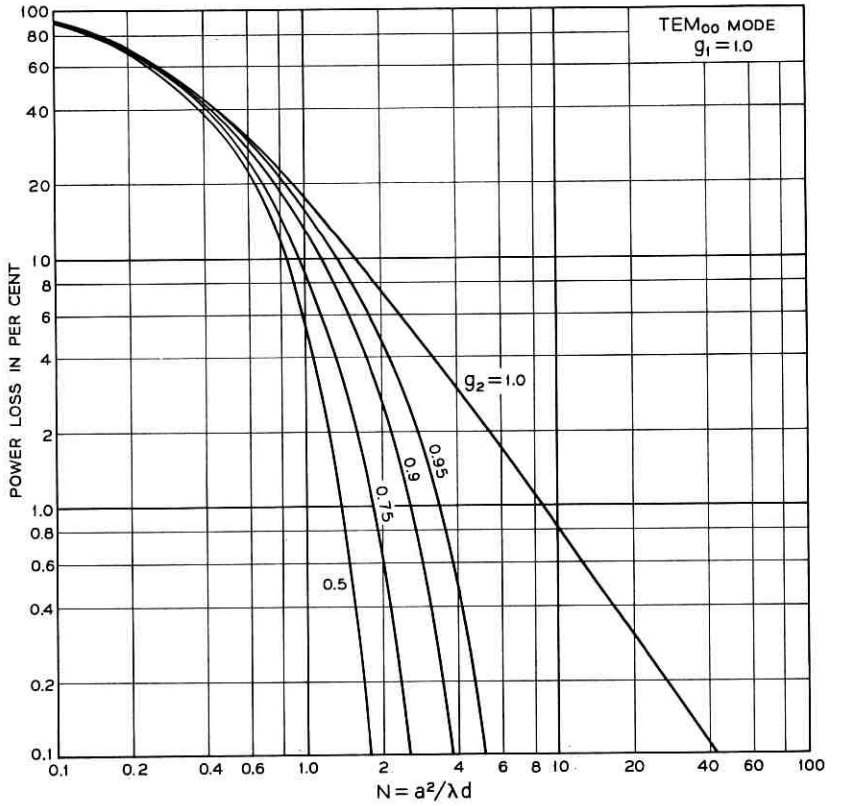

Fig. 9 — *Average* power loss per transit of the fundamental (TEM_{00}) mode for the half-symmetric geometry.

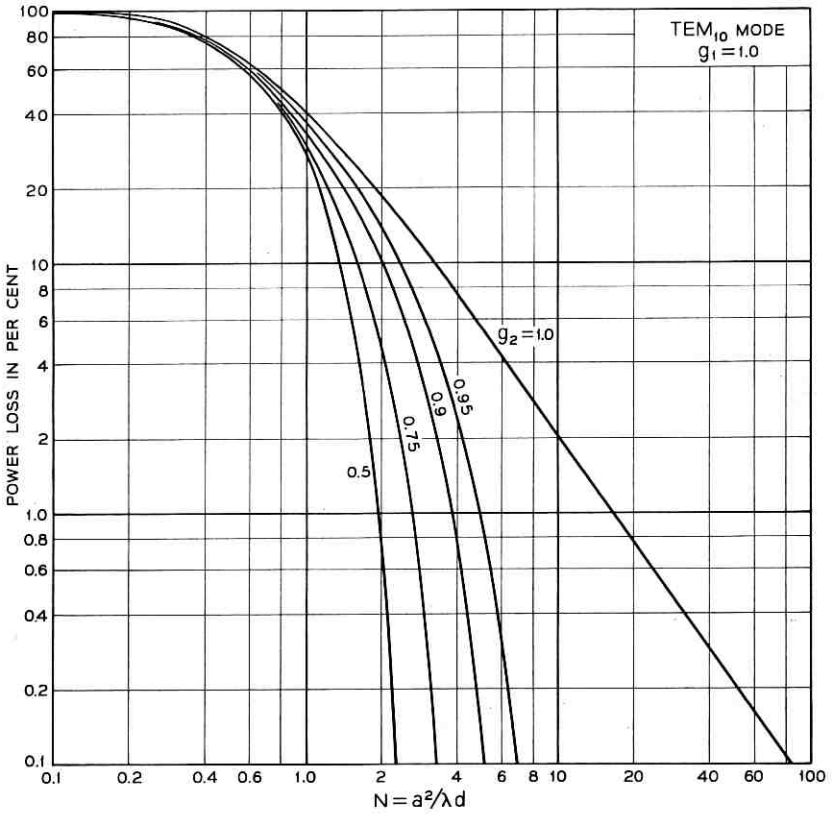

Fig. 10 — *Average* power loss per transit of the TEM_{10} mode for the half-symmetric geometry.

