THE BELL SYSTEM
TECHNICAL JOURNAL

An Improved High-Frequency Radiotelephone System Featuring Constant Net Loss Operation

By J. M. FRASER, H. H. HAAS and M. G. SCHACHTMAN

(Manuscript received December 9, 1966)

A system is described in which a special type of syllabic compandor and other features offer an improved solution to the problems of using HF radio facilities for overseas telephone trunks. A major disadvantage of existing systems is the use of voice-operated gates to prevent singing and echo effects. These gates interfere with the free flow of conversation. The new system allows stable full-duplex operation as in conventional telephone circuits because a nearly constant circuit loss is maintained between the trunk terminals.

Significant improvements in circuit quality result from full-duplex operation and from a reduction in the effects of radio noise and interference by compandor action. Preliminary results of a field trial between New York and Buenos Aires have been highly favorable.

I. INTRODUCTION

The first transatlantic telephone cable was laid in 1956. For almost thirty years prior to that time, transoceanic telephone communication was provided almost exclusively via high-frequency radiotelephone circuits. Numerous submarine cables have since been laid, and HF radio is no longer used for telephone service over routes of heaviest traffic, e.g., New York-London, where large numbers of cable and satellite circuits exist. Many long transoceanic routes, however, are

677

not served by submarine cables or satellites; direct telephone service between the United States and forty foreign countries and areas (see Fig. 1) is still provided by means of HF radio facilities exclusively.

Although a change in the composition of transoceanic telephone facilities is expected as satellites become more fully utilized, it is likely that the HF services will hold their own for economic reasons, particularly for small circuit groups.

For some time the HF radio spectrum, a rather limited resource to begin with, has been nearing the point of saturation. Thus, as new submarine cable and land* facilities make it possible to suspend existing HF routes, the frequencies are reassigned to provide service to new areas, typically to the developing nations of the world. Thus, the total number of HF circuits is expected to remain about the same well into the communication satellite era.

Since high frequencies propagate via the ionosphere, which is in a continual state of change, HF transmission is highly variable and requires special equipment and procedures to cope with the problems created by this variability. At times, HF signals (say, 4 to 27 MHz for overseas service) propagate easily over great distances; at other times, the medium will not propagate a given frequency at all. Any condition between these extremes is possible.

The ionospheric medium is a difficult and erratic one, subject to many propagation anomalies. Moreover, unique interconnection problems exist between point-to-point HF radiotelephone circuits and conventional land plant. Because of the unstable nature of the HF radio transmission medium, the combination of the 2-wire part of the land plant and a "4-wire" radio circuit presents problems that are not encountered, for example, in 4-wire cable or microwave radio circuits. Any interconnection of 4-wire and 2-wire circuits creates potential singing and echo problems. In the land plant, this problem is solved by operating the circuit at a net loss sufficient to prevent singing and to render echoes negligible.[1] Because such circuits are stable, any allowance for variations in loss is small, and the mode of operation may be characterized as "constant net loss." If HF radiotelephone circuits were operated in this mode without special controls, the allowance for variations would, in general, have to be rather large. If the design net loss were sufficient for stability with minimum transmission loss in the radio portion of the circuit, even the average loss would result in low

---

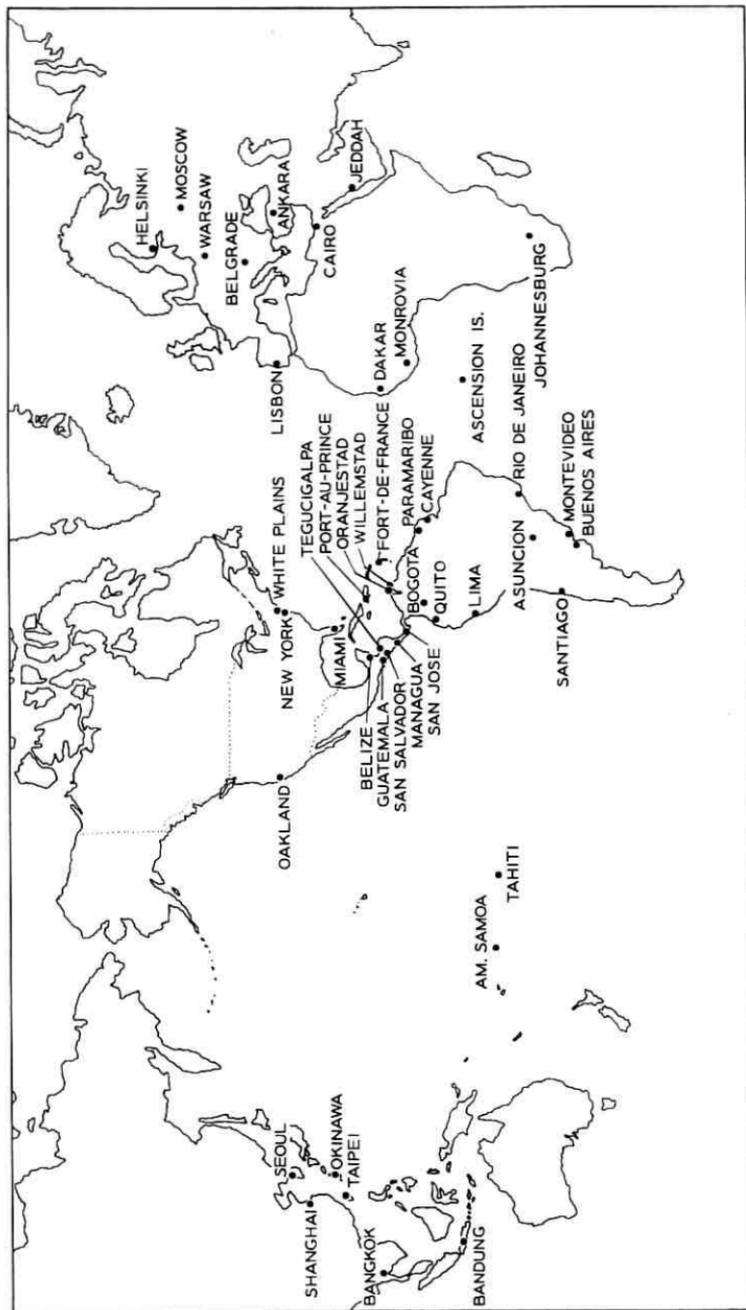* HF radio is widely used for intracontinental communication in some less developed parts of the world.

Fig. 1 — Bell System direct radiotelephone service (January 1, 1967).

received volume. The high loss occurring during deep fading would, at times, make the received volume inaudible. This would be a very inefficient mode of operation.

Input speech volumes from connecting plant vary over a wide range, at least 1000:1 in power. Linear transmission of this full range is an inefficient use of radio transmitter load capacity.

Because signal-to-noise and signal-to-interference ratios are frequently marginal, fairly high power is required for reasonable reliability. For example, at least a 4-kilowatt output might be required on a given path for adequate reception of the weakest talker. Then a 4-megawatt transmitter capacity would be required in a linear system for the strongest talker. The use of 4-megawatts perhaps one percent of the time would be uneconomical. In addition, the unnecessary radiation of excessive power would aggravate interference problems. As a counterexample, if 4 kilowatts were the maximum available to the strongest talker, the weakest talker would develop an output of only 4 watts and might be lost in the noise. These examples make it obvious that in order to minimize transmitter power and still accommodate the large range of speech volumes, some form of volume regulation is necessary. Volume regulation, which varies the gain of the circuit, compounds the problem of singing stability and echo effects discussed above.

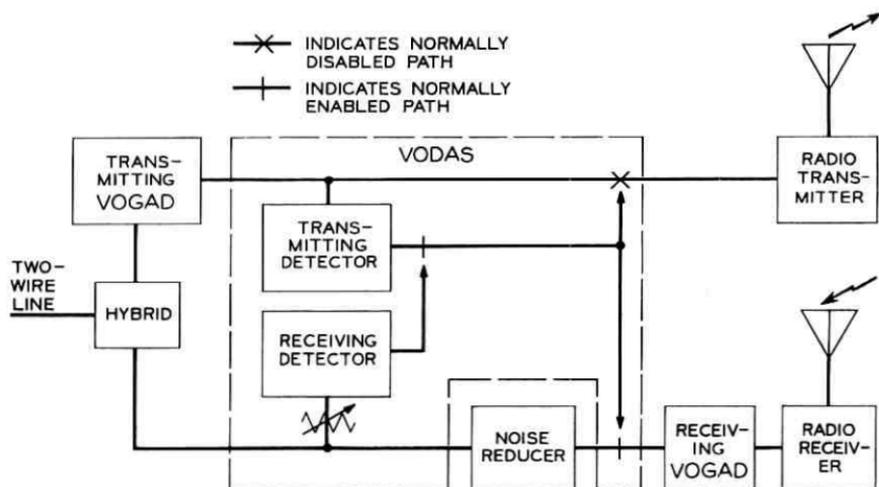The basic HF radiotelephone transmission problem may be sum-

Fig. 2 — Conventional VODAS terminal.

marized at this point as one of continuously maximizing the signal-to-noise ratio under varying conditions while at the same time preventing a net gain around the circuit that would allow singing to occur. The conventional solution to this problem[2] comprises at least the following two elements:

(*i*) VODAS, or Voice Operated Device Anti-Singing, which makes the circuit one-way at a time. With this arrangement, singing and talker echoes are avoided.

(*ii*) VOGAD, or Voice Operated Gain Adjusting Device, which loads the radio transmitter efficiently.

In one form or another, these elements have been widely used throughout the world from the beginning of the service up to the present time.

Fig. 2 is a simplified diagram of the radio terminal equipment used in overseas gateways of the Bell System. In addition to the transmitting VOGAD and VODAS, there is a receiving VOGAD and a "noise reducer." Briefly, the terminal operates as follows. The receiving path is normally enabled and the transmitting path disabled. Outgoing speech, regulated to constant volume by the transmitting VOGAD, operates the transmitting speech detector, which enables the transmit path and disables the receiving path. When the speech train ends, the VODAS reverts to the receiving mode. The receiving VOGAD provides fading regulation of incoming speech, which has already been coarsely compensated by automatic gain control in the receiver. The noise reducer is a voice-activated expandor capable of moderate noise and interference reduction in the speech gaps if radio conditions are at least fair. The receiving detector, by inhibiting the transmitting detector, prevents echoes of the received speech from switching the circuit to the transmitting mode. The receiving detector is provided with a sensitivity adjustment that may be set by a "technical operator."

Although its wide use for a long period attests to the soundness of the present terminal, a number of disadvantages have become apparent. Foremost is the fact that even under perfect transmission conditions, a one-way-at-a-time circuit inhibits the smooth flow of conversation; in fact, when double-talking occurs, a significant amount of speech may be lost. This effect is sometimes called "lock-out". Also, because of the difficulty in differentiating between speech and noise, some "clipping" by the speech detectors occurs under the best conditions, particularly of the initial portions of speech having low energy. VODAS designs employing a receiving detector for echo protection are particularly vulnerable to operation on high received noise or in-

terference. Such false operation is called "lock-up" and makes the circuit completely unusable as long as it persists. Frequent adjustment of the receiving detector sensitivity and other controls is required if transmission quality is to be optimized with changing radio conditions.

II. AN IMPROVED RADIOTELEPHONE SYSTEM

The HF radiotelephone transmission problems can be solved by a method other than the conventional one described. Experimental systems based on general principles known for some time[3, 4] recently have been investigated by the Bell Telephone Laboratories and others.[5, 6, 7, 8] The major disadvantages of the existing system are eliminated and significant improvements are realized in the new type of system. This paper describes the experimental system investigated by the Bell Telephone Laboratories, called the constant net loss (CNL) system.

The principles of the CNL system may be explained with reference to Fig. 3, which illustrates a circuit equipped at both ends with a new type of terminal. Input speech is applied to a "complete compressor," which gives constant output volume over a wide range of inputs. The compressor smooths out the syllabic changes in speech loudness in order to fully load the radio transmitter and to optimize the received signal-to-noise ratio on a syllable-to-syllable basis. Thus, it is more effective than the slow-acting VOGAD, which regulates differences between talkers but does little to the syllabic variations of a given talker. The control signal from the first stage of the compressor, a low-frequency (0-100 Hz) analog of the syllabic variations of the input speech, feeds the two stages to give complete compression as explained in Appendix A. In addition, the control signal frequency-modulates a subcarrier, which is transmitted over the radio link along with the compressed speech, but in a separate narrow-band channel. Fig. 4 shows the frequency allocations of a single voice channel with its control channel.

The control signal is used at the receiving end to control an expandor, which restores the original variations in loudness. The FM control channel has sufficient margin to assure that even under severe noise, interference, and fading conditions, the expandor will properly "track" the compressor. The blocks labeled LOG and ANTILOG in the control channel comprise an instantaneous compandor[9] (nonlinear compressor and expandor) whose function is to make the control signal less susceptible to noise and interference in the control channel.

Although restoration of the original speech loudness variations by
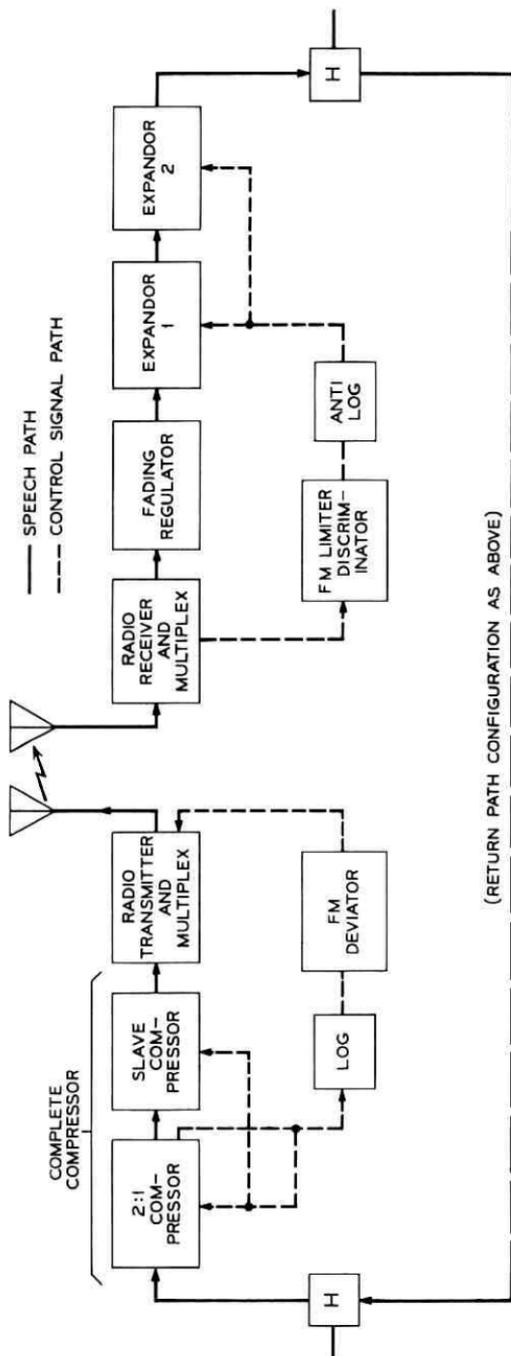
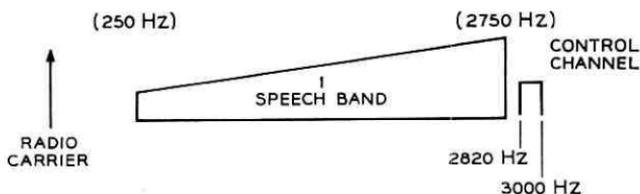Fig. 3 — Constant net loss radiotelephone system (simplified).

Fig. 4 — One-channel operation.

an expandor controlled from the sending end is of some benefit in itself, it has two other more essential purposes: (i) It maintains an overall constant net loss by making the expandor loss track the compressor gain; (ii) it very effectively mutes radio noise and interference during the gaps between syllables when the expandor loss is a maximum. A substantial subjective improvement accrues because the noise is reduced when it is most noticeable. Expandor tracking error caused by noise or interference in the control channel is minimized by the narrow bandwidth, FM carrier power, and other parameters of the control channel. The complementary compressor/expandor action via the control channel results in constant loss in the absence of fading. A stable loss around the outgoing and return paths allows the circuit to be set up "full-duplex" without a VODAS.

In the CNL terminal, fading variations that remain after the action of the receiver automatic gain control are absorbed by the fading regulator preceding the expandor. The fading regulator (which is also a syllabic complete compressor) operates independently of the control channel and somewhat more slowly than the transmitting compressor. The dynamic actions of the transmitting compressor, fading regulator, and expandor are such that, although gains and losses of these devices and of the transmission medium vary within the circuit, the net overall loss is approximately constant.

Since the CNL system has no VODAS with its inherent problems of clipping, lock-out, and receiving detector lock-up, a smoother flow of conversation results. Its compandor action and fading regulation are more effective in optimizing S/N (signal-to-noise) and S/I (signal-to-interference) ratios than the uncoordinated actions of the VOGADs and noise reducer of present terminals. An additional advantage is that no in-service adjustments of the terminal by technical operators are necessary, as is the case in the present terminals. A summary of the principal characteristics of experimental CNL systems is given in Table I.

TABLE I — SUMMARY OF CNL HF RADIOTELEPHONE
TERMINAL CHARACTERISTICS

System Capacity	Four speech channels plus four control channels can be accommodated on one radio carrier in conjunction with conventional 12-kHz radio and channelizing equipment.
Transmitting Compressor Operating Range	The range of input power over which compressor output is essentially constant is +8 dBm0 to −32 dBm0. Below −32 dBm0, the compressor acts as a linear amplifier; above +8 dBm0, some clipping of a sinewave occurs.
Speech Bandwidth	The nominal speech bandwidth is 250 to 2750 Hz for channels next to the carrier, 250 to 2950 Hz for outboard channels.
Control Signal Allocations and Power	(i) The FM control signal associated with an inboard channel occupies an 180-Hz band centered at 2910 Hz. With no system input, the subcarrier frequency is 2850 Hz; maximum test tone input (+8 dBm0) deviates the subcarrier upward 120 Hz to 2970 Hz.
	(ii) The outboard channel, 180 Hz wide, is centered on 3140 Hz, and the corresponding subcarrier frequency excursion is from 3200 Hz downward to 3080 Hz.
	(iii) The FM subcarrier power in the composite signal is applied to the radio transmitter at a level 16 dB below its peak envelope power rating and 6 dB below a test tone in the speech channel, the test tone being within the constant output range of the transmitting compressor.
Time Constants	The transmitting compressor has an attack time of 3 msec and a recovery time of 13.5 msec. The overall effective RC time constant of the compandor system is 20 msec.
Fading Regulator	The fading regulator maintains essentially constant output over an input range 20 dB below and 10 dB above the nominal no-fade level. The attack and recovery times are 12 msec and 54 msec, respectively. These values are 4 times the corresponding transmitting compressor time constants.
FM Control Signal Deviation/Loss Ratio	The deviation/loss ratio is the transfer constant that relates the deviation of the FM subcarrier to the expandor loss variations and is the best measure of the susceptability of the control channel to noise and frequency instability. The constant is 2 Hz per dB.

The CNL terminal is compatible with existing channelizing equipment, privacy devices, radio transmitters, and receivers. The system uses no more bandwidth overall than present systems. Although the FM channel shares bandwith with the speech, the loss of speech bandwidth is slight. This is possible because of more efficient use of the available bandwidth by means of sharper filter cut-offs. In conjunction

with conventional 12-kHz radio and channelizing equipment, a full complement of four speech channels plus four control channels can be accommodated (Fig. 5).

The control signal must share the available power of an existing transmitter with the compressed speech. Satisfactory performance is expected with control subcarrier magnitudes (see Table I) such that the total transmitter load is not significantly increased.

There are several reasons for using FM to transmit the control information. The foremost reason is that limiter action provides level compensation, making the control signal insensitive to fading (as regards purely level, rather than S/N variations). Also, bandwidth can be traded for S/N advantage. In addition, with FM, the dc com-
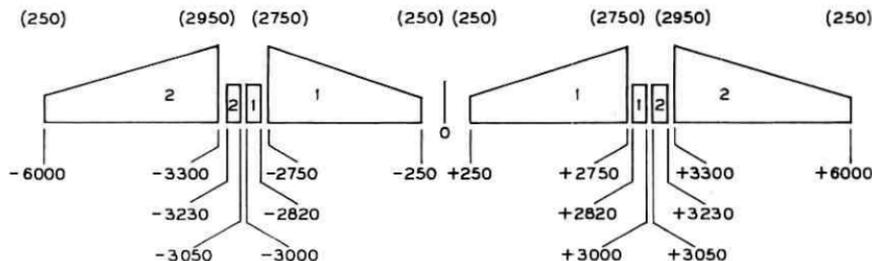
FIGURES IN ( ) ARE BASEBAND FREQUENCIES
OTHER FIGURES ARE FREQUENCIES WITH RESPECT TO RF CARRIER
ALL FREQUENCIES INDICATED ARE Hz

Fig. 5 — Four-channel operation.

ponent of the control signal may be preserved.* Finally, the narrow-band control signal is similar in many ways to narrow-band telegraph signals which have been transmitted successfully via HF radio for many years using FM or related methods. It is too early to predict the extent to which diversity would be useful or necessary in a CNL-type radiotelephone system, although it is widely used with HF radio telegraph systems.

A successful trial of experimental CNL terminals has been conducted between New York and Buenos Aires by the American Telephone and Telegraph Company with the cooperation of the foreign correspondent. Conventional terminals on one circuit of a regular four-channel system were replaced with CNL terminals. A large number of test and demonstration calls were made including some comparisons

---

* The control signal is derived by rectifying the speech output of the first compressor stage; it is therefore unipolar and has a dc component.

between the CNL circuit and a conventional circuit operating in the same radio system. Radio conditions on this path generally are fair to good, neglecting propagation outages. The participants in these tests were largely persons familiar with both cable and conventional radio overseas circuit performance. In the judgment of most of the participants, the CNL circuit quality approached that of a submarine cable circuit because of the compandor action and lack of VODAS impairments.

Following the above demonstrations in July, 1966, the CNL-equipped circuit began an extended period of commercial service. Traffic records indicate that the single CNL circuit carried one quarter of all calls on the New York-Buenos Aires route, which has a total of 12 circuits. The CNL circuit handled more calls than the next two most active conventional circuits combined. In the opinion of operating personnel, calls that would have encountered considerable impairments or operating difficulties with conventional terminals were handled without customer complaint. During periods of poor radio conditions, the CNL circuit was frequently "commercial" when some, if not all, of the conventional circuits were unusable.

Similar experimental terminals have been developed and tested by the British General Post Office (GPO) and French Postes Et Telecommunications (PTT). The British and American Administrations submitted reports of their work on the new type of terminal to the plenary meeting of the International Radio Consultative Committee (CCIR) in Oslo in the summer of 1966. This report included essential design parameters and operating conditions, and was submitted as an initial contribution toward adoption of international compatibility standards for the new system.

III. DESCRIPTION AND ANALYSIS OF IMPROVED SYSTEM

3.1 *System Application—Overall Block Diagram*

The block diagram Fig. 6 illustrates an arrangement of the experimental CNL equipment that was installed on a working radio circuit between Buenos Aires and New York. The equipment shown as shaded boxes is the same as is used with the present terminals. Other arrangements of signaling, privacy, channel shifters, and filters are possible; the best arrangements will depend on the characteristics of the specific equipment to be used on a given circuit.

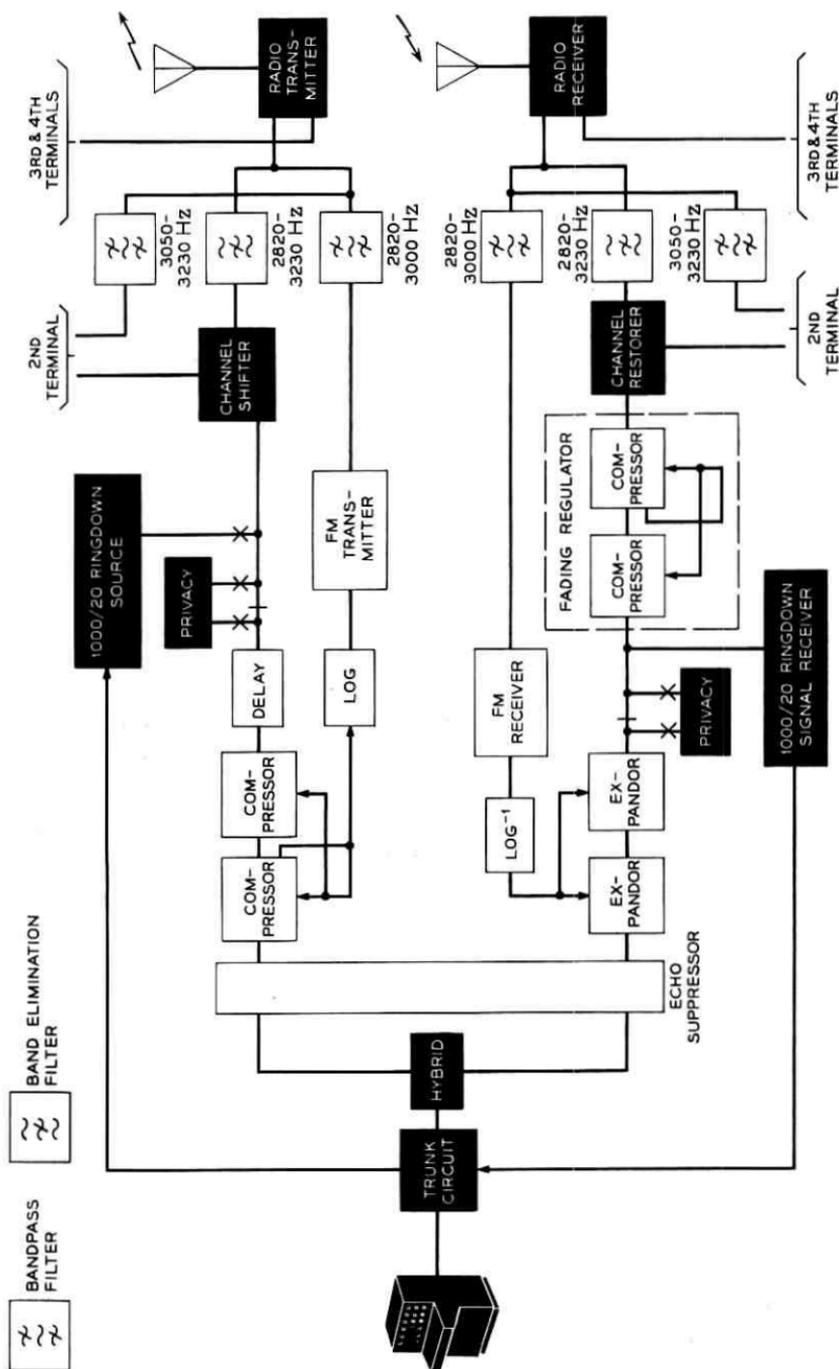One independent sideband (ISB) HF transmitter normally carries

Fig. 6 — System block diagram.

a maximum of four channels, two on each sideband. The particular arrangement of Fig. 6 was designed, among other reasons, to facilitate building up multichannel CNL systems. In such systems, the voice outputs of two separate terminals are combined in a channel shifter; a single band-elimination filter at the output of the shifter produces a slot to accommodate two control channels. The control subcarrier of each terminal is transmitted through a bandpass filter of the appropriate frequency; the subcarriers from two terminals are then combined with the output of the band-elimination filter. The combination of the two speech and two control channels is applied to one sideband of an ISB transmitter. The ISB transmitter has provision for two additional circuits on the other sideband, as shown in Fig. 5. Since only one circuit was equipped with CNL terminals in the trial installation, the second, third, and fourth terminals indicated on Fig. 6 were conventional VODAS terminals.

The filter and channel restorer arrangement of the receiving side is essentially a duplicate of the transmitting side, with a single band-elimination filter for one pair of speech channels and bandpass filters for each of two control channels. The attenuation required to control crosstalk between the speech and the control channels is provided by the combination of transmitting and receiving filters. It will be noted from Fig. 6 that the control channel was combined with the speech channel after the channel shifter and split before the channel restorer. This approach was used so that the characteristics of particular channel shifters and restorers would not affect the control channels.

Signaling for the trial installation was based on manual operation of the radio telephone circuit, with a ring-down signal of 1000 Hz modulated at 20 Hz inserted on the radio side of the terminal after the privacy. A 1000/20 signaling receiver, bridged after the fading regulator and before the privacy and expandor on the receive side, operated the appropriate switchboard equipment. Suitable trunk, monitoring, and control circuits similar to those found in VODAS terminals were provided.

The trial installation includes a flat delay unit to equalize the delay of the speech path with respect to the separate narrowband control path. (A discussion of delay equalization is given in Section 3.2.4.)

Terminals of the CNL type are inherently full-duplex and all equipment exclusive of the 4-wire terminating set, or hybrid, is arranged on a 4-wire basis. Fig. 6 shows connection to a 2-wire switchboard; 4-wire switchboards are also used. Privacy equipment must be 4-wire to preserve the two-way nature of the CNL principle. Either the inversion

or split-band type of privacy equipment may be used. In the split-band privacy, the speech is separated into five 550-Hz bands and transposed in frequency in any arrangement that provides reasonable privacy. With the CNL-type terminal, it may be necessary to restrict the possible transpositions to insure that the control channel slot is always taken out of the speech around 3000 Hz in order to minimize the effect of removing this slot from the speech band.

An echo suppressor is required on all long circuits because of the inherent propagation time. The receiving split-type echo suppressor used on the Buenos Aires-New York circuit has a speech gate only in the transmit leg; the gate blocks transmitted speech only when speech of greater magnitude is present in the receiving leg. This type of suppressor was preferred because of its break-in properties and also because it tends to reduce reradiation of echoes.

3.2 Analysis of Factors Affecting Performance

3.2.1 General Considerations

As mentioned previously, allocations of available bandwidth and transmitter power must be made in a CNL-type system, in which the compressed speech and the information to control the expandor are transmitted via separate channels. The objective in making this allocation is to obtain adequate margin against noise and interference in the control channel without encroaching substantially on the speech bandwidth and power. While the speech bandwidth is not critical, it is not a simple matter to determine what bandwidth should be allocated to the control signal. The procedure used in selecting the parameters of the experimental CNL system was to make a somewhat arbitrary allocation of bandwidth and power, then to investigate what further trade-offs could be made within these constraints.

The control channel bandwidth selected was 180 Hz. The net speech bandwidths obtained were as indicated on Fig. 5 and Table I. The impairment due to loss of about 200 Hz of high-frequency speech energy of the inboard channel is slight. There is no loss of speech bandwidth to the outboard channel because the control channel in this case lies in what was originally guard space (between 3 and 3.25 kHz).

In the experimental CNL system, the available power per channel was divided in the ratio 1:4, i.e., the FM subcarrier power was 6 dB below a 0-dBm0 voice-frequency test tone. The total load, including speech and subcarrier, was then only about 1 dB greater than the

speech power alone. It would be undesirable to put so much power in the subcarrier as to necessitate a reduction in the speech drive to the transmitter and a consequent reduction in signal-to-noise ratio at the receiver.

For further discussion of factors affecting performance of the improved system, attention is drawn to the fact that the transmitting compressor and receiving expandor together comprise a syllabic compandor, which is closely related to more conventional devices of this class. In these devices, as well as in the CNL system, the transmission gain (or loss) is varied at syllabic rates in order to gain advantage against noise and interference. The conventional syllabic compandor has a compression ratio of 2:1, and uses the residual speech envelope of the compressed signal to derive the information with which to control the expandor. Thus, it does not require a separate control channel, although there are circumstances where such an arrangement could be advantageous. For example, a control channel could improve tracking of such a compandor in the presence of high noise at the expandor input.

Speech can be transmitted through a conventional compandor with low distortion using no more bandwidth than that of the original speech as long as the rate of gain variation, or "speed," does not exceed syllabic rates. Hence, the bandwidth, within limits, is not a consideration in establishing the speed of a conventional compandor.

The following general requirements govern the speed of a syllabic compandor. (i) *Compressor speed:* a fast compressor is more effective than a slow compressor in raising weak speech syllables with respect to the noise before transmission. If this action is too abrupt, however, significant distortion is created. (ii) *Expandor speed:* the speed of the expandor must match that of the compressor. If it is too slow, the expandor may mutilate the initial parts of a syllable, or its loss may not be fully inserted during the gaps so that the noise reduction effect of the expandor suffers.

Within the range of syllabic rates, the speed of a conventional compandor may be varied over a fairly broad range and still satisfy the foregoing requirements. Furthermore, the compandor noise improvement, i.e., how well the compressor picks up weak speech before transmission and the expandor mutes the noise between syllables, is not highly sensitive to speed.

The use of a separate control channel to control the expandor in the CNL system introduces an additional factor governing speed of the

CNL compandor; otherwise, the criteria governing speed would be the same as a conventional compandor. The additional factor is the noise performance of the control channel, which is strongly dependent on compandor speed. The speed of the compandor in the CNL system may be regarded as an independent parameter upon which two main categories of system performance depend. These are: (i) the control channel noise performance and (ii) the compandor noise improvement. The first degrades with increasing speed and the second improves with increasing speed; therefore, it is necessary to select parameters such that a reasonable balance is achieved.

The principal elements of the CNL system affecting this balance can be represented by the model of Fig. 7. The control signal is derived from the output of the first compressor variolosser by rectifying a portion of the speech voltage at this point. The unfiltered output undergoes an initial smoothing, or prefiltering, in the network R_0C_0. It then branches into two paths, one feeding back through the low pass amplifier (R_1C_1) to control both variolossers of the compressor and the other feeding forward via the control channel and the low pass amplifier (R_2C_2) to control the variolossers of the expandor. R_0C_0 band limits the spectrum of the modulating signal at the input to the control channel. The primary role of R_2C_2 is that of a post-detection noise filter. In addition to their filtering functions, the networks R_0C_0 and R_2C_2 together establish the time response of the expandor. The FM channel itself can be made to have a negligible effect on the expandor response. A unique feature of this particular configuration is that, regardless of the overall speed of the compandor, the time response of the feedback path can be matched to the feed-forward path by simply making $R_1C_1 = R_2C_2$; dynamic tracking of the compressor and expandor is thus assured.*

R_1C_1 and R_2C_2 are single-pole feedback networks of identical operational amplifiers. The bandwidth of these elements, their time responses, and the equivalence between bandwidth and time response may be analyzed in straightforward manner. The 3-dB bandwidth frequency of an RC low-pass characteristic is

$$f = \frac{1}{2\pi \, RC} \, ,$$

---

* The representative spectra shown in Fig. 7 are illustrative of the significant bandwidths present at several points in the CNL terminal. Note that the noise bandwidth at the expandor variolosser is f_2 while the effective bandwidth of the control signal paths is less because of the influence of R_0C_0 in tandem with R_1C_1 or R_2C_2.
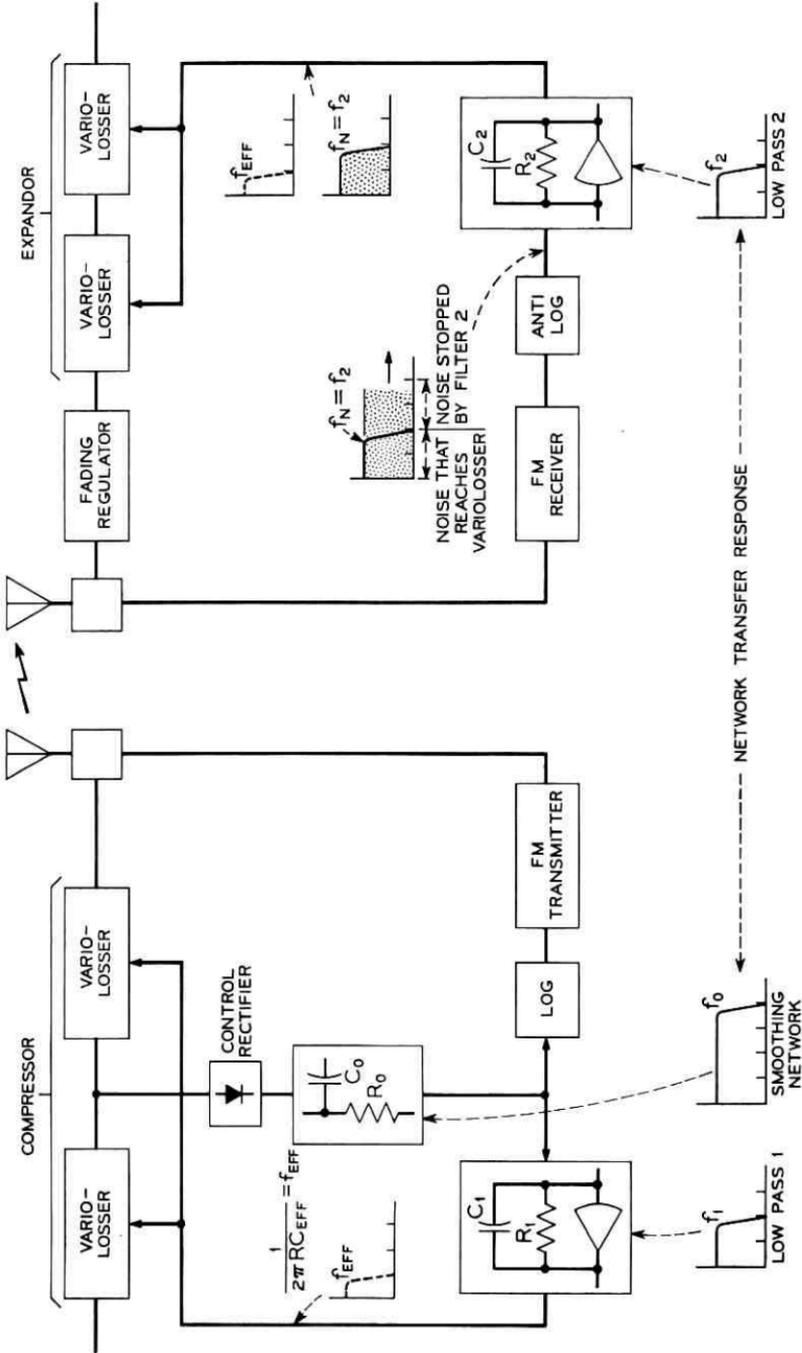
Fig. 7 — Time constant determining elements of CNL system.

from which a direct connection between speed of the compandor and noise performance of the control channel will now be apparent. That is, as the speed of the compandor is increased (by decreasing R_1C_1 and R_2C_2), f_2 widens to admit more noise to the control lead of the expander. The fluctuation of expander loss about its nominal value produced by random noise is, in effect, low-frequency amplitude modulation of the speech. This imparts a unique fluttering or "gritty" quality to the speech, which is a source of impairment over and above the effect of additive noise falling in the speech band. Interfering signals falling in the control channel also produce loss deviations, but of a less random nature. A limitation on the noise reaching the expander via the control channel imposes a corresponding limitation on $f_2 = 1/(2\pi R_2C_2)$. R_2C_2 strongly influences the resulting speed of the compandor, and therefore the compandor noise improvement. Both the effect of noise in the control channel and the compandor noise improvement are for the most part ultimately subjective. The task of finding an objective basis for evaluating these effects and analyzing them in detail is taken up in the following sections.

In Section 3.2.2 and related Appendix B, control channel noise performance is analyzed, a criterion of performance is established, and subjective limits are given.

In Section 3.2.3 the speed of the CNL compandor is discussed on a quantitative basis; relationships between time response and bandwidths are given. The related problem of dynamic tracking and the choice of operating ranges are dealt with in Section 3.2.4.

In Sections 3.2.5 and 3.2.6, respectively, the expander noise improvement and the efficacy of the complete compressor in loading the transmitter are discussed quantitatively, both as a function of compandor speed.

3.2.2 *Noise Performance of the Control Channel*

The performance of the control channel has two aspects: (*i*) "normal" operation when the subcarrier is above a certain noise threshold and (*ii*) the breaking threshold below which the channel breaks down.

The breaking threshold is reached in the FM channel of the CNL system when noise or interference at the discriminator unit is so strong relative to the subcarrier that the polarity of the signal is too often reversed. Under these conditions, impulsive noise currents are generated at the discriminator output; they drive the expander to its extreme loss or gain. As the threshold is approached, the impulsive noise

currents punch "holes" in the speech or produce annoying intervals of excessive loudness.

The threshold occurs with random noise when the subcarrier-to-noise ratio is in the vicinity of 10 dB on an rms basis. This ratio is a function of the line, or predetection, bandwidth and the subcarrier power. In the case of the experimental CNL system, the line bandwidth is 180 Hz, and the subcarrier power is 6 dB below the mean compressed speech power. Therefore, with the assumption that the noise is flat across the band, breaking occurs when the ratio of compressed speech to noise in the 2500-Hz speech band is about 5 dB.

The bandwidth occupied by the FM sidebands must not exceed the 180-Hz bandwidth allocated. It is a complex function of the amplitude and spectral distributions of the modulating signal and the frequency swing of the FM subcarrier. To further complicate matters, the modulating voltage is not symmetrically distributed and its spectrum is modified by the nonlinear LOG circuit preceding the frequency modulator. It was experimentally determined that the FM spectrum resulting from a control signal rolloff at 100 Hz (3 dB down) and a peak-to-peak swing of 120 Hz was in accord with the 180-Hz line bandwidth. The 100-Hz roll-off occurs in the control rectifier filter, represented symbolically by $R_0 C_0$ on Fig. 7. The 120-Hz swing divided by the expandor loss range (60 dB, as discussed in Section 3.2.4) establishes the deviation/loss ratio. It relates frequency errors of the FM subcarrier to the system loss error. These errors are due to noise and other mechanisms. The ratio enters into the calculations of Appendix B.

The magnitude of system loss fluctuations above the breaking threshold due to noise in the control channel has been calculated by means of an expression derived in Appendix B. The derivation assumes that the noise is random and flat across the band common to both the compressed speech and the subcarrier. Although a more usual situation in HF reception is a mixture of noise and tone-like interfering signals having a variety of characteristics together with selective fading, the simple model nevertheless gives useful results.

Equation (40) from Appendix B is

$$\text{loss fluctuation magnitude} \atop \text{(dB peak-to-peak)} = \frac{3(f_2)^{\frac{3}{2}}}{25\ S/N} \left[\frac{90}{f_2} - \tan^{-1}\left(\frac{90}{f_2}\right) \right]^{\frac{1}{2}}.$$

Fig. 8 is a plot of the loss fluctuation as a function of S/N with f_2 as a parameter. S/N is the ratio (numeric) of the speech voltage to the

noise voltage in the 2500-Hz speech band and is an indirect measure of the noise falling in the control channel. Expressing results in this manner makes it possible to correlate quality judgments based on the speech-to-noise ratio of the circuit with the degree of impairment caused by fluctuating system loss. By subjective tests, the approximate magnitudes were determined at which the loss fluctuations were judged (*i*) to have just noticeable impairment and (*ii*) to be so severe as to render the circuit uncommercial. These magnitudes are indicated on Fig. 8 as cross-hatched bands.

The abcissa of Fig. 8 refers to the S/N at the input to the receiving terminal. In a circuit equipped with CNL-type terminals, a S/N of about 10 dB would be judged uncommercial and 15 dB would be judged poor but usable. By the standards normally applied to HF radiotele-
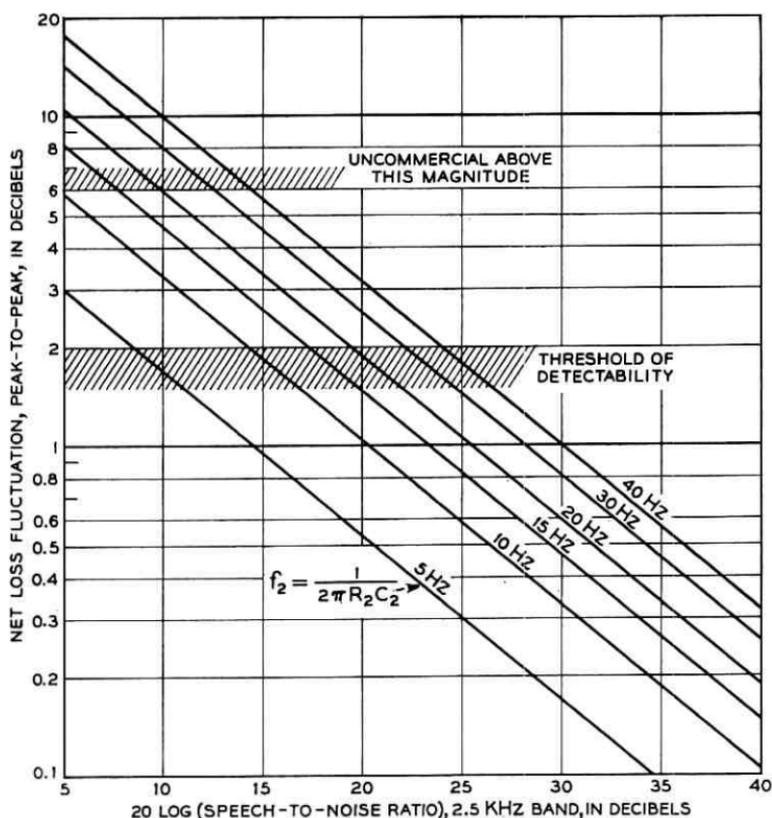

Fig. 8 — Net loss fluctuation as a function of speech-to-noise ratio with control channel bandwidth as a parameter.

phone circuits, 25 dB would be considered good. Fig. 8 shows that if f_2 becomes large, the effect of noise in the control channel would be detectable even under otherwise good conditions. On the other hand, an f_2 of 5 Hz or less provides more margin than is necessary. With an f_2 of 20 Hz, when the S/N drops to 10 dB, the loss fluctuations reach an uncommercial magnitude. This same S/N would be judged uncommercial with a noise-free control channel. Thus, with an f_2 of 20 Hz, the limit of commercial quality would be reached more or less simultaneously due to noise mixed with the speech and noise falling in the control channel. The value of f_2 adopted for the experimental CNL system was 15 Hz. It tends to be conservative in respect to loss fluctuation at the expense of compandor speed. The other speed-dependent performance factors discussed in Sections 3.2.5 and 3.2.6 change slowly with speed.