of the TEM_{00} mode, which is a measure of mode selectivity for the low-gain case, is plotted in Fig. 8 as functions of the Fresnel number. The dotted curves are contours of constant loss for the TEM_{00} mode. It is clear from this plot that the confocal or near-confocal geometry possesses good mode-selective properties, and that the plane-parallel or the concentric geometry is rather poor. However, if an aperture is placed in the midplane of a concentric resonator, its mode-selective properties have been shown to improve and to approach the confocal resonator.[9]

One of the commonly used resonator configurations is the half-symmetric geometry consisting of one plane and one curved mirror. If the spot size of the mode pattern at the plane mirror is very much smaller than the effective mirror aperture, such as in the case of the half-con-

centric or near half-concentric configuration, the half-symmetric resonator can be regarded as equivalent to a symmetric resonator of twice the length. However, if the mirror configuration is such that it lies between the half-confocal and the plane-parallel ($g_1 = 1.0$ and $0.5 \leqq g_2 \leqq 1.0$) and the effective mirror apertures are the same, the equivalence of half-symmetric and symmetric resonators is not valid and the modes and their eigenvalues must be recomputed. Computations were carried out for $g_1 = 1.0$ and $g_2 = 0.5$, 0.75, 0.9 and 0.95 corresponding to $g = 0$, 0.5, 0.8 and 0.9 of the symmetric case. Since the mode pattern at the plane mirror is now different from that at the curved mirror, the losses and the phase shifts incurred at the mirrors are also different. As before,[4] we define the *average* loss per transit as $1 - |\gamma^{(1)}\gamma^{(2)}|$ and the *average* phase

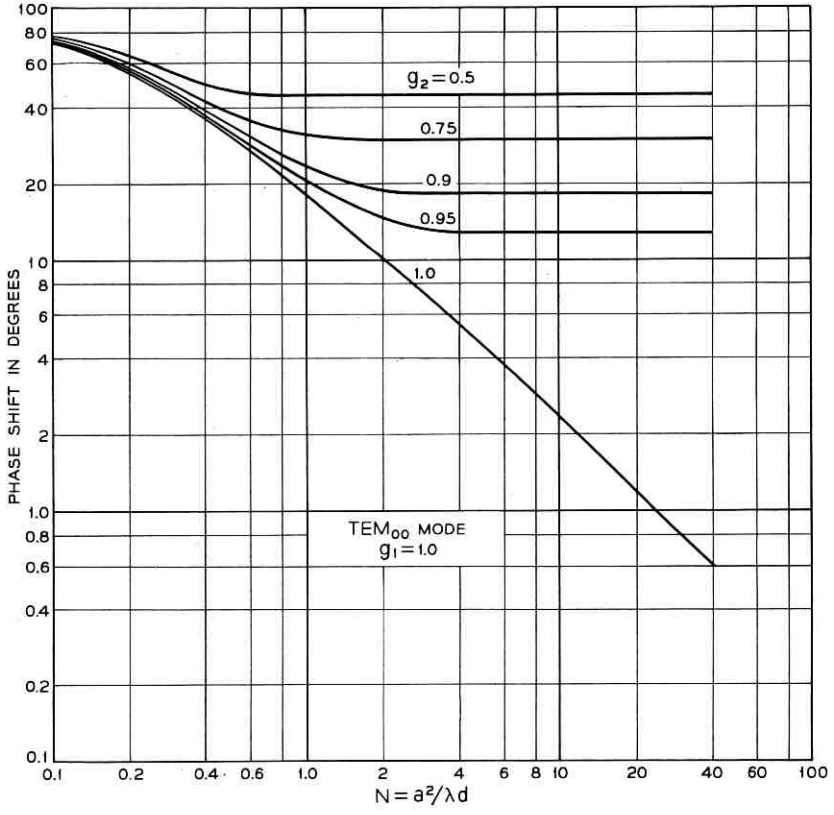

Fig. 11 — *Average* phase shift per transit (leading relative to the geometrical phase shift) of the fundamental (TEM_{00}) mode for the half-symmetric geometry.

shift per transit as (phase of $\gamma^{(1)}$ + phase of $\gamma^{(2)}$)/2. The computed average losses and average phase shifts per transit for the two lowest-order modes are given in Figs. 9 to 12, which are to be compared with those of the symmetric geometry shown in Figs. 3 to 6. The ratio of the loss of the TEM_{10} mode to the loss of the TEM_{00} mode for the half-symmetric resonator is plotted in Fig. 13. Comparing Figs. 8 and 13, it is seen that the mode-selective properties of the half-symmetric resonator are very similar to those of the corresponding symmetric resonator.