The CNL system employs nonlinear signal processing, or "instantaneous companding,"[4, 9] within the control channel. This is the function of the LOG and ANTILOG circuits shown in Fig. 7. Such processing reduces the effect of noise on weak control signals. With weak talkers or during idle circuit conditions the control channel can tolerate more noise for a given requirement on expandor loss fluctuations if instantaneous companding is used. While instantaneous companding increases the bandwidth of the analog control signal, a net improvement in control channel performance results in the CNL application. The response of the LOG circuit itself is plotted on normalized linear scales on Fig. 9. The ANTILOG characteristic must be sufficiently complementary to the LOG characteristic to insure overall linearity of the control channel.

3.2.3   *Time Constants*

Before considering CNL compandor speeds, it is necessary to have a precise definition of time constants. Also, an accepted definition will permit comparison of the CNL with other types of compandors, in particular, results obtained by other organizations (British GPO, etc.) currently working on CNL-type systems. The CCITT* has developed a definition for message compandor time constants that has found wide acceptance. This definition has the virtue of being readily implemented with simple measuring techniques and is suitable for specifying time constants in the CNL compandors described herein.

The definition of attack and recovery time constants of a compandor

---

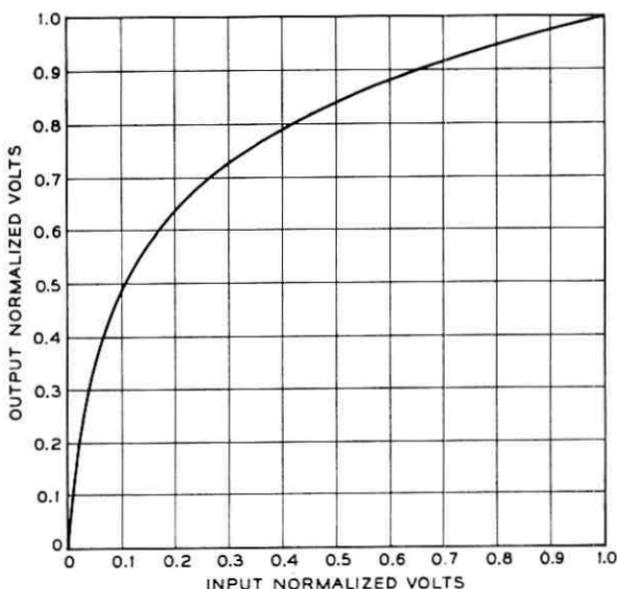* International Telegraph and Telephone Consultative Committee.

Fig. 9 — Log circuit input-output characteristic.

involves a measurement of the transient response of a 2:1 compressor. In the CNL system, the first stage is identical to the conventional compressor and the second variolosser is slaved to the first. This enables the CNL compressor time constants to be measured using the first compressor in accordance with the methods specified by the CCITT. The attack time is measured by increasing the level of the input tone by 12 dB from −16 dBm0 to −4 dBm0 and measuring the time interval between the occurrence of the step and the time when the output envelope reaches 1.5 times the final steady-state value. The recovery time is the time interval between the occurrence of a downward step of 12 dB (from −4 dBm0 to −16 dBm0) and the time when the output waveform reaches 0.75 times the final steady-state value.

For message compandors used on wire and radio relay circuits of moderate length, the recommended CCITT values are:[10,11]

$$\text{Attack time} = 3 \pm 2 \text{ msec}$$

$$\text{Recovery time} = 13.5 \pm 6.5 \text{ msec.}$$

In order to relate measured attack and recovery times of a compandor to circuit elements as an aid to design, R. O. Carter[12] analyzed a simple model of a compressor (Fig. 10) and developed formulas

for calculating attack and recovery times. In his model, the control rectifier consisted of a peak detector (diode) and a single RC smoothing network. He showed that the CCITT attack and recovery times were given in terms of the product RC (which has the dimensions of time) as follows:

$$\text{Attack time} \quad = 0.15 \text{ RC} \tag{1}$$

$$\text{Recovery time} = 0.675 \text{ RC.} \tag{2}$$

The value of RC, since it is common to both the attack and recovery time constants, provides a single measure of the speed of the compres-
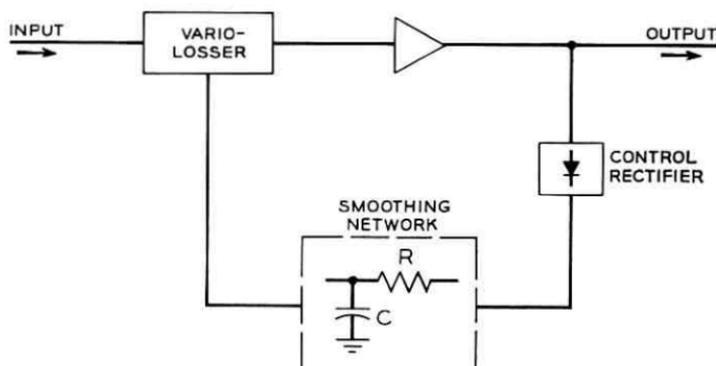
Fig. 10 — Simple model of 2:1 compressor.

sor. The bandwidth of the control signal into the variolosser is also a function of RC:

$$f = 1/(2\pi \text{ RC}). * \tag{3}$$

Using relations (1) and (2), the CCITT recommended time constants of 3-msec attack and 13.5-msec recovery correspond to a speed of RC = 20 msec. The objective of the following analysis is to examine the CNL system model (Fig. 7) and find an $RC_{\text{effective}}$ for the somewhat more complicated CNL compandor.

The transient response for the compressor configuration with R_0C_0 and R_1C_1 networks was calculated approximately by divorcing the

---

* This relationship permits a ready transformation between factors best considered on a time (or speed) basis and factors best considered on a frequency (or bandwidth) basis. It is exact only when a simple RC network is used in the model. The bandwidth, f, corresponds to the frequency at which the control path gain is 3 dB down from the value at dc.

time-constant determining components from the nonlinear feedback circuit and solving their response to a step input. A similar step response was determined for the simple RC network of Carter's model for the time $t = 0.150$ RC, the CCITT attack time. The two responses were made equal and the attack time for the CNL compressor, t_A, was obtained by solution of

$$1 - \exp\left(-\frac{0.150 \text{ RC}}{\text{RC}}\right) = 0.139$$

$$= 1 + \frac{R_0C_0 \exp\left(-\frac{1}{R_0C_0} t_A\right) - R_1C_1 \exp\left(-\frac{1}{R_1C_1} t_A\right)}{R_1C_1 - R_0C_0}. \quad (4)$$

The use of the step response of the networks alone to obtain an engineering estimate of the attack time was justified because the input to the speed-determining element under CCITT test conditions is essentially a step function in the time scales considered here. The speed of the CNL compressor (given by an $RC_{\text{effective}}$) is determined from this calculated attack time through use of (1). Similarly, the recovery time was obtained by solution of the following equation for t_R:

$$\exp\left(-\frac{0.675 \text{ RC}}{\text{RC}}\right) = 0.509$$

$$= \frac{R_1C_1 \exp\left(-\frac{1}{R_1C_1} t_R\right) - R_0C_0 \exp\left(-\frac{1}{R_0C_0} t_R\right)}{R_1C_1 - R_0C_0}. \quad (5)$$

This calculated recovery time can also be used to obtain $RC_{\text{effective}}$ using (2). The validity of expressions (4) and (5) was confirmed by direct measurement of attack and recovery times. Expressions (4) and (5) reduce to the form of (1) and (2) when either RC is much greater than the other. When the two time constants are equal,

$$t_A(\text{for } R_1C_1 = R_0C_0) = 0.652 \ R_0C_0 \quad (6)$$

$$t_R(\text{for } R_1C_1 = R_0C_0) = 1.649 \ R_0C_0 , \quad (7)$$

an attack time increase of over 4 times and a recovery time increase of almost $2\frac{1}{2}$ times over the values of a single R_0C_0, rather than a twofold increase as might be surmised at first glance. Plots of both the attack and recovery equations (4) and (5) for values of R_0C_0 and R_1C_1 in the range of interest are given in Figs. 11(a) and 11(b). These curves illustrate how the values of R_0C_0 can be adjusted to yield a desired attack or recovery time when R_1C_1 is specified.
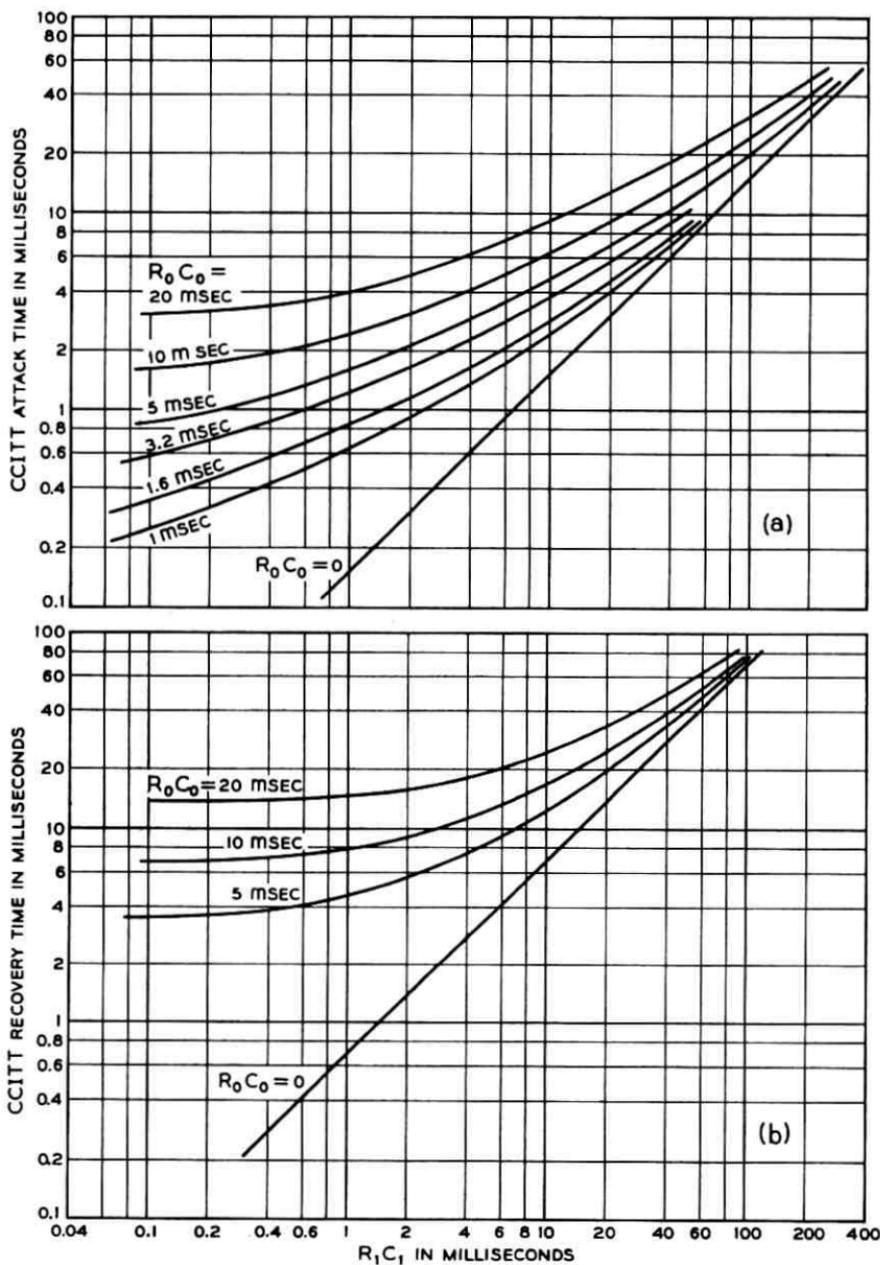
Fig. 11 — (a) Calculated attack time; (b) calculated recovery time.

In Section 3.2.2, the frequency determined by R_2C_2 was given as 15 Hz; substituting this value in (3) gives an $R_1C_1 = R_2C_2$ of 10.6 msec. The objective in selecting R_0C_0 was to yield compressor attack and recovery times recommended by the CCITT. This resulted in a choice of

$$R_0C_0 \text{(attack)} = 1.6 \text{ msec}$$

$$R_0C_0 \text{(recovery)} = 5.0 \text{ msec.}^*$$

From Figs. 11(a) and 11(b) these values give overall attack and recovery times as follows:

$$\text{Attack time} = 3 \text{ msec}$$

$$\text{Recovery time} = 13.5 \text{ msec.}$$

The corresponding $RC_{\text{effective}}$ given by (1) and (2) was 20 msec.

Fig. 12 shows the relationship between overall speed ($RC_{\text{effective}}$) and the characteristics of the lowpass networks (R_1C_1 and R_2C_2).

3.2.4 Dynamic Tracking and Operating Ranges

As mentioned in Section 3.2.1, the control paths from control rectifier to variolossers pass through essentially identical smoothing networks. Therefore, the input control signals to the transmitting compressor variolossers and receiving expandor variolossers have the same transient characteristic. Matching the control signals into the variolossers insures that the expandor loss matches the compressor gain on a dynamic basis. However, an imperfect reconstruction of the speech waveform at the system output can occur if:

(i) the transmission time delay of the control signal differs from that of the corresponding speech signal, causing a mismatch between the instantaneous expandor loss and the envelope of the speech signal, and/or

(ii) the fade regulator is too fast, thus interpreting the residual amplitude variations in the output of the transmitting complete-compressor as fades and introducing further compression (without corresponding expansion).

These effects result in speech that sounds distorted and possible circuit instability if round-trip propagation delay is short.

---

* The two values of R_0C_0 require that the simple model of Fig. 7 be modified to provide a control rectifier with different charge and discharge times.
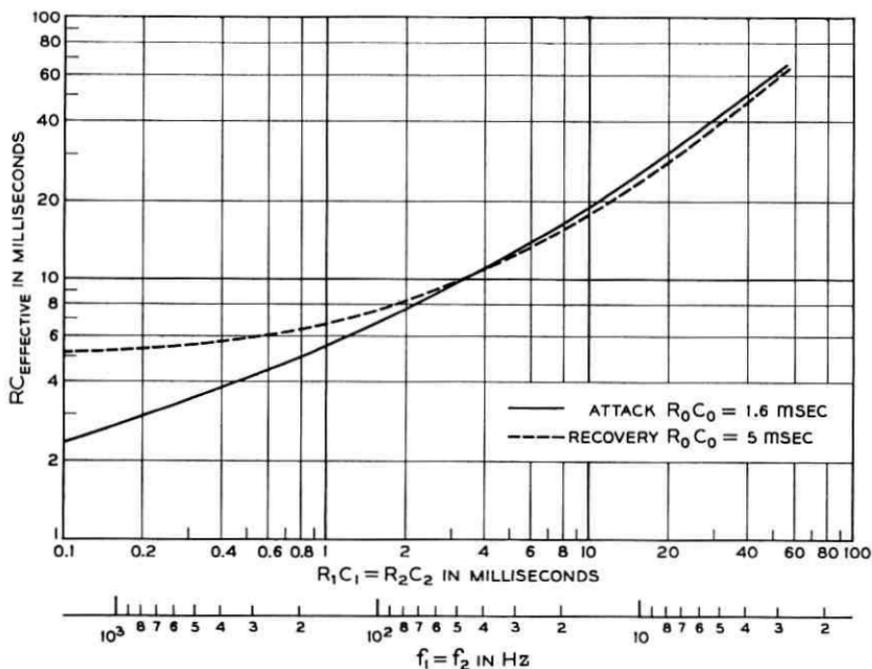
Fig. 12 — Overall speed versus low pass network time constant.

It has been found that a fade regulator RC time constant approximately 4 times $RC_{\text{effective}}$ reduces the additional compression distortion to acceptable levels without affecting performance during fading.

Flat voice-frequency delay equalization in the speech or pilot channel is used to correct the transmission time differences due to filter and privacy flat delays. For speech service, it is desirable to reduce the time difference to no more than 5 to 10 msec. For signaling and similar pulse transmission where pulses are amplitude modulated on a signaling tone, the distortion due to time differences is more severe; a 1- or 2-msec difference produces noticeable changes in pulse shape. Equalization requirements for actual signaling systems of this type would depend upon system sensitivity to pulse shape.

The need for time delay equalization is reduced if the CNL compandor is slow. If $R_1 C_1 \equiv R_2 C_2$ is large compared with the time difference, the rate of change of control current into the variolossers is slow in any interval corresponding to the transmission time difference. The expandor loss mistracking errors will therefore be small since the slope of the loss curve as a function of time is low.

The operating ranges of the variable gain elements of the CNL system are interdependent since the sum of the variable gains and losses, including medium variations, results in an approximately constant value under both dynamic and static conditions. This is essential for system stability and tracking. The transmitting compressor range is 40 dB; the fading regulator downward range is 20 dB and the upward range is 10 dB; the expandor range is 60 dB, and is the sum of the transmitting compressor range and the fading regulator downward range (see Fig. 13). The operating ranges chosen were a reasonable compromise based upon expected speech volume and fading ranges, as well as practical considerations.

3.2.5 Syllabic Compandor Quieting Versus Speed

The speed of the compressor affects the degree of compandor quieting of system noise heard by the subscriber. A limited subjective com-
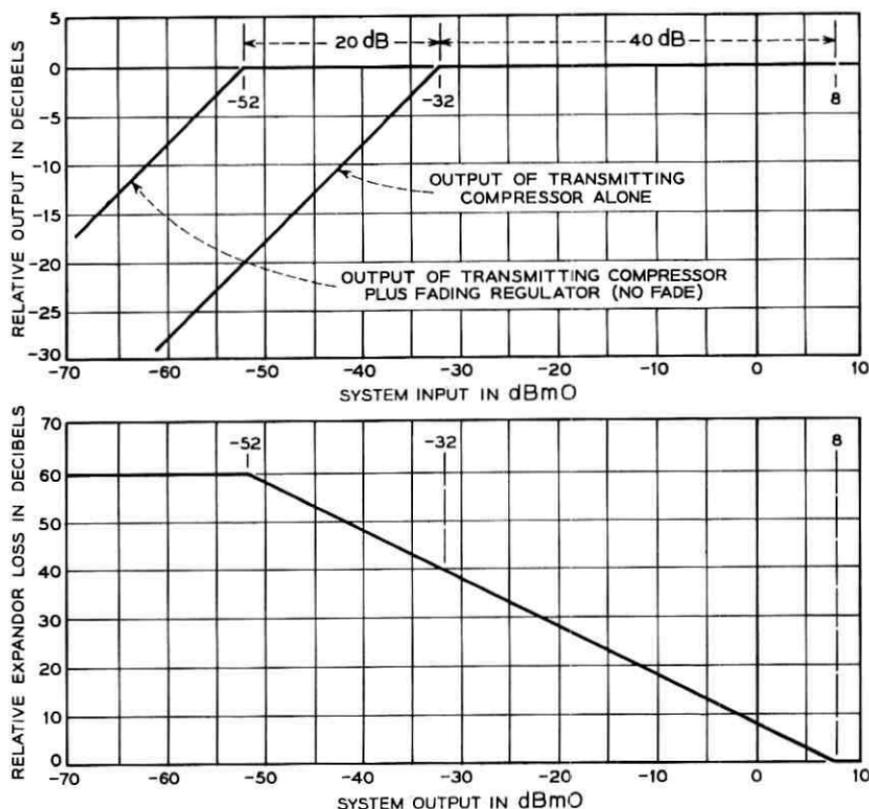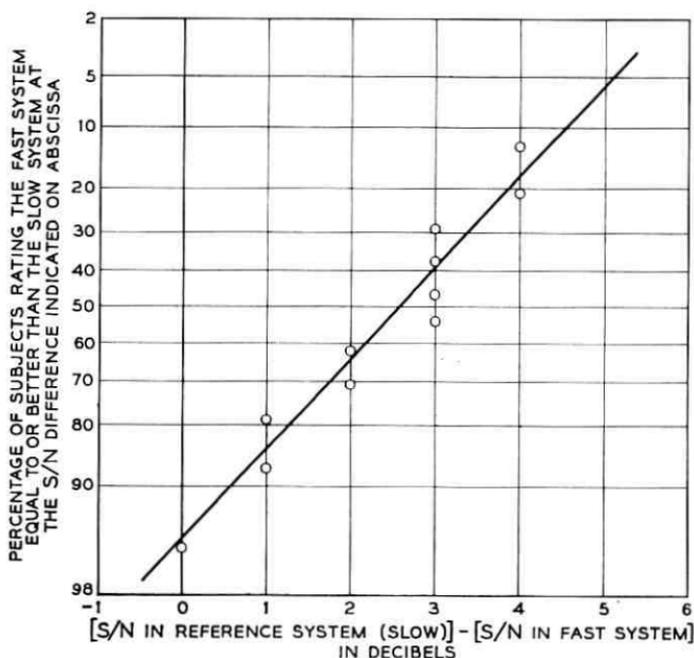

Fig. 13 — Static operating ranges.

Fig. 14 — Distribution of S/N differences for equal subject preference.

parison of a fast and slow CNL compandor (RC = 5 msec; RC = 75 msec) was made to determine the extent of the change in noise quieting solely related to speed. To this end, the FM control channel was made perfect (replaced by a dc connection), noise was added to the speech path, and recordings were made. The slow system was taken as the reference system, and an approximate 10-dB compressed speech-to-noise ratio maintained. The fast system was recorded with a variable speech-to-noise ratio over a range of 6 to 12 dB. Direct comparisons were made between the two sets of recordings by 12 test subjects. Each subject determined the value of variable noise (in the fast system) which in his judgment made the fast and slow systems equivalent in quality. The reference value of 10 dB was chosen because it is the approximate point at which the CNL system quieting would be most effective on a circuit that was noisy but still usable. Fig. 14 shows the resulting distribution of S/N differences between fast and slow CNL compandors where the subjects indicated equal preference. For the median subject, the fast system gave a $2\frac{1}{2}$-dB improvement. Since the speed-related difference in quieting is strongly affected by the amount of noise present, the $2\frac{1}{2}$-dB advantage will decrease when the S/N improves from the 10-dB test value.
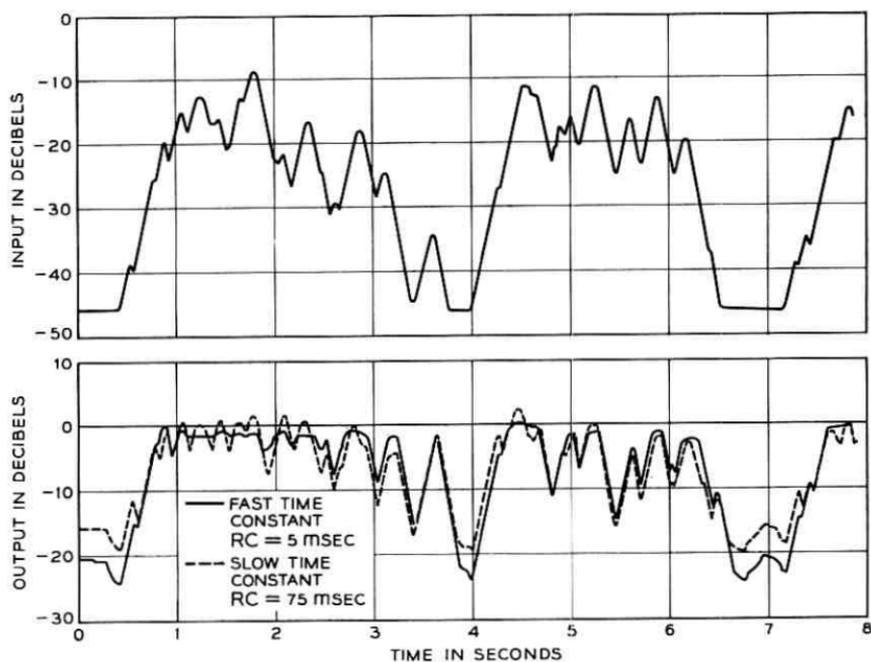
Fig. 15 — Chart record of compressor input and output for fast and slow time constants; recorder integrating time—approximately 100 milliseconds.

The conclusions drawn from the tests were that changes in compandor speed in the range of RC = 5 to 75 msec do not produce a sufficient change in the overall noise improvement to warrant consideration as a major factor in system "trade offs." Since the total noise improvement for the average talker is 20 to 30 dB, the change in performance with speed is of minor consequence.

3.2.6 *Transmitter Loading*

The complete compressor serves to reduce the variations in the speech signal into the radio transmitter. However, residual variations do remain and these variations become greater as the speed of the compandor is reduced. The regulated speech output from the transmitting compressor was examined to determine the extent of the regulation improvement as a result of increasing the speed. The method used in this examination was based upon recognition of current CCIR recommendations.[13] In these recommendations the *mean power* of the speech signal with smoothly read text is used as a criterion of ISB trans-

mitter loading. The *mean power* is defined: "The power supplied to
the antenna transmission line by a transmitter during normal opera-
tion, averaged over a time sufficiently long compared with the period
of the lowest frequency encountered in the modulation. A time of 0.1
sec during which the mean power is greatest will normally be selected."
For purposes of comparative measurements, the *mean power* was in-
terpreted as that value obtained by a running measurement of the
speech signal power with about a 0.1-sec integrating time and by se-
lection of the value of signal power when the running measurement is
greatest, i.e., at the crest of the speech envelope.

CNL output signals were examined using a power measuring chart
recorder with about 100-msec averaging time to determine how the
mean power and the distribution of signal peaks changed as the speed
of the transmitting compressor was adjusted to a very fast time con-
stant (RC = 5 msec) and then to a very slow time constant (RC =
75 msec). The input speech and the regulated speech output from the
compressor were recorded on a strip chart (see Fig. 15) for each time
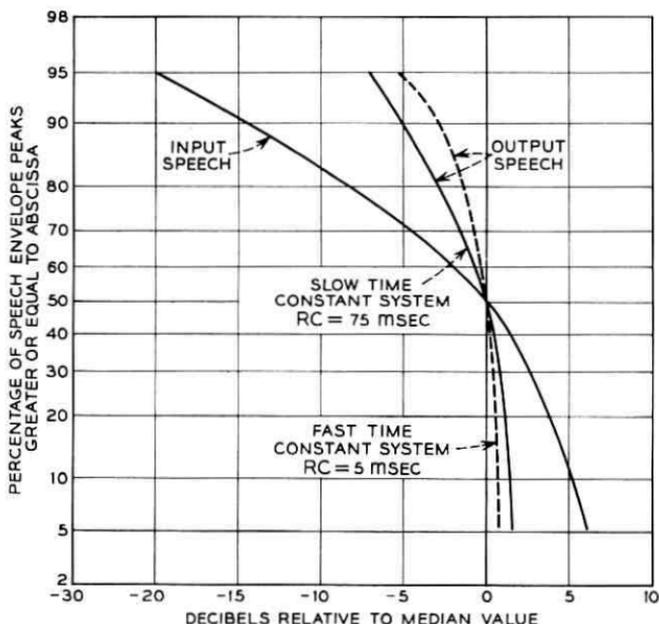constant extreme. The strip chart shows the power on a dB scale,

Fig. 16 — Input and output speech envelope peak distribution.

where an output of 0 dB corresponds to the steady state output of the compressor with an input 0 dBm0 1000-Hz tone.

When speech is present, the regulated output signals with both fast and slow time constants are no more than 5 dB apart. More importantly, the slow compressor (with more overshoots expected due to more sluggish regulation) has a mean power that is at most 2 dB higher than the fast compressor. Following accepted practice, the level into the transmitter for RC = 75 msec would then have to be dropped only 2 dB compared to the level for RC = 5 msec. If the peak values for each output are plotted to give a probability distribution (see Fig. 16), the slight difference between the outputs as a function of compressor speed is evident. In Fig. 16 the spread of the speech envelope peak distribution of the input speech is 26 dB. This spread of 26 dB is reduced to 9 dB at the output of the slow compressor and to 6 dB at the output of the fast compressor.

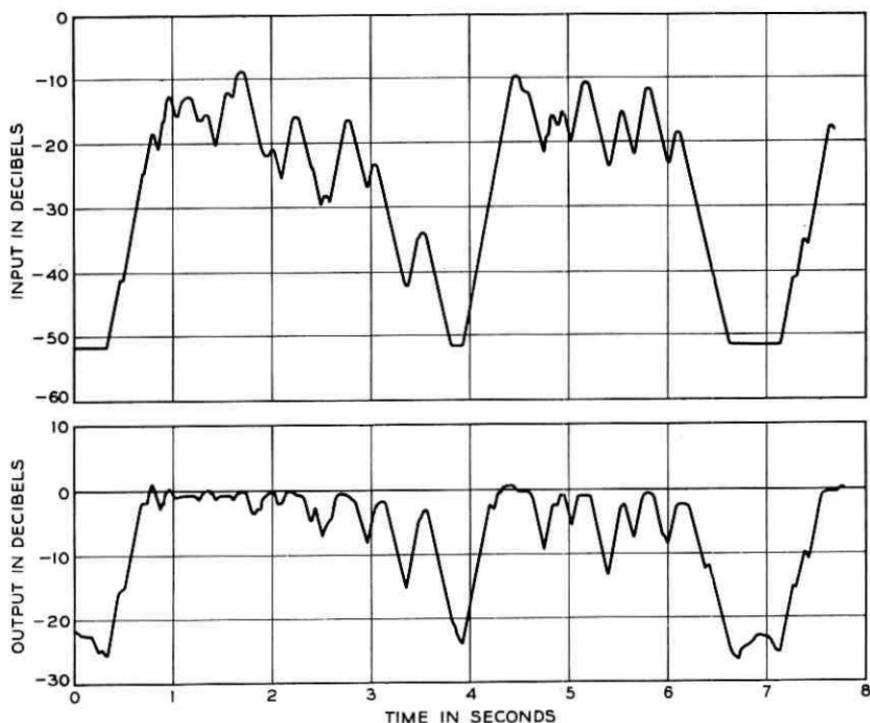Based on the above results, it was concluded that any otherwise

Fig. 17 — Chart record of compressor input and output—RC = 20 milliseconds; recorder integrating time—approximately 100 milliseconds.
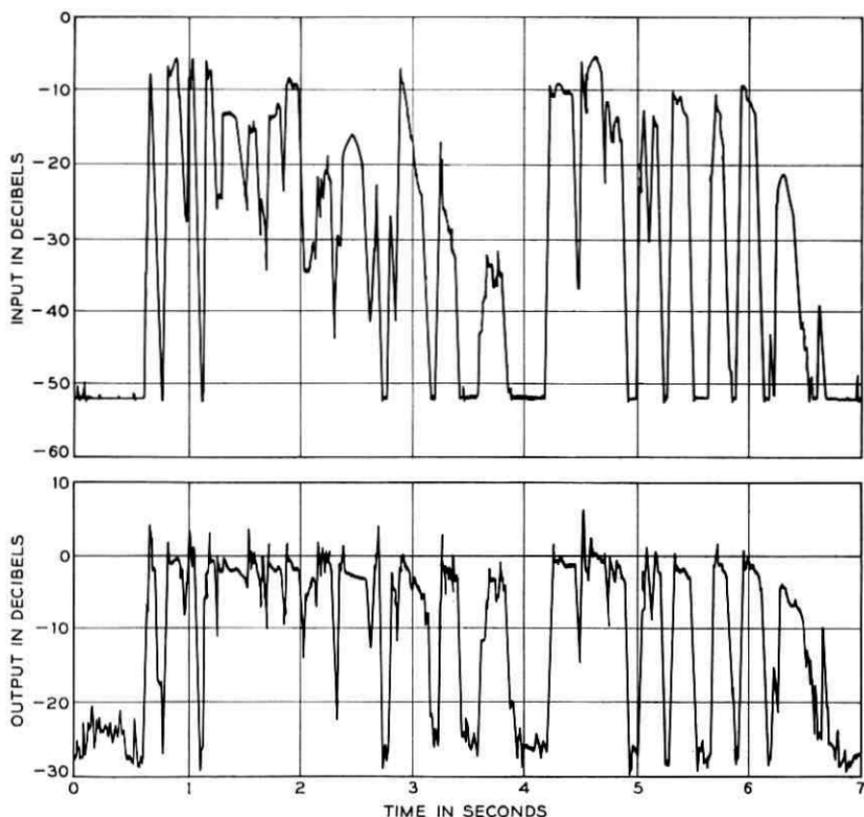
Fig. 18 — Chart record of compressor input and output—RC = 20 milliseconds; recorder integrating time—approximately 10 milliseconds.

proper RC time constant in the range 5 to 75 msec would produce nearly equivalent transmitter loading. The value of $RC_{effective} = 20$ msec used in the trial equipment falls within the above limits and, in fact, produces a load that has almost as little variation as the fast $RC = 5$ msec compandor. Fig. 17 shows an example of the input and output waveforms for the $RC_{effective} = 20$-msec system.

To verify that the accepted 100-msec averaging time used in this investigation has not obscured any very short term effects, an additional chart recording was made of system input and output with a 10-msec integrating time. This record, a portion of which is reproduced in Fig. 18, shows that while some overshoots remain, their magnitude and duration are not expected to produce unwanted effects in the radio transmitter.

IV. ACKNOWLEDGMENT

The authors are grateful for the contributions of many persons toward successful implementation and trial of the experimental CNL system, particularly those of Mr. A. E. Donkin.

APPENDIX A

Gain Relations in Single- and Two-Stage Compandors

Application of conventional compandors[14] to conventional wire, cable, and radio relay circuits produces level characteristics such as those shown in Fig. 19. At the compressor, low-level inputs are amplified
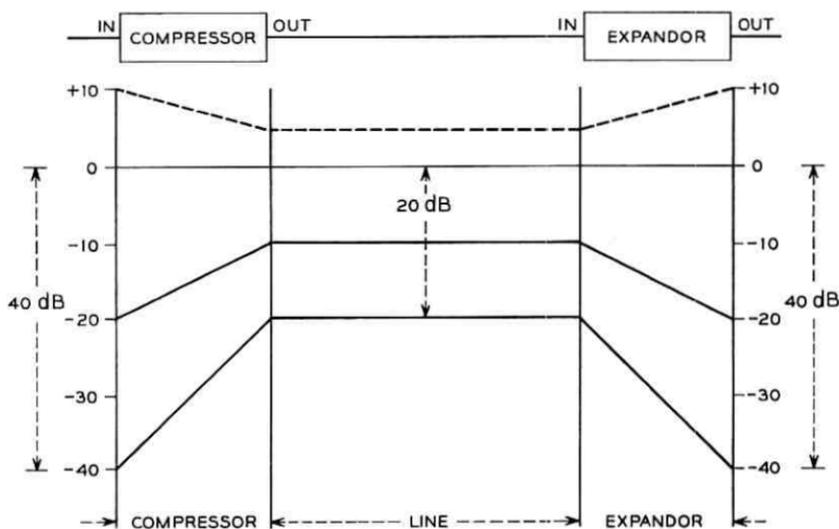

Fig. 19 — Example of 2:1 compandor action.

more than high-level inputs; correspondingly, at the expandor, low-level outputs result from more attenuation than that given to high-level outputs. (Note: certain very high inputs are attenuated to reduce overload.) As shown in the example of Fig. 19, conventional syllabic compandors halve the volume range between compressor and expandor; thus, a 0 to −40-dB range becomes a 0 to −20-dB range in the transmission portion. These compandors are known as 2:1 compandors;* the 2:1 characteristic is easily obtained as will be shown below.

The compressor portion of the compandor consists of only four simple elements (Fig. 20). The variolosser is a circuit configuration where

---

* In CCITT documents, this is stated as a compression ratio of 2.

the attenuation ratio in the speech path is *inversely* proportional to the unidirectional control current. The variolosser equation is

$$e_1 = k_1 \frac{e_{\text{in}}}{i_c} , \tag{8}$$

where e_{in} is the amplitude of the input envelope. The control rectifier acts as an envelope detector, producing control signals with a spectrum extending from 0 Hz to approximately 100 Hz.* The feedback path directs the control signal to control the variolosser attenuation. The 2:1 compression effect results because the control signal developed
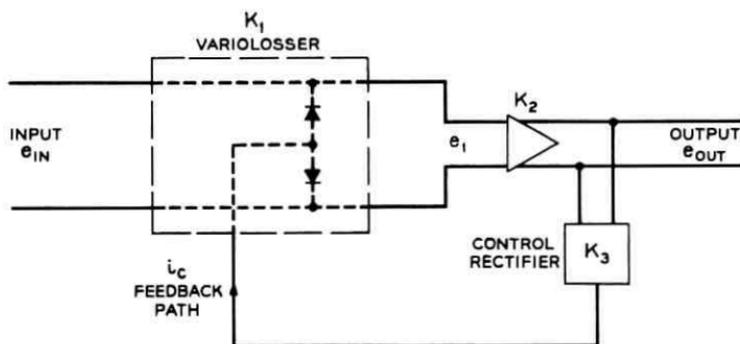

Fig. 20 — Elements of 2:1 compressor.

from the output is fed back to control a gain adjusting variolosser, which in turn changes the output. The fixed gain amplifier is added to the speech path for level adjusting purposes. The amplitude of the output envelope is

$$e_{\text{out}} = k_2 e_1 . \tag{9}$$

The control current is related to the output envelope by

$$i_c = k_3 e_{\text{out}} \tag{10}$$

from which

$$e_{\text{out}} = \frac{k_1 k_2 e_{\text{in}}}{i_c} \tag{11}$$

$$e_{\text{out}} = \frac{k_1 k_2 e_{\text{in}}}{k_3 e_{\text{out}}} \tag{12}$$

$$e_{\text{out}}^2 = \frac{k_1 k_2}{k_3} e_{\text{in}} . \tag{13}$$

* The upper cutoff frequency is not a clearly defined point, depending on the characteristics of the speech signal and the type of rectifier smoothing.

Taking logarithms

$$20 \log e_{\text{out}} = 10 \log e_{\text{in}} + 10 \log \frac{k_1 k_2}{k_3} \tag{14}$$

$$10 \log e_{\text{out}} = \tfrac{1}{2}(10 \log e_{\text{in}}) + K \tag{15}$$

$$E_{\text{out}}(\text{dB}) = \tfrac{1}{2}E_{\text{in}}(\text{dB}) + K'. \tag{16}$$

The output volume changes 1 dB for every 2-dB change in input volume. The foregoing analysis is based upon steady state conditions. The transient response (or time constant) of the compressor is determined by an RC network in the control rectifier. The transient response of compandors and the CNL system is discussed in the analysis section of the body of the paper.

The expandor portion of the compandor consists of five simple elements (Fig. 21). In this case, the variolosser is a circuit configuration where the attenuation ratio in the speech path is *directly* proportional to the unidirectional control current. The control rectifier acts as an envelope detector in the same way as in the compressor, but in this case, the control signal is applied in a forward acting manner rather than in a feedback manner. The variolosser equations

$$e_1 = k_6 i_e e_{\text{in}} \tag{17}$$

$$e_{\text{out}} = k_5 e_1 = k_5 k_6 i_e e_{\text{in}} \tag{18}$$

are then manipulated as in the case of the compressor and the following result is obtained:

$$e_{\text{out}} = k_3 k_4 k_5 k_6 e_{\text{in}}^2 \tag{19}$$

$$E_{\text{out}}(\text{dB}) = 2E_{\text{in}}(\text{dB}) + K''. \tag{20}$$
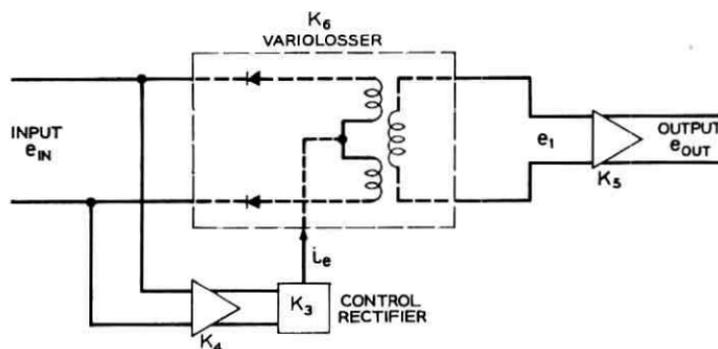

Fig. 21 — Elements of 2:1 expandor.

The output of the expandor changes 2 dB for every dB change in the input, thus undoing the effect of the compressor.

Idealized input-output characteristics of the conventional compandor are shown in Fig. 22. When installed at each end of a circuit, the output will be a true replica of the input if the circuit is distortionless. Since the expandor control signal is derived from the output of the transmission path at the input to the expandor, examination of the
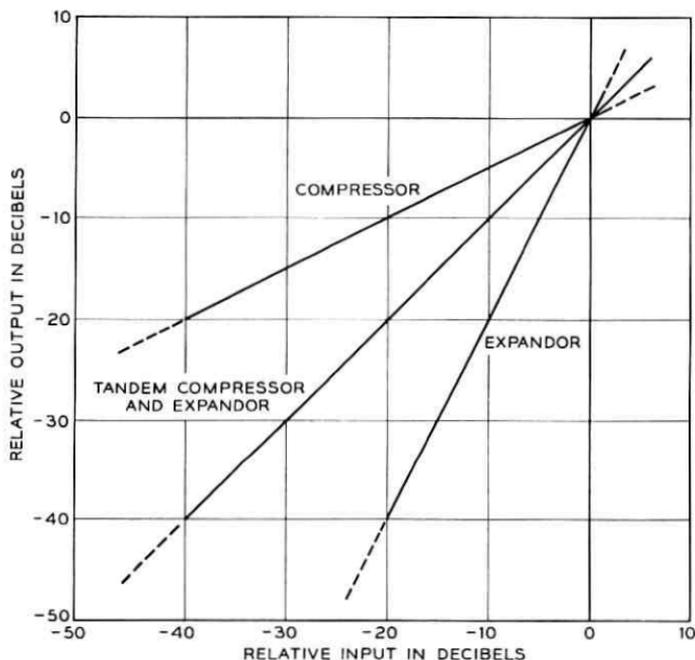

Fig. 22 — Input/output characteristic of 2:1 compandor.

characteristics of Fig. 22 verifies that gain variations (with time) of the medium will be magnified in dB by a factor of two. For this reason, conventional compandors have not found application in HF radio service.*

The CNL compandor arrangement (Fig. 23) solves the problem of net loss variations while providing essentially constant output from the transmitting compressor. Constant output is obtained by adding a second slave variolosser controlled by a replica of the regular con-

---

* Compandors were used for the first transatlantic radiotelephone circuit[15] (long wave) where the net loss was relatively stable.

o	0	o	0	o		o	0	o	0	o	0	o	0	o
-40	↑20	-20	↑20	o	NO FADE	o	0	o	0	o	↑20	-20	↑20	-40
o	0	o	0	o		-10	↑5	-5	↑5	o	0	o	0	o
-40	↑20	-20	↑20	o	10 dB FADE	-10	↑5	-5	↑5	o	↑20	-20	↑20	-40

FADING REGULATOR

COMP 1 — COMP 2 —→ antenna —→ antenna — COMP 3 — COMP 4 — EXP 1 — EXP 2

FM TRANS-MITTER   FM RECEIVER

RELATIVE LEVELS SHOWN IN LIGHT FACE TYPE
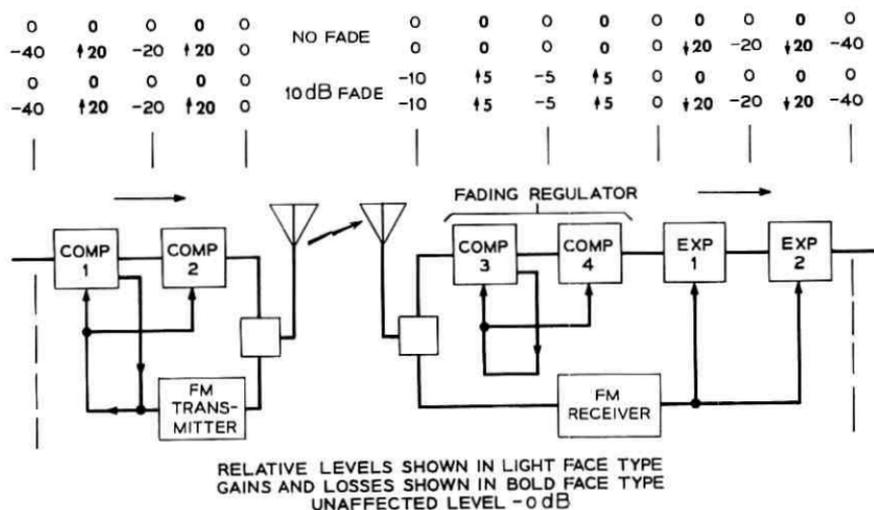GAINS AND LOSSES SHOWN IN BOLD FACE TYPE
UNAFFECTED LEVEL – o dB

Fig. 23 — Example of CNL system gains and losses.

trol signal present in the feedback path of the normal first stage compressor. Examination of Fig. 22 shows that for any input in the range of compression, a conventional compressor provides a variolosser control signal so as to insert an amount of gain equal to one-half the difference (in dB) between the "unaffected level" (or input level at which no gain or loss occurs in the compressor or expandor) and the input signal level. The slave variolosser gain is also controlled to be one-half the difference between the unaffected level and the input signal level, so the sum of the gains of two stages is sufficient to raise the level of the input signal to the unaffected level for any value of input signal in the range of compression. An example showing the CNL two-stage compressor gains and levels for two inputs is given in Fig. 23.

At the receiving terminal, expandors are used to restore the speech distribution. Two normal expandor variolossers are used, since each can accomplish the inverse of the action of one of the compressor variolossers. However, the local control arrangements at the expandors can not obtain control information from the input signal (which is nominally constant for all speech *in the absence of fading*). Reexamining the compressor and expandor equations

$$\text{compressor gain} = \frac{e_{\text{out}}}{e_{\text{in}}} = \frac{k_1 k_2}{i_c} \qquad (21)$$

$$\text{expandor gain} = \frac{e_{\text{out}}}{e_{\text{in}}} = k_5 k_6 i_e . \qquad (22)$$

Thus, if a compressor and expandor are in tandem, their net gain is constant if

$$i_c \equiv i_e . \qquad (23)$$

A suitable separate control channel, which supplies each expandor with an i_e that is a replica of i_c, assures this identity.

To maintain a constant net loss when there is fading in the transmission path, the CNL system uses a fading regulator that is similar to the transmitting compressor, except that the fading regulator has no corresponding expandor. The fading regulator action is based on the recognition of the requirement that the output of the radio receiver be constant since the input to the radio transmitter is controlled to a constant value by the transmitting compressor. If the output of the radio receiver is less than the nominally constant value, it is assumed that a fade has occurred.

The fading regulator inserts sufficient gain to raise the output to the nominal nonfade value required by the expandor. Fig. 23 shows the gains and losses with a test tone input for an example without fading and for an example with an assumed flat fade in the medium. In this simple model the overall effect is to provide a nominally constant overall net loss even with a time varying transmission medium.

APPENDIX B

Net Loss Fluctuation Caused by Noise in Control Channel

The effect of noise in the control channel can be obtained by use of the model shown in Fig. 24 (not to scale). A flat band of random noise extends across the speech and subcarrier bands. The bandwidth of the noise affecting the speech directly, $f_b - f_a$, is 2500 Hz. The bandwidth
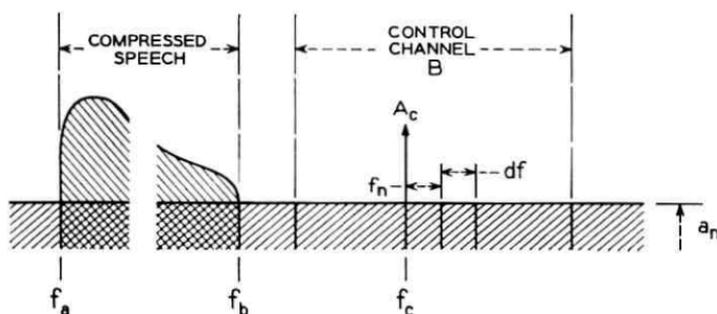

Fig. 24 — Signal and noise spectra.

B is the line, or predetection bandwidth, of the control channel, which is assumed to be defined by an ideal rectangular filter. It is assumed that the peak noise voltage in this band is always less than the peak subcarrier voltage A_c.

In the following analysis, an expression is derived that relates the magnitude of fluctuations in the speech channel loss, to the noise level. The noise level enters this expression implicity in terms of S/N, the speech-to-noise voltage ratio in a 2500-Hz band. It is thus possible to relate judgments concerning S/N and loss fluctuations, respectively.

The flat band of noise of Fig. 24 can be represented by an infinite number of equal-amplitude sinusoids having incommensurable but approximately uniformly spaced frequencies and incoherent phases. The noise in an incremental bandwidth df is represented by a single sinusoid whose mean-square voltage is equal to that of the noise. The mean-square noise voltage in df is just $a_n^2 df$, where a_n is the noise voltage density constant in rms volts per unit square root bandwidth. Thus, if A_n denotes the peak amplitude of the noise component,

$$\frac{A_n^2}{2} = a_n^2 \, df. \tag{24}$$

The superposition of the subcarrier and a noise component at a frequency $f_n = \omega_n/2\pi$ relative to the subcarrier can be represented by the phasor diagram (Fig. 25). The peaks of the subcarrier and noise
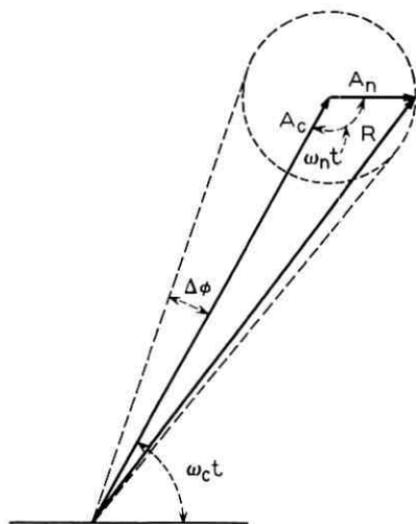

Fig. 25 — Phasor diagram, carrier and noise.

components are denoted by A_c and A_n, respectively. It is assumed that for $A_n \ll A_c$ the resultant angle modulation of the subcarrier produced by the noise will be equal to the summation of the modulations that would be produced by each noise component separately. Except for noise-produced modulation, it will be assumed that the subcarrier is otherwise unmodulated and centered in the band. Amplitude variations may be ignored on the basis that the subcarrier frequency demodulator is ideal.

Rotation of the phasor A_n about A_c causes the resultant R to oscillate about A_c with a peak phase deviation

$$\Delta\varphi = \tan^{-1} \frac{A_n}{A_c} \cong \frac{A_n}{A_c}. \qquad (25)$$

It can be shown[16] that if $A_n \ll A_c$ as assumed, $\varphi(t)$ is essentially sinusoidal and equivalent to pure phase modulation of the form

$$M(t) = A_c \sin(\omega_c t + \Delta\varphi \sin pt + \psi),$$

where the modulating signal is a sinusoid of arbitrary radian frequency p. By differentiation of arg $M(t)$, the instantaneous frequency is obtained:

$$\omega_i = \frac{d}{dt}(\omega_c t + \Delta\varphi \sin pt + \psi)$$
$$= \omega_c + p(\Delta\varphi) \cos pt. \qquad (26)$$

The instantaneous frequency deviation produced by the noise component at the frequency ω_n relative to the subcarrier can be deduced from (26) by using (25) and setting $p = \omega_n$.

$$\text{Inst. freq. dev.} = \omega_n \frac{A_n}{A_c} \cos \omega_n t, \qquad \text{radians per sec.} \qquad (27)$$

The mean-square frequency deviation due to a single noise component is thus

$$\langle(\text{freq. dev.})^2\rangle_{av} = \frac{1}{2}\left(\omega_n \frac{A_n}{A_c}\right)^2, \qquad (\text{radians/sec})^2$$
$$= \left(\frac{A_n^2}{2}\right)\left(\frac{f_n}{A_c}\right)^2, \qquad \text{Hz}^2. \qquad (28)$$

To obtain the incremental mean-square frequency deviation $\langle d\delta^2\rangle_{av}$ due to noise in the band df at a frequency f_n relative to the carrier, (24) is substituted into (28):

$$\langle d\delta^2\rangle_{av} = \left(\frac{a_n f_n}{A_c}\right)^2 df, \qquad \text{Hz}^2. \qquad (29)$$

If a discriminator sensitivity of 1 volt per Hz is assumed for convenience, the integral of $\langle d\delta^2 \rangle_{\mathrm{av}}$ over the line bandwidth may be equated to the positive-frequency discriminator output mean-square noise voltage spectrum integral:

$$\int_{f_c - B/2}^{f_c + B/2} \langle d\delta^2 \rangle_{\mathrm{av}} = \int_0^{B/2} G_n(f) \, df, \tag{30}$$

where

$$G_n(f) = \begin{cases} 2\left(\dfrac{a_n f}{A_c}\right)^2, & \mathrm{V}^2/\mathrm{Hz}, \quad 0 \leqq f \leqq B/2 \\ 0, & f > B/2. \end{cases} \tag{31}$$

The noise spectrum at the output of the single-pole post detection filter is

$$F_n(f) = |Y(f)|^2 G_n(f), \qquad \mathrm{V}^2/\mathrm{Hz}, \tag{32}$$

where

$$|Y(f)| = [1 + (f/f_2)^2]^{-\frac{1}{2}} \tag{33}$$

is the voltage transfer function of the filter and $f_2 = 1/2\pi R_2 C_2$ is the frequency at which the response is down 3 dB. Substituting (31) and (33) into (32) and integrating over half the line bandwidth gives the total mean-square noise voltage output of the control channel that acts on the expandor variolosser:

$$\int_0^{B/2} F_n(f) \, df = 2\left(\frac{a_n}{A_c}\right)^2 \int_0^{B/2} \left[\frac{f^2}{1 + (f/f_2)^2}\right] df, \qquad \mathrm{V}^2. \tag{34}$$

The effect of the ANTILOG circuit (see Fig. 7) on the noise output of the control channel involves a nonlinear transformation of a random process that is difficult to handle mathematically, hence the ANTILOG circuit has thus far been ignored. In spite of this omission, the approximate results derived herein were in good agreement with measurements.

It is to be noted, however, that because of the ANTILOG circuit the translation from a frequency deviation at the FM receiver input to a change in expandor variolosser loss in dB is linear with a slope (or deviation/loss ratio) of 2 Hz per dB. By referring the control channel noise output voltage, (34), back to the discriminator input, an equivalent frequency deviation can be obtained; this deviation, divided by the deviation/loss ratio, gives the fluctuation in expandor loss directly in dB. Accordingly, when this mean square deviation, denoted by $\langle \delta^2 \rangle_{\mathrm{av}}$, is equated to the right side of (34) (on the basis of the assumed dis-

criminator sensitivity of 1 volt per Hz) and the integral is evaluated, we have:

$$\langle \delta^2 \rangle_{av} = 2\left(\frac{a_n}{A_c}\right)^2 (f_2)^3 \left[\frac{B}{2f_2} - \tan^{-1}\left(\frac{B}{2f_2}\right)\right], \qquad Hz^2. \qquad (35)$$

In the CNL system, B is 180 Hz. Substituting this value into (35) and taking the square root to get rms, the result is

$$\delta_{rms} = \frac{a_n}{A_c/\sqrt{2}} (f_2)^{\frac{3}{2}} \left[\frac{90}{f_2} - \tan^{-1}\frac{90}{f_2}\right]^{\frac{1}{2}}, \qquad Hz. \qquad (36)$$

This result is next recast in terms of S/N, the ratio of the rms compressed speech voltage to the rms noise voltage in the 2500-Hz speech band. First, note that

$$\frac{A_c}{\sqrt{2}} = rms \; subcarrier \; voltage = \frac{S}{2}, \qquad (37)$$

since the subcarrier is transmitted 6 dB below the speech in the CNL system; note also that the rms noise voltage

$$N = \left[\int_0^{2500} a_n^2 \, df\right]^{\frac{1}{2}} \qquad (38)$$
$$= 50 \, a_n \, .$$

Then substitution of (37) and (38) into (36) yields

$$\delta_{rms} = \frac{(f_2)^{\frac{3}{2}}}{25 \; S/N} \left[\frac{90}{f_2} - \tan^{-1}\left(\frac{90}{f_2}\right)\right]^{\frac{1}{2}}, \qquad Hz. \qquad (39)$$

Dividing (39) by the deviation/loss ratio, 2 Hz/dB gives the expandor loss error in dB rms. The result can be expressed in dB peak-to-peak (by ignoring peaks in excess of 3 times rms) if (39) is multiplied by 6. When (39) is modified by these two factors, the result is

$$Loss \; error = (\tfrac{1}{2})(6) \; \delta_{rms}$$
$$\qquad (40)$$
$$= \frac{3(f_2)^{\frac{3}{2}}}{25 \; S/N} \left[\frac{90}{f_2} - \tan^{-1}\left(\frac{90}{f_2}\right)\right]^{\frac{1}{2}}, \; dB \; peak\text{-}to\text{-}peak.$$

Fig. 8 is a plot of the loss error versus 20 log (S/N) with f_2 as a parameter.

REFERENCES

1. Members of Technical Staff, Bell Telephone Laboratories, *Transmission Systems for Communications*, Bell Telephone Laboratories, New York, 1964, Chapter 2.

2. Wright, S. B. and Mitchell, D., Two-Way Radio Telephone Circuits, B.S.T.J., *11*, July, 1932, pp. 368–382; also Proc. IRE, *20*, July, 1932, pp. 1117–1130.
3. Herman, J., U. S. Patent No. 1,662,888, issued March 20, 1928.
4. Wright, S. B., Amplitude Range Control, B.S.T.J., *17*, October, 1938, pp. 520–538 (see especially pp. 537–538).
5. Deman, P., Circuits Téléphonique sur Liaison de Faible Qualitié, L. Onde Electrique, *41*, July-August, 1961, p. 656.
6. Law, H. B. and Thompson, D., Some Experiments with Syllabic Constant-Volume Amplifiers, IEE Convention on HF Communication, Savoy Place, London, March 25–27, 1963, pp. 283–285.
7. Wheeler, L. K., Taking the Noise Out of Radio Circuits, Post Office Telecommunications Journal, *18*, No. 2, Summer, 1965.
8. Carter, R. O. and Wheeler, L. K., LINCOMPEX—A System for High Frequency Radio-Telephone Circuits, British Communications and Electronics, *12*, No. 8, August, 1965.
9. Crisson, G., U. S. Patent No. 1,737,830, issued Dec. 3, 1929.
10. International Telegraph and Telephone Consultative Committee (CCITT), *IIIrd Plenary Assembly Blue Book*, International Telecommunication Union, Geneva, May-June, 1964, Vol. III, Recommendation G162, pp. 58–59.
11. *Ibid.*, Annex 5, pp. 445–451.
12. Carter, R. O., Theory of Syllabic Compandors, Proc. IEE, *111*, No. 3, March, 1964, pp. 503–513.
13. International Radio Consultative Committee (CCIR), *Documents of the Xth Plenary Assembly*, International Telecommunication Union, Geneva, 1963, Vol. I, Recommendation 326.
14. Carter, C. W., Jr., Dickieson, A. C., and Mitchell, D., Application of Compandors to Telephone Circuits, Trans. AIEE, *65*, 1946, pp. 1029–1086.
15. Mathes, R. C. and Wright, S. B., The Compandor—An Aid Against Static in Radio Telephony, B.S.T.J., *13*, July, 1934, pp. 313–332; also Electrical Engineering, *53*, June, 1934, pp. 860–866.
16. Black, H. S., *Modulation Theory*, D. Van Nostrand Company, New York, 1953, Chapter 14, p. 223.

Experiments with an Underground Lens Waveguide

By D. GLOGE

A laser beam was transmitted over a distance of 1/2 mile in an underground iron pipe using glass lenses 400 feet apart. The beam deflection and the temperature gradient in the air-filled pipe was measured simultaneously over periods of several weeks. It was found that gradients up to 0.02°C/cm, changing with the season, accounted for beam displacements up to 2 cm on the lenses. Since all these deviations are slow and no other severe disturbance was noticed, the conclusion is that the transmission of optical beams underground is possible without evacuation of the conduit. Suggestions for the construction of such a beam waveguide are made which could reduce the mentioned temperature influence by a factor of 100.

I. INTRODUCTION

If a beam of coherent light is to be transmitted along the surface of the earth it will be necessary to redirect and focus it at intervals by means of lenses or mirrors to follow the terrain.[1,2] Furthermore, it must be shielded from atmospheric temperature fluctuations which result in variations of the index of refraction. A laboratory experiment which, in a 100-m metal pipe, folded a transmission path back upon itself gave the impression that the atmospheric effects could be overcome by choosing the proper beam enclosure.[3,4] Other experimenters, using a 1-km long pipe above ground, considered evacuation as the most reasonable means to avoid temperature effects.[5] In any case, the transmission path would most likely be installed underground where temperature variations are much smaller than in the open air.

To gain information about the transmission characteristics in this case a ½-mile underground iron pipe was used to build a lens waveguide with focusing lenses 400 feet apart. The temperature field and the beam displacement in the tube were measured simultaneously and compared with theoretical estimates. This gave valuable indications for the construction of underground lens waveguides.

II. GROUND TEMPERATURE

The temperature distribution in the ground depends very much on the type of soil, on the moisture content, and on the character of the surface. The common feature of any soil, however, is a high heat capacity c and comparatively low conductivity κ which leads to very slow heat diffusion processes. To give a representative picture, a half space with constant c and κ is considered. Then heat diffuses only along the vertical direction y and the temperature T can be found from the equation

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial y^2}, \qquad (1)$$

where the ratio

$$\alpha = \frac{\kappa}{c} \qquad (2)$$

is called the thermal diffusivity.

Since the temperature changes at the surface are mainly periodic it is convenient to consider the Fourier amplitudes T_ω of components with frequency ω. The solution of (1) in complex form is then

$$T(y,\omega) = T_\omega \exp\left(i\omega t - y\sqrt{i\omega/\alpha}\right). \qquad (3)$$

The temperature is retarded and attenuated with increasing depth and increasing frequency.

To find the diffusivity α involved in this experiment, thermocouples were installed 0, 1, and 3 feet below the surface. A recording over 2 weeks is shown in Figs. 1(a) and (b). A comparison of the temperatures at the surface and 1 foot below show that a 1-foot layer of soil reduces the amplitude of the diurnal temperature cycle by a factor of 10 and introduces a phase lag of 10 hours. This corresponds to a diffusivity of $\alpha = 5.5 \ 10^{-3}$ cm^2/s.

Data gathered at different locations in the United States yield an average diffusivity of $5.4 \ 10^{-3}$ cm^2/s.[4] Handbook data for clay soil are shown in Table I. According to these data the diurnal temperature cycle at a 3-foot depth is reduced by three orders of magnitude and therefore not noticeable in Fig. 1(b).

The recording of Fig. 1(c) shows the temperature gradient in the iron pipe 5 feet below the surface, which was used for the transmission experiment. This gradient was measured by a digital quartz thermometer whose sensors were installed at the top and the bottom of the tube. These sensors are connected with the equipment by 100 feet of cable so that the measurements could be taken far inside the tube and remote
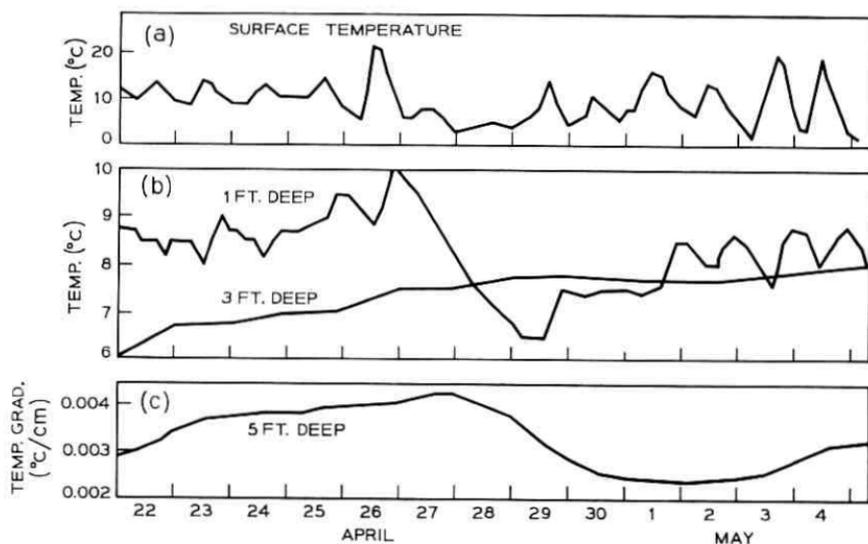
Fig. 1 — Temperature recordings at Bell Telephone Laboratories, Holmdel, N. J. from April 22 to May 4, 1966. (a) Temperature at the surface. (b) Temperature at a depth of 1 and 3 ft. (c) Thermal gradient in an iron pipe with a wall thickness of ¼″ and an i.d. of 3.5″ at a depth of 5 ft.

from any disturbance. No diurnal temperature changes were noticeable.

From (3) one finds the temperature gradient in the soil to be

$$\vartheta_e = \frac{\partial T}{\partial y} = -T_\omega \sqrt{i\frac{\omega}{\alpha}} \exp\left(i\omega t - y\sqrt{i\omega/\alpha}\right). \qquad (4)$$

Gradients of very low and very high frequency disappear at a large enough depth. If the Fourier spectrum of the temperature on the surface is assumed to be flat then Fig. 2 gives the amplitude spectrum of the gradient at various depths.

TABLE I

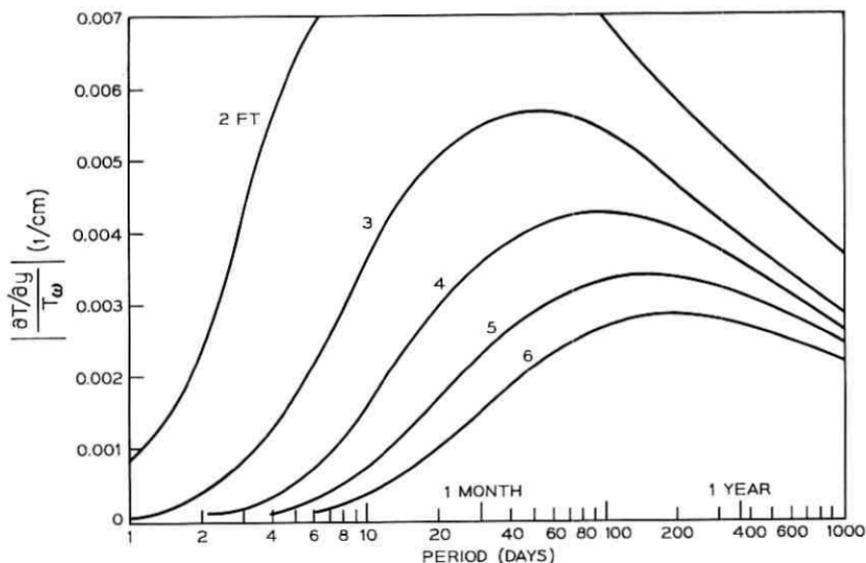|		Still air	Clay soil	Steel	Aluminum	Fiberglass
Conductivity κ	$\dfrac{\text{cal}}{\text{cm s}\,°\text{C}}$	$5.5\ 10^{-5}$	$2.8\ 10^{-3}$	0.11	0.55	10^{-4}
Heat capacity c	$\dfrac{\text{cal}}{\text{cm}^3\,°\text{C}}$	$3\ 10^{-4}$	0.5	1	0.65	—
Diffusivity α	$\dfrac{\text{cm}^2}{\text{s}}$	0.18	$5.4\ 10^{-3}$	0.1	0.85	—

Fig. 2 — Amplitude spectrum of the thermal gradient at various depths related to the amplitude at the surface of the ground.

Actually, the temperature spectrum at the surface is not constant but has peaks with periods of 1 day and 1 year. About 1 year is also the period which causes maximum response of the temperature gradient at a depth of 5 feet.

The mean temperature behavior valid for the area of New Jersey, as shown in Fig. 3, is based on data gathered in Ref. 6. The corresponding temperature cycle at a depth of 5 feet was calculated from (2). The dots show measured values. The presence of an air-filled iron pipe distorts the temperature field considerably. As is shown in the appendix, a function describing this field can be derived from an equivalent problem in the theory of electrostatic fields. Assuming that ϑ_e, the gradient in the absence of a tube, is approximately constant over an area comparable to the tube crosssection, the field inside the tube is found to be homogeneous having a thermal gradient ϑ_i that is given by (18).

For the tube used the inner radius $r_1 = 1.75$ inches, the outer radius $r_2 = 2$ inches and κ_0, κ_1, and κ_2 are the thermal conductivities of air, iron, and soil, respectively (see Fig. 4). With the numbers given in Table I, one finds that the internal gradient ϑ_i is reduced by a factor of 0.4 in comparison with ϑ_e. Therefore,

$$\vartheta_i = 0.4\sqrt{i\frac{\omega}{\alpha}} \exp\left(i\omega t - y\sqrt{i\omega/\alpha}\right). \tag{5}$$
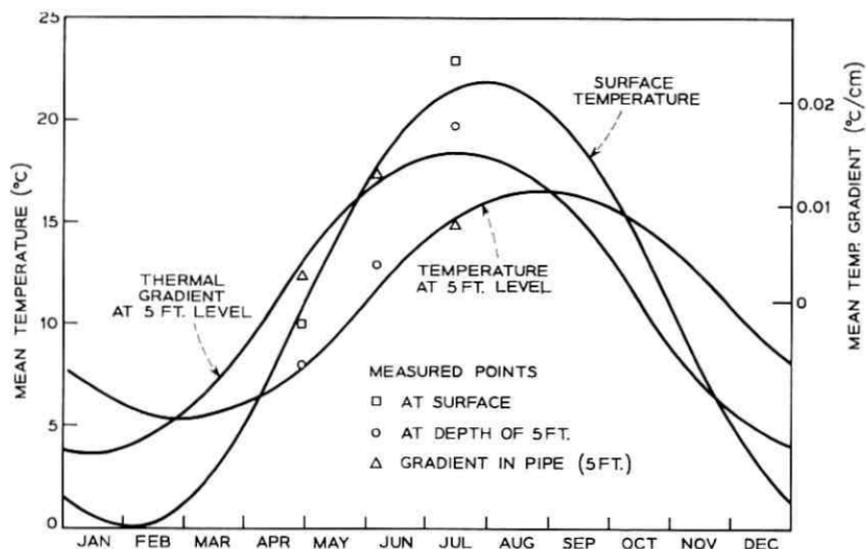
Fig. 3 — Mean annual change of the temperature at the surface and at a depth of 5 ft and the corresponding thermal gradient in an iron pipe at a depth of 5 ft as expected from data given in Ref. 7. The dots show values measured at Bell Telephone Laboratories, Holmdel, N. J. in 1966.
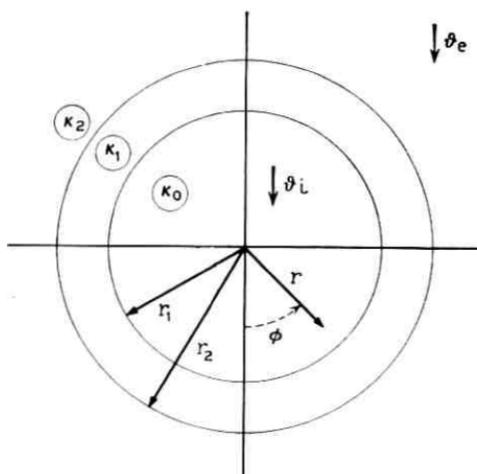

Fig. 4 — Sketch of the pipe arrangement used to derive the internal temperature field.

As long as the thermal gradient has no horizontal component and its value is below a certain threshold, convection cannot develop in a cylindrical container.[7] For the air-filled tube with the given dimensions, this threshold is 0.02°C/cm.

III. THE LENS WAVEGUIDE

The experimental setup as shown in Fig. 5(a) makes use of a straight iron pipe of $\frac{1}{2}$-mile length with an inner diameter of 3.5 inches at a mean depth of 5 feet. Roughly every 140 m there is access to the pipe through manholes. Here biconvex thin lenses with a focal length of 70 m are installed.

The fundamental mode of this confocal lens waveguide is a beam with Gaussian field distribution. Operating at the laser wavelength of 0.63 μ, the $1/e$ width of the field at the lenses is $2\sqrt{140}$ m 0.63 μ/π = 10.4 mm. The lenses are supported by the pipe itself and mounted close to the wall of the manhole where the pipe comes out of the ground. No vibrations were noticed. The conduit is airtight over the whole transmission length.

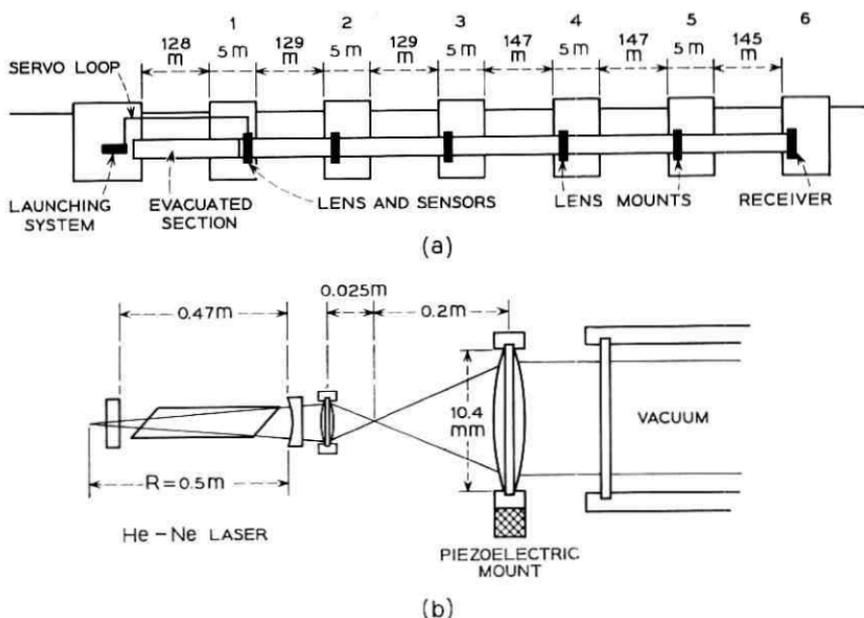The lenses have a diameter of 60 mm and can be adjusted for about

Fig. 5 — Sketch of the experimental setup: (a) The transmission path. (b) The launching system.
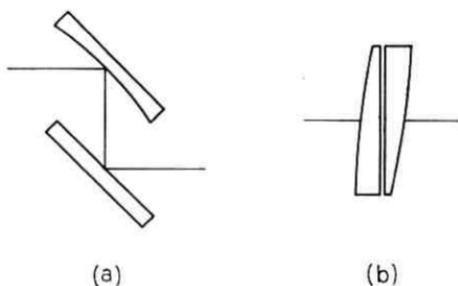
Fig. 6 — Possible redirecting devices in a lens waveguide. (a) Mirror arrangement as proposed by Kompfner.[1] (b) Rotatable split lens which contains a variable prism.

±15 mm from the outside of the pipe. This is sufficient to guide the beam in the conduit which is straight over the total length of $\frac{1}{2}$ mile. Of course, aligned sections cannot be expected in an actual field installation. Redirecting devices must therefore be provided where the described model has only focusing lenses.

In order to gain a feeling for the usefulness of the described model, some thought will be devoted to the features of redirecting devices. Fig. 6 shows two examples. In Fig. 6(a) a plain and a curved mirror are mounted opposite to each other.[1] Fig. 6(b) shows two eccentric parts of semiconvex lenses which can be rotated independently. The space between the adjacent surfaces is filled with an oil film which has the same refractive index as the lenses. The complete device acts therefore as a lens plus a variable prism.

The device in Fig. 6(a), as well as the one in Fig. 6(b), uses two (but not more than two) reflecting or refracting surfaces both to direct and to focus the beam. Those surfaces determine essentially the transmission loss in an optical waveguide.[3,4,5] Therefore, redirection of the beam may be achieved without a large increase of the transmission loss, no matter if reflection or refraction is used. The experiment described here was started with refracting devices since they are easier to mount and to adjust, but an experiment with reflecting devices is planned as well.

The lenses used in the experiment have a surface quality of $\lambda/10$ and are furnished with an antireflection coating which guarantees less than 0.2 percent reflection at 0.63 μ wavelength. The bandwidth is about 500 Å. The absorption loss of the lens material is in the order of 0.02 percent for a lens thickness of 5 mm. Assuming a total absorption loss of 0.1 percent in the glass and the layers of the two coatings, an

overall loss of 0.5 percent must be expected per lens. (This varies by about ±0.05 percent since the reflections from both surfaces interfere with each other and the total reflection therefore depends on the lens thickness.)

A loss of the same order for mirrors was reported in Ref. 3. No significant variations of the loss are therefore to be expected in this case. For the mean lens spacing of 140 m a waveguide loss of 0.15 dB/km would occur. This, of course, holds only for perfectly clean surfaces. A considerable increase of the loss must be expected if dust accumulates on the lenses. No loss measurements have been performed up to the present time.

The Gaussian field distribution is generated in the almost concentric resonator of a He-Ne laser and magnified by a factor of 8 in the lens arrangement shown in Fig. 5(b). These lenses are polished to a quality of $\lambda/10$ but are not corrected for aberrations.

The launching system consisting of laser and magnifier as shown in Fig. 5(b) is mounted on a solid concrete table resting on a 1-foot thick concrete floor. Although the air temperature around the launching system is kept constant within 2°C the temperature changes in the environment are large enough to cause deviations of the proper launching angle of $1.5 \ 10^{-5}$ rad or 2 mm at the first lens. To eliminate this effect a servo-loop was built which holds the beam center at the first lens within an area of $30 \times 30 \ \mu$. Of course no temperature effect can now be measured in the first pipe section, but by evacuating this section it is possible to attain a deviation angle of the forthgoing beam of no more than $30 \ \mu/128 \ m = 2.3 \ 10^{-7}$ rad.

To measure the displacement of the beam the circuit in Fig. 7(a) was used. If the shown CdS-cells have the resistances R_1 and R_2 the output voltage is

$$V = \frac{V_o}{2} \frac{R_1 - R_2}{R_1 + R_2}. \tag{6}$$

Since these resistances, in the region used, are inversely proportional to the illuminating intensity I, (6) can also be written in the form

$$V = \frac{V_o}{2} \frac{I_2 - I_1}{I_2 + I_1}. \tag{7}$$

This shows that the output voltage does not depend on the absolute laser intensity but only on the ratio of the intensities absorbed by the resistors, which is a function of the displacement d and the beam width. A typical sensitivity curve of the device is shown in Fig. 7(b). The
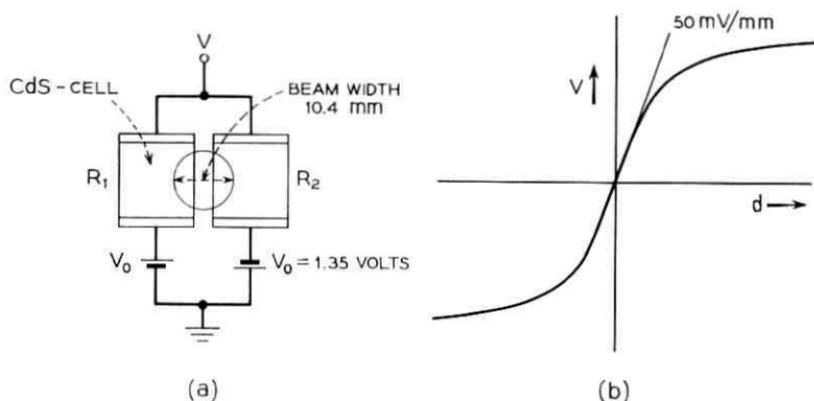
Fig. 7 — Device to measure the beam displacement with photoresistors. (a) The circuit. (b) The sensitivity curve.

same circuits were used to produce the horizontal and the vertical feedback signals and to measure the beam movement at the end of the transmission path.

IV. BEAM DISPLACEMENT

Assume only small deviation of the beam from the ideal propagation axis z. Then the displacement d in a transverse direction y can be calculated from the equation

$$\frac{\partial^2 d}{\partial z^2} = \frac{1}{n}\frac{\partial n}{\partial y}. \tag{8}$$

Here n is the refractive index which for air at atmospheric pressure has the temperature dependence

$$n = 1 + \Delta n\, \frac{T_o}{T} \tag{9}$$

with $\Delta n = 3\ 10^{-4}$. This yields

$$\frac{\partial^2 d}{\partial z^2} = -\frac{\Delta n}{n}\frac{1}{T_o}\frac{\partial T}{\partial y}. \tag{10}$$

If $\partial T/\partial y$ is assumed constant over a transmission length L a simple integration of (10) gives the displacement

$$d = -\frac{\Delta n}{2n}\frac{L^2}{T_o}\frac{\partial T}{\partial y} \tag{11}$$

at the end of the path L, provided the beam is injected on axis at the beginning. To measure this effect the stabilized beam coming out of the evacuated pipe section was sent through the next section and its displacement recorded in the second manhole. At the same time the temperature gradient was measured in another pipe section. Inserting these values into (11) the expected beam displacement was calculated. The thermal gradient and the measured and calculated displacement are shown in Fig. 8. There are short-term movements superimposed on the main slope which are caused by diurnal temperature changes in the manholes. Here the pipe warms up over the day and cools down at night. This introduces convection currents along the pipe which cause a transverse temperature gradient and a deflection of the beam.[7]

Recordings of the beam position over the full transmission length again showed diurnal deviations (up to 4 mm) introduced by the temperature change in the manholes.

So far only changes of the temperature gradient have been considered but nothing has been said about the absolute gradient. Fig. 3 shows that a maximum gradient of $\pm 0.02°C/cm$ must be expected. In this case, if the beam enters the waveguide on axis, the displacement on the second lens according to (11) is 2 cm. The third lens is passed at
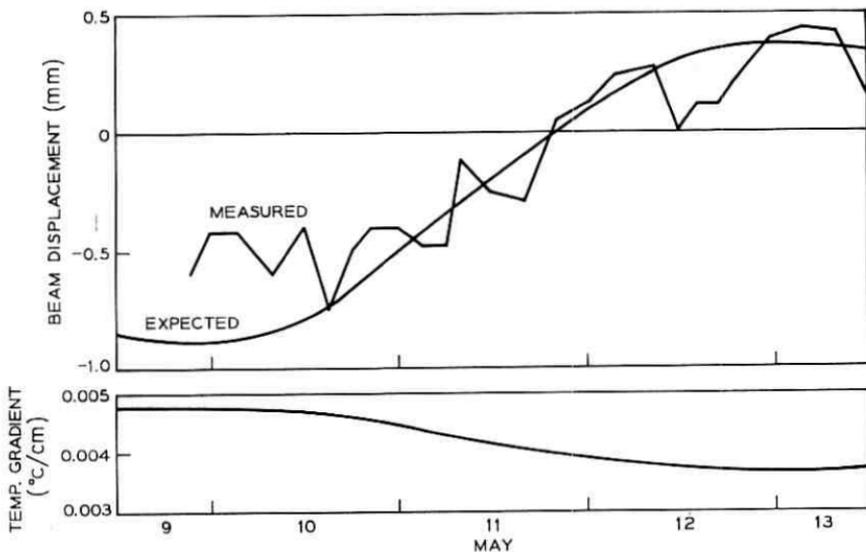

Fig. 8 — Measured and expected beam displacement as calculated from the thermal gradient shown below for a laser beam transmitted over 128 m through an iron pipe at a depth of 5 feet in the ground.
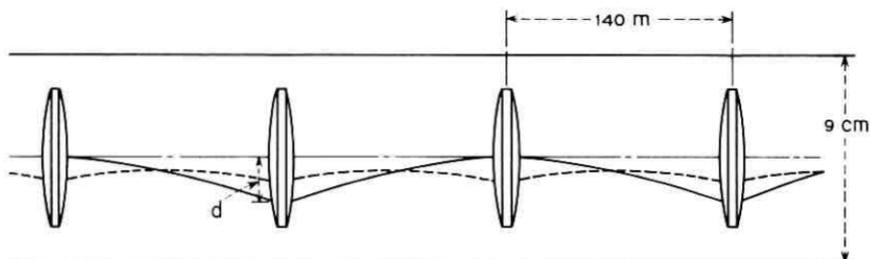
Fig. 9 — Beam path in the experimental lens waveguide at the time of maximum temperature gradient.

the center, but at the fourth lens again 2-cm displacement occurs. The beam undulates about the dotted line shown in Fig. 9. A proper launching of the beam would avoid the undulations and the beam would follow the dotted line. This behavior was indeed found quantitatively during measurements in June when a high temperature gradient occurred.

Since the beam center has to pass a lens at least 1 cm from the lens edge to avoid noticeable diffraction, the described lens waveguide can just handle the occurring temperature effect. Of course, this waveguide runs below a uniform grass surface and the temperature gradient is fairly constant everywhere. An actual transmission path, however, will be subject to changes of the temperature gradient from section to section and that will cause undulations of the beam which increase with the transmission length. For this reason a thermal gradient as great as that measured would be a serious drawback for the waveguide and a considerable reduction of the temperature influence in the conduit seems necessary.

This can be accomplished by employing a pipe material of higher conductivity. The reduction factor ϑ_i/ϑ_e calculated from (18) is shown in Fig. 10 for iron and aluminum tubes with various diameter ratios. Using an aluminum pipe instead of the present iron pipe would reduce the inside thermal gradient by a factor of 5.

The shielding effect can be improved even more by using several coaxial layers of strongly alternating conductivity. As a practical example for a triple-layer pipe, the combination aluminum-concrete-aluminum was calculated using a concrete volume three times larger than the aluminum volume. The thickness of the layers is chosen in such a way that the shielding effect is largest. As Fig. 10 shows, such a pipe can be more effective than a solid aluminum pipe if the layers
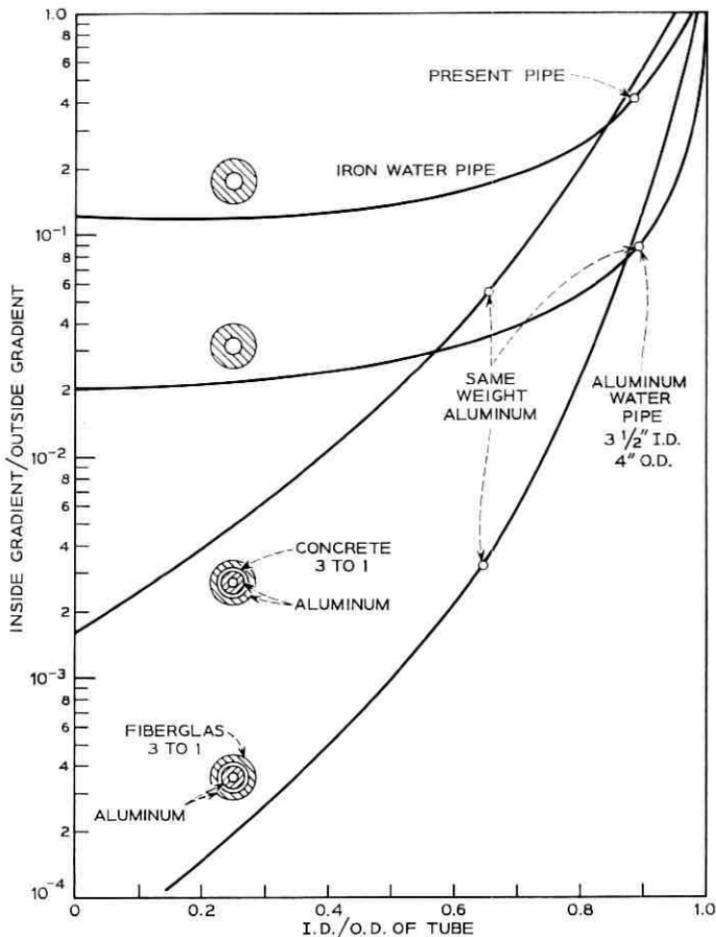
Fig. 10 — The reduction factor versus the ratio of inside to outside diameter for a solid iron pipe, a solid aluminum pipe and aluminum pipes stacked with concrete and fiberglass using a volume ratio 1:3.

are very thick. For the same amount of aluminum as in the solid tube, however, it is only a factor of $\frac{2}{3}$ better.

Using fiberglass instead of concrete improves the shielding effect considerably. A pipe with the same aluminum weight as before yields a reduction factor of $3 \cdot 10^{-3}$. Compared with the iron water pipe now in use, the shielding effect of the latter arrangement is two orders of magnitude better, and the maximum expected gradient in such a pipe would therefore be $2 \cdot 10^{-4} °C/cm$.

V. CONCLUSIONS

The results of the described experiment show that the main disturbance of a coherent light beam transmitted in an air-filled underground conduit is a slowly varying displacement of the beam center. This displacement can be coordinated with thermal gradients in the atmosphere of the pipe and predicted theoretically. No other time dependent sources of disturbance were noticed in the experiment. The mounting, as well as the adjustment of the lenses, was not critical.

To reduce the beam displacement to a tolerable amount some simple provisions must be made: (*i*) the conduit should be imbedded in the soil everywhere and exposure to temperature changes should be minimized; (*ii*) the conduit should be airtight to avoid longitudinal air currents; (*iii*) a tube of high thermal conductivity or a multilayer tube with a good shielding effect should be chosen. With the proper design no larger gradients than 0.001°C/cm would occur in the pipe. If, for example, the gradient is not constant over the transmission path but varies with a rms value of 0.001°C/cm then, if the lens spacing is 140 m and the beam enters on axis, beam undulations occur with a mean amplitude of 0.5 mm at the beginning of the waveguide, increasing to 5 mm after 100 lens sections. There the beam could be realigned by an electronic servo system.[8]

Drift and vibrations of the ground were not investigated until now, but these might very well cause larger displacements than the ones just mentioned. This would necessitate beam aligning servo systems at even shorter intervals.

VI. ACKNOWLEDGMENTS

APPENDIX

To find the temperature field in an around several coaxial cylinders of different material as shown in Fig. 4, the equivalent problem in the theory of electrostatic fields is considered. It is known that a circular dielectric rod with its axis perpendicular to a homogeneous electric field develops a homogeneous internal field determined by its polarization and an external field that consists of the superposition of the original field plus a field generated by the polarization of the rod.

In polar coordinates, a homogeneous temperature field with gradient ϑ_e has the form $\vartheta_e r \cos \varphi$ (see Fig. 4). If now a circular rod is brought into the field, an additional term of the form $(\theta/r^2) \cos \varphi$ arises originated by the "polarization" θ of the rod. In a more complicated configuration of several coaxial cylinders, the field in the various materials can still be described by terms of the above type.