Although the computed results are for equal mirror apertures ($a_1 = a_2$), they are also applicable to certain "equivalent" geometries with unequal mirror apertures. The equivalence relations are discussed and given in a recent paper by Gordon and Kogelnik.[16]

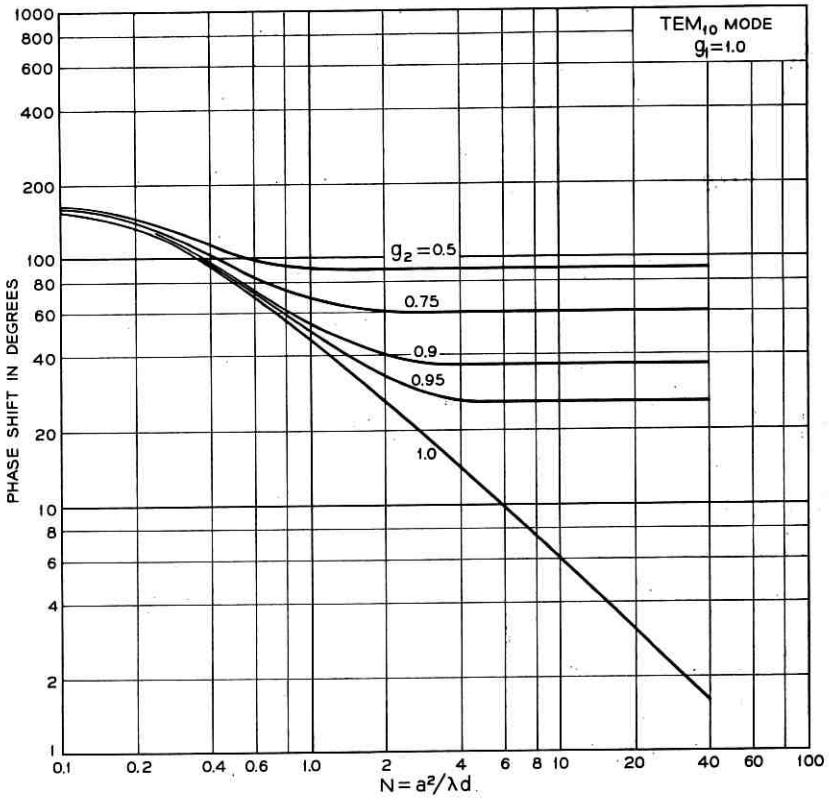

Fig. 12 — *Average* phase shift per transit (leading relative to the geometrical phase shift) of the TEM_{10} mode for the half-symmetric geometry.

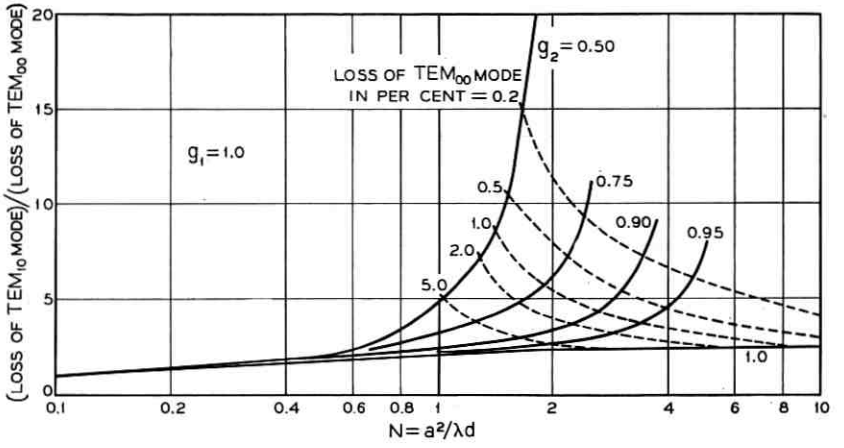

FIG. 13 — Ratio of the *average* losses per transit of the two lowest-order modes for the half-symmetric geometry. The dotted curves are contours of constant *average* loss for the TEM_{00} mode.