Labeling the n layers starting with $\nu = 0$ for the center rod, one has for the field T_ν in the νth layer

$$T_\nu = \vartheta_\nu r \cos \varphi + \frac{\theta_\nu}{r^2} \cos \varphi \tag{12}$$

for

$$\nu = 0, 1 \cdots n.$$

T_n is the field in the surrounding and $\vartheta_n \equiv \vartheta_e$, the original thermal gradient. $\theta_0 \equiv 0$ since the field in the center bore is homogeneous. This leaves $2n$ constants ϑ_ν and θ_ν. They can be found by matching the fields at the boundaries which yields $2n$ linear algebraic equations of the form

$$\vartheta_{\nu-1} - \frac{\theta_{\nu-1}}{r_\nu^2} = \vartheta_\nu - \frac{\theta_\nu}{r_\nu^2} \tag{13}$$

and

$$\kappa_{\nu-1}\left(\vartheta_{\nu-1} + \frac{\theta_{\nu-1}}{r_\nu^2}\right) = \kappa_\nu\left(\vartheta_\nu + \frac{\theta_\nu}{r_\nu^2}\right). \tag{14}$$

To calculate the internal gradient $\vartheta_i \equiv \vartheta_o$ it is convenient to define the ratios

$$p_\nu^+ = \frac{\kappa_{\nu-1} + \kappa_\nu}{2\kappa_\nu} \tag{15}$$

$$p_\nu^- = \frac{\kappa_{\nu-1} - \kappa_\nu}{2\kappa_\nu} \tag{16}$$

and a determinant

$$D_n = \begin{vmatrix} p_0^+ & \frac{r_1}{r_2} p_1^- & \cdots & -(-1)^n \frac{r_1}{r_n} p_n^- \\ -\frac{r_1}{r_2} p_0^- & p_1^+ & & -(-1)^{n-1} \frac{r_2}{r_n} p_n^- \\ \vdots & & & \\ (-1)^n \frac{r_1}{r_n} p_0^- & (-1)^{n-1} \frac{r_2}{r_n} p_1^- & & p_n^+ \end{vmatrix}. \tag{17}$$

Then

$$\frac{\vartheta_0}{\vartheta_e} = \frac{1}{D_n}.$$

This is the ratio by which the gradient inside a multilayer tube is reduced in comparison with the gradient in the homogeneous soil (in absence of that pipe). For a solid tube as shown in Fig. 4, $n = 2$ and one finds

$$\frac{\vartheta_0}{\vartheta_e} = \frac{1}{\dfrac{\kappa_0 + \kappa_1}{2\kappa_1}\dfrac{\kappa_1 + \kappa_2}{2\kappa_2} + \left(\dfrac{r_1}{r_2}\right)^2 \dfrac{\kappa_0 - \kappa_1}{2\kappa_1}\dfrac{\kappa_1 - \kappa_2}{2\kappa_2}}. \tag{18}$$

REFERENCES

1. Kompfner, R., unpublished work.
2. Goubau, G. and Christian, F. R., On the Guided Propagation of Electromagnetic Wave Beams, IRE Trans. *AP-9*, May, 1961, pp. 248–256.
3. DeLange, O. E., Losses Suffered by Coherent Light Redirected and Refocussed in an Enclosed Medium, B.S.T.J., *44*, February, 1965, pp. 283–302.
4. Seifert, G. and Gloge, D., Messungen an einer Optischen Pendelimpulsstrecke, Archiv El. Übertragung, *19*, November, 1965, pp. 633–635.
5. Goubau, G. and Christian, F. R., Loss Measurements with a Beam Waveguide for Long-Distance Transmission at Optical Frequencies, Proc. IEEE, *52*, December, 1964, p. 1739.
6. Grisby, C. M. and Shuhart, F. M., Earth Temperatures, unpublished work.
7. Gloge, D., Thermal Instabilities in Optical Waveguides, unpublished work.
8. Marcatili, E. A. J., Ray Propagation in Beam-Waveguides with Redirectors, B.S.T.J., *45*, January, 1966, pp. 105–115.

Axis-Crossings of the Phase of Sine Wave Plus Noise

By A. J. RAINAL

This paper is concerned with the axis-crossings of the resultant phase $-\pi \leqq \theta(t,a) \leqq \pi$ of a sinusoidal signal of amplitude $\sqrt{2a}$ and frequency f_0 plus Gaussian noise of unit variance having a narrow-band power spectral density which is symmetrical about f_0. The discontinuous phase process $\theta(t,a)$ is present at the output of the IF amplifier of a radio or radar receiver during the reception of a sinusoidal signal immersed in Gaussian noise. Also, the phase process $\theta(t,a)$ is basic in Rice's recent analysis of noise in FM receivers. The following theoretical results are presented concerning the axis-crossings (level-crossings) of $\theta(t,a)$ at an arbitrary level θ:

(i) The average number of upward (or downward) axis-crossings per second.

(ii) The conditional probability that an upward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given a downward axis-crossing at t.

(iii) The conditional probability that a downward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t.

(iv) The conditional probability that an upward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t.

(v) The variance of the number of axis-crossings observed in a time τ.

The theoretical probability functions are presented in graphs as a continuous function of τ for various values of θ and "a" for the case when the Gaussian noise has a Gaussian power spectral density.

I. INTRODUCTION

Consider the stationary random process $I(t,a)$ consisting of a sinusoidal signal of amplitude $\sqrt{2a}$ and frequency f_0 plus Gaussian noise $I_N(t)$, of unit variance, having a narrow-band power spectral density $W_b(f - f_0)$ which is symmetric about f_0. Rice's[1] graphical representation of $I(t,a)$ is illustrated in Fig. 1 in order to define the Rayleigh envelope process $R(t,a)$ and the resultant phase $-\pi \leqq \theta(t,a) \leqq \pi$. The purpose of this
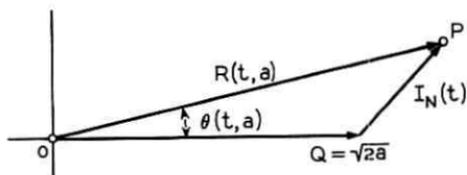
737

Fig. 1 — Graphical representation of $I(t,a) = \sqrt{2a} \cos 2\pi f_0 t + I_N(t) = R(t,a)$ $\cos [2\pi f_0 t + \theta(t,a)]$. The point P wanders around, as time goes by, in the plane of the figure and generates the phase process $\theta(t,a)$.

paper is to present some theoretical results concerning the axis-crossing points of the stationary, discontinuous phase process $\theta(t,a)$. In the literature, these same points are also called level-crossings. The axis-crossing points and the axis-crossing intervals of $\theta(t,a)$ are defined in Fig. 2. The axis-crossing points and the axis-crossing intervals of $R(t,a)$ are defined in a similar manner and were discussed by Rice[2] and Rainal.[3,4] The Rayleigh process $R(t,a)$ and the phase process $\theta(t,a)$ are present at the output of the IF amplifier of a typical radio or radar receiver during the reception of a sinusoidal signal immersed in Gaussian noise. Also, the phase process $\theta(t,a)$ is basic in Rice's[5] recent analysis of noise in FM receivers.

Using a notation consistent with Refs. 3 and 6, we shall present the following theoretical results, in terms of well-known tabulated functions, concerning the axis-crossings of $\theta(t,a)$ at an arbitrary level θ and arbitrary signal-to-noise power ratio "a":

   (i) N_θ, the average number of upward (or downward) axis-crossings per second.

   (ii) $Q_1^-(\tau,\theta,a) \, d\tau$, the conditional probability that an upward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given a downward axis-crossing at t.

   (iii) $Q_1^+(\tau,\theta,a) \, d\tau$, the conditional probability that a downward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t.

   (iv) $[U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)] \, d\tau$, the conditional probability that an upward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t.

II. AVERAGE NUMBER OF AXIS-CROSSINGS PER SECOND

N_θ, the average number of upward axis-crossings per second of the level θ by the phase process $\theta(t,a)$, follows directly from some results due to Rice. Rice[1] showed that

$$N_\theta = \int_0^\infty dR \int_0^\infty d\theta' \; \theta' P(R,\theta,\theta'), \tag{1}$$

where

$$P(R,\theta,\theta') = \frac{R^2}{2\pi\sqrt{2\pi\beta}} \exp\left[-\frac{R^2}{2} - \frac{(\theta'R)^2}{2\beta} + QR \cos\theta - \frac{Q^2}{2} \right]$$

$$Q = \sqrt{2a}$$

$$\beta = 4\pi^2 \int_0^\infty W_b(f - f_0)(f - f_0)^2 \, df$$

$$-\pi \le \theta \le \pi.$$

$W_b(f - f_0)$ = one-sided narrow-band power spectral density of $I_N(t)$. Performing the integrations we find that

$$N_\theta = \frac{\sqrt{\beta}}{2\pi} \exp\left[-a \sin^2 \theta\right] \Phi(Q \cos \theta), \tag{2}$$

where

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} \, dy.$$

Equation (2) was also derived by Tikhonov[7].

Since $\theta = 0$ is a level of symmetry we have that $N_\theta = N_{-\theta}$. Also, the average number of downward axis-crossings per second is given by the right-hand side of (1) with the upper limit of integration of θ' set to $-\infty$. Thus, the average number of downward axis-crossings per second is also equal to N_θ.
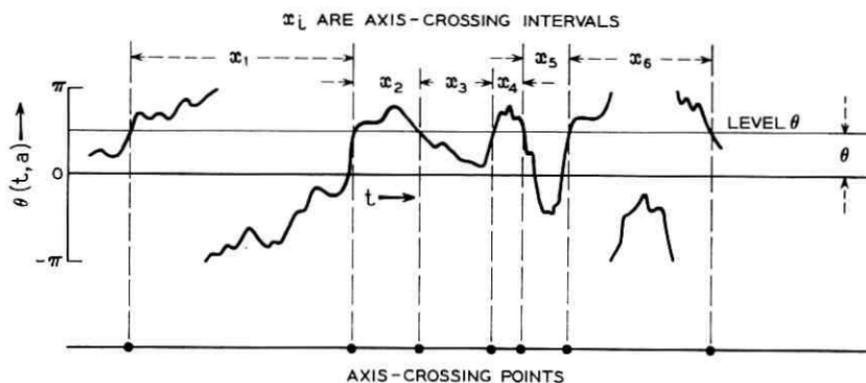

Fig. 2 — The level θ defines the axis-crossing points and the axis-crossing intervals of the discontinuous phase process $\theta(t,a)$.

When the level $\theta = \pm\pi$ and "a" is large, $2N_\theta\tau$ represents the average number of clicks observed in a time τ at the output of an ideal FM receiver[5,8] during the reception of a unmodulated carrier in the presence of receiver noise. The variance of the number of clicks observed in a time τ is discussed in Section IV.

III. CONDITIONAL PROBABILITY FUNCTIONS

The reader should refer to Rice[2] for the definition of all notation which is not defined in this paper. For the phase process $\theta(t,a)$, the conditional probability $Q_1^-(\tau,\theta,a)\,d\tau$, the conditional probability that an upward axis-crossing of the level θ occurs between $t + \tau$ and $t + \tau + d\tau$ given a downward axis-crossing of the level θ at t, is given by an equation analogous to Rice's[2] equation (86):

$$
Q_1^-(\tau,\theta,a)\,d\tau = -d\tau\,N_\theta^{-1}\int_{-\infty}^{\infty} dR_1' \int_{-\infty}^{\infty} dR_2' \int_0^{\infty} dR_1 \int_0^{\infty} dR_2
$$
$$
\cdot \int_{-\infty}^0 d\theta_1' \int_0^{\infty} d\theta_2'\,\theta_1'\theta_2' p(R_1\,,\,R_1'\,,\,\theta,\,\theta_1'\,,\,R_2\,,\,R_2'\,,\,\theta,\,\theta_2'),
\tag{3}
$$

where

$$
p(R_1\,,\,R_1'\,,\,\theta,\,\theta_1'\,,\,R_2\,,\,R_2'\,,\,\theta,\,\theta_2')
$$
$$
= \frac{R_1^2 R_2^2}{(2\pi)^4 M}\exp\left\{-\frac{1}{2M}\left[M_{11}[R_1^2 + R_2^2 - 2Q(R_1 + R_2)\cos\theta + 2Q^2]\right.\right.
$$
$$
+ 2M_{12}[R_1 R_1' - R_2 R_2' - Q(R_1' - R_2')\cos\theta + Q(R_1\theta_1' - R_2\theta_2')\sin\theta]
$$
$$
+ 2M_{13}[R_1 R_2' - R_2 R_1' - Q(R_2' - R_1')\cos\theta + Q(R_2\theta_2' - R_1\theta_1')\sin\theta]
$$
$$
+ 2M_{14}[R_1 R_2 - Q(R_2 + R_1)\cos\theta + Q^2]
$$
$$
\left.\left.+ M_{22}[R_1'^2 + R_2'^2 + R_1^2\theta_1'^2 + R_2^2\theta_2'^2] + 2M_{23}[R_1' R_2' + R_1 R_2\theta_1'\theta_2']\right]\right\}.
$$

The M's are given in Rice's[2] Appendix I with

$$
m(\tau) = \int_0^{\infty} W_b(f - f_0)\cos 2\pi(f - f_0)\tau\,df.
\tag{4}
$$

By performing the integrations with respect to R_1' and R_2' we find that

$$
Q_1^-(\tau,\theta,a) = -N_\theta^{-1}\int_0^{\infty} dR_1 \int_0^{\infty} dR_2
$$
$$
\cdot \int_{-\infty}^0 d\theta_1' \int_0^{\infty} d\theta_2'\,\theta_1'\theta_2' p(R_1\,,\,\theta,\,\theta_1'\,,\,R_2\,,\,\theta,\,\theta_2'),
\tag{5}
$$

where

$$p(R_1, \theta, \theta_1', R_2, \theta, \theta_2') = \frac{R_1^2 R_2^2}{(2\pi)^3 \sqrt{M_{22}^2 - M_{23}^2}}$$

$$\cdot \exp\left\{-\frac{1}{2M}\left[M_{22}(R_1^2 \theta_1'^2 + R_2^2 \theta_2'^2) + 2M_{23}R_1 R_2 \theta_1' \theta_2' \right.\right.$$

$$\left.\left. + 2Q\sin\theta[M_{12} - M_{13}][R_1\theta_1' - R_2\theta_2']\right]\right\} \cdot \exp\left(-G_0/2M\right)$$

and

$$G_0 = \left\{ M_{11}(R_1^2 + R_2^2) + 2M_{14}R_1 R_2 \right.$$

$$+ 2Q(Q - R_1\cos\theta - R_2\cos\theta)(M_{11} + M_{14})$$

$$+ \frac{(-M_{12}^2 M_{22} - M_{13}^2 M_{22} + 2M_{12}M_{13}M_{23})}{(M_{22}^2 - M_{23}^2)}$$

$$\cdot[(R_1 - Q\cos\theta)^2 + (R_2 - Q\cos\theta)^2]$$

$$+ \frac{(-M_{12}^2 M_{23} - M_{13}^2 M_{23} + 2M_{12}M_{13}M_{22})}{(M_{22}^2 - M_{23}^2)}$$

$$\left. \cdot[2(R_1 - Q\cos\theta)(R_2 - Q\cos\theta)]\right\}.$$

By introducing the variables x, y, in place of θ_1', θ_2' with the following transformation

$$R_1 \theta_1' = -\left[\frac{M_{22}}{1 - m^2}\right]^{\frac{1}{2}} x - Q\left[\frac{M_{12} - M_{13}}{M_{22} - M_{23}}\right]\sin\theta \qquad (6)$$

$$R_2 \theta_2' = \left[\frac{M_{22}}{1 - m^2}\right]^{\frac{1}{2}} y + Q\left[\frac{M_{12} - M_{13}}{M_{22} - M_{23}}\right]\sin\theta, \qquad (7)$$

we find that

$$Q_1^-(\tau,\theta,a) = \frac{N_\theta^{-1} M_{22}}{(2\pi)^2 (1 - m^2)^2} J(r_1, h_1)$$

$$\cdot \exp\left\{\frac{Q^2 \sin^2\theta(M_{12} - M_{13})^2}{M(M_{22} - M_{23})}\right\} \int_0^\infty dR_1 \int_0^\infty dR_2 \exp\left(-G_0/2M\right), \qquad (8)$$

where

$$J(r_1, h_1) = \frac{1}{2\pi\sqrt{1 - r_1^2}} \int_{h_1}^\infty dx \int_{h_1}^\infty dy \,(x - h_1)(y - h_1)e^z$$

$$z = -\frac{x^2 + y^2 - 2r_1 xy}{2(1 - r_1^2)}$$

$$r_1 = \frac{M_{23}}{M_{22}}$$

$$h_1 = -Q\left[\frac{M_{12} - M_{13}}{M_{22} - M_{23}}\right]\left[\frac{1 - m^2}{M_{22}}\right]^{\frac{1}{2}} \sin \theta.$$

Finally, by introducing the new variables x_0, y_0 in place of R_1, R_2 with the following transformation

$$R_1 = x_0 + Q \cos \theta \qquad (9)$$

$$R_2 = y_0 + Q \cos \theta, \qquad (10)$$

we find, after some simplifications using Jacobi's[9] theorem, that

$$Q_1^-(\tau,\theta,a) \qquad (11)$$
$$= [2\pi N_\theta]^{-1}[1 - m^2]^{-\frac{1}{2}}M_{22}J(r_1, h_1) \exp\left[\frac{-2a \sin^2 \theta}{1 + m}\right]K(m, h_0),$$

where

$$K(m, h_0) = \frac{1}{2\pi\sqrt{1 - m^2}} \int_{h_0}^\infty dx_0 \int_{h_0}^\infty dy_0 \, e^{z_0}$$

$$z_0 = -\frac{x_0^2 + y_0^2 - 2mx_0y_0}{2(1 - m^2)}$$

$$h_0 = -Q \cos \theta$$

$$h_1 = -Q\left[\frac{m'}{1 + m}\right]\left[\frac{1 - m^2}{M_{22}}\right]^{\frac{1}{2}} \sin \theta.$$

The conditional probability $Q_1^+(\tau,\theta,a) \, d\tau$, the conditional probability that a downward axis-crossing of the level θ occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t, is obtained from (3) by changing the signs of the ∞'s in the limits of integration of θ_1' and θ_2'. We find that $Q_1^+(\tau,\theta,a)$ is equal to the right-hand side of (11) with h_1 replaced by $-h_1$. This latter result also follows from the symmetry relation $Q_1^+(\tau,\theta,a) = Q_1^-(\tau,-\theta,a)$.

The conditional probability $[U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)] \, d\tau$, the conditional probability that an upward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an upward axis-crossing at t, is obtained from (3) by changing the lower limit of integration of θ_1' to $+\infty$. We find that $U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)$ is equal to the right-hand side of (11) with the

function $J(r_1, h_1)$ replaced by the function $J_1(r_1, h_1)$, where

$$J_1(r_1, h_1) = \frac{1}{2\pi \sqrt{1 - r_1^2}} \int_{h_1}^{-\infty} dx \int_{h_1}^{\infty} dy \ (x - h_1)(y - h_1)e^z. \quad (12)$$

The conditional probability that a downward axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given a downward axis-crossing at t is obtained from (3) by changing the upper limit of integration of θ_2' to $-\infty$. The result is that this conditional probability function is equal to the conditional probability function $U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)$ as one would expect from symmetry.

The functions $J(r_1, h_1)$, $K(m,h_0)$, and $J_1(r_1, h_1)$ are expressed in terms of Karl Pearson's [10,11,12] tabulated function (d/N) in Refs. 2 and 3. Thus, the conditional probability functions $Q_1^-(\tau,\theta,a)$, $Q_1^+(\tau,\theta,a)$, and $U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)$ are expressed in terms of well-known tabulated functions.

Since $\theta = 0$ is a level of symmetry, we need only discuss the conditional probabilities when θ is restricted to the interval $0 \leq \theta \leq \pi$. The corresponding results when θ is in the remaining interval $-\pi \leq \theta < 0$ can be deduced from the following symmetry conditions:

$$Q_1^-(\tau,\theta,a) = Q_1^+(\tau, -\theta,a) \quad (13)$$

$$U_1(\tau,\theta,a) - Q_1(\tau,\theta,a) = U_1(\tau, -\theta,a) - Q_1(\tau, -\theta,a). \quad (14)$$

IV. VARIANCE OF THE NUMBER OF AXIS-CROSSINGS IN A TIME τ

For an arbitrary level θ and arbitrary signal-to-noise ratio "a," let $N(\tau,\theta,a)$ denote the number of axis-crossings observed in a time τ. Then, we have that

$$EN(\tau,\theta,a) = 2N_\theta\tau \quad (15)$$

and

$$\text{Var } N(\tau,\theta,a) \equiv EN^2(\tau,\theta,a) - [2N_\theta\tau]^2, \quad (16)$$

where

$$E = \text{Expectation}$$

$$\text{Var} = \text{Variance}.$$

Using McFadden's[13] general result, also see Rice's derivation in Bendat,[14] we have that

$$EN^2(\tau,\theta,a) = 2N_\theta\tau + 4N_\theta \int_0^\tau (\tau - x)U_1(x,\theta,a) \ dx. \quad (17)$$

In this latter equation, $U_1(\tau,\theta,a)\ d\tau$ denotes the conditional probability that an axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ given an axis-crossing at time t. Since the joint probability that an axis-crossing occurs between $t + \tau$ and $t + \tau + d\tau$ and an axis-crossing occurs between t and $t + dt$ can be expressed as

$$2N_\theta U_1(\tau,\theta,a)\ dt\ d\tau = 2N_\theta[U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)]\ dt\ d\tau \\ + N_\theta Q_1^+(\tau,\theta,a)\ dt\ d\tau + N_\theta Q_1^-(\tau,\theta,a)\ dt\ d\tau, \tag{18}$$

we have that

$$U_1(\tau,\theta,a) = [U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)] + \tfrac{1}{2}Q_1^+(\tau,\theta,a) + \tfrac{1}{2}Q_1^-(\tau,\theta,a). \tag{19}$$

Thus, Var $N(\tau,\theta,a)$ can be computed by using (16), (2), (17), and (19)

$$\text{Var } N(\tau,\theta,a) = 2N_\theta\tau + 4N_\theta \int_0^\tau (\tau - x)U_1(x,\theta,a)\ dx - [2N_\theta\tau]^2 \tag{20}$$

$$= 2N_\theta\tau\Big\{1 + 2 \int_0^\tau \Big[1 - \frac{x}{\tau}\Big][U_1(x,\theta,a) - 2N_\theta]\ dx\Big\}. \tag{21}$$

For large τ, (21) becomes

$$\text{Var } N(\tau,\theta,a) \doteq 2N_\theta\tau\Big\{1 + 2 \int_0^\tau [U_1(x,\theta,a) - 2N_\theta]\ dx\Big\}. \tag{22}$$

When twice the value of the integral in (22) is small compared with unity we have that

$$\text{Var } N(\tau,\theta,a) \doteq 2N_\theta\tau. \tag{23}$$

This is the relation one would expect if the axis-crossing points represent a poisson point process for which $U_1(\tau,\theta,a) \equiv 2N_\theta$ for all τ.

Rice[5] assumed a poisson point process for the case $\theta = \pi$ and "a" large in order to use (23) in his analysis of noise in FM receivers. Indeed, for the case of a Gaussian autocorrelation function (22) serves to justify Rice's use of (23) for large τ, $\theta = \pi$, and $a \geqq 4$. For this case, with $a = 4$, numerical integration showed that the value of the integral in (22) is approximately 0.05.

Notice that (22) not only applies to the point process defined by $\theta(t,a)$ but also applies to more general stationary point processes.

Incidentally, the probability function $U_1(\tau,\theta,a)$ can also be used to compute, approximately, the probability density $p_0(\tau,\theta,a)$ of the axis-crossing intervals x_i by using the following basic integral equation of renewal theory:

$$p_0(\tau,\theta,a) = U_1(\tau,\theta,a) - p_0(\tau,\theta,a) * U_1(\tau,\theta,a). \tag{24}$$

The symbol * denotes the convolution operation, that is,

$$f * g \equiv \int_{-\infty}^{\infty} f(t)g(\tau - t) \, dt. \tag{25}$$

Equation (24) is based on the assumption that a given axis-crossing interval is statistically independent of the sum of the previous $(m + 1)$ axis-crossing intervals for all non-negative integral m. A theorem in Paragraph 5.2 shows that the assumption is false when $m = 0$. Thus, (24) can only yield approximate results.

The exact probability density of the axis-crossing intervals x_i is at present unknown. However, the first moment of this probability density is equal to $[2N_\theta]^{-1}$.

V. SOME SPECIAL CASES AND A THEOREM

In this section we shall state some special cases of the conditional probability functions. We shall also present a theorem concerning the dependence of two successive axis-crossing intervals.

5.1 Large τ and Fixed θ, a

As τ becomes large we find that $Q_1^-(\tau,\theta,a)$, $Q_1^+(\tau,\theta,a)$, and $U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)$ approach the value N_θ as one would expect.

5.2 Small τ and Fixed θ, a

By expanding $m(\tau)$ as

$$m(\tau) = 1 - \frac{\beta}{2} \tau^2 + \frac{b_3 \mid \tau^3 \mid}{3!} + \frac{b_4 \tau^4}{4!} + \frac{b_5 \mid \tau^5 \mid}{5!} + \cdots, \tag{26}$$

we find that as $\tau \to 0$ from the right with $b_3 \neq 0$

$$Q_1^-(\tau,\theta,a) \to Q_1^+(\tau,\theta,a) \to \frac{2b_3}{3\beta} \left[\frac{3\sqrt{3} + 2\pi}{12\pi} \right] \tag{27}$$

$$U_1(\tau,\theta,a) - Q_1(\tau,\theta,a) \to \frac{2b_3}{3\beta} \left[\frac{3\sqrt{3} - \pi}{12\pi} \right]. \tag{28}$$

Equation (28) suggests that wiggles having infinite rapidity and infinitesimal amplitude are associated with the phase process $\theta(t,a)$ when $b_3 \neq 0$ or $W_b(f - f_0) = O(f^{-4})$ as $f \to \infty$.

We also find that for small τ with $b_3 = 0$:

$$Q_1^-(\tau,\theta,a) \doteq \frac{b_4 - \beta^2}{4\beta} J(1,h_1)\tau \exp\left[-\frac{\beta}{8} (\tau Q \sin \theta)^2 \right] \tag{29}$$

$$Q_1^+(\tau,\theta,a) \doteq \frac{b_4 - \beta^2}{4\beta} \, J(1,-h_1)\tau \, \exp\left[-\frac{\beta}{8} (\tau Q \sin \, \theta)^2 \right] \quad (30)$$

$$U_1(\tau,\theta,a) - Q_1(\tau,\theta,a) \doteq \frac{b_4 - \beta^2}{4\beta} \, J_1(r_1 \,\, , h_1)\tau \, \exp\left[-\frac{\beta}{8} (\tau Q \sin \, \theta)^2 \right], \quad (31)$$

where

$$h_1 \doteq \frac{\beta Q \sin \, \theta}{\sqrt{b_4 - \beta^2}}.$$

It is interesting to compare the above results with the corresponding results at the level I of a Gaussian process $I(t)$ having the normalized autocorrelation function $m(\tau)$. That is, compare the above results with Rice's[2] equation (63) or Rainal's[6] equations (44), (52), (53), and (54) when $I = Q \sin \, \theta$. The results are identical.

Thus, a theorem[6] concerning the dependence of two successive axis-crossing intervals of the Gaussian process $I(t)$ also applies to the phase process $\theta(t,a)$. That is, if $\theta(t,a)$ is a phase process, defined in paragraph one, having a finite expected number of axis-crossing points per unit time at any level θ, then two successive axis-crossing intervals at that level θ are statistically dependent.

The theorem implies that successive axis-crossing points do not form a Markov or Poisson point process.

5.3 $Q_1^+(\tau,\theta,a)$ for small τ, $b_3 = 0$, and large $Q \sin \, \theta$

For small τ and large $Q \sin \, \theta$ with $b_3 = 0$ or $W_b(f - f_0) \neq O(f^{-4})$ as $f \rightarrow \infty$, we find from (30) that

$$Q_1^+(\tau,\theta,a) \doteq \frac{\beta}{4} \, \tau(Q \sin \, \theta)^2 \, \exp\left[-\frac{\beta}{8} (\tau Q \sin \, \theta)^2 \right]. \quad (32)$$

Thus, $Q_1^+(\tau,\theta,a)$ is approximated by a Rayleigh probability density identical to Rice's[2] equation (65) when $I = Q \sin \, \theta$.

5.4 $a = 0$ and arbitrary θ, τ

When $a = 0$ we find that

$$Q_1^-(\tau,\theta,0) = Q_1^+(\tau,\theta,0)$$

$$= 2\beta^{-\frac{1}{2}}[1 - m^2]^{-\frac{3}{2}} \frac{M_{22}}{(2\pi)^2} \, [r_1(\pi - \cos^{-1} \, r_1) \quad (33)$$

$$+ \sqrt{1 - r_1^2}][\pi - \cos^{-1} \, m]$$

$$U_1(\tau,\theta,0) - Q_1(\tau,\theta,0) = 2\beta^{-\frac{1}{2}}[1 - m^2]^{-\frac{3}{2}} \frac{M_{22}}{(2\pi)^2} [-r_1 \cos^{-1} r_1$$
$$+ \sqrt{1 - r_1^2}][\pi - \cos^{-1} m], \tag{34}$$

where

$$0 \leqq \cos^{-1} r_1 \leqq \pi$$
$$0 \leqq \cos^{-1} m \leqq \pi.$$

Thus, when $a = 0$ the conditional probabilities are independent of the level θ as one would expect.

5.5 *Large a, $\theta = 0$, and arbitrary τ*

When "a" is large and $\theta = 0$ we find that

$$Q_1^-(\tau,0,a) = Q_1^+(\tau,0,a) \tag{35}$$
$$\doteq \beta^{-\frac{1}{2}}[1 - m^2]^{-\frac{1}{2}} \frac{M_{22}}{2\pi} [r_1(\pi - \cos^{-1} r_1) + \sqrt{1 - r_1^2}]$$

$$U_1(\tau,0,a) - Q_1(\tau,0,a) \tag{36}$$
$$\doteq \beta^{-\frac{1}{2}}[1 - m^2]^{-\frac{1}{2}} \frac{M_{22}}{2\pi} [-r_1 \cos^{-1} r_1 + \sqrt{1 - r_1^2}].$$

Thus, when "a" is large and $\theta = 0$, the conditional probabilities are independent of "a".

Again, it is interesting to compare the above results with the corresponding results at the level $I = 0$ of a Gaussian process $I(t)$ having the normalized autocorrelation function $m(\tau)$. That is compare the above results with Rice's[2] equations (62) and (85a). The results are identical. One would expect identical results from Rice's[1] equation (3.6).

5.6 *$\theta = \pi$ and arbitrary a, τ*

When $\theta = \pi$ we find that

$$Q_1^-(\tau,\pi,a) = Q_1^+(\tau,\pi,a)$$
$$= [\beta^{\frac{1}{2}}(1 - m^2)^{\frac{3}{2}}\Phi(-Q)]^{-1} \frac{M_{22}}{2\pi} [r_1(\pi - \cos^{-1} r_1)$$
$$+ \sqrt{1 - r_1^2}]K(m,Q) \tag{37}$$

$$U_1(\tau,\pi,a) - Q_1(\tau,\pi,a)$$

$$= [\beta^{\frac{1}{2}}(1 - m^2)^{\frac{1}{2}}\Phi(-Q)]^{-1} \frac{M_{22}}{2\pi} [-r_1 \cos^{-1} r_1 + \sqrt{1 - r_1^2}]K(m,Q). \quad (38)$$

VI. RESULTS FOR A GAUSSIAN AUTOCORRELATION FUNCTION

For purposes of computation we shall take $W_b(f - f_0)$ and $m(\tau)$ as follows:

$$W_b(f - f_0) = \frac{1}{\sigma_b \sqrt{2\pi}} \exp\left[\frac{-(f - f_0)^2}{2\sigma_b^2}\right] \quad (39)$$

and

$$m(\tau) = \exp\left[\frac{-(2\pi\sigma_b\tau)^2}{2}\right]. \quad (40)$$

This particular selection was also made by Rice[2] and Rainal[4] in their study of the duration of fades associated with the Rayleigh process $R(t,a)$.

From (40) we see that it is convenient to define normalized time as $u_b = 2\pi\sigma_b\tau$. All our results are plotted with respect to normalized time u_b. The units of N_θ are now "crossings per unit of normalized time."

Figs. 3 through 11 present the resulting conditional probability functions for various values of the level θ and for various values of signal-to-noise power ratio "a". For large values of u_b all of the conditional probability functions approach the value of N_θ in accordance with Paragraph 5.1.

Figs. 9 and 11 compare $Q^+(\tau,\theta,a)$ for $\theta = \pi/2$ and $a = 4$, 10 with a corresponding Rayleigh density in accordance with (32). Thus, we conclude that the Rayleigh probability density is a good approximation when τ is small and $Q \sin \theta = \sqrt{2a} \sin \theta \geq 2\sqrt{2}$.

Fig. 7 compares well with Figs. 2 and 3 of Ref. 6. Thus, we conclude that (35) and (36) are good approximations when $a \geq 4$.

VII. CONCLUSIONS

The theoretical probability functions $Q_1^-(\tau,\theta,a)$, $Q_1^+(\tau,\theta,a)$, and $U_1(\tau,\theta,a) - Q_1(\tau,\theta,a)$ are expressible in terms of well-known tabulated functions. These results can be used to compute Var $N(\tau,\theta,a)$, the variance of the number of axis-crossing points observed in a time τ. These results can also be used to compute, approximately, the probability density of axis-crossing intervals x_i via renewal theory. The exact probability density is at present unknown.

Fig. 3 — Plots of the probability functions $Q_1^-(u_b, \theta, a)$, $Q_1^+(u_b, \theta, a)$ and $U_1(u_b, \theta, a) - Q_1(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.

Fig. 4 — Plots of the probability function $Q_1^-(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.

Fig. 5 — Plots of the probability function $Q_1^+(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.

Fig. 6 — Plots of the probability function $U_1(u_b, \theta, a) - Q_1(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having auto-correlation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.
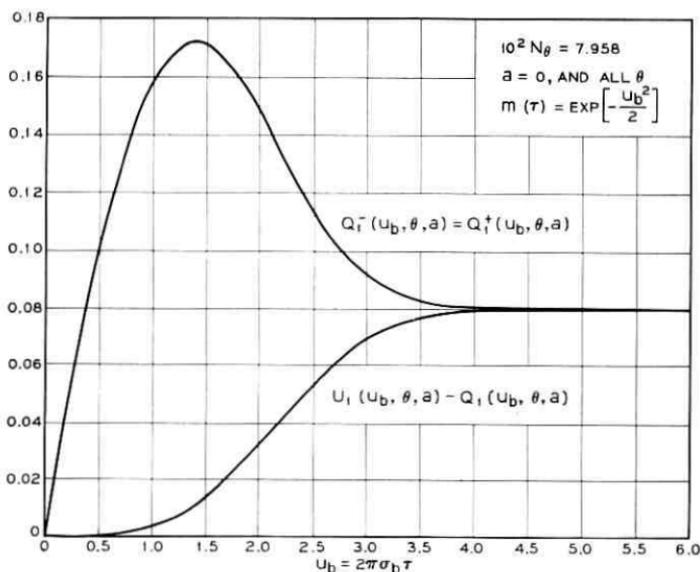
750
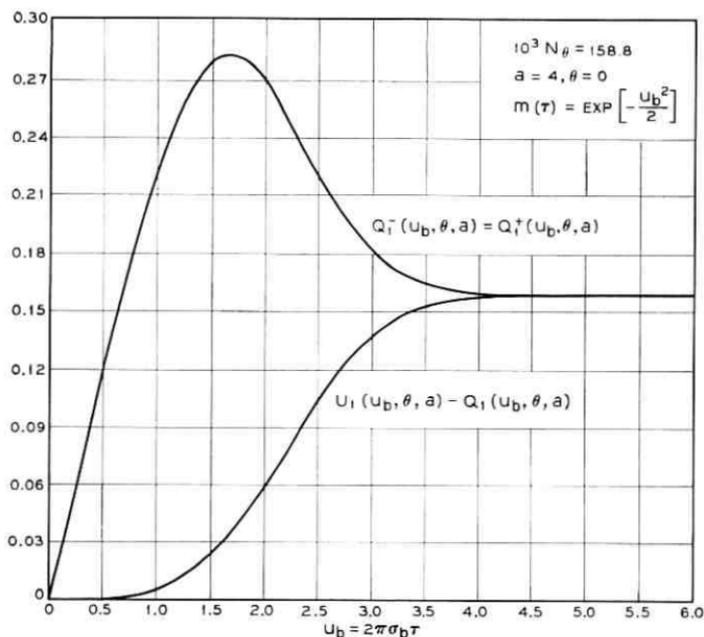
Fig. 7 — Plots of the probability functions $Q_1^-(u_b, \theta, a)$, $Q_1^+(u_b, \theta, a)$ and $U_1(u_b, \theta, a) - Q_1(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.
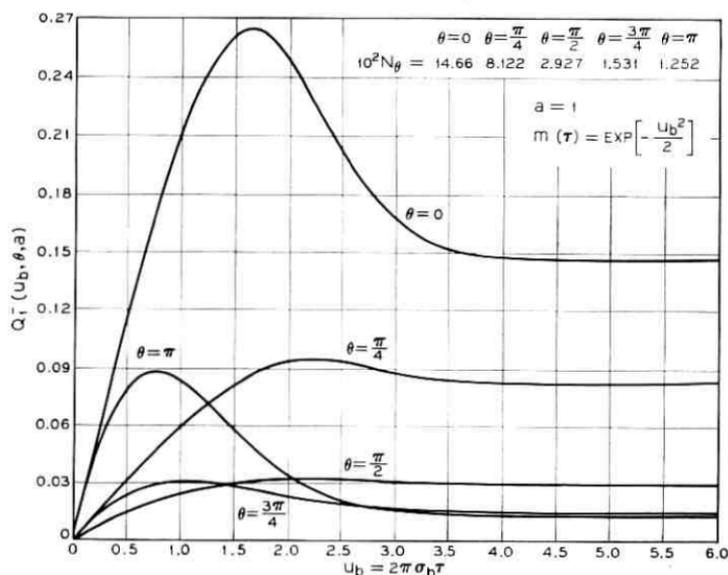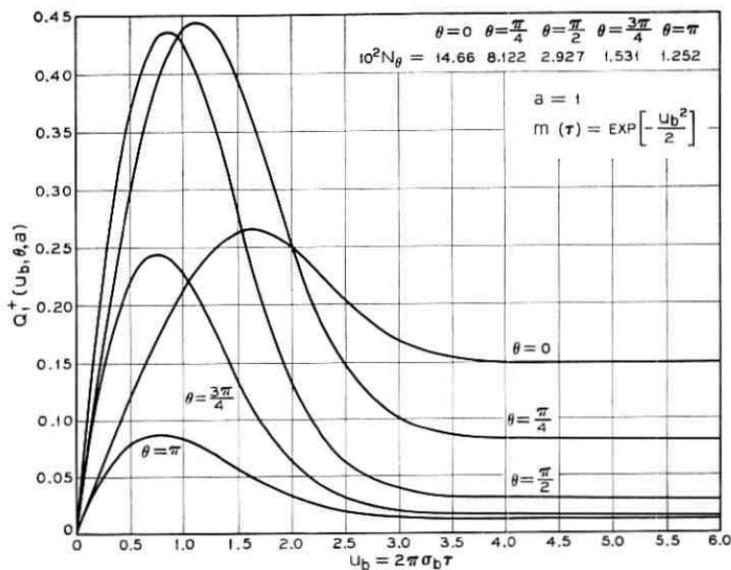

Fig. 8 — Plots of the probability function $Q_1^-(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.

751
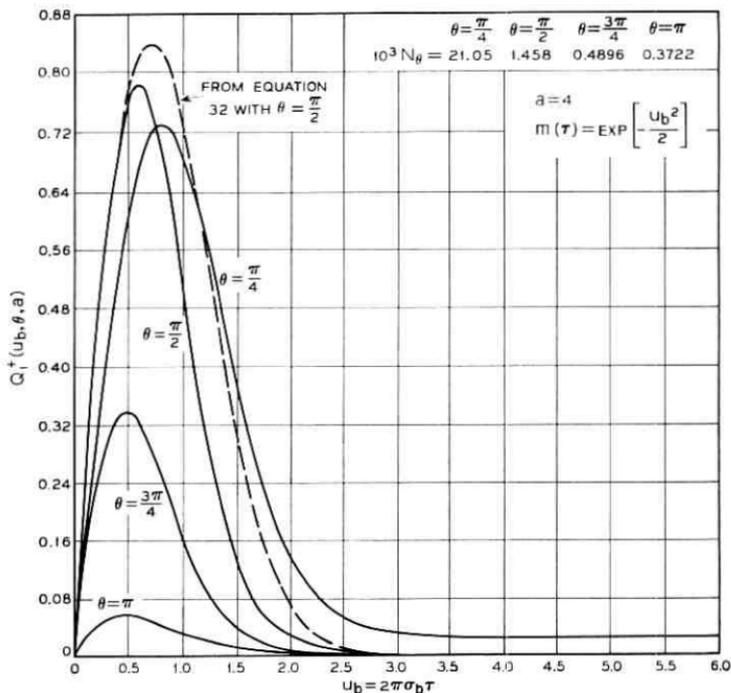
Fig. 9 — Plots of the probability function $Q_1^+(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.
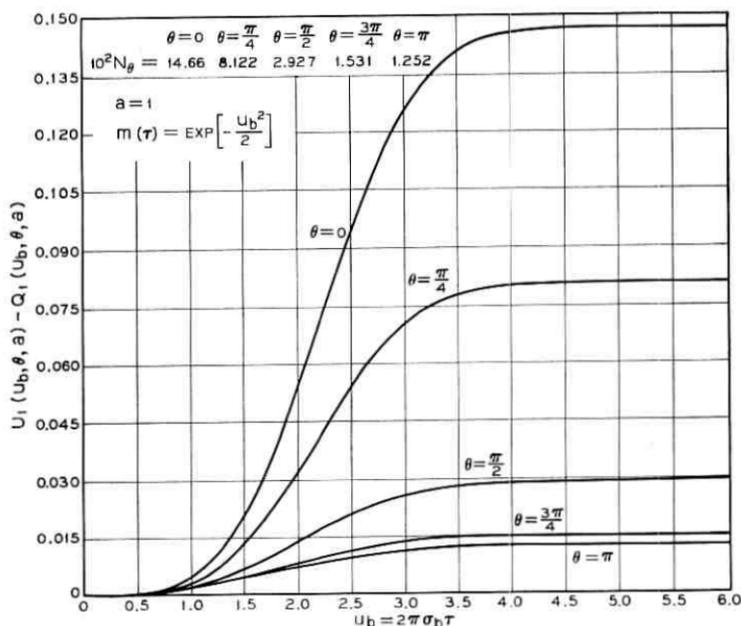

Fig. 10 — Plots of the probability function $U_1(u_b, \theta, a) - Q_1(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.

752
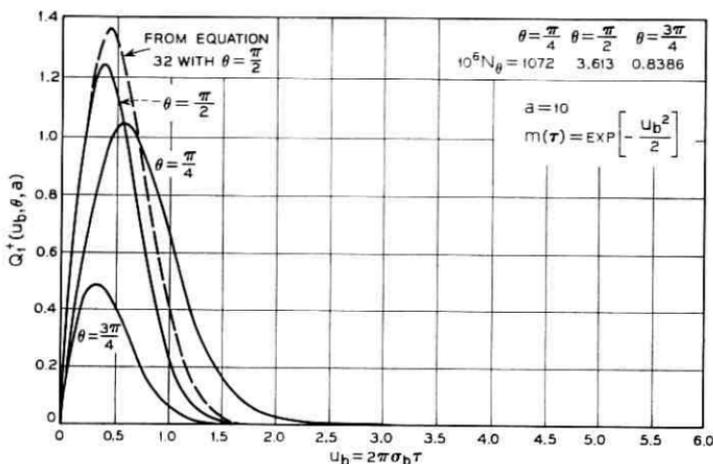
Fig. 11 — Plots of the probability function $Q_1^+(u_b, \theta, a)$ associated with the axis-crossing points defined by the level θ of $\theta(t,a)$. $\theta(t,a)$ denotes the resultant phase of a sinusoidal signal plus narrow-band Gaussian noise having autocorrelation function $m(\tau)$. "a" denotes the signal-to-noise power ratio.
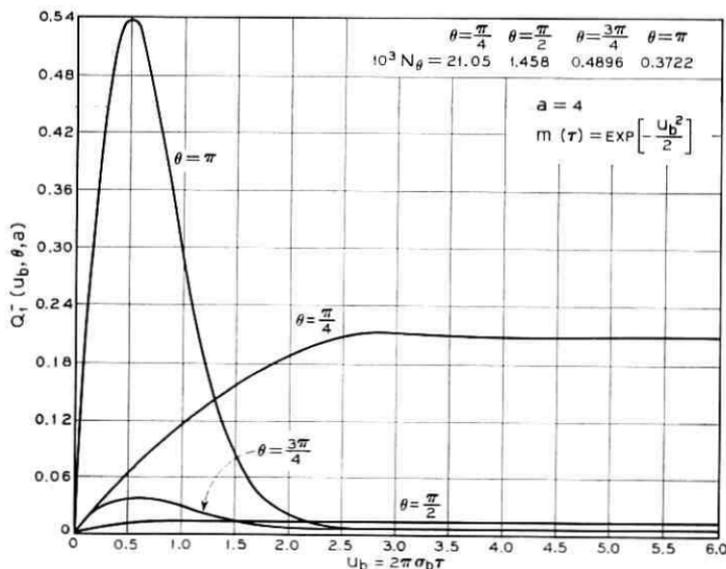
Because the level $\theta = 0$ is a level of symmetry, results for $0 \leqq \theta \leqq \pi$ imply results for $-\pi \leqq \theta < 0$.

When $W_b(f - f_0) = O(f^{-4})$ as $f \to \infty$, wiggles having infinite rapidity and infinitesimal amplitude are associated with the phase process $\theta(t,a)$.

When $\theta = 0$ with the signal-to-noise power ratio $a \geqq 4$, the conditional probability functions associated with the phase process $\theta(t,a)$ are equal, approximately, to the corresponding results for a certain Gaussian process.

When $W_b(f - f_0) \neq O(f^{-4})$ as $f \to \infty$, and $Q \sin \theta$ is large, $Q_1^+(\tau,\theta,a)$ for small τ is approximated by a Rayleigh probability density.

When $\theta = \pi$ and $a \geqq 4$, Var $N(\tau,\theta,a)$ for large τ is equal, approximately, to $2N_\theta \tau$, the variance resulting from a poisson point process.

When N_θ is finite, two successive axis-crossing intervals of $\theta(t,a)$ are statistically dependent. Thus, the axis-crossing points do not represent exactly a Markov point process or a poisson point process.

VIII. ACKNOWLEDGMENT

REFERENCES

1. Rice, S. O., Statistical Properties of a Sine Wave Plus Random Noise, B.S.T.J., *27*, January, 1948, pp. 109–157.
2. Rice, S. O., Distribution of the Duration of Fades in Radio Transmission, B.S.T.J., *37*, May, 1958, pp. 581–635.
3. Rainal, A. J., Axis-Crossing Intervals of Rayleigh Processes, B.S.T.J., *44*, July-August, 1965, pp. 1219–1224.
4. Rainal, A. J., Duration of Fades Associated with Radar Clutter, B.S.T.J., *45*, October, 1966, pp. 1285–1298.
5. Rice, S. O., Noise in FM Receivers, in *Time Series Analysis*, M. Rosenblatt, Editor, John Wiley and Sons, Inc., New York, N. Y., 1963, Chapter 25, pp. 395–422.
6. Rainal, A. J., Zero Crossing Intervals of Envelopes of Gaussian Processes, Technical Report, No. AF-110, DDC No. AD-601-231, The Johns Hopkins University, Carlyle Barton Laboratory, Baltimore, Maryland, June, 1964. Abstracted in IEEE Trans. Inform. Theor., *IT-11*, January, 1965, p. 159.
7. Tikhonov, V. I., Mean Number of Frequency and Phase Surges, Radio Eng. Elec. Phys., No. 6, June, 1962, p. 888, Equation 23.
8. Blachman, N. M., FM Reception and the Zeros of Narrowband Gaussian Noise, IEEE Trans. Inform. Theor., *IT-10*, July, 1964, pp. 235–241.
9. Aitken, A. C., *Determinants and Matrices*, Interscience Publishers, Inc., New York, N. Y., 1958, p. 97.
10. Pearson, Karl, ed., *Tables for Statisticians and Biometricians*, Cambridge University Press, 1931, Part II, Table VIII, Vols. of Normal Bivariate Surface, pp. 78–109.
11. National Bureau of Standards (1959), Tables of the Bivariate Normal Distribution Function and Related Functions, Applied Math. Series 50, U. S. Government Printing Office, Washington 25, D. C.
12. Gupta, S. S., Probability Integrals of Multivariate Normal and Multivariate t, Ann. Math. Stat., *34*, No. 3, September, 1963, p. 792.
13. McFadden, J. A., On the Lengths of Intervals in a Stationary Point Process, J. Royal Stat. Soc., Series B, *24*, No. 2, p. 370, 1962, Equation 3.4.
14. Bendat, J. S., *Random Noise Theory*, John Wiley and Sons, Inc., New York, N. Y., 1958, p. 396.

Operating Characteristics for a Linear Detector of CW Signals in Narrow-Band Gaussian Noise

By G. H. ROBERTSON

This article presents a set of curves that can be used to evaluate in detail over a wide range of operating conditions the performance of systems using envelope detectors to search for CW signals in narrow-band Gaussian noise. The charts discussed relate the probability, P_D, of detecting such signals to (i) the signal-power/noise-power ratio, S/N, (ii) the proportion of false detections, P_{FA}, and (iii) the number, M, of independent samples of the envelope of the combined signal and noise that are averaged in making one attempt at detection. The curves and scales were calculated and drawn entirely by computer. The computation program was designed so that nearly linear curves could be produced, thereby increasing the accuracy and ease of interpolation.

I. INTRODUCTION

The curves* given in this article relate statistical properties of the output of an envelope detector to the signal-power/noise-power ratio, S/N, of a CW signal in narrow-band Gaussian noise at the input.

The signal will be represented here by a finite section of a sine wave which may be divided further into segments, for convenience in processing, so that most of the signal energy will be contained in a relatively narrow band, comparable in width to the reciprocal of the segment length. A filter will be used to select the part of the spectrum in which the signal may be found. The output of this filter will be applied to an envelope detector, and the amplitude of the output of the detector will be measured. Only if the magnitude measured at this

---

755

point exceeds a chosen threshold value will a signal be presumed present.

For other detection tests two or more samples of the output of the detector, taken far enough apart in time that the noise values are independent, will be averaged. Only if the value of such an average exceeds a chosen threshold will a signal be presumed present. The curves given here for the averages of multiple samples depend on the assumption that a detected signal remains steady while the samples are taken, and that the noise background is weakly stationary.

This form of detection criterion has been discussed by several authors,[1, 2, 3, 4] who give major emphasis to the use of a square-law detector because it is easier to analyze. A linear detector, however, is much more practical and is extensively used, so results that specifically apply to it are desirable.

The curves available in the literature are difficult to use for checking system performance at more than a few scattered points. Those given here, on the other hand, are believed to be suitable for defining system performance in detail over a considerable range of threshold values and, furthermore, are easy to use.

II. DISCUSSION OF THE MATHEMATICAL MODEL

We assume that the spectrum of noise in the vicinity of the signal is smooth enough so that the output of the filter can be considered a narrowband Gaussian process with the possible addition of a sinusoidal signal. S. O. Rice[5, 6] derived an expression for the probability density of the envelope of such a waveform. Rice's expression, paraphrased, is

$$P(\rho) \, d\rho = 2h^2\rho \exp\left[-h^2(r^2 + \rho^2)\right]I_0(2h^2r\rho) \, d\rho \qquad (1)$$

which gives the probability that the envelope lies within an interval, $d\rho$, of ρ. $P(\rho)$ is thus the probability density. $I_0(z)$ is the modified Bessel function of zero order defined by

$$I_0(z) = \sum_0^\infty \frac{(\frac{1}{2}z)^{2n}}{(n!)^2}. \qquad (2)$$

Note that in (1) the term h^2r^2, which represents the signal/noise power ratio, and $1/\sqrt{2}h$, which represents the rms noise level, can be identified. In this article S/N always means the ratio of the signal power to the noise power accompanying it in the specified narrow band.

When there is no signal (1) becomes

$$P(\rho) \, d\rho = 2h^2 \rho \exp\left(-h^2 \rho^2\right) d\rho \tag{3}$$

which represents a Rayleigh distribution having

$$2\sigma^2 = 1/h^2 \tag{4}$$

for the second moment about zero.

The distribution of the envelope of the filter output is thus given by (1) when a signal is present, and by (3) when there is no signal. From these expressions distribution curves can be drawn for cases where only noise is present, and where a known signal occurs with the noise. To generate the performance charts given here, one has to be able to calculate the area lying above some threshold for distribution curves corresponding to no signal, and for those corresponding to all the values of S/N of interest.

Only in the case of noise alone, for single-sample detection, is it possible to get an explicit formula for the area under portions of such a curve. Moments of the distribution given in (1) can be calculated with the help of Ref. 7. We give here the nth moment about zero:

$$\nu_n = \int_0^\infty P(\rho)\rho^n \, d\rho$$

$$= 2 \exp\left(-h^2 r^2\right) \int_0^\infty h^2 \rho^{n+1} \exp\left(-h^2 \rho^2\right) I_0(2h^2 r\rho) \, d\rho \tag{5}$$

$$= (\sqrt{2}\sigma)^n \Gamma\left(\frac{n+2}{2}\right) {}_1F_1\left(-\frac{n}{2}\, ; 1;\, -h^2 r^2\right).$$

In this $h^2 r^2$ is the signal/noise power and

$$2\sigma^2 = 1/h^2$$

as stated earlier.

$\Gamma(z)$ is the Gamma function and ${}_1F_1\,(a;b;z)$ is the Confluent Hypergeometric function defined by

$$
\begin{aligned}
{}_1F_1(a;b;z) = 1 &+ \frac{a}{b}\cdot\frac{z}{1} + \frac{a(a+1)}{b(b+1)}\cdot\frac{z^2}{2\cdot 1} \\
&+ \frac{a(a+1)(a+2)}{b(b+1)(b+2)}\cdot\frac{z^3}{3\cdot 2\cdot 1} + \cdots .
\end{aligned}
\tag{6}
$$

When the signal/noise power becomes zero we have

$$\nu_{0_n} = (\sqrt{2}\sigma)^n \Gamma\left(\frac{n+2}{2}\right). \tag{7}$$

This gives the nth moment about zero for the distribution due to noise alone. We therefore have

$$\nu_n = \nu_{0_n} {}_1F_1\left(-\frac{n}{2}; 1; -h^2r^2\right) \tag{8}$$

which gives the nth zero moment of the distribution when $S/N = h^2r^2$. We can thus derive a type-A Gram-Charlier[8] (G-C) series approximation to the distribution curve for any S/N, and, since the series can be integrated, the area under any desired part of the curve can be found. Knowing the moments of the primary distribution, we can calculate the moments of a secondary distribution formed by averaging M samples of the primary distribution. We thus form a G-C series which can be integrated to get the area under any desired part of the secondary distribution curve.

Charts are included that cover cases in which, when only noise is present, the probability is as low as 10^{-6} that the threshold will be exceeded, even with the use of single samples in detection attempts.

Producing a G-C series that would give results within 1 percent of the true value when the threshold is great enough to give such a low probability required evaluating many more moments than is usual in G-C series. In fact, 31 moments were calculated, which resulted in a series that oscillates about the true value of the distribution as the threshold is varied. A method that sometimes reduces these oscillations is to use special groupings of the terms, forming a so-called Edgeworth series.[8] It occurred to the author, however, that a better fit might be obtained by averaging the approximations given by several G-C series carried out to different numbers of terms since their oscillations as the threshold varies are not in phase. This involves very little more work than deriving the series that has the most terms, since it merely requires that the terms be weighted to represent the number of times they occur in forming the average. Fig. 1 shows the results obtained by forming G-C series that use up to 31 moments and averaging over the highest 23 and 13 terms, respectively. Both show less oscillation over restricted—though appreciable—ranges of threshold than does the Edgeworth series that uses up to 30 moments. The curve showing

CURVE ABOVE IS LOG₁₀ Q (READ RIGHT SCALE)

CURVES BELOW GIVE THE PER-CENT DEVIATION FROM Q (READ LEFT SCALE)

Q IS THE PROBABILITY OF EXCEEDING THRESHOLD (RAYLEIGH DISTRIBUTION)

A AVERAGING SERIES FOR 9 THROUGH 31 MOMENTS
B AVERAGING SERIES FOR 19 THROUGH 31 MOMENTS
C EDGEWORTH SERIES TO MOMENT 30

THRESHOLD ⟶
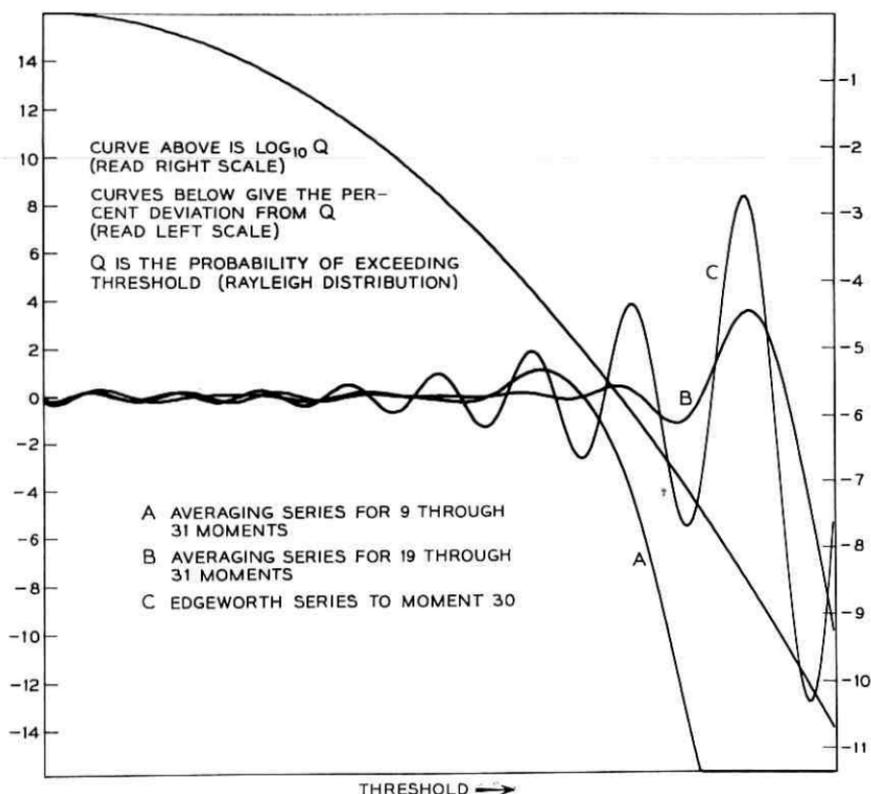
Fig. 1 — Comparison of methods for evaluating G-C series.

the logarithm of Q, the function approximated by the series, allows the useful range to be assessed. It makes possible the calibration of the horizontal scale, which is the same for all four curves, in terms of the probability that the threshold will be exceeded. The logarithm to the base ten of this probability is given by the right scale of Fig. 1.

III. DESCRIPTION OF THE CHARTS

We will now describe briefly how the charts (Figs. 2 through 15) were produced and point out their main features. Some of these are believed novel and show the advantage of being able to use a digital computer to generate and draw charts of this type.

For each value of S/N at which a curve was desired, the detection threshold, T, was expressed in two ways: (i) measured with respect

to the mean of the distribution for noise only, and standardized by dividing by the square root of the variance, and (*ii*) measured with respect to the mean of the distribution appropriate to the S/N, and standardized with respect to this distribution.

The form in (*i*) is appropriate for calculating the probability of false alarm (i.e., the probability that noise alone will cause the threshold to be exceeded), and the form in (*ii*) is appropriate for calculating the probability of detection (the probability that a signal at the chosen
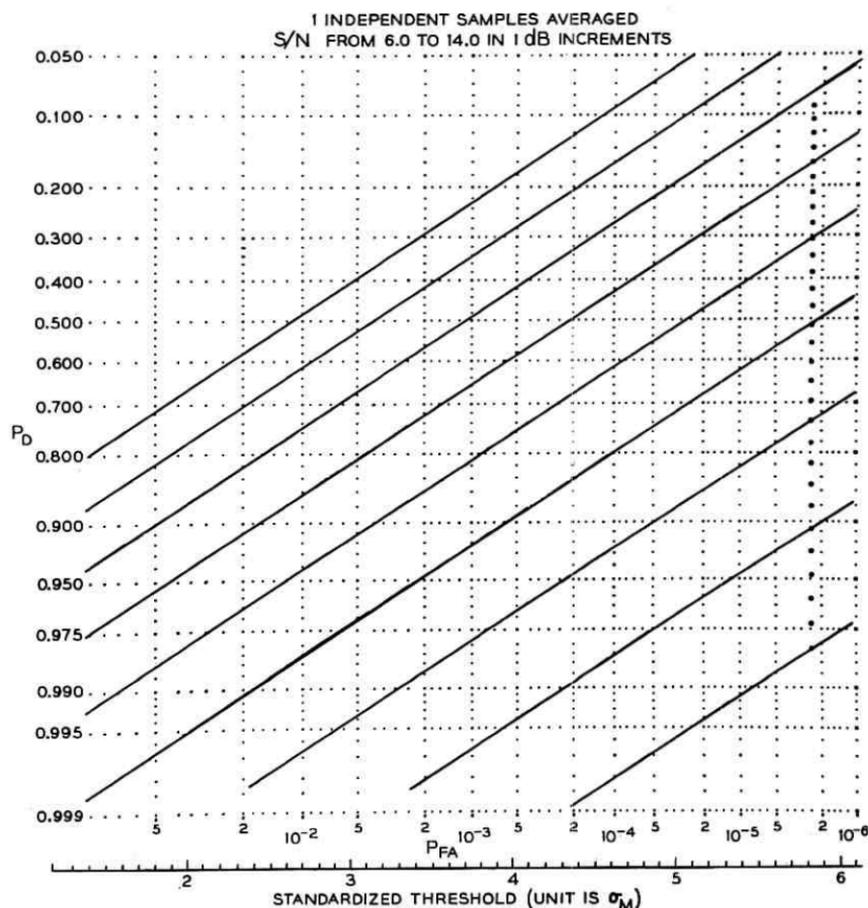
Fig. 2 — ROC curves.
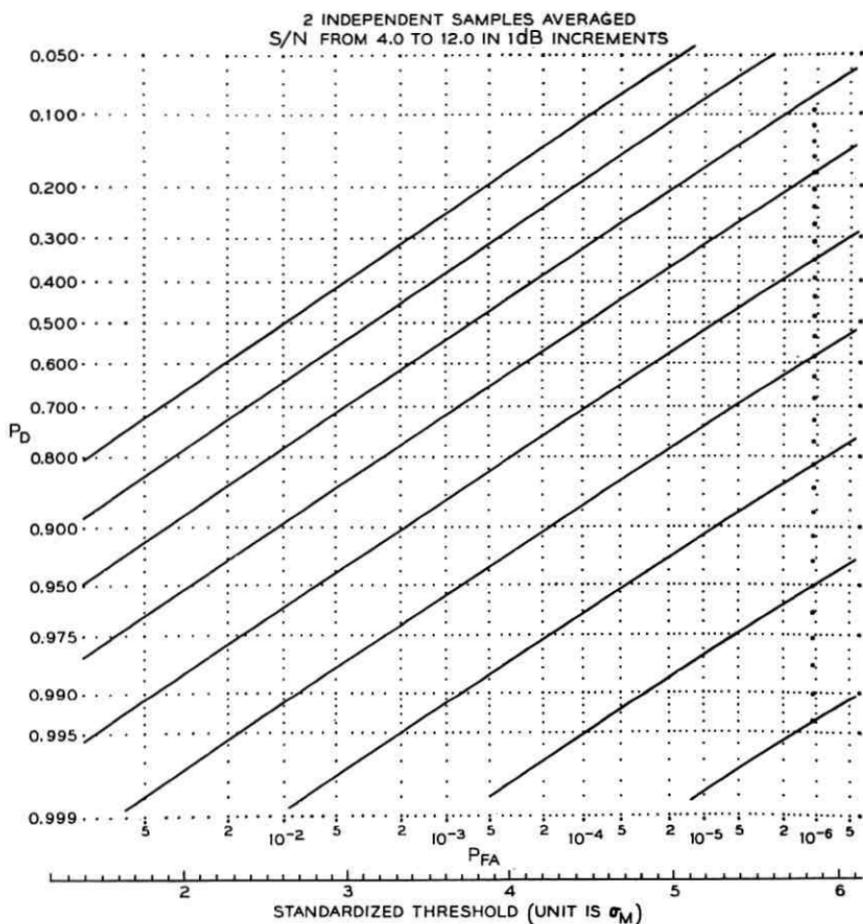
Fig. 3 — ROC curves.

S/N will cause the threshold to be exceeded). The relationship between these two forms is linear, so that if the threshold were varied a straight line would give the relationship between a scale for measuring the variation based on (i) and another based on (ii). If these scales were calibrated in terms of probability of false alarm, P_{FA}, and probability of detection, P_D, useful charts could be obtained that use these probability parameters as horizontal and vertical axes, respectively.

But, for any P_{FA} scale, the corresponding P_D scale depends on S/N. Fortunately, the P_D scale does not change rapidly with S/N for the range we are interested in, and, to allow several S/N values to be included on each chart, a constant normal probability scale was chosen for P_D. The penalty incurred by doing this is the slight curvature of some of the S/N lines plotted in Figs. 2 through 15.

It can be seen that the horizontal scale, P_{FA}, changes as the number of samples averaged changes. This is because the probability dis-
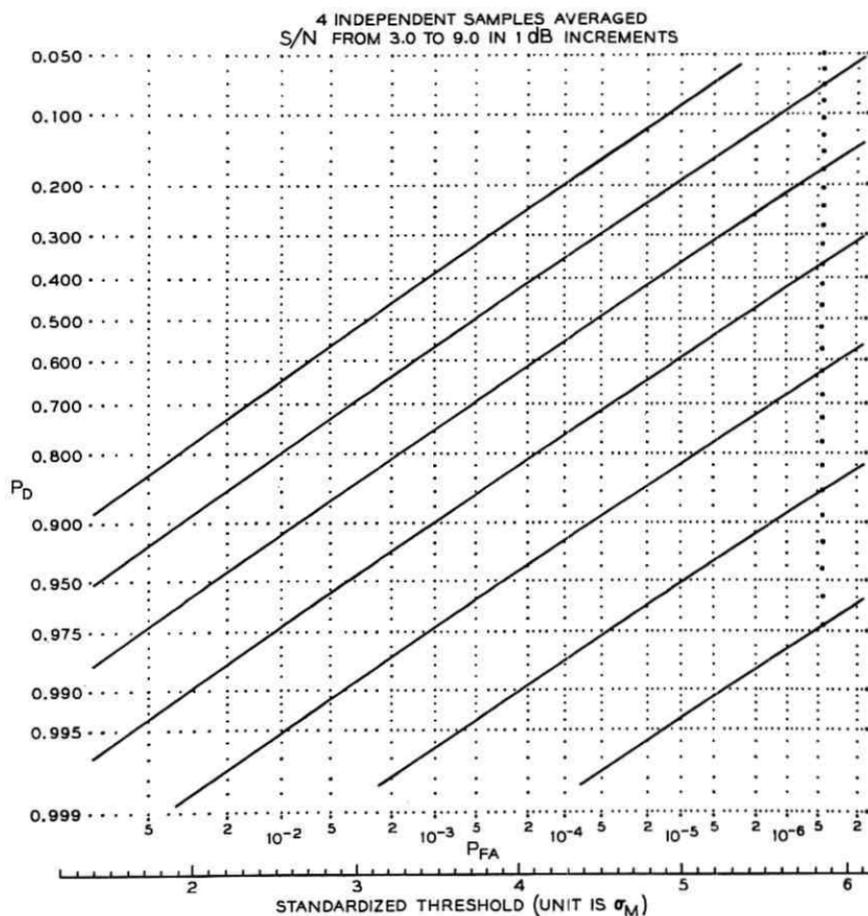
Fig. 4 — ROC curves.
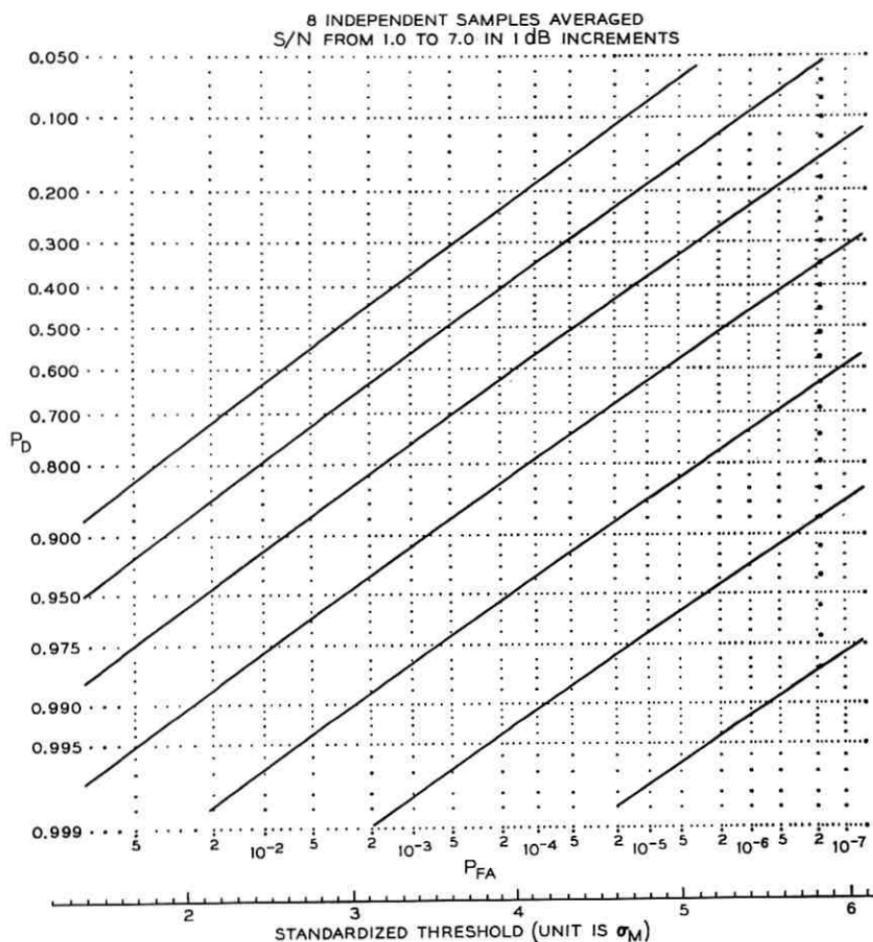
Fig. 5 — ROC curves.

tribution for the average of a number of samples depends on the number averaged as well as on the basic distribution. Since an explicit formula for the cumulative probability is known in the case of the basic (Rayleigh) distribution, the Rayleigh distribution was used to get the P_{FA} scale for the chart based on single-sample decisions. All the other scales were calculated from G-C series approximations to the appropriate distributions.

To make interpolation easier, a column of heavy dots was put near the right edge of each chart. These dots mark 0.2 dB increments in S/N between the lines. Since any two adjacent S/N lines are very nearly straight and parallel it is easy to interpolate to within 0.1 dB between the lines using a parallel ruler guided by the heavy dots.

The horizontal and vertical grids were plotted as rows and columns of dots in order to carry the accuracy of the scales into all parts of the chart without also producing a confusing mesh of lines.
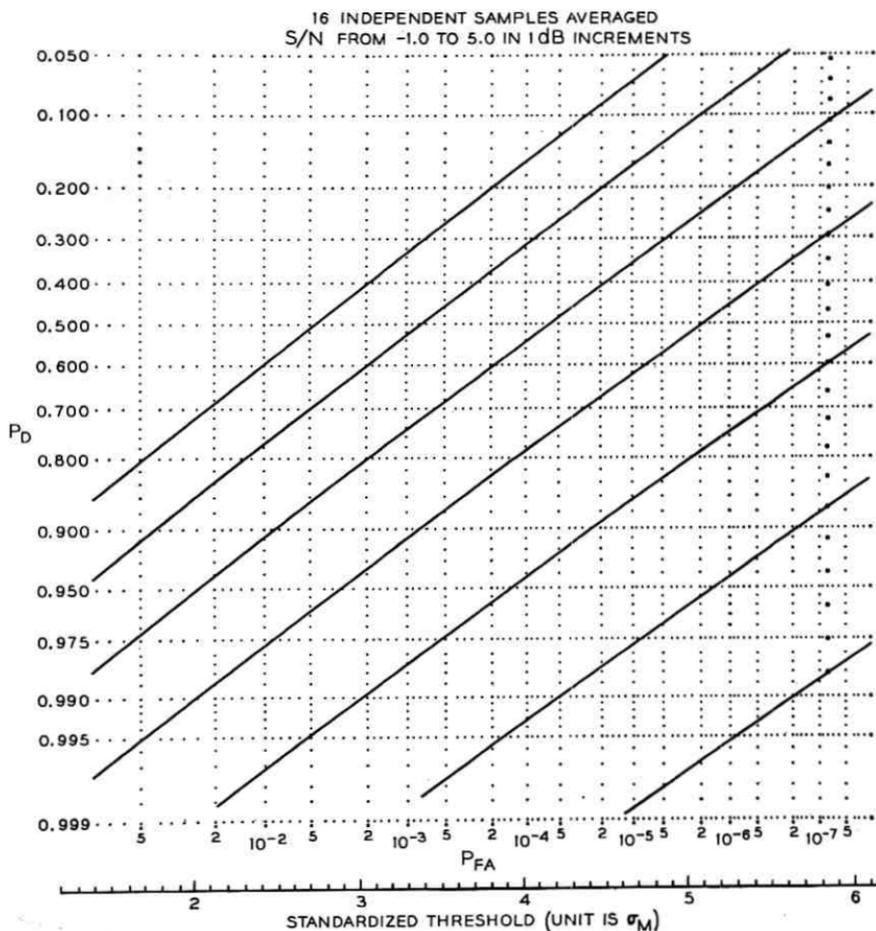
Fig. 6 — ROC curves.
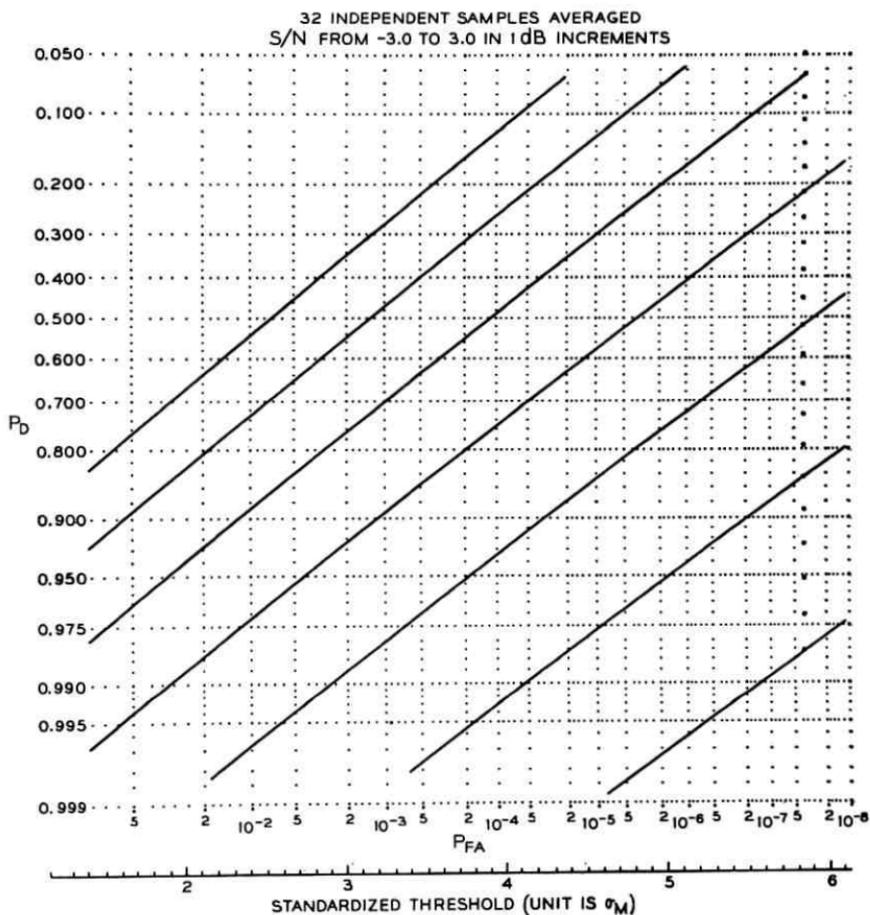
Fig. 7 — ROC curves.

Each decade of the horizontal, P_{FA}, scale, reading from left to right, is given by

$$10.0\,(1.0)\,4.0\,(0.5)\,2.0\,(0.2)\,1.0,$$

where the increments are in parentheses.

The vertical, P_D, scale, reading from top to bottom, is

$$0.05\,(0.01)\,0.10\,(0.02)\,0.90\,(0.01)\,0.950\,(0.005)\,0.980$$

$$0.980\,(0.0025)\,0.990\,(0.001)\,0.998\,(0.0005)\,0.999,$$

the increments again being in parentheses.

On Fig. 16 are curves for two pairs of P_D, P_{FA} values that show how the S/N for each pair varies with the logarithm to the base 2 of the
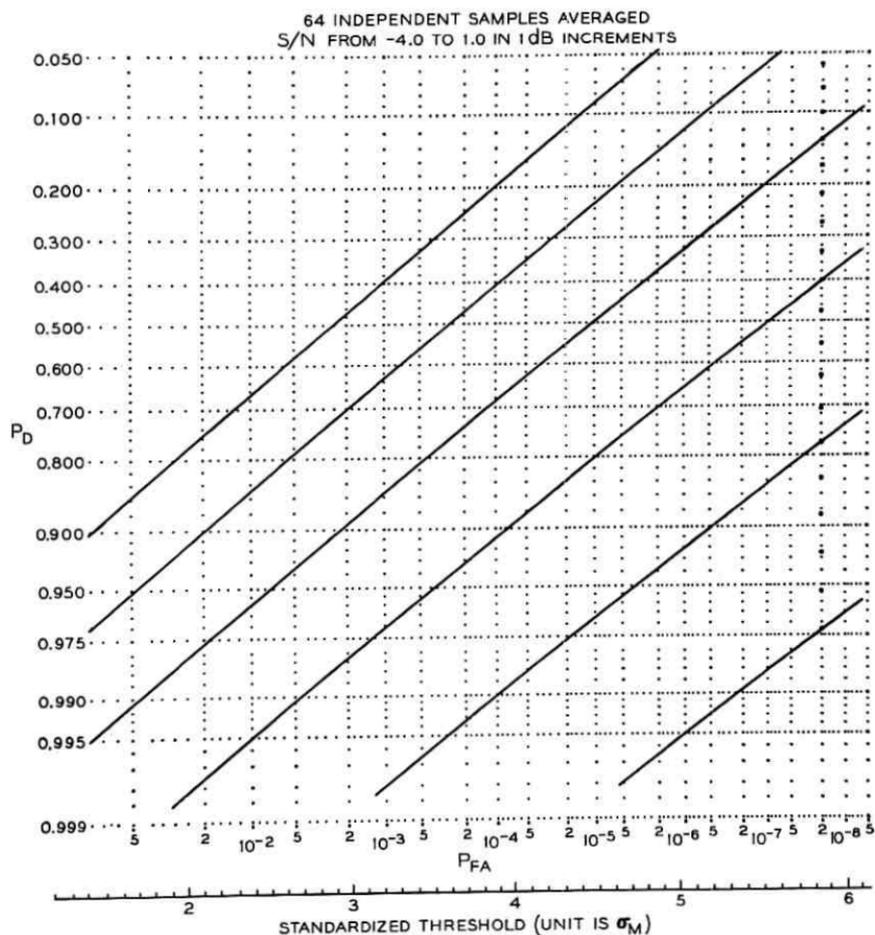
Fig. 8 — ROC curves.
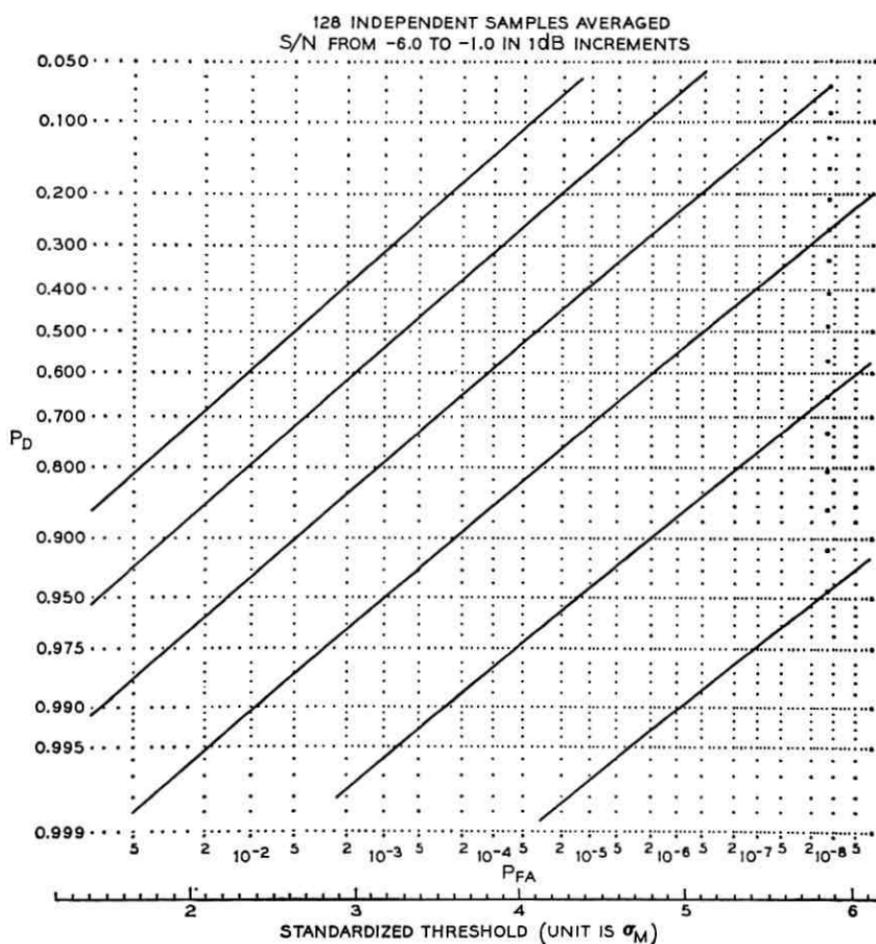
Fig. 9 — ROC curves.

number of samples averaged. Over any restricted range in which the number of samples is doubled, either curve could be approximated quite satisfactorily by a straight line.

The slopes of these two curves are almost the same at the same S/N, although the P_D, P_{FA} values are quite different. This means that, within any small range of P_D, P_{FA} values, curves like those shown in Fig. 16 can be assumed parallel with little sacrifice in accuracy. Con-

sequently, the change in S/N corresponding to a small change in $\log_2 M$ indicated by such a curve will hold quite accurately for a small range of P_D, P_{FA} values around the pair for which the curve is actually drawn.

The S/N corresponding to any pair of P_D, P_{FA} values can therefore be calculated for a different number of samples averaged by assuming a linear relationship between S/N and the logarithm to the base 2 of the number of samples.
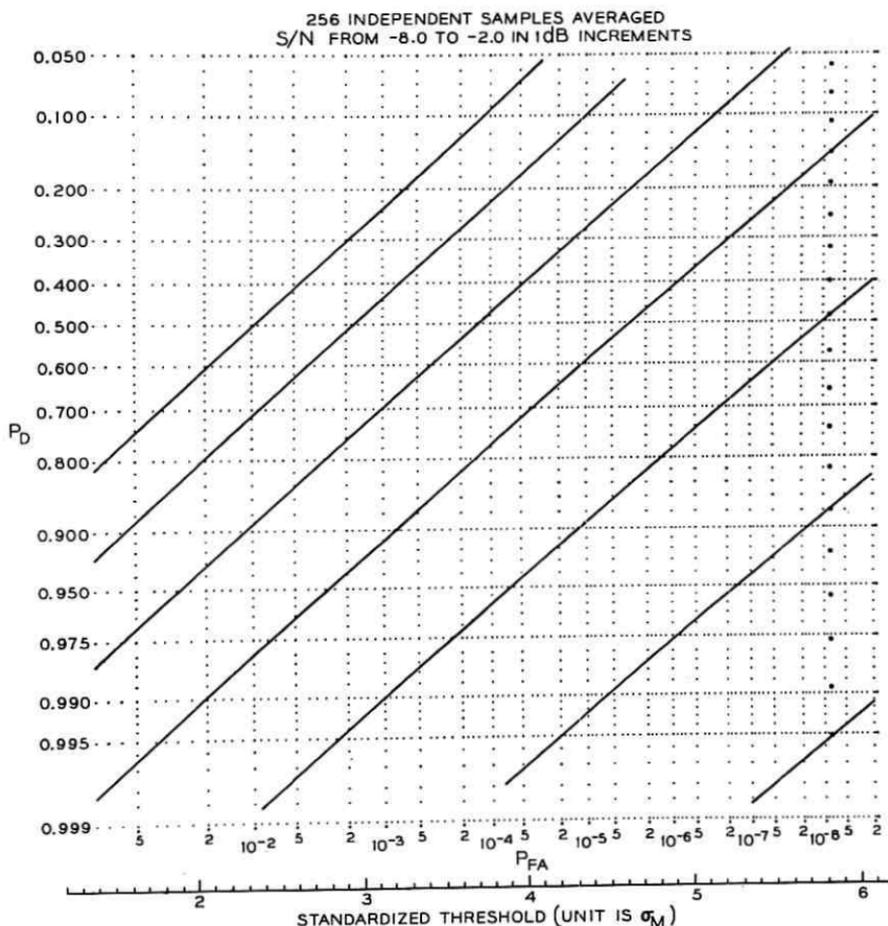
Fig. 10 — ROC curves.
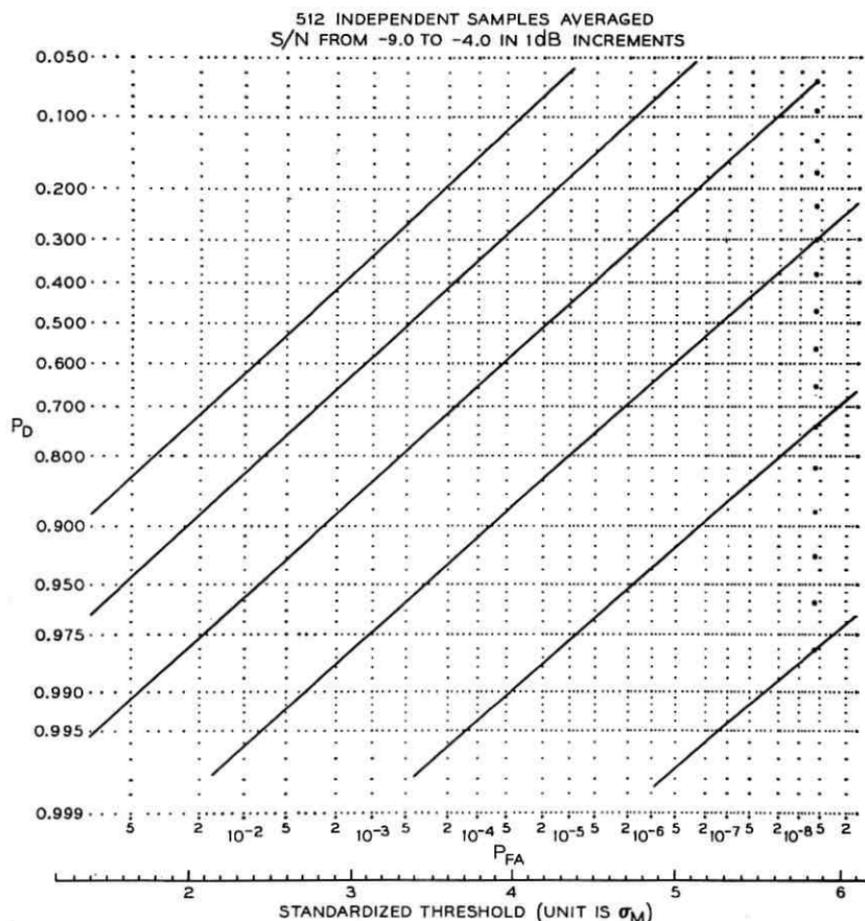
Fig. 11 — ROC curves.

IV. CONCLUSION

4.1 *Efficiency of Post-Detector Integration*

When more than one sample can be averaged to improve detection sensitivity the process is usually called post-detector integration. The curves of Fig. 16 show how efficient this process is at different values of S/N and P_D, P_{FA} operating points. The gradients of these curves show the rate at which detectability is improved, measured in dB/

double-the-number-of-samples, ds, for two operating points. This ratio is convenient to use in interpolating between adjacent charts, since the number of samples doubles from one chart to the next.