IV. CONCLUSIONS

The diffraction losses of the modes of interferometer-type maser resonators can be utilized for mode discrimination. Results of the computation show that the confocal or near-confocal resonator has good mode-selective properties; but its mode volume is small compared with those of the plane-parallel or concentric resonator of the same Fresnel number.[2,17,18] Since it is desirable to utilize as much of the active material as possible in order to obtain the maximum output power from a maser, a geometry that is as close to plane-parallel as is consistent with the requirements of mode discrimination and mechanical stability is to be preferred. In practice, the perturbing effects of the maser tube wall, the mirror irregularities, the nonlinearity of the active medium, etc. will modify the idealized modes and their losses and therefore should be taken into consideration.

V. ACKNOWLEDGMENT

The computational assistance of Mrs. C. L. Beattie is gratefully acknowledged.

REFERENCES

1. Gordon, E. I., and White, A. D., Single-Frequency Gas Lasers at 6328A, Proc. IEEE, *52*, Feb., 1964, p. 206.

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. Fox, A. G., and Li, Tingye, Modes in a Maser Interferometer with Curved and Tilted Mirrors, Proc. IEEE, *51*, Jan., 1963, p. 80.
5. Streifer, W., Modes in Spherical Resonators with Rectangular Mirrors, J. Opt. Soc. Am., *54*, Nov., 1964, p. 1399.
6. Gloge, D., Calculations of Fabry-Perot Laser Resonators by Scattering Matrices, Arch. Elekt. Ubertragung, *18*, March, 1964, p. 197.
7. Heurtley, J. C., Optical Resonators with Circular Mirrors, J. Opt. Soc. Am., *54*, Nov., 1964, p. 1400.
8. Fox, A. G., and Li, Tingye, Resonant Modes in a Maser Interferometer, B.S.T.J., *40*, March, 1961, p. 453.
9. Li, Tingye, Mode Selection in an Aperture-Limited Concentric Maser Interferomteter, B.S.T.J., *42*, Nov., 1963, p. 2609.
10. Stratton, J. A., *Electromagnetic Theory*, McGraw-Hill, New York, 1941, p. 372.
11. Slepian, D., Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty — IV: Extensions to Many Dimensions; Generalized Prolate Spheroidal Functions, B.S.T.J., *43*, Nov. 1964, p. 3009.
12. Vainshtein, L. A., Open Resonators for Lasers, JETP (USSR), *44*, March, 1963, p. 1050; Sov. Phys. JETP, *17*, Sept. 1963, p. 709.
13. LaTourette, J. T., Jacobs, S. F., and Rabinowitz, P., Improved Laser Angular Brightness Through Diffraction Coupling, Appl. Opt., *3*, August, 1964, p. 981.
14. Barone, S. R., and Newstein, M. C., Fabry-Perot Resonances at Small Fresnel Numbers, Appl. Opt., *3*, Oct., 1964, p. 1194.
15. Siegman, A. E., Unstable Optical Resonators for Laser Applications, Proc. IEEE, *53*, March, 1965, p. 277.
16. Gordon, J. P., and Kogelnik, H., Equivalence Relations among Spherical Mirror Optical Resonators, B.S.T.J., *43*, Nov., 1964, p. 2873.
17. Sinclair, D. C., Choice of Mirror Curvatures for Gas Laser Cavities, Appl. Opt., *3*, Sept., 1964, p. 1067.
18. Clark, P. O., A Self-Consistent Field Analysis of Spherical-Mirror Fabry-Perot Resonators, Proc. IEEE, *53*, Jan., 1965, p. 36.

Contributors to This Issue

DELAMAR T. BELL, B.S. in Eng., 1926, Georgia Institute of Technology; M.S.E.E., 1928, University of Cincinnati; Bell Telephone Laboratories, 1928—. He was first engaged in the development of audio-frequency amplifiers, oscillators, and other apparatus for use in motion picture and recording systems. He later worked on the development of transmission system networks and, during World War II and for a time thereafter, worked on networks for military fire-control and ballistic missile simulation applications. He has since worked on carrier transmission systems equipment engineering, and he is presently supervisor of a voice-frequency applications group in the exchange transmission equipment design area. Member, IEEE.