It can be seen that the gradients are almost the same where the S/N is the same. The rate of improvement is better than 2 dB/ds for S/N above 10 dB, presumably reaching a maximum of 3 dB/ds for high S/N. The rate apparently becomes constant at 1.5 dB/ds for S/N values less than −10 dB.
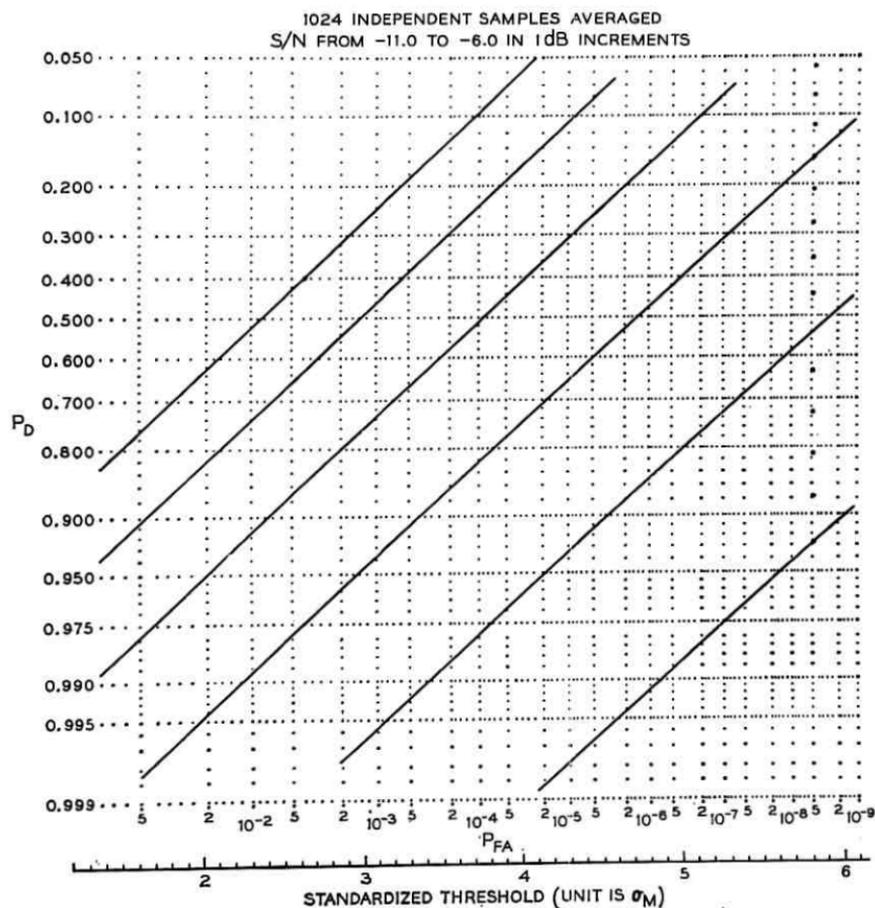
Fig. 12 — ROC curves.
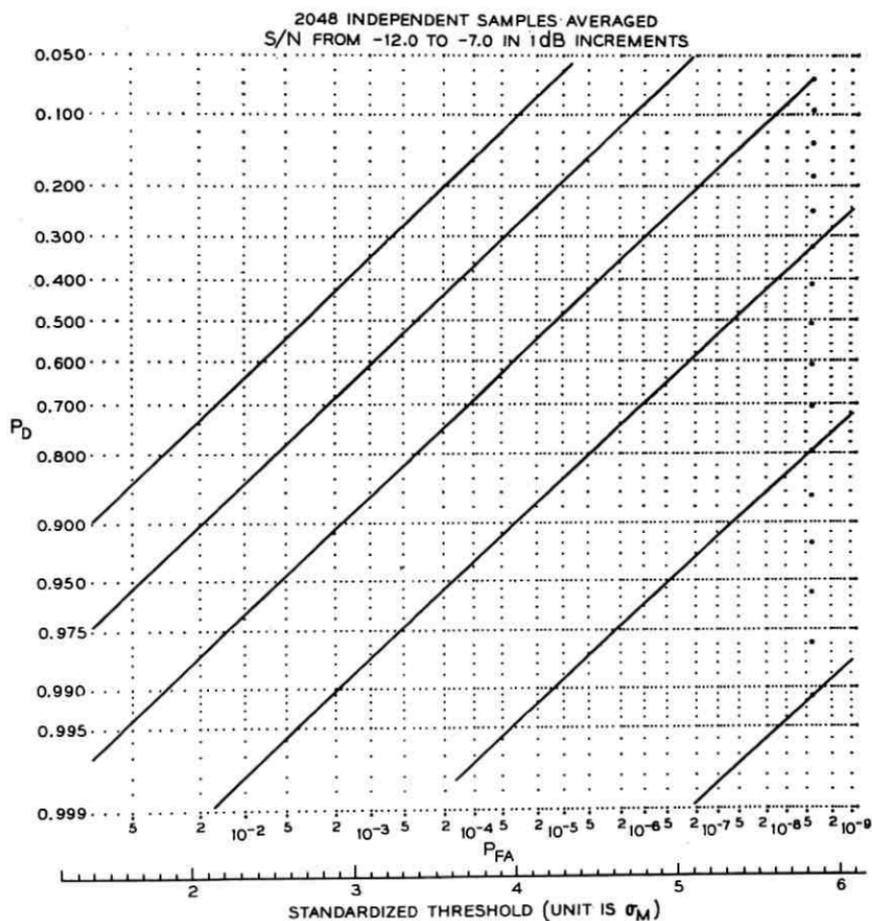
Fig. 13 — ROC curves.

4.2 *Extending the Scope of the Charts*

The detector input is restricted to a relatively narrow band of frequencies to ensure that the envelope can be measured accurately after detection. By bandshifting high enough, this result can be obtained even for bands of quite appreciable width. The charts can therefore be used in situations where the significant signals may be quite short bursts of nearly sinusoidal waveform.

The same standardized threshold scale applies to all the charts. It is given in units of

$$\sigma_M = \sigma/\sqrt{M},\qquad(9)$$

where σ is the rms noise into the detector and M independent samples are averaged.

Although the curves given here refer strictly to envelope detectors they can be applied when square-law-detectors are used if a small
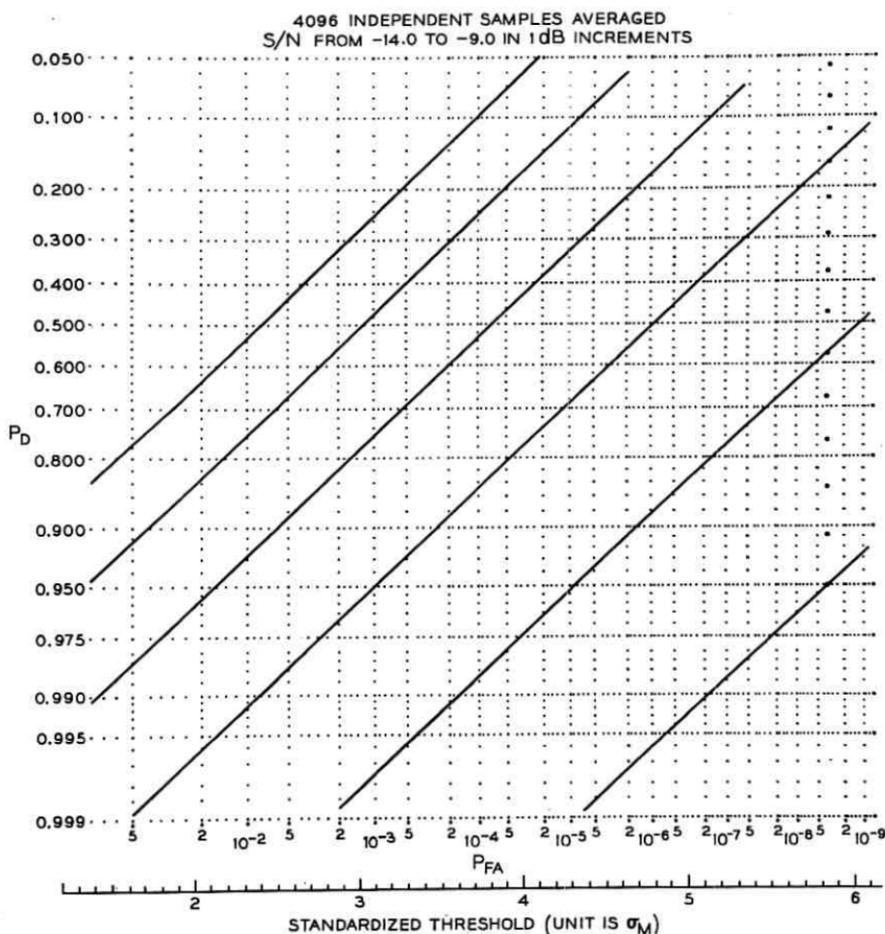
Fig. 14 — ROC curves.
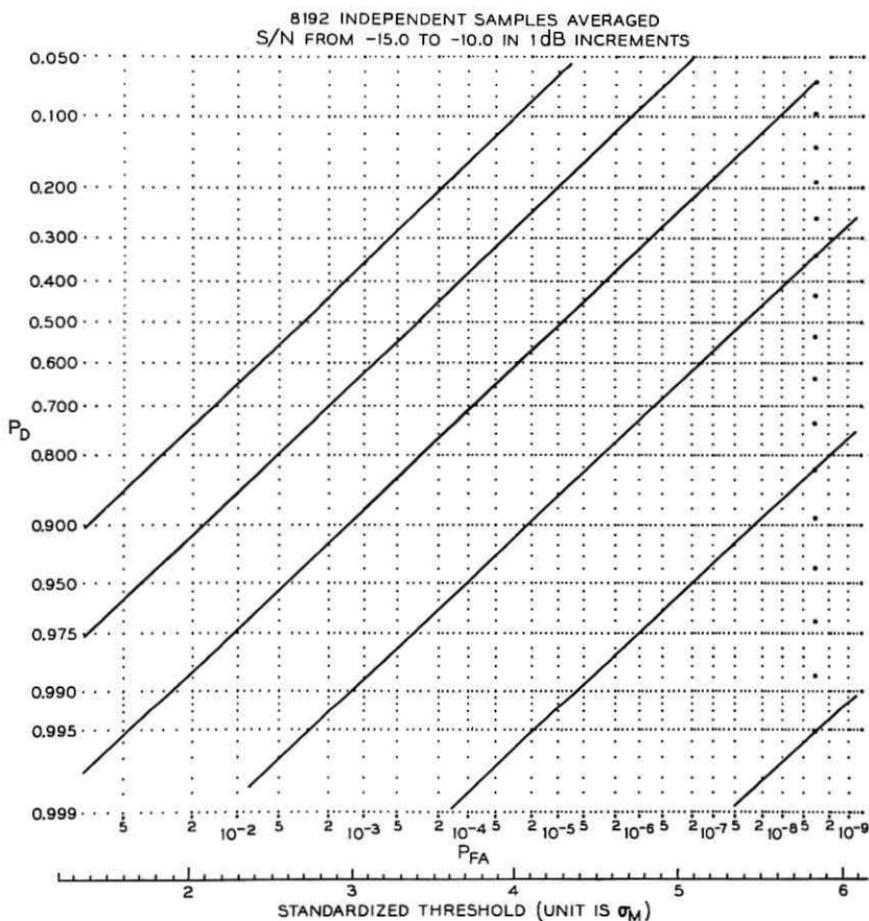
Fig. 15 — ROC curves.

error (always less than 0.2 dB) in S/N is taken into account, and the standardized threshold scale is changed. A calibration curve for this error is given in Ref. 1. Reasonable agreement with this curve was found by generating ROC curves for a square-law detector after suitably modifying the computer program. It was possible thus to get ROC curves for 1 sample and for averaging 128 or more independent samples.
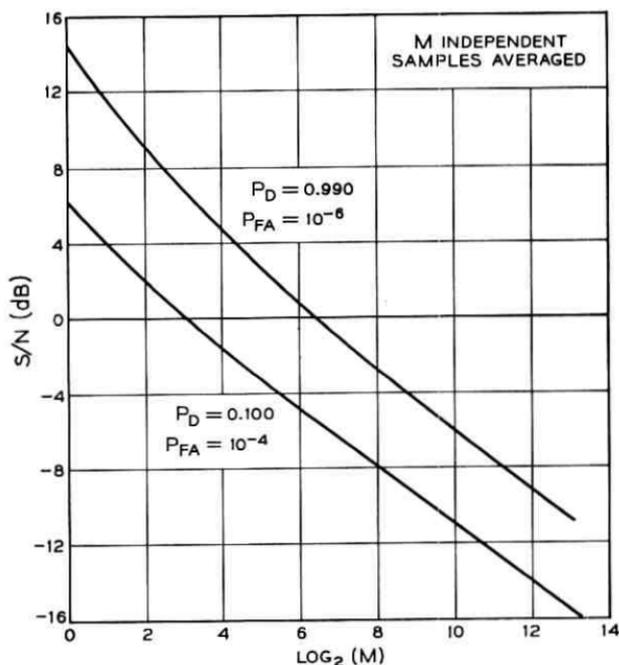
Fig. 16 — Sensitivity related to number of samples averaged.

REFERENCES

1. Marcum, J. I., A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix, *Research Memorandum RM-753*, The Rand Corporation, July 1, 1946.
2. Schwartz, M., A Statistical Approach to the Automatic Search Problem, Harvard University Dissertation, May 1, 1951.
3. Peterson, W. W., Birdsall, T. G., and Fox, W. C., The Theory of Signal Detectability, 1954 Symposium on Information Theory, IRE Trans. PGIT, *IT-4*, September, 1954, pp. 171–212.
4. Kaplan, E. L., Signal-Detection Studies with Applications, B.S.T.J., *34*, March, 1955, pp. 403–437.
5. Rice, S. O., Mathematical Analysis of Random Noise, B.S.T.J., *24*, January, 1945, pp. 46–156, Section 3.10.
6. Rice, S. O., Properties of a Sine Wave Plus Random Noise, B.S.T.J., *27*, January, 1948, pp. 109–157, Section 3.
7. Ryshik, I. M. and Gradstein, I. S., *Tables of Series Products and Integrals*, 2nd Ed., Plenum Press, 1963, Formula 4.434.
8. Fry, T. C., *Probability and Its Engineering Uses*, Van Nostrand, 2nd Ed., 1965, Sections 124 through 127.

Phase and Amplitude Modulation in High-Efficiency Varactor Frequency Multipliers – General Scattering Properties

By C. DRAGONE

The presence of phase and amplitude modulation in the signals of a varactor frequency multiplier is analyzed, and some general multiplier properties are derived. The following summarizes one of the most important results of this paper.

Consider a frequency multiplier which has the following characteristics: (i) a varactor which has a square-law characteristic, (ii) the order of multiplication is $N = 2^n = 2, 4$, etc., (iii) the minimum number of idlers, and (iv) it is lossless and tuned. It is shown that for this multiplier there is no conversion between small-index, low-frequency amplitude and phase modulation. Therefore, since narrow-band noise from external sources will be present at the input and output of the multiplier, the noise components corresponding to phase modulation of the carriers can be treated independently of the noise components corresponding to amplitude modulation.

Consider now the input and output noise sidebands corresponding to phase modulation (PM). It is shown that the multiplier behaves with respect to these sidebands as an amplifier with the following properties: (i) a forward voltage transmission equal to N, (ii) unity reverse transmission, (iii) an output reflection of $1, -1, 3$, respectively, for $N = 2, 4, 8$, and (iv) no input reflections. As a consequence of these properties the multiplier is "potentially" unstable with respect to PM.

The utility of the multiplier properties derived in this paper will be illustrated by the discussion in a companion paper in this issue which shows how, in practical cases, instability arises and how it can be avoided.

I. SUMMARY OF RESULTS

The frequency multipliers to be considered are harmonic generators which use varactor diodes as nonlinear elements. Noise produces un-

wanted amplitude and phase modulation in the signals of a frequency multiplier. In this paper, the presence of these modulations is analyzed and some general multiplier properties are derived. The following summarizes one of the most important results of this paper. Consider a frequency multiplier which has the following characteristics:

(i)   a varactor which has a square-law characteristic,
(ii) the order of multiplication is $N = 2^n$,
(iii) the minimum number of idlers, and
(iv) it is lossless and tuned.

It is shown that for this multiplier there is no conversion between small-index, low-frequency amplitude, and phase modulation. Therefore, since narrow-band noise from external sources will be present at the input and output of the multiplier, the noise components corresponding to phase modulation of the carriers can be treated independently of the noise components corresponding to amplitude modulation.

Consider now the input and output noise sidebands corresponding to phase modulation (PM). It is shown that the multiplier behaves with respect to these sidebands as an amplifier with the following properties:

(i)   a forward voltage transmission equal to N,
(ii) unity reverse transmission,
(iii) an output reflection of 1, -1, 3, respectively, for $N = 2, 4, 8$, and
(iv) it has no input reflections.

As a consequence of these properties the multiplier is "potentially" unstable with respect to PM. This summarizes one of the most important results of this paper. Now let us consider the results of this paper in more detail.

This paper is concerned with the presence of amplitude and phase modulation in the multiplier signals. Suppose, for the moment, that these two types of modulation are independent of each other. That is, suppose that the multiplier does not produce AM-to-PM conversion, and vice-versa. Suppose, furthermore, that only PM is present. Then each signal will consist of a carrier and of a pair of sidebands in quadrature with respect to the carrier. Since either sideband can be obtained from a knowledge of the other, then one may consider only one of the two sidebands and ignore the other one. For example, let the upper sideband be chosen as the variable, and let it be described in terms of propagating waves. Then the input variables of the multiplier are the

two waves* $\vec{v}_{\omega_0+p}$, $\overset{\leftarrow}{v}_{\omega_0+p}$ which constitute the upper sideband of the input carrier ω_0 . Similarly, the two output waves $\vec{v}_{N\omega_0+p}$, $\overset{\leftarrow}{v}_{N\omega_0+p}$ represent the output variables. ω_0 is the input "carrier" frequency of the multiplier and p is the frequency of the fluctuations. At this point, the scattering formalism furnishes a convenient way of describing the properties of the multiplier. More precisely, one may define the PM scattering parameters of the multiplier as the reflection and transmission coefficients which relate the "scattered" waves $\overset{\leftarrow}{v}_{\omega_0+p}$, $\vec{v}_{N\omega_0+p}$ to the "incident" ones $\vec{v}_{\omega_0+p}$, $\overset{\leftarrow}{v}_{N\omega_0+p}$. In this way one obtains the PM scattering matrix $\hat{S}_\theta$ defined by

$$\frac{\overset{\leftarrow}{v}_{\omega_0+p}}{V_{\omega_0}} = \overset{\rightarrow}{\rho_\theta}\,\frac{\vec{v}_{\omega_0+p}}{V_{\omega_0}} + \overset{\leftarrow}{T_\theta}\,\frac{\overset{\leftarrow}{v}_{N\omega_0+p}}{V_{N\omega_0}}$$

$$\frac{\vec{v}_{N\omega_0+p}}{V_{N\omega_0}} = \overset{\rightarrow}{T_\theta}\,\frac{\vec{v}_{\omega_0+p}}{V_{\omega_0}} + \overset{\leftarrow}{\rho_\theta}\,\frac{\overset{\leftarrow}{v}_{N\omega_0+p}}{V_{N\omega_0}} ,$$

(1)

where V_{ω_0} is the Fourier coefficient of the input carrier ω_0 , and $V_{N\omega_0}$ is the Fourier coefficient of the output carrier $N\omega_0$. Note that the sidebands have been normalized with respect to the relative carriers. Equations (1) are equivalent to (8).†

In a completely similar way one defines the AM scattering matrix $\hat{S}_a$.

It has been found that a multiplier with characteristics (i) through (iv) above does not produce AM $\rightleftarrows$ PM conversion and that it has the following scattering matrices:

$$\hat{S}_\theta = \begin{vmatrix} \overset{\rightarrow}{\rho_\theta} & \overset{\leftarrow}{T_\theta} \\ \overset{\rightarrow}{T_\theta} & \overset{\leftarrow}{\rho_\theta} \end{vmatrix} = \begin{vmatrix} 0 & (-1)^n \\ N & \dfrac{1-(-1)^n N}{3} \end{vmatrix}$$

(2a)

$$\hat{S}_a = \begin{vmatrix} \overset{\rightarrow}{\rho_a} & \overset{\leftarrow}{T_a} \\ \overset{\leftarrow}{T_a} & \overset{\leftarrow}{\rho_a} \end{vmatrix} = \begin{vmatrix} \dfrac{N-(-1)^n}{3N} & \dfrac{(-1)^n}{N} \\ 1 & 0 \end{vmatrix}$$

(2b)

if the modulation frequency p is small enough.

Notice that (2a) gives the properties stated at the beginning of this summary. As a consequence of (2a), the multiplier is potentially unstable with respect to PM. AM instabilities cannot occur. This follows from (2b).

---

* $\overset{\leftarrow}{v}_\omega$, $\vec{v}_\omega$ designate the Fourier coefficients of the voltage components of frequency ω. They propagate in the directions indicated by the arrows.

† Note that $\vec{v}_{\omega_0+p}/V_{\omega_0}$, etc. correspond to the modulation indexes $j\overset{\rightarrow}{\theta_1}$, etc. defined in the next section.

Equations (2a), (2b) also have some other interesting consequences. For instance, if one injects a single tone $N\omega_0 + p$ into the output port of a multiplier, then the multiplier will reflect a pure PM wave. That is, the multiplier behaves as an ideal reflection-type limiter in the vicinity of the output carrier. This property is common to all types of multipliers [see (19)] and may be useful if one wants to modulate the output phase of a multiplier without generating AM. Note that $| \rho_{\bar{\theta}} | > 1$ for $N > 4$. This means that, if $N > 4$, then the output port of the multiplier reflects amplifying the PM components of the output sidebands.

It is also shown that some of the properties expressed by (2a), (2b) are common to all lossless multipliers. A consequence of this is that most of the results obtained for a multiplier with the characteristics listed above can be qualitatively extended to all efficient multipliers.

II. INTRODUCTORY REMARKS

This paper is concerned with the presence of small amplitude and phase fluctuations in the electrical variables of the networks to be considered. By using the familiar terminology and concepts of modulation theory, the problem can be defined as follows.

The electrical signals will consist of sinusoidal carriers which are both amplitude- and phase-modulated. The amplitudes of the modulating waves* will be supposed to be small enough to guarantee superposition to hold, and to allow the modulated waves to be approximated by the sum of the carriers and their first-order sidebands. Furthermore, since superposition holds, consideration will be limited to sinusoidal modulating waves. Then, if p is the frequency of the modulating waves (i.e., of the fluctuations), each carrier frequency ω will be surrounded by two small side-frequencies $\omega + p$, $\omega - p$ representing the sidebands of ω. One can say that the object of this analysis is to study the presence of these side-frequencies around the carriers. A pair of sidebands is completely specified either by its spectrum or by the spectrum of the modulating waves associated with it. More precisely, consider a small voltage $v(t)$ consisting of two side-frequencies† $\omega + p$, $\omega - p$

---

* The "modulating waves" are the "fluctuations" and will be clarified later.

† Through this analysis a real function $g(t)$ consisting of sinusoids will be represented by the complex Fourier series

$$g(t) = \sum_{-\infty}^{\infty} C(\omega_i) \exp [j\omega_i t] = 2(\text{Re}) \sum_{0}^{\infty} C(\omega_i) \exp [j\omega_i t].$$

The complex coefficients C will be called "the Fourier coefficients" of $g(t)$. Furthermore, only positive frequencies will be considered, as illustrated by the second

$$v(t) = 2(Re)(v_{\omega+p} \exp [j(\omega + p)t] + v_{\omega-p} \exp [j(\omega - p)t]) \qquad (3)$$

and let $V_c(t)$ be the carrier associated with $v(t)$. Let ω, V_c be the carrier frequency and the Fourier coefficient of $V_c(t)$, respectively. Then the Fourier coefficients of the modulating waves $a(t)$, $\theta(t)$ associated with $v(t)$ are given by

$$a = \frac{1}{2}\left(\frac{v_{\omega+p}}{V_c} + \frac{v_{\omega-p}^*}{V_c^*}\right)$$

$$\theta = \frac{1}{2j}\left(\frac{v_{\omega+p}}{V_c} - \frac{v_{\omega-p}^*}{V_c^*}\right). \qquad (4)$$

In fact, if one uses (4) to express $v(t)$ in terms of a, θ, one obtains*

$$v(t) = 2(\mathrm{Re})[(a + j\theta)V_c \exp [j(\omega + p)t]$$

$$+ (a - j\theta)^* V_c \exp [j(\omega - p)t]] \qquad (5)$$

$$= a(t)V_c(t) + \frac{\theta(t)}{\omega} V_c'(t).$$

Since it is supposed $| v_{\omega \pm p} | \ll | V_c |$, one has $| a | \ll 1$, $| \theta | \ll 1$. Therefore, the last equality of (5) gives

$$v(t) + V_c(t) = [1 + a(t)]V_c\left[t + \frac{\theta(t)}{\omega}\right]. \qquad (6)$$

Equation (6) shows that $a(t)$, $\theta(t)$ are the modulating waves which produce the sidebands $v(t)$.† $a(t)$ is the amplitude modulating wave, and $\theta(t)$ is the phase modulating wave.

a, θ, the Fourier coefficients of $a(t)$, $\theta(t)$, will be called, respectively, the AM *index* and the PM *index* of $v(t)$. Extensive use will be made of the AM, PM indexes to represent a pair of sidebands.‡

---

summation. This convention will not be followed in the Appendix where (and only there) both positive and negative frequencies will have to be considered for convenience. Note that $C(\omega_i) = C^*(-\omega_i)$ because $g(t)$ is real. The asterisk ( )* will always indicate the complex conjugate of a complex quantity. (Re) means "real part of".

* $a(t) = 2(\mathrm{Re})(a \exp jpt)$, $\theta(t) = 2(\mathrm{Re})(\theta \exp jpt)$, $V_c(t) = 2(\mathrm{Re})(V_c \exp j\omega t)$ according to the preceding footnote.

† When $a(t)$, $\theta(t)$ are applied to the carrier $V_c(t)$.

‡ As already noted, a, θ, represent the spectra of the modulating waves relative to $v(t)$. It should also be noted that the representation of $v(t)$ by means of a, θ corresponds to the familiar representation of a pair of sidebands by means of the so-called "in-phase" and "quadrature" components about a given carrier. This is shown by (5) in which aV_c, a^*V_c are the "in-phase" components and $j\theta V_c$, $j\theta^*V_c$ are the "quadrature" components. Therefore (a) can also be regarded as the normalized (with respect to jV_c) amplitude of the "quadrature" upper-sideband.

2.1   *The Variables of the Interconnection Between Two Stages*

With reference to Fig. 1 consider now the interconnection between two general stages. Let $V(t)$ be the interstage voltage and assume that $V(t)$ consists of a sinusoidal carrier $V_c(t)$ and a pair of sidebands $v(t)$. Now suppose that a short piece of line of characteristic impedance Z_c is inserted between the two stages. Its electrical length is assumed to be so short that it may be inserted without altering the electrical properties of the cascade connection of the two stages. This should be regarded as an artifice made to allow the signals of the interconnection to be described in terms of propagating waves.* The sidebands $v(t)$
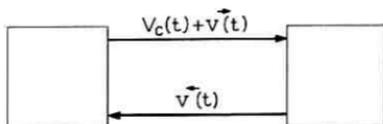
$$V(t) = V_c(t) + v(t) = V_c(t) + \overrightarrow{v}(t) + \overleftarrow{v}(t)$$

Fig. 1 — Description of the interstage voltage in terms of propagating waves.

now consist of two waves $\overrightarrow{v}(t)$, $\overleftarrow{v}(t)$ propagating in opposite directions, as indicated by the arrows →, ←, and

$$v(t) = \overrightarrow{v}(t) + \overleftarrow{v}(t). \tag{7}$$

Finally, consider the respresentation of $\overrightarrow{v}(t)$, $\overleftarrow{v}(t)$ by means of their PM, AM indexes with respect to the carrier V_c. By using (4) one obtains† two pairs of modulation indexes

$$\overrightarrow{a}, \ \overrightarrow{\theta}, \quad \text{the indexes of} \quad \overrightarrow{v}(t)$$

$$\overleftarrow{a}, \ \overleftarrow{\theta}, \quad \text{the indexes of} \quad \overleftarrow{v}(t)$$

($\overrightarrow{a}$, etc.) provide a complete description of the fluctuations present at the generical interconnection, as illustrated in Fig. 2.

2.2   *Definition of the PM, AM Scattering Matrices of a Stage*

In most cases to be analyzed‡ each stage will have two basic properties:

---

* Z_c will always be chosen such that only one propagating wave exists at the carrier frequency. Therefore, the voltage of the interconnection consists of a modulated wave propagating in one direction, and of a pair of sidebands without carriers propagating in the other direction, as it is illustrated in Fig. 1. The carrier $V_c(t)$ is used as a reference for deriving the modulation indexes of both sets of sidebands.

† By replacing $v(t)$, $v_{\omega+p}$, $v_{\omega-p}$ with $\overrightarrow{v}(t)$, $\overrightarrow{v}_{\omega+p}$, $\overrightarrow{v}_{\omega-p}$ one obtains $\overrightarrow{a}$, $\overrightarrow{\theta}$. In a similar way one obtains $\overleftarrow{a}$, $\overleftarrow{\theta}$.

‡ In Ref. 1, AM ⇌ PM conversion is considered.

(i) it will be linear with respect to the modulating signals (since they are supposed to be small)

(ii) it will not generate any AM-to-PM (or vice versa) conversion.

As a result of property (ii) the PM indexes of a stage are independent from the AM indexes, and vice versa. Therefore, the two cases AM, PM can be treated separately.

Consider for instance the PM case. With reference to the stage of Fig. 3, let $\theta_1^\rightarrow$, $\theta_i^\leftarrow$ be the input $(i = 1)$ and output $(i = 2)$ PM indexes. Then, because of (i), the scattered indexes $\theta_1^\leftarrow$, $\theta_2^\rightarrow$ can be related to the incident ones $\theta_1^\rightarrow$, $\theta_2^\leftarrow$ through a set of linear equations of the type

$$\theta_1^\leftarrow = \rho_\theta^\rightarrow \theta_1^\rightarrow + T_\theta^\leftarrow \theta_2^\leftarrow$$

$$\theta_2^\rightarrow = T_\theta^\rightarrow \theta_1^\rightarrow + \rho_\theta^\leftarrow \theta_2^\leftarrow. \tag{8}$$
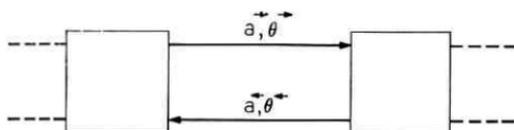
Fig. 2 — Description of the interstage sidebands in terms of modulation indexes.

Notice that (8) are equivalent to (1). The scattering matrix $\hat{S}_\theta$

$$\hat{S}_\theta = \begin{vmatrix} \rho_\theta^\rightarrow & T_\theta^\leftarrow \\ T_\theta^\rightarrow & \rho_\theta^\leftarrow \end{vmatrix} \tag{9}$$

will be called the *PM scattering matrix** of the stage. $T_\theta^\rightarrow$ will be called the *PM forward transmission coefficient*. $T_\theta^\leftarrow$ is the *PM reverse transmission coefficient*. $\rho_\theta^\rightarrow$ is the *PM input reflection coefficient*. $\rho_\theta^\leftarrow$ is the *PM output reflection coefficient*.

In a completely similar way the AM scattering matrix $\hat{S}_a$ of a stage is defined†

$$\hat{S}_a = \begin{vmatrix} \rho_a^\rightarrow & T_a^\leftarrow \\ T_a^\rightarrow & \rho_a^\leftarrow \end{vmatrix}. \tag{10}$$

---

* Note that the complex coefficients $\rho^\rightarrow$, $T^\rightarrow$, $T^\leftarrow$, $\rho^\leftarrow$ correspond to the familiar coefficients S_{11}, S_{12}, S_{21}, S_{22} which are usually employed to represent scattering parameters. (See Ref. 2). As can be directly verified from (8), if a unit index is incident on either port, T gives the index transmitted out of the other port, and ρ gives the index reflected at the same port. The arrows $\rightarrow$, $\leftarrow$ are applied to T, ρ to indicate the directions of the incident waves to which T, ρ refer.

† The convention of using the letters a, θ for designating the AM, PM cases will always be followed through the analysis.
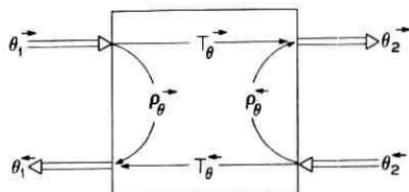
Fig. 3 — Description of the PM properties of a stage by means of the PM scattering coefficients.

In conclusion in this analysis: the *variables* are the PM, AM indexes; the *frequency* of the variables is p; the *parameters* specifying the behavior of the various networks are the *PM, AM scattering coefficients*.

III. GENERAL PROPERTIES OF A LOSSLESS MULTIPLIER WHICH DOES NOT PRODUCE AM ⇄ PM CONVERSION

It is important to emphasize that this analysis applies both to a multiplier consisting of a single stage and to a chain of multipliers.

Furthermore, in this paper, the analysis will be limited to the case of a lossless multiplier which does not produce AM ⇄ PM conversion. This is the most important case for the following two reasons. The first reason is that, in general, a multiplier which is tuned does not produce AM ⇄ PM conversion if the sidebands are close enough to the carriers (see Appendix). Note that it is assumed that the bias network is properly designed so that effects such as those described in Ref. 3 are absent. Furthermore, results obtained for the case of no AM ⇄ PM conversion can be extended to the general case of AM ⇄ PM, as it is shown in the second paper. The second reason is that the question of stability generally arises only when the losses of the multiplier are small. Furthermore, if the losses are small, the results which are obtained by neglecting them can be readily extended to include them, as is pointed out in the second paper.

In this section it is demonstrated that a lossless multiplier of order N has the following properties:

$$\vec{T_\theta}\vec{T_a}^* = N(1 - \vec{\rho_\theta}\vec{\rho_a}^*)$$

$$\overleftarrow{T_\theta}\overleftarrow{T_a}^* = \frac{1}{N}(1 - \overleftarrow{\rho_\theta}\overleftarrow{\rho_a}^*)$$ (11)

$$\frac{\overleftarrow{\rho_a}^*}{\overleftarrow{T_a}^*} + N\frac{\vec{\rho_\theta}}{\vec{T_\theta}} = 0, \qquad N\frac{\vec{\rho_a}^*}{\vec{T_a}^*} + \frac{\overleftarrow{\rho_\theta}}{\overleftarrow{T_\theta}} = 0.$$

The equations of (11) have three important properties.

One is that they depend on only one multiplier parameter: the order of multiplication N. They do not depend in any way on the actual circuit configuration or on the type of varactor, etc.

Another property is that they are frequency independent. That is, they do not contain p explicitly even though $T_\theta^\rightarrow$ etc. will in general be frequency dependent.

The last important property of (11) is that they represent 4 independent relations among the 8 unknowns $T_\theta^\rightarrow$, etc. For instance, if the AM coefficients are known, then the PM coefficients can be calculated by means of (11), and vice versa.

3.1 *Demonstration of (11)*

Consider the multiplier of order N shown in Fig. 4(a) connected to a load R_N and driven by a voltage generator having impedance R_1. Let $V_r(t)$ $(r = 1, N)$ be the voltage of the input port $(r = 1)$ and output port $(r = N)$ of the multiplier in the absence of fluctuations. It will be supposed that the input of the multiplier is matched to the generator. Therefore, the input impedance of the multiplier is R_1. The multiplier is assumed to be lossless and without AM $\rightleftarrows$ PM. The characteristic impedance associated with the rth $(r = 1, N)$ port is R_r, according to the footnote on p. 780.

Assume now that two small waves $v_1^\rightarrow(t)$, $v_N^\leftarrow(t)$ are arriving towards the multiplier, and that they consist of the frequencies $\omega_0 \pm p$, $N\omega_0 \pm p$, respectively. The problem to be considered is to find the scattered
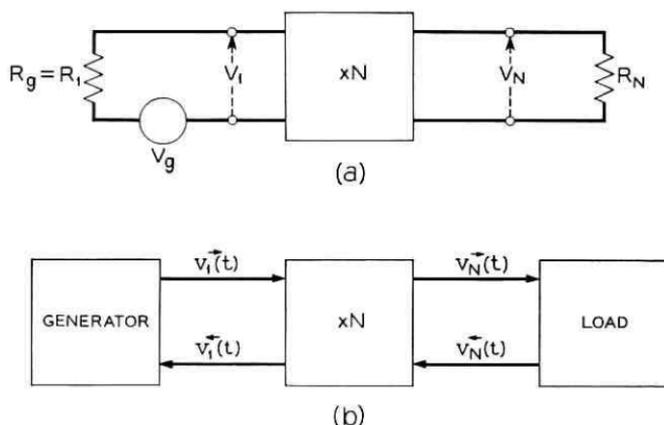

Fig. 4 — (a) Multiplier of order N. (b) Input and output sidebands of the multiplier.

waves $v_1^{\leftarrow}(t)$, $v_N^{\leftarrow}(t)$, as illustrated in Fig. 4 (in which the carriers $V_1(t)$, $V_N(t)$ are not shown).

Let $(\theta_r^{\rightarrow}, \theta_r^{\leftarrow})$, $(a_r^{\rightarrow}, a_r^{\leftarrow})$ be the indexes of $v_r(t)$ with respect to the carrier $V_r(t)$. Then, according to (5),

$$v_r^{\rightarrow}(t) = 2(\mathrm{Re})\{(a_r^{\rightarrow} + j\theta_r^{\rightarrow})V_r \exp [j(r\omega_0 + p)t]$$
$$+ (a_r^{\rightarrow *} + j\theta_r^{\rightarrow *})V_r \exp [j(r\omega_0 - p)t]\} \qquad (12)$$

and similarly for $v_r^{\leftarrow}(t)$ (just change the arrows directions). V_r is the Fourier coefficient of $V_r(t)$ $(r = 1, N)$.

Now let $P_{r\omega_0+p}^{\rightarrow}$ be the power carried by the forward wave of frequency $r\omega_0 + p$, and let $P_{r\omega_0-p}^{\leftarrow}$, $P_{r\omega_0-p}^{\rightarrow}$, $P_{r\omega_0+p}^{\leftarrow}$ have similar meanings. Then the total power at the frequency ω $(\omega = r\omega_0 \pm p)$ is the sum of the two components $(\rightarrow)$, $(\leftarrow)$:

$$P_\omega = P_\omega^{\rightarrow} + P_\omega^{\leftarrow} . \qquad (13)$$

Consider now the fact that the multiplier is lossless, and that therefore Manley–Rowe relations hold.* That is,

$$\left(\frac{P_{\omega_0+p}}{\omega_0 + p} - \frac{P_{\omega_0-p}}{\omega_0 - p}\right) - \left(\frac{P_{N\omega_0+p}}{N\omega_0 + p} - \frac{P_{N\omega_0-p}}{N\omega_0 - p}\right) = 0. \qquad (14)$$

From (12) one can calculate $P_\omega^{\rightarrow}$, $P_\omega^{\leftarrow}$ $(\omega = r\omega_0 \pm p)$. For instance,

$$P_{r\omega_0+p}^{\rightarrow} = |a_r^{\rightarrow} + j\theta_r^{\rightarrow}|^2 \frac{|V_r|^2}{2R_r} = |a_r^{\rightarrow} + j\theta_r^{\rightarrow}|^2 P_0 \qquad (15)$$

in which $| \ |$ indicates the absolute value. P_0 is the power delivered by the generator to the multiplier at ω_0 and transferred by the multiplier to the load R_N. (The input carrier power is equal to the output carrier power because the multiplier is lossless.)

After calculating $P_\omega^{\rightarrow}$, $P_\omega^{\leftarrow}$, one obtains P_ω $(\omega = r\omega_0 \pm p)$ by means of (13). Then, by substituting P_ω in (14), one obtains†

$$(\mathrm{IM})[N(\theta_1^{\rightarrow}a_1^{\rightarrow *} - \theta_1^{\leftarrow}a_1^{\leftarrow *}) + \theta_N^{\leftarrow}a_N^{\leftarrow *} - \theta_N^{\rightarrow}a_N^{\rightarrow *}] = 0, \qquad (16)$$

where (IM) indicates the imaginary part.

Now if one considers the multiplier PM, AM scattering matrices

---

* The minus sign in front of the second addend of (14) occurs because power is assumed positive when it is flowing towards the right. This, at the output port, is opposite to the usual convention of assuming positive the power absorbed by the varactor. See Ref. 4.

† In deriving (16) the approximation $r\omega_0 \pm p \cong r\omega_0$ has been made so that (14) becomes:

$$NP_{\omega_0+p} - NP_{\omega_0-p} - P_{N\omega_0+p} + P_{N\omega_0-p} = 0.$$

$\hat{S}_a$, $\hat{S}_\theta$ defined in the previous section

$$\begin{vmatrix} \overleftarrow{\theta_1} \\ \overrightarrow{\theta_N} \end{vmatrix} = \hat{S}_\theta \begin{vmatrix} \overrightarrow{\theta_1} \\ \overleftarrow{\theta_N} \end{vmatrix} , \qquad \begin{vmatrix} \overleftarrow{a_1} \\ \overrightarrow{a_1} \end{vmatrix} = \hat{S}_a \begin{vmatrix} \overrightarrow{a_1} \\ \overleftarrow{a_N} \end{vmatrix} \qquad (17)$$

and substitutes (17) in (16), one obtains

$$\begin{aligned} (\text{IM})[\overrightarrow{\theta_1}\overrightarrow{a_1}*(N &- N\overrightarrow{\rho_\theta}\overrightarrow{\rho_a}* - \overrightarrow{T_\theta}\overrightarrow{T_a}*) \\ &+ \overleftarrow{\theta_N}\overleftarrow{a_N}*(1 - \overleftarrow{\rho_\theta}\overleftarrow{\rho_a}* - N\overleftarrow{T_\theta}\overleftarrow{T_a}*) \\ &+ \overrightarrow{\theta_1}\overleftarrow{a_N}*(-N\overrightarrow{\rho_\theta}\overleftarrow{T_a}* - \overleftarrow{\rho_a}*\overrightarrow{T_\theta}) \\ &+ \overleftarrow{\theta_N}\overrightarrow{a_1}*(-N\overleftarrow{T_\theta}\overrightarrow{\rho_a}* - \overleftarrow{\rho_\theta}\overrightarrow{T_a}*)] = 0. \end{aligned} \qquad (18)$$

A relation which has to hold for any choice of ($\overrightarrow{\theta_1}$, $\overleftarrow{\theta_1}$, etc.) and therefore it has to hold for any choice of ($\overrightarrow{\theta_1}\overrightarrow{a_1}*,\overleftarrow{\theta_N}\overleftarrow{a_N}*$,etc.). This is possible only if (11) holds. Note that the approximation $r\omega_0 \pm p \cong r\omega_0$ has been made in deriving (18).

IV. SLOW-VARYING FLUCTUATIONS

Suppose now that the input generator V_g is phase modulated and that its amplitude is kept constant. If the modulation frequency p is small enough, then the output phase deviation will be just N times that of the input generator. Furthermore, all amplitudes will remain constant. These properties are common to all multipliers and are well known.[5,6] Therefore, they will not be demonstrated here. As a consequence of these properties one has*

$$\lim_{p\to 0} \overrightarrow{T_\theta} = N \qquad \lim_{p\to 0} \overleftarrow{\rho_\theta} = 0.$$

By combining these two results with (11) one obtains

$$\begin{array}{ll} \overrightarrow{T_\theta} = N & \overrightarrow{\rho_\theta} = 0 \\ \overrightarrow{T_a} = 1 & \overleftarrow{\rho_a} = 0 \\ \overleftarrow{T_a}* = \dfrac{1}{N\overleftarrow{T_\theta}} & \overleftarrow{\rho_a}* = -\dfrac{\overleftarrow{\rho_\theta}}{N\overleftarrow{T_\theta}} \end{array} \qquad (19)$$

which apply to the limiting case of slowly varying fluctuations. From (19) one can see that $\overrightarrow{T_a} = 1$. This has the meaning that if one amplitude modulates the input generator, then the output voltage will have an identical (percentage-wise) modulation.

---

* The multiplier is matched at the carrier frequency ω_0 . Therefore, there is no reflection present at the input of the multiplier if modulation is absent. Consequently, input reflections will remain absent if the input phase is slowly modulated. Therefore, $\overrightarrow{\rho_\theta} = 0$ if $p \cong 0$.

Note that (19) leaves only two coefficients to be determined: two of $(\rho_a^{\rightarrow *}, \rho_\theta^{\leftarrow}, T_\theta^{\leftarrow})$ or two of $(\rho_a^{\rightarrow *}, \rho_\theta^{\leftarrow}, T_a^{\leftarrow *})$. Both (11) and (19) represent quite general properties of ideal (i.e., lossless, tuned) multipliers. They do not depend on the actual circuit configuration of the multiplier, on the type of varactor characteristic, on the power level, etc.* Note that from (19) one has

$$(T_a^{\rightarrow *} T_a^{\leftarrow *})(T_\theta^{\rightarrow} T_\theta^{\leftarrow}) = 1.$$

This means that either the "PM round-trip transmission" $(T_\theta^{\rightarrow} T_\theta^{\leftarrow})$ or the "AM round-trip transmission" $(T_a^{\rightarrow} T_a^{\leftarrow})$ is larger than unity.† Therefore, reflections of the external circuit (the output load and the input generator) can produce instabilities.[1]

V. MULTIPLIER OF ORDER $N = 2^n$

In this section the following type of multiplier will be considered:

  (*i*) the nonlinear capacitance has a square-law Q–V characteristic,
  (*ii*) the order of multiplication is $N = 2^n$, and
  (*iii*) it has the least number of idlers[7] $(2\omega_0, \cdots, 2^{n-1}\omega_0)$.

This type of multiplier is most important because it can be *exactly* treated with little difficulty and it is *realistic* at the same time. In fact, abrupt-junction varactors exactly satisfy (*i*); graded-junctions approximately satisfy (*i*); most practical designs are based on (*iii*) for reasons of simplicity. Condition (*ii*) excludes from this treatment two important cases: $N = 3$, $N = 5$. On the other hand, the (exact) results which will be obtained for $N = 2$, $N = 4$, $N = 8$, $N = 16$, $N = 32$, etc. may be qualitatively extended to the remaining cases $N = 3$, $N = 5$, etc.

In the Appendix, it is shown that, because of (*i*), (*ii*), and (*iii*), one has $T_\theta^{\leftarrow} = (-1)^n$, $\rho_a^{\rightarrow} = [N - (-1)^n]/3N$. Therefore, by combining these results with (19), one has

$$\hat{S}_\theta = \begin{vmatrix} \rho_\theta^{\rightarrow} & T_\theta^{\leftarrow} \\ T_\theta^{\rightarrow} & \rho_\theta^{\leftarrow} \end{vmatrix} = \begin{vmatrix} 0 & (-1)^n \\ N & \dfrac{1 - (-1)^n N}{3} \end{vmatrix} \tag{2a}$$

$$\hat{S}_a = \begin{vmatrix} \rho_a^{\rightarrow} & T_a^{\leftarrow} \\ T_a^{\rightarrow} & \rho_a^{\leftarrow} \end{vmatrix} = \begin{vmatrix} \dfrac{N - (-1)^n}{3N} & \dfrac{(-1)^n}{N} \\ 1 & 0 \end{vmatrix}. \tag{2b}$$

---

* They are valid both for a single stage and a multiplier consisting of many stages.
† Neglecting the possibility that they are both unity. See the next section.

Therefore, the multiplier scattering matrices are *uniquely determined as functions of only the order of multiplication* N.

The physical meaning of the scattering coefficients (2) may be illustrated by the following two examples.

5.1 *First example.*

Let a small signal $N\omega_0 + p$ be *injected* into R_N by means of a directional coupler connected as shown in Fig. 5(a). Then the voltage across R_N will be both amplitude and phase modulated, and the AM, PM indexes will have equal amplitudes according to (4). Let α be their amplitude. Next, reverse the connection of the directional coupler, as shown in Fig. 5(b). Now the following facts will be observed:

(*i*) *The voltage of R_N contains only PM.* That is, the AM component of $\omega + p$ *is absorbed without reflection by the output of the multiplier* (because $\rho_a^- = 0$). The PM component, on the contrary, is reflected back into the load R_N. The PM index across R_N is $\rho_\theta^- \alpha$ with ρ_θ^- given by (2). Note that $|\rho_\theta^-| \geqq 1$ for all values of N, and that $|\rho_\theta^-| > 1$ for $N > 4$. Therefore, the PM component is reflected with amplification for $N > 4$. For instance, if $N = 8$ the gain is 3. All this indicates the circuit of Fig. 5(b) as a possible scheme for modulating the output phase of a multiplier without producing any AM (and with gain if $N > 4$).

(*ii*) The voltage across R_1 has both AM and PM. The PM index is $T_\theta^{--}\alpha = (-1)^n \alpha$ and the AM index is $T_a^{--}\alpha = (-1)^n/N \cdot \alpha$.
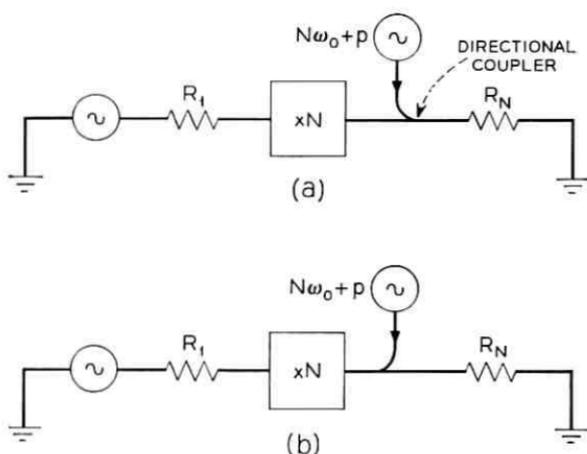

Fig. 5 — Injection of a tone at the output port of the multiplier.

5.2 Second example.

Now let the previous experiment be repeated at the input side of the multiplier, as shown by Fig. 6. Let α be the amplitude of the indexes generated across R_1 when the generator $\omega + p$ is connected as shown in Fig. 6(a). Then, after connecting $\omega + p$ as shown in Fig. 6(b), one has:

(i) The AM index across R_N is $T_a^{\rightarrow}\alpha = \alpha$ and the PM index is $T_\theta^{\rightarrow}\alpha = N\alpha$. These are general properties which have already been found and discussed in the more general case of the preceeding section [see (19)].

(ii) Across R_1 there is only AM and the AM index is $\rho_a^{\rightarrow}\alpha$ with $\rho_a^{\rightarrow}$ given by (2).

In practical applications it is important to bear in mind the following meaning of (2). If only phase modulation is present, then one can calculate the upper sidebands by simply replacing the multiplier with a linear and time-invariant amplifier whose scattering properties are given by (2a). Similarly, if only AM is present then one can calculate the upper sidebands by using (2b). All this is true provided the upper sidebands are normalized with respect to V_e, in the AM case, and with respect to jV_e, in the PM case.

Consider now the PM case. From (2a) one obtains that the round-trip PM transmission $T_\theta^{\rightarrow}T_\theta^{\leftarrow}$ is

$$| T_\theta^{\rightarrow}T_\theta^{\leftarrow} | = N. \tag{20}$$
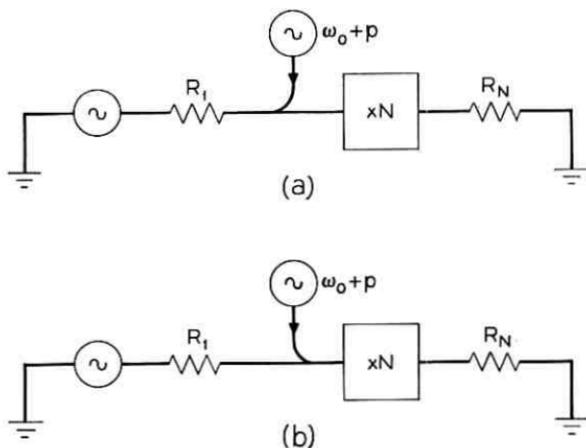

Fig. 6 — Injection of a tone at the input port of the multiplier.

Therefore, interactions between the multiplier and the other stages can cause phase instabilities.*Furthermore, from (20) one can see that the possibility of instabilities increases with the order of multiplication N. However, the presence of loss will decrease the round-trip transmission. More precisely, if η is the multiplier efficiency,† then the round-trip power transmission will become approximately‡

$$| T_{\theta}^{\rightarrow}T_{\theta}^{\leftarrow} |^{2} = N^{2}\cdot\eta^{2} \tag{21}$$

and the possibility that interactions of the external circuit cause PM instability exists when approximately

$$\eta > \frac{1}{N}. \tag{22}$$

Note that amplitude instabilities cannot occur.§ In fact, if the output port of the multiplier is connected to an arbitrary passive stage, then the input AM reflection of the multiplier is always less than¶

$$| \rho_{a}^{\rightarrow} | + | T_{a}^{\rightarrow}T_{a}^{\leftarrow} | , \tag{23}$$

which is never greater than unity, as one may verify from (2b).

APPENDIX

Demonstrations for the multiplier of order 2^{n} which has the minimum number of idlers and uses a "square-law" varactor. Slowly varying modulations.

Let $S_{t}(t)$ be the total elastance of the varactor. It can be separated into the sum of a time-varying component $S(t)$ and an average component S_{0}

$$S_{t}(t) = S(t) + S_{0} .$$

The "external circuit" connected to the varactor can be represented by an impedance Z_{e} in series with a voltage generator V_{g} , as it is shown in Fig. 7.

Let $Z(\omega)$ be the total impedance in series with $S(t)$. Then $Z = Z_{e} + S_{0}/j\omega$.

---

* This is discussed in detail in the second paper.
† Output power divided by input power.
‡ η^{2} is the round-trip transmission through an attenuator which has a forward (power) transmission η. Equation (21) has been obtained by representing the multiplier as the cascade connection of such an attenuator and a lossless multiplier.
§ It is important to note that it is assumed that the "bias circuit" is properly designed, so that low-frequency fluctuations of the average varactor capacitance C_{0} are avoided.
¶ This can be obtained by using standard techniques.

The hypotheses are:

$$Z = 0 \quad \text{if} \quad \omega \cong 2^s\omega_0 \quad (s = 1, \cdots , n - 1) \tag{24a}$$

$$Z = R_1 \quad \text{if} \quad \omega \cong \omega_0 \tag{24b}$$

$$Z = R_N \quad \text{if} \quad \omega \cong N\omega_0 = 2^n\omega_0 \tag{24c}$$

$$Z = \infty \quad \text{if} \quad \omega \quad \text{is far from} \quad 2^s\omega_0 \quad (s = 0, \cdots , n). \tag{24d}$$

Equation (24a) requires $Z = 0$ in the neighborhood of the idler frequencies $2\omega_0 , \cdots , 2^{n-1}\omega_0$. Equations (24b), (24c) require Z to be equal to the input impedance R_1 for $\omega \cong \omega_0$, and to be equal to the output impedance R_N for $\omega \cong N\omega_0$. In (24d) it is required that current flow be limited to the frequencies of (24a), (24b), and (24c). Because
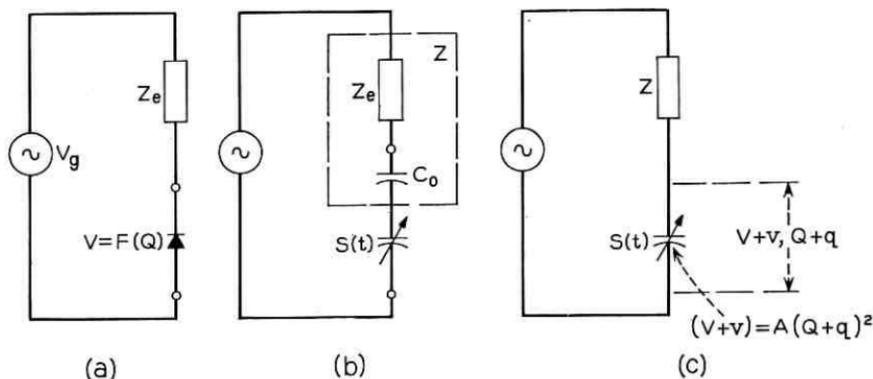
Fig. 7 — Equivalent circuit of the frequency multiplier.

of this last requirement, the variable component $Q(t)$ of the charge of the varactor (*when the sidebands are absent*) is of the type

$$Q(t) = \sum_{\substack{s=0, \cdots, n \\ r=\pm 2^s}} Q_r \exp (jr\omega_0 t). \tag{25}$$

Since the varactor has a square-law Q–V characteristic $F(Q)$, the total voltage* $V_t(t)$ across the varactor is related to $Q(t)$ through a relation of the type

$$F(Q(t)) = V_t(t) = AQ(t)^2 + Q(t)S_0 + V_0 . \tag{26}$$

Therefore, the time-varying part of V_t will consist of two components: one component $V(t)$ due to $AQ(t)^2$

---

* The sign convention in this appendix is that power is positive when flowing into the varactor.

$$V(t) = A \sum_{i \neq r} Q_i Q_r \exp [j(r + i)\omega_0 t] = \sum V_l \exp (jl\omega_0 t)$$

$$(i = \pm 2^s; \quad r = \pm 2^h; \quad s, h = 0, \cdots, n) \tag{27}$$

and another component $V' = Q(t)S_0$ which represents the contribution of a constant capacitance $C_0 = 1/S_0$. The total elastance is obtained by taking the derivative of (26) with respect to $Q(t)$:

$$S_t(t) = S(t) + S_0 = 2AQ(t) + \frac{1}{C_0} \tag{28}$$

and one finds a constant capacitance C_0 in series with a time-varying elastance component $S(t)$ originating from $AQ(t)$. Therefore, the equivalent circuit of the multiplier becomes that of Fig. 7(b) in which the nonlinear element has a Q–V characteristic of the type AQ^2, and $V(t)$, $Q(t)$, $S(t)$ represent its voltage, charge, and elastance, respectively, when the sidebands are absent. From (28), (25) one has

$$S(t) = 2AQ(t) = \sum_{\substack{s=0,\cdots,n \\ r=\pm 2^s}} S_r \exp (jr\omega_0 t) \tag{29}$$

with

$$S_r = 2AQ_r . \tag{30}$$

Let now

$$q(t) = \sum_{\substack{s=0,\cdots,n \\ r=\pm 2^s \\ i=\pm 1}} q_{r,i} \exp [j(r\omega_0 + ip)t] \tag{31}$$

be the sideband components of the charge of the varactor. Then the voltage sidebands across $S(t)$ are

$$v(t) = S(t)q(t) = \sum v_{r,i} \exp [j(r\omega_0 + ip)t]. \tag{32}$$

By substituting (29), (31), in (32) one obtains* $v_{r,i}$

$$v_{r,\pm 1} = \sum_{\substack{s=0 \\ u=\pm 2^s}}^{n} S_u q_{(r-u),\pm 1} \qquad (r = 1, \cdots, 2^n). \tag{33}$$

But from (24d) and from the fact that $Z(\omega + p) \cong Z(\omega)$, one has

$$q_{(r-u),\pm 1} = 0 \qquad \text{for} \qquad (r - u) \neq \pm 1, \cdots, \pm 2^n. \tag{34}$$

In (33), therefore, one has to consider only those values of u which satisfy (put: $r = 2^h$)

$$2^h - u = \pm 1, \cdots, \pm 2^n. \tag{35}$$

---

* Note that $q_{r,i} = q^*_{r,-i}$, $v_{r,i} = v^*_{r,-i}$, (see footnote on p. 780). Therefore, consideration can be limited to the case $r > 0$.

Then, the only possible values of u are

$$u = 2^{h-1}, \qquad u = 2^{h+1}, \qquad u = -2^h, \tag{36}$$

because all other values of u (remember that u has to be of the type $u = \pm 2^s$) would cause $|r - u|$ to be either zero or odd. Therefore, remembering that $r = 2^h$, one finds that (33) consists of only three terms:

$$v_{r,\pm 1} = q_{r/2,\pm 1}S_{r/2} + q_{2r,\pm 1}S_{-r} + q^*_{r,\mp 1}S_{2r} \quad (r = 1, \cdots, 2^n). \tag{37}*$$

In conclusion, in order that $q(t)$, $v(t)$ represent sideband components produced by sources located only at the input and the output of the multiplier, it has to be (*necessary* and *sufficient* condition)

$$v_{r,\pm 1} = 0 \quad \text{for} \quad r = \pm 2, \cdots, \pm 2^{n-1} \tag{38a}$$

$$q_{r,\pm 1} \neq 0 \quad \text{only for} \quad r = \pm 1, \cdots, \pm 2^n \tag{38b}$$

$$v_{r,\pm 1} = q_{r/2,\pm 1}S_{r/2} + q_{2r,\pm 1}S_{-r} + q^*_{r,\mp 1}S_{2r} \quad \text{for} \quad r = 1, \cdots, 2^n. \tag{38c}$$

Equation (38a) follows from (24a); (38b) follows from (24d); (38c) is (37).

Note that (38a) and (38b), are the constraints given by the "external circuit" (which includes C_0), and (38c) is the constraint given by the elastance $S(t)$ of the varactor.

The "equilibrium equations" for the "carriers" Q_r, V_r are obtained in a similar way. They are

$$V_r = 0 \quad \text{for} \quad r = \pm 2, \cdots, \pm 2^{n-1} \tag{39a}$$

$$Q_r \neq 0 \quad \text{only for} \quad r = \pm 1, \cdots, \pm 2^n \tag{39b}$$

$$V_1 = j\omega_0 Q_1 R_1 \tag{39c}$$

$$V_N = -jN\omega_0 Q_n R_n \tag{39d}$$

$$V_r = A[Q_{r/2}Q_{r/2} + 2Q_{2r}Q^*_r] \quad \text{for} \quad r = 1, \cdots, 2^n. \tag{39e}$$

Equations (39a) and (39b) follow from (24a) and (24d), respectively. (39c) follows from (24b) and the hypothesis that the input generator is matched to the multiplier. Equation (39d) follows from (24c). Equation (39e) can be derived from (27) by using the *same* procedure used to derive (38c) from (32).

---

* Note, if $r = 1$, the first term of the second member is zero; if $r = N$, only the first term of the second member is nonzero.

A.1 *Demonstration of* $T_a^- = (-1)^n/N$

Theorem 1: If $\hat{q}$,$\hat{v}$ *satisfies* (38) *and if*

$$\bar{q}_{r,\pm 1} = \hat{q}_{r,\pm 1} \exp\left[j\frac{\pi}{2}(-1)^s\right]$$

$$\bar{v}_{r,\pm 1} = -\hat{v}_{r,\pm 1} \exp\left[j\frac{\pi}{2}(-1)^s\right] \qquad (r = 2^s; s = 0, \cdots, n)$$

(40)

then also $\bar{q}$, $\bar{v}$ *satisfies* (38), *i.e.*, *it gives sidebands produced by sources at* $\omega_0 \pm p$, $N\omega_0 \pm p$.

Proof: $\bar{q}$, $\bar{v}$ clearly satisfies (38a) and (38b) because $\hat{q}$, $\hat{v}$ does. Therefore, it remains to be demonstrated that (38c) is satisfied.

By substituting (40) in (38c) one obtains

$$-\hat{v}_{r,\pm 1} \exp\left[j\frac{\pi}{2}(-1)^s\right] = \hat{q}_{r/2,\pm 1}S_{r/2} \exp\left[j\frac{\pi}{2}(-1)^{s-1}\right]$$

$$+ \hat{q}_{2r,\pm 1}S_{-r} \exp\left[j\frac{\pi}{2}(-1)^{s+1}\right] + \hat{q}_{r,\mp 1}^* S_{2r} \exp\left[-j\frac{\pi}{2}(-1)^s\right],$$

which is satisfied because $\hat{v}$, $\hat{q}$ satisfies (38), and because

$$-\exp\left[j\frac{\pi}{2}(-1)^s\right] = \exp\left[j\frac{\pi}{2}(-1)^{s\mp 1}\right] - \exp\left[-j\frac{\pi}{2}(-1)^s\right].$$

Suppose now that $\hat{q}$, $\hat{v}$ is produced by a PM source located at the input of the multiplier. Therefore, remembering that $\rho_\theta^- = 0$ because of (19), the input components of $\hat{q}$, $\hat{v}$ represent a PM forward wave, and also the output components constitute a PM forward wave. Consider now the input $(r = 1)$ components of $\bar{q}$, $\bar{v}$. From (40) one has

$$\bar{v}_{1,\pm 1} = -j\hat{v}_{1,\pm 1} \qquad \bar{q}_{1,\pm 1} = j\hat{q}_{1,\pm 1} \qquad (41)$$

from which one can see that, since $\hat{v}_{1,\pm 1}$, $\hat{q}_{1,\pm 1}$ is an FM "forward" wave, then $\bar{v}_{1,\pm 1}$, $\bar{q}_{1,\pm 1}$ is an AM "backward" wave. In a completely similar way one finds that also the output components $\bar{v}_{N,\pm 1}$, $\bar{q}_{N,\pm 1}$ represent an AM "backward wave". Finally, remembering (4) and the definitions of T_θ^-, T_a^-, one has

$$T_\theta^- = \left(\frac{\theta_N^-}{\theta_1^-}\right)_{\theta_N = 0} = \frac{\hat{v}_{N,1}}{V_N}\frac{V_1}{\hat{v}_{1,1}}$$

$$T_a^- = \left(\frac{a_1^-}{a_N^-}\right)_{a_1 = 0} = \frac{\bar{v}_{1,1}}{V_1}\frac{V_N}{\bar{v}_{N,1}}.$$

By combining this result with (40) one has

$$\frac{\hat{v}_{N,1}}{\hat{v}_{1,1}}\frac{\hat{v}_{1,1}}{\hat{v}_{N,1}} = T_\theta^\to T_a^\gets = (-1)^N.$$

But $T_\theta^\to = N$. Therefore, one obtains the desired result

$$T_a^\gets = \frac{(-1)^n}{N}.$$

A.2 *Demonstration of* $\rho_a^\to = [N - (-1)^n]/N \cdot 3$

$$\rho_a^\to = \frac{N - (-1)^n}{N \cdot 3}. \tag{42}$$

First it will be shown that the charge $\overset{\circ}{q}(t)$ defined by

$$\overset{\circ}{q}_{r,1} = \overset{\circ}{q}_{r,-1} = (1 - \rho(s))Q_r \qquad (r = 2^s, s = 0, \cdots, n) \tag{43}$$

with

$$\rho(s) = \frac{2^{n-s} - (-1)^{n-s}}{2^{n-s} \cdot 3} \tag{44}$$

gives the charge sidebands produced by an AM wave of amplitude $a_1^\to = 1$ arriving at the input of the multiplier.

Let first the fact that $\overset{\circ}{q}(t)$ satisfies (38) be demonstrated. Substitute (43) in (37) and take into account (30). One obtains $\overset{\circ}{v}$, the voltage associated with $\overset{\circ}{q}$,

$$\overset{\circ}{v}_{r,\pm 1} = 2A\{Q_{r/2}Q_{r/2}(1 - \rho(s - 1))$$
$$+ Q_{2r}Q_{-r}[(1 - \rho(s)) + (1 - \rho(s + 1))]\}$$
$$(r = 2^s; s = 1, \cdots, n - 1). \tag{45}$$

From (44) one has

$$2 - \rho(s) - \rho(s + 1) = 2(1 - \rho(s - 1)).$$

Therefore, by using (39e), (45) gives

$$\overset{\circ}{v}_{r,\pm 1} = 4(1 - \rho(s - 1))V_r = 0 \quad \text{for} \quad r = 2, \cdots, 2^{n-1}$$

because of (39a). Therefore, $\overset{\circ}{q}, \overset{\circ}{v}$ satisfies (38). Furthermore, it represents AM because the sidebands $\overset{\circ}{q}_{r,\pm 1}$ are "in-phase" with the carriers Q_r. This is shown by (43).

Now let the output indexes $a_N^\to$, $a_N^\gets$ of $\overset{\circ}{q}, \overset{\circ}{v}$ be calculated. Consider the following expressions: (38c) with $r = N$; (30) with $r = N/2$; (43)

with $r = N/2$; (39e) with $r = N$. From them one obtains

$$\overset{\circ}{v}_{N,1} = V_N .$$

From this first relation, by using (39d) and (43) with $r = N$, one obtains

$$\overset{\circ}{v}_{N,1} = -R_N(jN\omega_0\overset{\circ}{q}_{N,1}).$$

This second relation gives $a_N^{\leftarrow} = 0$, because $jN\omega_0\overset{\circ}{q}_{N,1}$ is the current at $N\omega_0 + p$. Then the first relation gives at once $a_N^{\rightarrow} = 1$.

From the fact that $a_N^{\leftarrow} = 0$, $a_N^{\rightarrow} = 1$ and from the fact that (19) gives $T_a^{\rightarrow} = 1$, the input indexes must be

$$a_1^{\rightarrow} = 1, \qquad a_1^{\leftarrow} = \rho_a^{\rightarrow} .$$

Therefore, the input voltage sidebands must be

$$\overset{\circ}{v}_{1,\pm1} = (1 + \rho_a^{\rightarrow})V_1 . \tag{46}$$

But, by substituting (43) in (37) with $r = 1$, one has

$$\overset{\circ}{v}_{1,1} = (1 - \rho(1))Q_2 S_{-1} + (1 - \rho(0))Q_1^* S_2$$

from which, remembering that $Q_2 S_{-1} = Q_1^* S_2 = V_1$ because of (39e) and (30), one has

$$\overset{\circ}{v}_{1,1} = (2 - \rho(1) - \rho(0))V_1 = (1 + \rho(0))V_1 \tag{47}$$

in which use has been made of (44). Finally, by combining (47) with (46), one obtains

$$\rho_a^{\rightarrow} = \rho(0) = \frac{N - (-1)^n}{3N}$$

which is the desired result.

A.3 *Demonstration of the fact that the multiplier does not produce AM $\leftrightarrows$ PM conversion, if p is small enough.*

An input PM source does not cause AM if p is small enough. In fact, one may verify that the sidebands produced by an input PM forward wave of amplitude $\theta_1^{\rightarrow} = \theta$ are

$$q_{r,1} = jrQ_r\theta, \qquad q_{r,-1} = jrQ_r\theta^* \tag{48}$$

which are in "quadrature" with respect to the carriers Q_r. To demonstrate that there is no PM $\rightarrow$ AM conversion it therefore remains to be shown that an output PM source does not produce AM. This can be done by applying transformation (40) to the AM sidebands given by (43). In fact, in this way one obtains PM sidebands. They are

produced by PM sources located both at the input and the output of the multiplier. Therefore, one concludes that there is no PM $\rightarrow$ AM, if $p \cong 0$.

By using the same procedure one can show that there is no AM $\rightarrow$ PM conversion. In fact, the discussion following (43) shows that an input AM source does not produce PM. Then, by applying (40) to (48), one finds that an output AM source does not produce PM. This concludes the demonstration.

REFERENCES

1. Dragone, C., Phase and Amplitude Modulation in High-Efficiency Varactor Frequency Multipliers of Order $N = 2^n$—Stability and Noise, B.S.T.J., this issue, pp. 797–834.
2. Carlin, H. F., The Scattering Matrix in Network Theory, Trans. IRE, *CT-3*, June, 1956, pp. 88–96, (see his extensive bibliography).
3. Prabhu, Vasant, Noise Performance of Abrupt-Junction Varactor Frequency Multipliers, Proc. IEEE, February, 1966, pp. 285–287. Maunsell, H. I., private communication.
4. Manley, J. M., and Rowe, H. E., Some General Properties of Nonlinear Elements, Part 1, General Energy Relations, Proc. IRE, July, 1956, pp. 904–913.
5. See the use of frequency multiplication in the Armstrong type indirect FM transmitter (see for instance, Schwartz, M., *Information, Transmission, Modulation, and Noise*, McGraw-Hill Co., New York, 1955).
6. Burckhardt, C. B., Spurious Signals in Varactor Multipliers, Proc. IEEE, *53*, April, 1965, pp. 389–390.
7. Penfield, P. and Rafuse, R. P., *Varactor Applications*, MIT Press, 1962.

Phase and Amplitude Modulation in High-Efficiency Varactor Frequency Multipliers of Order $N = 2^n$ – Stability and Noise

By C. DRAGONE

A general analysis of the stability conditions of frequency multipliers of order $N = 2^n$ is presented. The frequency multipliers to be considered are harmonic generators which use varactor diodes as nonlinear elements. The type of instability investigated is that which causes spurious tones to appear at the output of a multiplier in the vicinity of the desired harmonic. It arises because an efficient multiplier is potentially unstable with respect to the quadrature components of its sidebands.

This paper shows how instability arises and how it can be avoided. One of the main results is that, to obtain stability in practical cases, it is sufficient that the bandwidths of the various resonant circuits satisfy some simple conditions.

I. SUMMARY OF RESULTS

A general analysis of the stability conditions of frequency multipliers of order $N = 2^n$ is presented. The frequency multipliers to be considered are harmonic generators which use varactor diodes as nonlinear elements.

That stability is one of the most serious problems in high-efficiency multipliers is a widely known experimental fact.[1,2]

At the present time little is known of the restrictions placed by the condition of stability on the available circuit configurations. Consequently, present design procedures leave the problem of stability to be solved experimentally, and this is often done at the expense of efficiency. Furthermore, multipliers which are individually stable may become unstable when connected together to form a chain. As a result, isolators are often needed. The isolators will guarantee stability but will lower the overall efficiency.

The practical importance of the problem to be analyzed is illustrated by the following example.

Suppose that one has designed a stable and very efficient octupler, and that one wants to reduce the output noise by using a bandpass filter consisting of a high Q resonant circuit connected in series with the load. In general, one will not obtain the desired result. In fact, the output noise will, in general, increase rather than decrease. Furthermore, if the bandwidth of the filter is too narrow, then the multiplier will become unstable.*

This example illustrates the important fact that, as the circuit approaches an unstable condition, the output noise level increases indefinitely. Then, when the multiplier is on the point of becoming unstable, the output noise becomes very large at some frequencies. Furthermore, when instability arises, spurious tones appear in the vicinity of the carriers. Therefore, not only is it important that a multiplier be stable, but it is also important that it be far from instability, if one wants a low-noise multiplier.

This paper shows how instability arises, and how to avoid it. It also shows how to derive the output noise from a knowledge of the various noise sources. Some of the results which have been obtained are summarized by the following statements.

1.1 *Doubler*

Consider the doubler first.

(*i*) The stability of the doubler will not, in general, depend on the impedance presented by the input circuit (to the varactor) in the "vicinity" of the input carrier ω_0. More precisely, in the design of the input circuit, consideration can be limited to those frequencies whose distance from the carrier ω_0 is larger than half the bandwidth of the output circuit [roughly; see (44), (45), (46)]. At those input frequencies which are not in the "vicinity" of the input carrier,

(*ii*) the impedance presented by the input circuit to the varactor should be large rather than small, as compared to R_1 [see (13), Theorem 2, Theorem 3], where R_1 is the impedance presented by the input circuit at ω_0. Furthermore:

(*iii*) The output bandwidth should be large compared to the input bandwidth [see (35), etc.]. Furthermore:

(*iv*) An efficient chain of more than two doublers which are individually stable will, in general, be stable if each doubler has been designed according to (*iii*) and if, in addition, each doubler is sufficiently broadband with respect to the preceding one [see (58)]. Furthermore:

---

* This example is derived at the end of this paper. See (51), (52), etc.

(v) Those circuit configurations should be preferred which produce PM $\rightleftarrows$ AM conversion [see (55)]. Therefore, low-pass circuit configurations at the input, and high-pass circuit configurations at the output, are, in general, preferable to bandpass circuit configurations.

1.2 *Multiplier of Order $N = 2^n$*

Consider a multiplier of order $N = 2^n > 2$ which has the minimum number of idlers. This memorandum shows that such a multiplier is equivalent to a chain of doublers.* Therefore, results obtained for the doubler can be extended to this type of multiplier. For instance (iv) gives:

An efficient multiplier of order 2^n which has the minimum number of idlers should have

$$B_1 \ll B_2 \ll \cdots \ll B_{n-1} \ll B_n \,,$$

where B_1 is the bandwidth of the input circuit, B_2 is the "equivalent" bandwidth of the first idler, etc.

Note that the above results† apply to the case of a lossless multiplier. The presence of losses reduces the restrictions placed by the condition of stability. If $\eta < 1/N$, then the question of stability does not arise any more, in general. This has been shown by Ref. 3.

II. PRELIMINARY CONSIDERATIONS

2.1 *Method of approach of the mathematical description of the multiplier and its signals*

This paper is concerned with the presence of amplitude and phase fluctuations in the multiplier signals. Suppose, for the moment, that these two types of fluctuations are independent from each other. That is, suppose that the multiplier does not produce AM-to-PM conversion, and vice-versa. Suppose, furthermore, that only PM is present. Then each signal will consist of a carrier and of a pair of sidebands in quadrature with respect to the carrier. Since either sideband can be obtained from a knowledge of the other, then one may consider only one of the two sidebands and ignore the other one. Let then, for instance, the

---

* See Theorem 1 of the second section. Note that the equivalence is exact only if the varactor has a square-law Q–V characteristic.

† These results implied that the impedance presented by the circuit (to the variable capacitance) is "large" at the frequencies which are far from the various carriers. Therefore, subharmonic oscillations and bias instabilities are not included in this analysis.

upper sideband be chosen as the variable, and let it be described in terms of propagating waves.

Then the input variables of the multiplier will be the two waves* $v_{\omega_0+p}^{\rightarrow}$, $v_{\omega_0+p}^{\leftarrow}$ which constitute the upper sideband of the input carrier ω_0. Similarly, the output waves $v_{N\omega_0+p}^{\rightarrow}$, $v_{N\omega_0+p}^{\leftarrow}$ represent the output variables. ω_0 is the input "carrier" frequency of the multiplier, and p is the frequency of the fluctuations. At this point the scattering formalism furnishes a convenient way of describing the properties of the multiplier. More precisely, one may define the PM scattering parameters of the multiplier as the reflection and transmission coefficients which relate the "scattered" waves $v_{\omega_0+p}^{\leftarrow}$, $v_{N\omega_0+p}^{\rightarrow}$ to the "incident" ones $v_{\omega_0+p}^{\rightarrow}$, $v_{N\omega_0+p}^{\leftarrow}$. One obtains in this way the PM scattering matrix $\hat{S}_\theta$ of the multiplier.

Through the analysis $v_{\omega_0+p}^{\rightarrow}$, $v_{\omega_0+p}^{\leftarrow}$, $v_{N\omega_0+p}^{\rightarrow}$, $v_{N\omega_0+p}^{\leftarrow}$, have been normalized with respect to the carriers. In this way the variables become the dimensionless coefficients

$$j\theta_1^{\rightarrow} = \frac{v_{\omega_0+p}^{\rightarrow}}{V_{\omega_0}}, \qquad j\theta_1^{\leftarrow} = \frac{v_{\omega_0+p}^{\leftarrow}}{V_{\omega_0}}, \qquad j\theta_N^{\rightarrow} = \frac{v_{N\omega_0+p}^{\rightarrow}}{V_{N\omega_0}}, \qquad j\theta_N^{\leftarrow} = \frac{v_{N\omega_0+p}^{\leftarrow}}{V_{N\omega_0}},$$

which represent "modulation indexes". Note that V_{ω_0} is the Fourier coefficient of the input carrier ω_0, and $V_{N\omega_0}$ is the Fourier coefficient of the output carrier $N\omega_0$.

In a completely similar way the AM case is treated. By normalizing the AM upper sidebands with respect to V_{ω_0}, $V_{N\omega_0}$, one obtains four modulation indexes $a_1^{\rightarrow}$, $a_1^{\leftarrow}$, $a_N^{\rightarrow}$, $a_N^{\leftarrow}$ which represent the AM variables. Furthermore, by considering how these waves are scattered by the multiplier, one finds the AM scattering matrix $\hat{S}_a$ which describes the multiplier AM properties.

2.2 General Scattering Properties of an "Ideal" Multiplier

Consider a multiplier of order $N = 2^n$ which has the following properties: it is lossless, it has the minimum number of idlers $(n - 1)$, and it uses a varactor having a square-law Q–V characteristic. The first paper (See Ref. 3) has shown that, if the fluctuations are slow enough $(p \cong 0)$, then such a multiplier does not produce AM $\leftrightarrows$ PM conversion† and it has the following scattering matrices:

---

* $v_\omega^{\leftarrow}$, $v_\omega^{\rightarrow}$ designate the Fourier coefficients of the voltage components of frequency ω. They propagate in the directions indicated by the arrows. For more details on the mathematical description of the multiplier and its signals see Ref. 3.

† It is important to note that it has been assumed that the "bias circuit" of the multiplier is properly designed, so that low-frequency fluctuations of the average capacitance C_0 are avoided. See Ref. 4.

$$\hat{S}_a = \begin{vmatrix} \vec{\rho_a} & \overleftarrow{T_a} \\ \vec{T_a} & \overleftarrow{\rho_a} \end{vmatrix} = \begin{vmatrix} \dfrac{N - (-1)^n}{3N} & \dfrac{(-1)^n}{N} \\ 1 & 0 \end{vmatrix} \tag{1}$$

$$\hat{S}_\theta = \begin{vmatrix} \vec{\rho_\theta} & \overleftarrow{T_\theta} \\ \vec{T_\theta} & \overleftarrow{\rho_\theta} \end{vmatrix} = \begin{vmatrix} 0 & (-1)^n \\ N & \dfrac{1 - (-1)^n N}{3} \end{vmatrix}. \tag{2}$$

A multiplier for which (1), (2) apply will be called "ideal". According to the preceding discussion, one has that, if only PM is present and if a multiplier is "ideal", then

$$\frac{\overleftarrow{v_{\omega_0+p}}}{V_{\omega_0}} = \overleftarrow{\rho_\theta}\,\frac{\vec{v}_{\omega_0+p}}{V_{\omega_0}} + \overleftarrow{T_\theta}\,\frac{\overleftarrow{v}_{N\omega_0+p}}{V_{N\omega_0}}$$

$$\frac{\vec{v}_{N\omega_0+p}}{V_{N\omega_0}} = \vec{T_\theta}\,\frac{\vec{v}_{\omega_0+p}}{V_{\omega_0}} + \overleftarrow{\rho_\theta}\,\frac{\overleftarrow{v}_{N\omega_0+p}}{V_{N\omega_0}}. \tag{3}$$

If only AM is present, the relations existing among the upper sidebands are similar to (3) (Replace $\vec{\rho_\theta}$, $\overleftarrow{T_\theta}$, etc. with $\vec{\rho_a}$, $\overleftarrow{T_a}$, etc.).

In conclusion, one can calculate the sidebands present in an "ideal" multiplier in the following way. First, separate the PM components from the AM components. Next, solve separately the two cases AM, PM. The PM case is solved by means of (3). That is, one replaces the multiplier with a linear amplifier which has

    (i) a forward transmission coefficient $\vec{T_\theta}$,
    (ii) a reverse transmission coefficient $\overleftarrow{T_\theta}$,
    (iii) an input reflection coefficient $\vec{\rho_\theta}$, and
    (iv) an output reflection coefficient $\overleftarrow{\rho_\theta}$.

Then one supposes that the variables consist of the PM upper sidebands only, as it is illustrated by (3), and one readily calculates them. In a completely similar way one calculates the AM upper sidebands.

2.3 *General Considerations on Stability*

The frequency response of an unstable circuit which is on the point of becoming stable is infinite.* Hence, if the parameters† of a circuit are continuously varied, a circuit can go from a stable to an unstable situation if and only if its frequency response becomes infinite. Therefore, a certain situation A is stable if‡

---

* See Ref. 5, p. 316, or Ref. 6, p. 112.
† These parameters are discussed in the next sections.
‡ Note that there are stable circuits for which (4) is not satisfied (conditionally

one can cause the circuit to pass from situation A to a stable
situation B by continuously varying its parameters, *without*   (4)
causing its frequency response to become infinite.

For instance, consider a multiplier which is individually stable and
which is connected to a generator and a load which have reflections.
Suppose, furthermore, that one is concerned about the possibility that
these reflections cause instability. Then, according to (4) one may find
out whether or not the circuit is unstable (or conditionally stable) by
examining whether or not the circuit response becomes infinite as the
reflections are decreased in amplitude. This can be done by inserting
attenuators between the multiplier and the other stages (the load and
the generator). In this example, situation B occurs when all reflections
become zero. The importance of (4) and of this example will be better
understood in the next sections.

Consider now the case of a multiplier which does not produce AM $\rightleftarrows$
PM conversion. Then only phase instabilities can occur, as has been
shown by Ref. 3. Therefore, one is concerned about its behavior with
respect to PM only.

In general, when the multiplier is on the point of becoming unstable,
all four PM scattering coefficients $(T^{\rightarrow}, T^{\leftarrow}, \rho^{\rightarrow}, \rho^{\leftarrow})$ become infinite si-
multaneously.* Therefore, one may choose any one of them to study
the stability of the multiplier. However, since the forward transmission
$T^{\rightarrow}$ is of special interest because it gives the output response to input
PM signals, it will be selected as the multiplier frequency response to
be analyzed.

The stability of a multiplier depends on the values of certain param-
eters which will be discussed in the following sections. Those values of
these parameters for which the circuit is stable constitute the so called
"region of stability". At the boundary of the region of stability the PM
forward transmission becomes infinite at one or more frequencies, as
has already been pointed out. Therefore, in order to find the stability
region, one must find those "critical" values of the parameters for
which $T^{\rightarrow} = \infty$. Then, once the boundary is found, one must identify
which one of the two regions separated by the boundary is stable. This
can be done by applying the stability test (4) or any of the usual stability

---

stable circuits. See Ref. 6, p. 162). However, for practical reasons, one is interested
in designing a multiplier which does not become unstable if the losses are increased.
(i.e., which is unconditionally stable.) For such a multiplier (4) is necessary and
sufficient.

* See Ref. 6, p. 164. Since the discussion will be mostly confined to the PM case,
the subscripts a, θ used in (1), (2), and (3) to distinguish the AM case from the
PM case will be omitted unless necessary.

tests (such as the "Nyquist plot", etc.) to one point of either one of the two regions.

As will be seen, a multiplier can be represented by means of a chain of stages which are individually stable. The analysis of such a chain may be carried on as follows. Consider the cascade connection of the two stages (b), (c), shown in Fig. 1. If the two stages are individually stable the overall forward transmission becomes infinite if and only if

$$\overleftarrow{\rho_b}\overrightarrow{\rho_c} = 1, \tag{5}$$

where $\overleftarrow{\rho_b}$ is the output reflection of the first stage* and $\overrightarrow{\rho_c}$ is the input reflection of the second stage. Equation (5) may be obtained in the
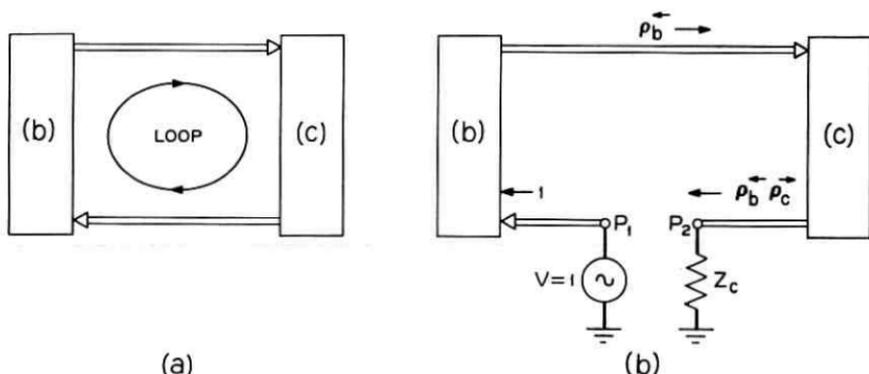
Fig. 1 — "Loop transmission" at the interconnection between two stages.

following way. Consider the interconnection between the two stages (b), (c), and suppose that the reverse path is separated from the forward path, as illustrated in Fig. 1(a). Next, break the reverse-path connection as it is shown in Fig. 1(b). In addition, terminate P_2 in the characteristic impedance Z_c of the interconnection, and apply a unit voltage to P_1, as illustrated in Fig. 1(b). Then the voltage which appears at P_2 is $\overleftarrow{\rho_b}\overrightarrow{\rho_c}$. Therefore, $\overleftarrow{\rho_b}\overrightarrow{\rho_c}$ represents the "return voltage" (i.e., the familiar loop transmission $\mu\beta$)† of the loop indicated in Fig. 1(a), and (5) is demonstrated.

In some cases, one will be interested in knowing the reflection $\rho^{\leftarrow\prime}$ presented by the output of a stage when its input port is connected to a generator which has an output reflection $\overleftarrow{\rho_g} \neq 0$ (see Fig. 2). The output reflection $\rho^{\leftarrow\prime}$ is given by

---

* Which occurs when the other port of the stage is terminated in its characteristic impedance.

† See Ref. 6, p. 44.

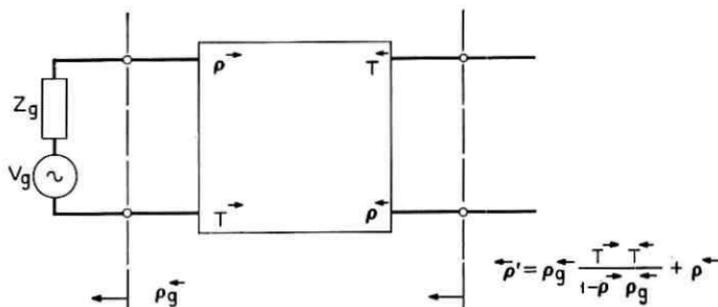Fig. 2 — Effect of $\rho_g^{\leftarrow}$ on the output reflection of a stage.

$$\rho^{\leftarrow\prime} = \rho_g^{\leftarrow} \frac{T^{\rightarrow}T^{\leftarrow}}{1 - \rho^{\rightarrow}\rho_g^{\leftarrow}} + \rho^{\leftarrow}, \tag{6}$$

where $\rho^{\rightarrow}$, $\rho^{\leftarrow}$, $T^{\rightarrow}$, $T^{\leftarrow}$ designate the scattering coefficients of the stage when its ports are terminated in the respective characteristic impedances. Equation (6) may be obtained by using standard techniques.[7]

III. EFFECT OF INTERACTIONS BETWEEN A MULTIPLIER AND TWO PASSIVE STAGES

Consider a multiplier connected between two stages as shown in Fig. 3(a). Such a problem is encountered in two practical cases. One is the design of the input and output networks of a multiplier, as will be seen in the next section. The other case occurs when the load and the generator have reflections and one is concerned about the possibility that these reflections may cause instability. This is the case considered here and it is illustrated in Fig. 3(b). It should be clear, however, that results obtained for either case can be extended to the other one.

In this section it is supposed that the multiplier is lossless. Furthermore, it is assumed that interaction only occurs in the vicinity of the carriers* so that one can approximate the multiplier properties by means of (2) (which is valid if losses are absent and if $p \cong 0$). It is assumed, in addition, that the generator and the load do not produce AM $\rightleftarrows$ PM conversion. This is the worst case, as is explained in the next section.

Since there is no AM $\rightleftarrows$ PM conversion, one is only concerned about phase instabilities and, therefore, it will be assumed that there is only

---

* This will be justified by the results obtained in the next section. In fact, the next section shows that, if the multiplier is "properly" designed, then its PM transmissions ($T^{\rightarrow}$, $T^{\leftarrow}$) are maximum for $p \cong 0$ and decrease monotonically with p. Therefore, interaction is in general more likely to occur for $p \cong 0$ (i.e., for those values of p for which (2) holds), in practical applications.