R. C. BOYD, B.S.E.E., 1947, M.S.E.E., 1948, University of Michigan; Bell Telephone Laboratories, 1948—. He was first concerned with systems engineering studies of Bell System and military transmission systems and with initial installations of N1 carrier. He later worked on systems engineering of exchange and short-haul carrier systems, and is presently Head, Carrier Engineering Department. Member, IEEE, Tau Beta Pi and Phi Kappa Phi.

FRED J. HERR, B.S.E.E., 1942, Cooper Union Institute of Technology; M.S., 1952, Stevens Institute of Technology; Bell Telephone Laboratories, 1936—. He was first engaged in the development of measuring equipment for coaxial transmission systems and during World War II worked on the development and testing of proximity fuses. Later he was concerned with broad-band coaxial systems and long- and short-haul carrier, participated in the laying of submarine cables, and did system design analysis and terminal maintenance planning for the SD submarine cable system. Mr. Herr has also been concerned with analysis of video transmission equipment, and he presently supervises a coaxial systems analysis group. Member, IEEE and Tau Beta Pi.

TINGYE, LI, B.Sc., 1953, University of Witwatersrand (South Africa); M.S., 1955, and Ph.D., 1958, Northwestern University; Bell Telephone Laboratories, 1957—. He has been engaged in studies of microwave

antennas and propagation, and in optical maser research. Member, IEEE, Eta Kappa Nu and Sigma Xi.

W. Ralph Lundry, B.S.E.E., 1928, Iowa State University; M.A. (Physics), 1934, Columbia University; Bell Telephone Laboratories, 1928—. He has worked on the design and development of transmission networks, filters, and equalizers for carrier telephone and television systems. He presently supervises a group responsible for carrier circuits development for exchange transmission. Member, IEEE.

George S. Moschytz, Dipl. El.-Ing., 1958, Dr. Sc. Techn., 1960, Institute of Telecommunications, Swiss Federal Institute of Technology, Zurich; Laboratories RCA, Ltd., Zurich, 1960–1962; Bell Telephone Laboratories, 1963—. Since joining Bell Laboratories, he has worked on the synthesis of circuits suitable for miniaturized thin-film components for data transmission equipment. Member, IEEE and Swiss Electrical Society.

W. W. Rigrod, B.S. in E.E., 1934, Cooper Union Institute of Technology; M.S. in Eng., 1941, Cornell University; D.E.E, 1950, Polytechnic Institute of Brooklyn; State Electrotechnical Institute (USSR), 1935–1939; Westinghouse Electric Corporation, 1941–1951; Bell Telephone Laboratories, 1951—. He has worked on physical problems connected with gaseous-discharge and vacuum-electronic devices at microwave frequencies. Since 1961 he has been engaged in research studies of lasers and coherent optics. Member, IEEE, American Physical Society and Sigma Xi.

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, SIAM, Eta Kappa Nu, Sigma Xi and Tau Beta Pi.

William H. Steier, B.S.E.E., 1955, Evansville College; M.S.E.E., 1957 and Ph.D. (E.E.), 1960, University of Illinois; Bell Telephone Laboratories, 1962—. He has been engaged in high-frequency measurements of waveguide systems and more recently has worked on the

design and measurement of optical waveguide systems. Member, American Physical Society and IEEE.

LEON H. STEIFF, B.S.M.E., 1946, Northeastern University; Bell Telephone Laboratories, 1956—. His first assignment was equipment design for P1 carrier, the first Bell System fully transistorized carrier system. He has since worked on wideband data transmission system development and the 150-mc pocket radio receiver. He currently supervises a group engaged in mechanical design of short-haul carrier systems; the group is also responsible for introducing new materials and techniques for Bell System use. Member, A.S.M.E.; Registered Professional Engineer, Massachusetts.

E. R. TAYLOR, B.S. in E.E., 1921, Johns Hopkins University; American Telephone and Telegraph Co., 1921–1934; Bell Telephone Laboratories, 1934–1963. Throughout his Bell System career, he specialized in transmission systems engineering and development. He has been granted 30 U.S. patents in the field of transmission and related systems. At the time of his retirement he was supervisor of the carrier applications group at the Merrimack Valley Bell Laboratories. Member, IEEE.

L. F. WILLEY, B.S. in E.E., 1937, Pennsylvania State University; Bell Telephone Laboratories, 1937—. He first worked on apparatus development of quartz crystal filters for use in the K and L1 broadband carrier systems. He later worked on waveguide networks for military projects and broadband carrier systems, and has more recently worked on transistor circuit development for short-haul carrier systems. Member, Tau Beta Pi and Eta Kappa Nu.