PM. Therefore, consideration can be limited to the upper sidebands only as it is indicated in Fig. 3 where $Z_g(\omega_0 + p)$, $Z_L(N\omega_0 + p)$ are the impedances of the generator and the load at the upper-sidebands, respectively. Let ρ_g, $\rho_L{}^*$ be the reflection coefficients of the generator and the load, respectively. Then

$$\rho_g = \frac{Z_g(\omega_0 + p) - R_1}{Z_g(\omega_0 + p) + R_1} \qquad \rho_L = \frac{Z_L(N\omega_0 + p) - R_N}{Z_L(N\omega_0 + p) + R_N}, \qquad (7)$$

where R_1, R_N are the values of Z_G, Z_L for $p = 0$.† Now let ρ' indicate* the reflection presented by the multiplier to the load when the input

(a)

(b)
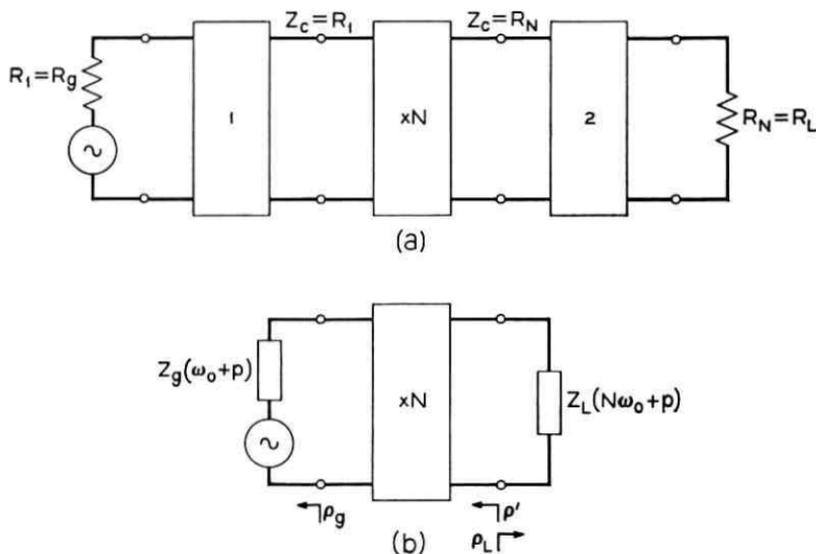
Fig. 3 — Multiplier connected between two stages.

port is connected to the generator. From (6), taking into account the fact that now $\rho^{\rightarrow} = 0$ [see (2)], one has

$$\rho' = \rho^{\leftarrow} + \rho_g T^{\rightarrow} T^{\leftarrow}, \qquad (8)$$

where $\rho^{\leftarrow}$, $\rho^{\rightarrow}$, $T^{\rightarrow}$, $T^{\leftarrow}$ designate the multiplier PM scattering coefficients. From (2) one has

$$\rho^{\leftarrow} = (-1)^n \frac{N - (-1)^n}{3}, \qquad T^{\rightarrow} T^{\leftarrow} = (-1)^n N. \qquad (9)$$

---

* The arrows will be omitted, unless necessary.
† R_1 is the input impedance of the multiplier at ω_0 (the multiplier is assumed to be matched: $Z_g(\omega_0) = R_1$).

According to (5), infinite transmission occurs if (and only if)

$$\rho_L \rho' = 1. \tag{10}$$

Equation (10) is obtained from (5) by considering the cascade connection of the generator and the multiplier as stage (b), and by considering the load as stage (c). By substituting (8) in (10), and by taking into account (9), one obtains

$$\rho_L \rho' = (-1)^n \rho_L \left(N \rho_o - \frac{N - (-1)^n}{3} \right) = 1. \tag{11}$$

According to the stability test given by (4), one has that the circuit is unstable depending on whether or not at some frequencies*

$$(\text{Re})(\rho_L \rho' - 1) \geqq 0 \tag{12}$$

$$(\text{IM})[\rho_L \rho'] = 0.$$

This can be shown by applying (4) as illustrated by the example of the preceding section. In fact, consider a circuit for which (12) occurs at some frequencies. Next, insert an attenuator† between the load and the multiplier. Then, as the attenuator is varied from zero to infinite attenuation, the forward transmission becomes infinite, because of (12), (11). Therefore, (12) guarantees instability. Next, consider a circuit for which (12) never occurs. In this case, as the attenuator is increased, the forward transmission never becomes infinite. Therefore, one concludes that the circuit is stable if and only if (12) never occurs.

3.1 Discussion of (11)

Let the effect of ρ_o on ρ' be considered first. Since both N and $[N - (-1)^n]/3$ are always positive, one has from (11) that the magnitude of ρ' is maximum when $\rho_o = -1$. Therefore,

The output reflection (ρ') is maximum when

$$|Z_o| \ll R_1. \tag{13}$$

Since $|\rho_o| \leqq 1$, the first relation of (12) gives that instability can occur only if

$$|\rho_L| \geqq \frac{3}{4N - (-1)^n} \equiv \rho_{LO}. \tag{14}$$

---

* (RE) means "Real part of". (IM) means "Imaginary part of".

† The attenuator is supposed to be ideal, i.e., without phase delay. Note that the circuit is stable when the attenuator provides infinite attenuation. In fact, in this case $|\rho_L| = 0$ and therefore (12) is never satisfied.

Therefore,

> In order to guarantee stability it is sufficient to require
> $|\rho_L| < |\rho_{Lo}|$. Note that, in a practical case, it will suffice to
> require (14) in the vicinity of the output carrier only.* That   (15)
> is, consideration can be limited to those frequencies which
> fall within the pass-band of the output circuit.

Now put (11) into the form

$$(-1)^n\left(N\rho_g - \frac{N - (-1)^n}{3}\right) - \frac{\rho_L^*}{|\rho_L|^2} = 0. \tag{16}$$

By considering the imaginary part of (16) one finds that (11) [and
therefore (12), also] is satisfied only if either

$$\rho_{LI} = \rho_{gI} = 0 \tag{17}$$

or

$$(-1)^{n+1}\rho_{LI}\rho_{gI} > 0, \tag{18}$$

where ρ_{LI}, ρ_{gI} are the imaginary components of ρ_L, ρ_g. According to
(17), (18) one can say

> Instability can occur only if at some frequencies p either
> one of the following two situations occur: (19)

> (i) both Z_g, Z_L are real.
> (ii) $X_L X_g(-1)^{n+1} > 0$ $(N = 2^n)$.

Consider now the imaginary part of (11). After some manipulations
one obtains

$$\rho_{LR}\left[N\rho_{gR} - \frac{N - (-1)^n}{3}\right] = \frac{-N\rho_{LR}^2}{\rho_{LI}^2}(\rho_{gI}\rho_{LI}). \tag{20}$$

Therefore, if ρ_{gI}, $\rho_{LI} \neq 0$, then (20), (18) give

$$(-1)^n\rho_{LR}\left[N\rho_{gR} - \frac{N - (-1)^n}{3}\right] > 0. \tag{21}$$

Note that, if $\rho_{gI} = \rho_{LI} = 0$, then (21) follows directly from the first
relation of (12). According to (21) one can say

> Instability can occur only if, at some frequencies,
> (21) is satisfied.   (22)

---

* See footnote on p. 806.

It is recalled that $\rho_{LR} > 0$ or $\rho_{LR} < 0$ depending on whether or not $|Z_L| > R_N$ (and similarly for ρ_{gR}, $|Z_g|$, R_1). The importance of (19), (22) follows from the fact that they allow the question of stability to be answered in many cases simply by looking at the signs of X_g, X_L and at the magnitudes of $|Z_g|$, $|Z_L|$ (more precisely, of $|Z_g|/R_1$, $|Z_L|/R_N$). This will be illustrated by the examples of the next section. Note that the presence of $(-1)^n$ in (11), (18), (21), follows from the fact that there is a reversal in the sign of both $\rho^{\rightarrow}$, $T^{\rightarrow}T^{\leftarrow}$ each time N is increased by a factor of two. This is explained by Theorem 1 of the next section which shows that a multiplier of order 2^n is equivalent to a chain of doublers. Since the round-trip transmission of a doubler is negative, each doubler gives a contribution of 180° to the phase of the overall round-trip transmission (from which $(-1)^n$ follows).

Consider now some special cases. If $N = 2, 4$, then (11) becomes

$$\rho_L \rho' = (-1)^n \rho_L (N \rho_g - 1) = 1.$$

Therefore,

if $N = 2$ or 4, instabilities can occur only if both the generator and the load are interacting (i.e., both ρ_L, $\rho_g \neq 0$).   (23)

If $N > 4$, on the other hand, then $|\rho'| > 1$ even if $|\rho_g| = 0$. Therefore,

if $N > 4$, then instability can arise even if the generator does not interact (i.e., $\rho_g = 0$).   (24)

This discussion of interactions will be concluded by emphasizing that

if $\rho_L = 0$ there is no interaction at all. Therefore, one may say that the load reflection ρ_L is the primary cause of in-   (25) stabilities.

IV. ANALYSIS OF THE FREQUENCY DEPENDENCE OF THE PROPERTIES OF A MULTIPLIER OF THE TYPE (26) AND DISCUSSION OF THE STABILITY CONDITIONS

Consider a multiplier which has the following characteristics:

The order of multiplication is $N = 2^n$. It has the minimum number $(n - 1)$ of idlers. It is tuned.   (26)

The analysis of the frequency dependence of the properties of such a multiplier is based on the following theorem which is demonstrated in the Appendix.

Theorem 1:* *A multiplier of the type* (26) *is equivalent to the chain of stages shown in Fig. 4. The doublers of Fig. 4 are "ideal". More precisely, they are frequency independent and lossless, and they do not produce* $AM \rightleftarrows PM$ *conversion. Their scattering matrices are given by* (2) *(with* $N = 2$). *That is,*

$$\hat{S} = \begin{vmatrix} 0 & -1 \\ 2 & 1 \end{vmatrix} \qquad (\textit{If only PM is present})$$

$$\hat{S} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{vmatrix} \qquad (\textit{If only AM is present}). \tag{27}$$
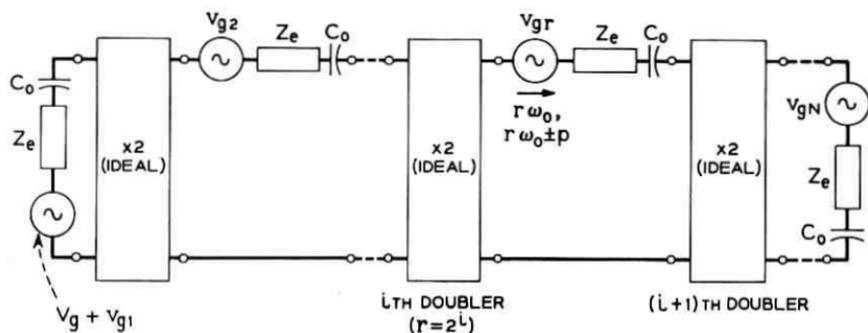
Fig. 4 — Chain of doublers equivalent to the multiplier of order $N = 2^n$ shown in Fig. 5.

Note that this theorem applies to the general case of a multiplier which is lossy and produces $AM \rightleftarrows PM$ conversion. Fig. 5 shows the equivalent circuit of the actual multiplier from which the chain is derived. Z_e is the impedance presented by the external circuit to the varactor. R_s is the series resistance of the varactor. C_0 is the varactor average capacitance. The generator V_G consists of a carrier voltage V_g of frequency ω_0, and of noise terms v_{g1}, v_{g2}, $\cdots$, v_{gN} which correspond to the various sidebands $\omega_0 \pm p$, $2\omega_0 \pm p$, etc.

The impedance Z_e resonates with C_0 at the "desired" carriers ω_0, $2\omega_0$, $\cdots$, $N\omega_0$. Furthermore, Z is so large for ω far from ω_0, $2\omega_0$, $\cdots$, $N\omega_0$ that current flows through the varactor only at the desired carriers and their sidebands.

Let now the chain of Fig. 4 be examined. Consider the impedance

---

* In deriving (27) the approximation $\omega_0 \pm p \cong \omega_0$ has been made. For more details, see the considerations made in the Appendix on the frequency dependence of the properties of the "ideal" doublers of Fig. 4.
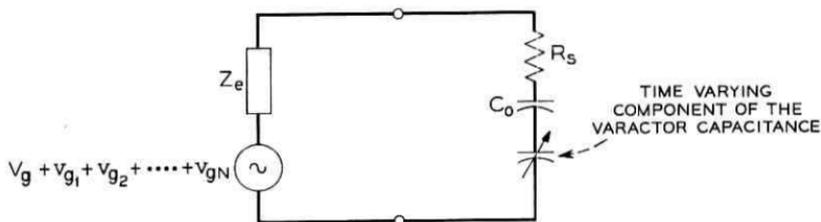
Fig. 5 — Multiplier equivalent circuit.

connecting the output of the ith doubler to the input of the $(i + 1)$th doubler $(i < n)$. Current is flowing only at the frequencies $r\omega_0$, $r\omega_0 \pm p$ $(r = 2^i)$, through this impedance. Furthermore, at $\omega \cong r\omega_0$ (i.e., $\omega = r\omega_0$, $r\omega_0 \pm p$) Z_e gives the impedance presented by the $r\omega_0$-idler circuit to the varactor. Therefore, the impedance connected to the output of the ith doubler corresponds to the ith idler and can be represented by means of a series resonant circuit (resonant at $r\omega_0$) as illustrated in Fig. 6. The resistance R_r shown in Fig. 6 represents the losses of the idler circuit and it is equal to R_s, the series resistance of the varactor, if the varactor is the only lossy element. The chain of Fig. 4 can be represented by the diagrammatic circuit of Fig. 7 in which, according to the preceding considerations, S_{i+1} represents the ith idler circuit and can be approximated as illustrated in Fig. 6. The impedance $Z_{in}^{(i+1)}$ is the characteristic impedance of the input port of the $(i + 1)$th doubler. It is equal to the impedance presented by the input port of the $(i + 1)$th doubler at the carrier frequency $r\omega_0$. $Z_{out}^{(i)}$ is the characteristic impedance of the output port of the ith doubler and it is related to $Z_{in}^{(i)}$ through the formula $Z_{out}^{(i+1)} = Z_{in}^{(i)} + R_r$ (with $r = 2^i$). $Z_{in}^{(i)}$ is calculated in the Appendix.

At $\omega \cong N\omega_0$, the impedance Z_e is the impedance presented to the
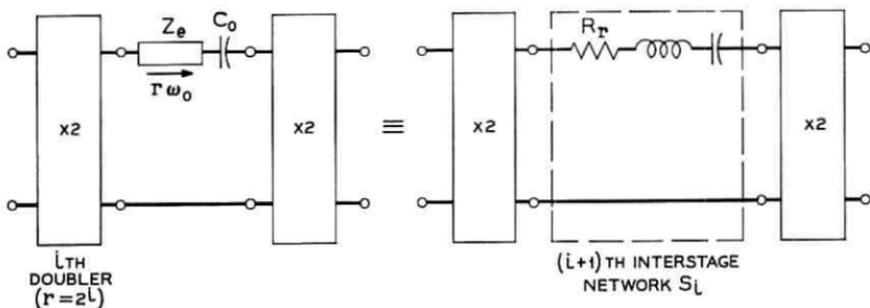
Fig. 6 — $(i + 1)$th interstage network S_{i+1}.

varactor by the output circuit of the multiplier. Accordingly, the impedance connected to the output of the last doubler of Fig. 4 can be represented as shown in Fig. 7, where S_{n+1} corresponds to the output circuit of the multiplier. The input circuit of the multiplier is represented in Fig. 7 by means of a network S_1 connected between the generator and the first doubler. Note that S_1 , S_{n+1} include R_a and C_0 . Furthermore, they are tuned at the carrier frequencies and therefore they provide unity transmission at the carriers if losses are absent.
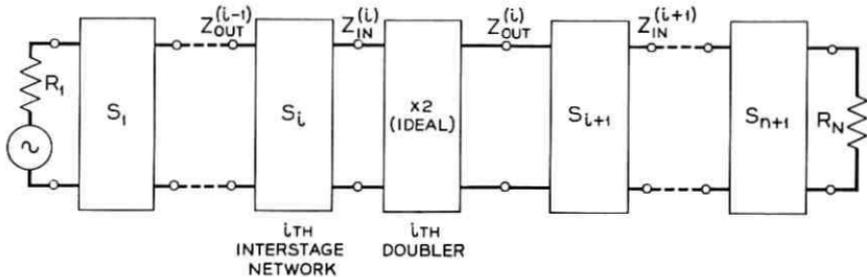
Fig. 7 — Schematic representation of the chain of doublers equivalent to the multiplier of order $N = 2^n$.

4.1 *Analysis of the Worst Case (No Losses, No AM $\rightleftarrows$ PM)*

Assume that the multiplier does not produce AM $\rightleftarrows$ PM conversion, and that it is lossless. (That is, S_1 , S_2 , $\cdots$, S_{n+1} are lossless and do not produce AM $\rightleftarrows$ PM conversion.)

Let the doubler be examined in detail. It is the most important multiplier because it represents the elementary constituent of any multiplier or chain of multipliers.

4.2 *Discussion of the Doubler*

According to the preceding discussion, the equivalent circuit of a doubler consists of three stages: an input stage S_1 , an "ideal" doubler, and an output stage S_2 . Therefore, the general considerations of the preceding section on the interactions between a multiplier and two other stages are applicable to the analysis of a doubler. Suppose that the scattering coefficients of the input and output stages are labeled 1 and 2, respectively. Then the overall forward PM transmission is given by*

* (28) can be obtained by means of standard techniques. See for instance Ref. 7.

$$\overrightarrow{T_i} = \frac{2\overleftarrow{T_1}\overrightarrow{T_2}}{1 - \overrightarrow{\rho_2} + 2\overleftarrow{\rho_1}\overrightarrow{\rho_2}} . \tag{28}$$

Since in this section it is assumed that S_i does not produce AM $\rightleftarrows$ PM conversion and is lossless, its PM scattering parameters are simply given by its reflection and transmission coefficients at the frequency $\omega + p$ ($\omega = \omega_0$ or $2\omega_0$ depending on whether $i = 1$ or $i = 2$). The characteristic impedances of the two ports of S_i are equal because S_i is lossless; they will be denoted $Z_c^{(i)}$. If, for instance, S_i is either one of the two simple circuits of Fig. 8, then

$$\hat{S}_i = \begin{vmatrix} \rho_i & T_i \\ T_i & \rho_i \end{vmatrix}$$
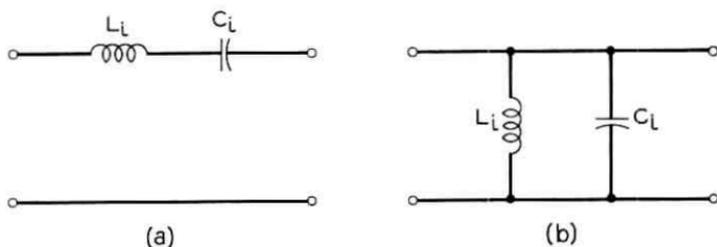
Fig. 8 — Single-tuned resonant circuits.

with†

$$T_i = \overrightarrow{T_i} = \overleftarrow{T_i} = \frac{B_i}{B_i + jp}$$

$$\rho_i = \overrightarrow{\rho_i} = \overleftarrow{\rho_i} = \pm(1 - T_i) \tag{29}$$

$$B_i = \frac{Z_c^{(i)}}{L_i} ,$$

where the negative sign applies to the case of Fig. 8(b). Note that B_i is the 3-dB bandwidth of T_i. Note, furthermore, that the only difference between the two circuits is that they have opposite reflections. If $p \neq 0$, ∞, the signs of the real and imaginary components of the reflections are

$$\begin{array}{lll} \text{circuit (a)} & \rho_{iR} > 0, & \rho_{iI} > 0 \\ \text{circuit (b)} & \rho_{iR} < 0, & \rho_{iI} < 0, \end{array} \tag{30}$$

---

† The approximation $\omega \cong \omega + p$ has been made in deriving (29). That is, the resonant circuit has been supposed to be very narrow band.

where ρ_{iR} is the real component of ρ_i, and ρ_{iI} is the imaginary component.

Suppose now that both S_1, S_2 can be approximated by either one of the two circuits (a), (b) of Fig. 8. From (18), (21) one has that instability may occur only if

$$\rho_{1I}\rho_{2I} > 0, \qquad \rho_{2R}(1 - 2\rho_{1R}) > 0. \qquad (31)$$

From the first of (31) and from (30) it follows that the circuit is unstable only if $\rho_{1R}\rho_{2R} > 0$. Therefore, by combining this result with the second inequality of (31), one has that instability requires

$$\rho_{2R} > 0, \qquad \rho_{1R} > 0. \qquad (32)$$

Therefore, the only case in which instability may occur, is that in which both S_1, S_2 are of the type (a). Let now this case be examined in detail (see Fig. 9).

4.3 *Discussion of the Case in Which Both S_1 and S_2 can be Approximated by Means of a Single-Tuned Series Resonant Circuit*

This case is most important for two main reasons. A first reason is that the equivalent circuit of a varactor includes the series connection of C_0 and the inductance of the diode mount. Therefore, in most practical cases it will be possible to represent both the input and output circuit of a multiplier by means of a series resonant circuit, by first approximation. The second reason is that, in the case of an idler circuit, it has already been pointed out that the equivalent circuit is a series resonant circuit, to a first approximation.

Let the frequency p be normalized with respect to the output bandwidth (that is, $B_2 = 2$) and let $\gamma = B_1/B_2$. From (28) and (29) one has

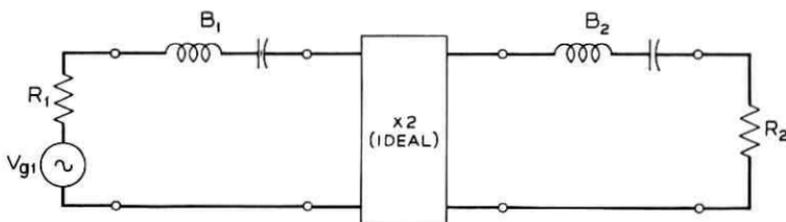$$T_i^{\rightarrow} = \frac{2\gamma}{(\gamma - 2p^2) + jp}. \qquad (33)$$

Fig. 9 — Case in which both the input and output circuit of the doubler can be approximated by means of single-tuned series resonant circuits.

By analyzing (33) one finds that the behavior of $T_t^{\rightarrow}$ is different depending on whether $\gamma < \frac{1}{4}$ or $\gamma > \frac{1}{4}$. More precisely,

- (i) if $\gamma < \frac{1}{4}$, then maximum transmission $(T_M^{\rightarrow})$ occurs at $p = 0$; therefore, $T_M^{\rightarrow} = 2$; and
- (ii) if $\gamma > \frac{1}{4}$, then maximum transmission occurs at $p = \sqrt{(4\gamma - 1)/8}$ and it is given by

$$T_M^{\rightarrow} = \frac{2\gamma}{1/4 + j\sqrt{(4\gamma - 1)/8}}.$$

All this is illustrated in Fig. 10, where $20 \log_{10} |T_t^{\rightarrow}|$ is plotted as a function of p for different values of $\gamma (= B_1/B_2)$. One can see from Fig. 10 that if the output bandwidth is enough greater than the input bandwidth (i.e., $\gamma < \frac{1}{4}$), then the transmission curve decreases monotonically with p. On the other hand, if $\gamma > \frac{1}{4}$, then

a peak appears in the transmission curve, and the maximum transmission increases indefinitely with γ.   (34)
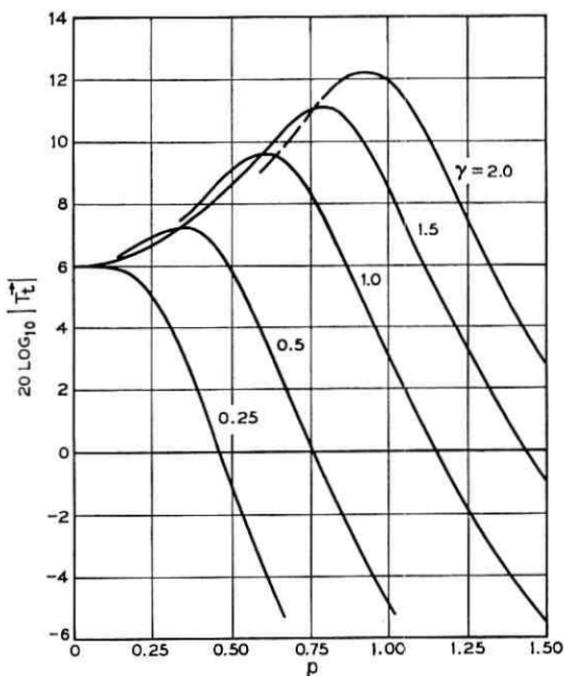

Fig. 10 — Effect of γ $(=B_1/B_2)$ on the forward PM transmission of the doubler of Fig. 9.
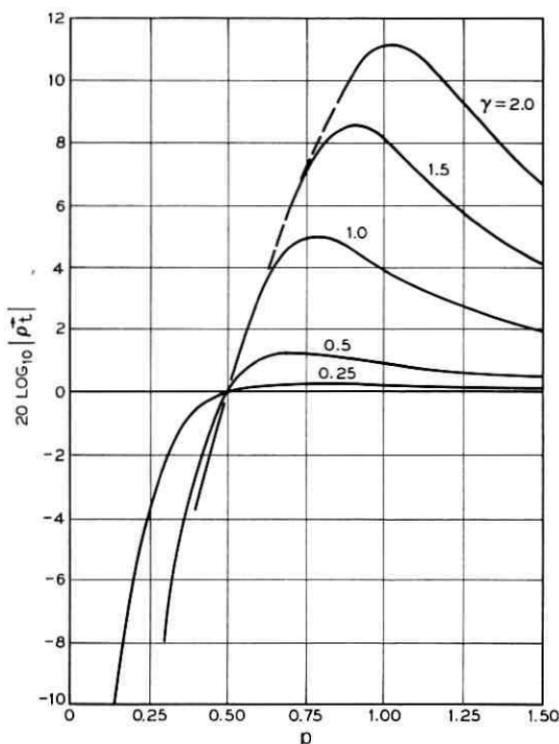
Fig. 11 — Effect of γ $(=B_1/B_2)$ on the input PM reflection of the doubler of Fig. 9.

This is in accordance with (2). One concludes from (34) that small values of B_1/B_2 are desirable if large values of $| T_t^{\rightarrow} |$ are to be avoided. Small values of B_1/B_2 are desirable also for the following stability considerations. If B_1/B_2 is large, then the input PM reflection $\rho_t^{\rightarrow}$ becomes larger than unity, as shown in Fig. 11. In addition, large values of B_1/B_2 enhance the reverse overall transmission $T_t^{\leftarrow}$, as shown in Fig. 12. Consequently, one may conclude

Large values of $\gamma(B_1/B_2)$ deteriorate the stability of the doubler and enhance the PM forward transmission (i.e., the output noise) of the multiplier in the vicinity of the carriers, if both S_1, S_2 are of the type shown in Fig. 8(a). (35)

It is important to note that the circuit becomes unstable only in the limiting case $\gamma = \infty$. Remember that it has been pointed out in the preliminary considerations of the first section that all scattering coeffi-
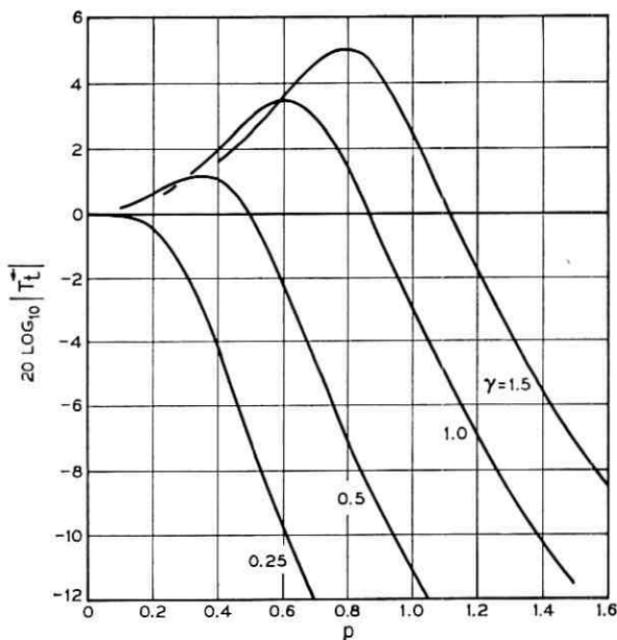
Fig. 12 — Effect of γ ($=B_1/B_2$) on the reverse PM transmission of the doubler of Fig. 9.

cients become infinite when the multiplier is on the point of becoming unstable.

4.3.1 Remaining cases

In the remaining cases, [in which both S_1, S_2 are of the type (a), (b) and at least one of S_1, S_2 is of the type (b)], the overall transmission T_t^- decreases monotonically with p for all values of B_1/B_2. This is in accordance with (32) and it can be directly verified by using (28), (29).

4.4 Discussion of the General Case

In most practical cases the output and input circuits of a multiplier are far more complicated than the simple cases considered in the preceding discussion. For instance, the impedance of the external circuit $Z_e + 1/j\omega C_0$ may have spurious resonances at the sidebands, etc. It is therefore important to consider the case in which the input and output circuits of a multiplier have more complicated configurations. Let first the input circuit be considered in detail.

4.5 *Discussion of the Case in which only the Output Circuit can be Approximated by Means of a Single-Tuned Series Resonant Circuit*

According to the terminology of the first section, let Z_g be the impedance presented by the input circuit (i.e., by the cascade connection of S_1 and R_1) to the "ideal" doubler of Fig. 13. Similarly, let Z_L be the impedance presented by the output circuit to the output port of the "ideal" doubler.

Consider now the two cases illustrated in Fig. 14. The parameters B_{1a}, B_{1b}, B_2 represent the bandwidths of the various resonant elements (when they are considered individually). In order to investigate the
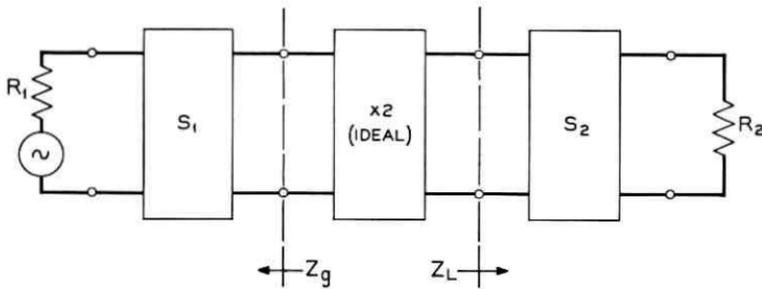

Fig. 13 — Impedances Z_g and Z_L presented by the input and output circuits to the "ideal" doubler.

possibility of instabilities, consider the reflections ρ_L, ρ_g, defined in the previous section ($\rho_g = \rho_1^-$, $\rho_L = \rho_2^-$). Then

$$\rho_L = \frac{jp}{B_2 + jp}. \tag{36}$$

ρ_g is different in the two cases (a), (b). In case (a), ρ_g is given by

$$\rho_g = \frac{-p^2(B_{1a} + B_{1b}) - jp(B_{1a}^2 - B_{1b} + 2p^2)}{[B_{1a}^2 B_{1b} - p^2(3B_{1a} + B_{1b}) + jp(2B_{1a} + B_{1a}^2 - 2p^2)]}. \tag{37}$$

Case (b) can be obtained simply by changing the sign of (37) and by interchanging B_{1a}, B_{1b}.

Consider now the infinite gain condition*

$$1 - \rho_L + 2\rho_L\rho_g = 0 \tag{38}$$

which is obtained by letting the denominator of (28) equal zero. Those values of B_{1a}, B_{1b}, B_2 for which (38) is satisfied at one or more fre-

* Note that (38) is equal to (11) ($n = 1$, $N = 2$).

quencies furnish the "boundary" of the region of stability (see the general considerations on stability of the first section).

Consider now the circuit of Fig. 14(a). By substituting (36), (37), in (38) one finds that the circuit becomes unstable under certain conditions. These conditions are illustrated in Fig. 15. One can see from Fig. 15 that, if both B_{1b}/B_{1a}, B_2/B_{1a} are too small, then the circuit is unstable. If $B_{1b} > B_{1a}$, then the input circuit does not have spurious resonances,
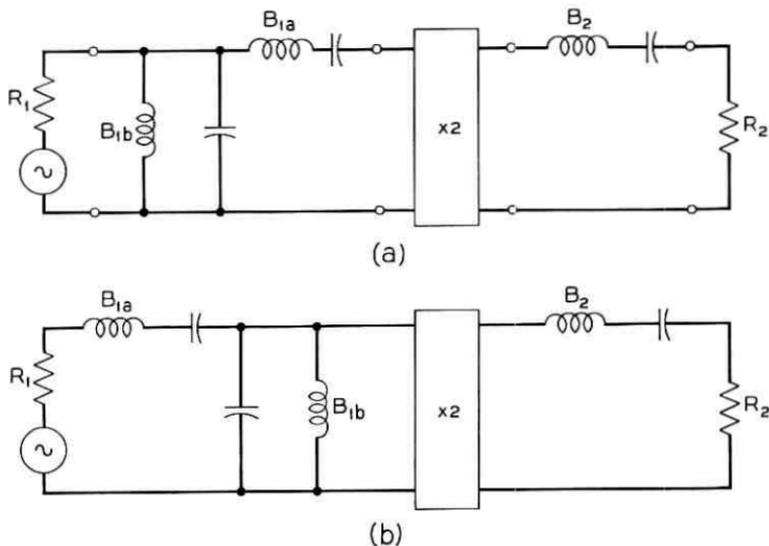

Fig. 14 — Two typical examples in which only the output circuit can be approximated by means of a single-tuned series resonant circuit.

as one may verify from (37).* But Fig. 15 shows that the circuit may still be unstable. Therefore, one can say

> instability may occur even if the circuit does not have spurious resonances. (That is, even if the circuit impedance is real only at the desired carriers ω_0, $2\omega_0$.)   (39)

Fig. 15 shows that if $B_{1b} \gg B_{1a}$, then stability is secured. This can be explained by noticing that for $B_{1b} \gg B_{1a}$ the circuit reduces to that of Fig. 9, which has already been found to be stable.

Fig. 15 shows also that stability is secured if $B_2 > 1.2B_{1a}$ (approximately). This can be explained as follows. Consider first the following theorem.

---

* If $B_{1b} > B_{1a}$, then $\rho_{gI} \neq 0$ for $0 < p < \infty$.

Theorem 2:   *If at all frequencies p either one of the following two conditions*

$$| \rho_L | < | \rho_{LO} | \tag{40}$$

$$| Z_o | \gg R_1 \tag{41}$$

is satisfied, then the doubler is stable. $| \rho_{LO} |$ *is* $\frac{1}{3}$ *if losses are absent, and it is greater than* $\frac{1}{3}$ *if losses are present.*

Proof: The preceding section has shown* that the first condition guarantees $T_i^{\rightarrow} \neq \infty$. Consider therefore, the second condition. Note
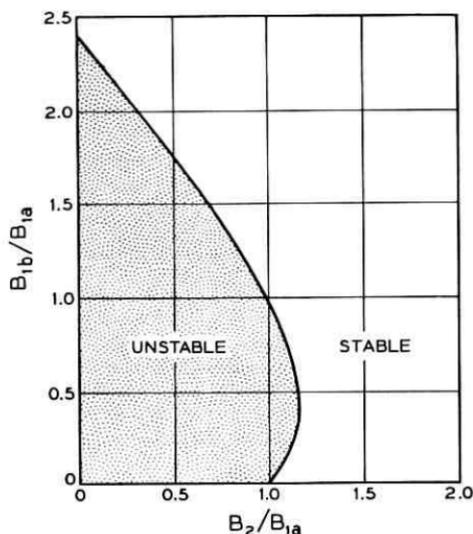

Fig. 15 — Stability region of the doubler of Fig. 14(a).

that $| Z_o | \gg R_1$ is equivalent to $\rho_o \cong 1$. Therefore, by substituting $\rho_o = 1$ in (38) and neglecting the exceptional case $\rho_L = -1$, one finds that (41) also guarantees $T_i^{\rightarrow} \neq \infty$. One concludes therefore, that either one of (40), (41) guarantees $T_i^{\rightarrow} \neq \infty$. Consider now a circuit which satisfies (40), (41). Next, decrease the amplitude of $| \rho_L |$. Since, as $| \rho_L |$ is decreased, the forward transmission never becomes infinite one concludes that the circuit is stable, according to (4).

Consider now the circuit of Fig. 14(a) and suppose $B_{1a} \ll B_2$. Note that condition (40) is satisfied for

$$p < \frac{1}{2\sqrt{2}} B_2 , \tag{42}$$

---

* See (14). Note that $| \rho_L | < \frac{1}{3}$ corresponds to the worst case (no losses; $\rho_o = 1$).

as one can verify from (36). Furthermore, since $B_{1a} \ll B_2$, one has that condition (41) is satisfied at the frequencies p

$$p > \frac{1}{2\sqrt{2}} B_2 \gg B_{1a} . \tag{43}$$

Therefore, the circuit is stable, as shown by Fig. 15.

The preceding demonstration used the fact that the condition of stability places restrictions on only those values of ρ_g (i.e., of Z_g) which occur at those frequencies p for which $|\rho_L| < |\rho_{LO}|$. This property is most important and will be further emphasized by stating

> the condition of stability places restrictions only at those frequencies p which are rejected by the output filter. There-
> fore, at those frequencies which are passed by the output   (44)
> filter, the impedance Z_g produced by the output filter can be
> quite arbitrary (all this applies to the case of a doubler).

In this statement, the frequencies which are "rejected" by the output filter are given by the complement of (42)

$$p > \frac{1}{2\sqrt{2}} B_2 \tag{45}$$

if losses are absent. The frequencies which "pass through" the output filter are given by (42)

$$p < \frac{1}{2\sqrt{2}} B_2 . \tag{46}$$

Consider now the circuit of Fig. 14(b). Consider first the following theorem.

Theorem 3:   If $R_g > R_1$, then infinite transmission never occurs.

Proof: Let $Z_g = R_g + jX_g$, and assume

$$R_g > R_1 . \tag{47}$$

Then, after some manipulations one finds

$$| 1 - 2\rho_g | = \left| \frac{(R_g + R_1) - 2(R_g - R_1) - jX_g}{(R_g + R_1) + jX_g} \right| \leqq 1.$$

Therefore, (38) is never satisfied for all values of ρ_L. This demonstrates Theorem 3.

Consider now the circuit of Fig. 14(b). Since $X_L > 0$, from (19)

one has that infinite gain occurs only if

$$X_g > 0. \tag{48}$$

But one can directly verify that the circuit of Fig. 14(b) has the following properties

$$R_g > R_1 \text{ when } X_g > 0.$$

Therefore, infinite transmission never occurs because of Theorem 3. This demonstrates that the circuit is always stable. Theorem 3 has a very important consequence.

> Spurious resonances which occur at the input sidebands do not produce instability, provided they produce high impedances $(R_g > R_1)$ at the varactor terminals (this is valid for a doubler). (49)

Note that (49) is valid for any arbitrary output circuit.

4.6 *Case in which the Input Circuit can be Approximated by Means of a Single-Tuned Series Resonant Circuit*

Suppose now that the input network consists of a series resonant circuit and that the output circuit is quite arbitrary (see Fig. 16). Then one has

$$\rho' = 1 - 2\rho_g = \frac{B_{1a} - jp}{B_{1a} + jp}.$$

Therefore, $|\rho'| = 1$. Consequently instability may arise in the limiting case $|\rho_L| = 1$. If one neglects this limiting case, one can say

> If the input circuit of the doubler can be approximated by a single-tuned resonant circuit in series with the generator, then stability is secured for any arbitrary output circuit configuration (in the case of a doubler). (50)
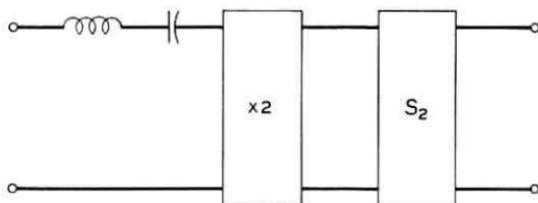

Fig. 16 — Case in which only the input circuit can be approximated by means of a single-tuned series resonant circuit.

Note that (50) is in agreement with (13), (41) which show that it is in general desirable that $|Z_s| > R_1$.

4.7 Discussion of a Frequency Multiplier of Order $N > 2$ (both a Single Stage and a Chain of Stages)

The following discussion is concerned with the problem of designing a stable chain of doublers. This problem arises both in the design of a single stage multiplier of order $2^n > 2$, and in the design of a multiplier of order 2^n consisting of many stages. That the chain may be unstable even if each doubler is individually stable, is a widely-known experimental fact. It will be shown here that if $N > 4$, then instability may arise even if: each doubler is individually stable and it is, in addition, of the type shown in Fig. 9 (which is the simplest case which may occur).

Consider an octupler and suppose that both its input and output circuits consist of single tuned resonant circuits connected in series to R_1 and R_L. Then, according to Fig. 6, its equivalent circuit is that shown in Fig. 17 and the bandwidths B_1, B_2, B_3, B_4 are the parameters on which the frequency dependence of T_i^- (and therefore, also the stability of the circuit) depends. To simplify the analysis assume the following conditions:

$$\frac{B_3}{B_4} = \frac{B_2}{B_3} = \frac{B_1}{B_2} = \gamma, \qquad (51)$$

and let the frequency p be normalized with respect to $B_4/2$ (i.e., $B_4 = 2$). In this way T_i^- depends on the parameter γ, only. T_i^- has been calculated for different values of γ and the results are shown in Fig. 18. One can see from Fig. 18 that the multiplier is unstable if

$$\gamma = 1.425 \text{ (approximately)}. \qquad (52)$$

This result is in agreement with (35), which pointed out that small values of γ are desirable. The conclusion from (52) is that
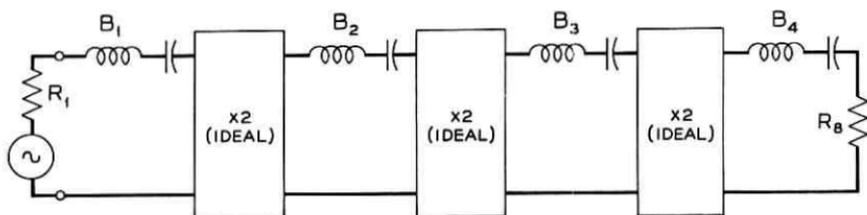

Fig. 17 — Equivalent circuit of an octupler in which both the input and output circuits consists of single-tuned series resonant circuits.

if three or more doublers of the type shown in Fig. 9 are connected in cascade, the chain will be unstable unless each (53) stage is broadband enough with respect to the preceding one.

Consider now a doubler which is individually stable, and suppose that the input circuit is made more narrowband than the output circuit. Then, if a stage is connected to the input port of the doubler, instability will not occur. In fact, according to (44), instability may arise only from interactions occurring at the frequencies which are rejected by
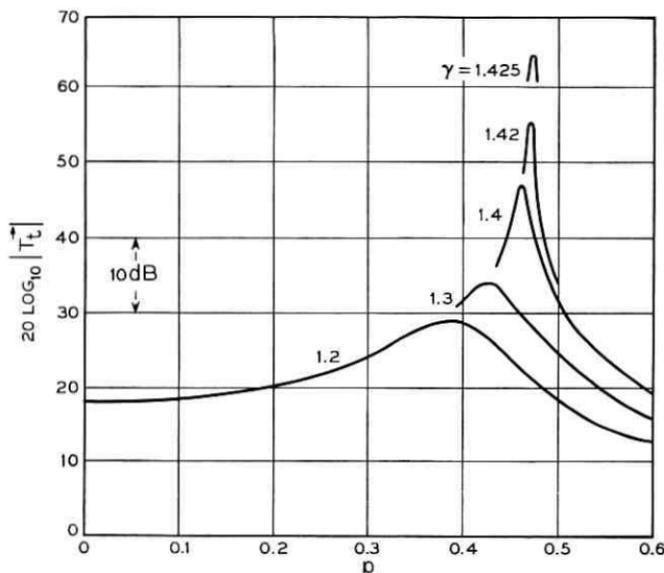

Fig. 18 — Effect of γ ($=B_i/B_{i+1}$) on the forward PM transmission of the octupler of Fig. 17.

the output circuit. But these frequencies are also rejected by the input circuit. Therefore, the input filter prevents these "dangerous" interactions. Note that if a broadband* stage is connected to the output of the doubler, then instability does not occur because the output circuit of the doubler will reject those frequencies at which the broadband stage presents reflections. The conclusion can be stated as follows.

Consider a chain consisting of doublers which are individually stable and which have $B_1 \ll B_2$ (where B_1, B_2 are the input and output bandwidths of the general doubler, respectively). (54) The chain will be stable, provided the input circuit of each

---

* Broadband with respect to the output circuit of the doubler.

doubler is broadband with respect to the output circuit of the preceding doubler.

This is also in accordance with (52) and with the conclusions derived from it.

4.8   *Discussion of the Effect of* $AM \rightleftarrows PM$ *conversion*

Consider a doubler and suppose, for instance that the output circuit S_2 produces $AM \rightleftarrows PM$ conversion. Then, if a unit PM wave is incident to the input port of S_2, two waves will be reflected back,

$$\rho_{(PM,PM)} \ , \ \rho_{(PM,AM)}$$

where the subscript (,) indicates the type of incident wave and the type of reflected wave (in this order). Conservation of energy requires

$$\mid \rho_{(PM,PM)} \mid \; \leqq 1 - \mid \rho_{L(PM,AM)} \mid$$

which shows that the amplitude of the reflected PM wave will in general be decreased by the presence of PM $\rightarrow$ AM conversion. *This is a desirable effect.* Note that the AM wave generated by S_2 may be reflected back and converted again to PM by S_2. However, for this to happen, the AM wave must travel through the "ideal" multiplier in the reverse direction, be reflected back by S_1, and travel through the multiplier in the forward direction. Through this path the AM wave is attenuated because the AM round-trip (power) transmission of the "ideal" doubler is $\frac{1}{4}$, (see (1)). The conclusion is that, when the AM wave is converted back to PM by S_2, it has small effect in most practical cases. Similar arguments can be applied to the case in which S_1 has $AM \rightleftarrows PM$ conversion. The conclusion is

input and output circuits which produce $AM \rightleftarrows PM$ conversion are in general preferable to those which do not produce $AM \rightleftarrows PM$ conversion.   (55)

Because of the qualitative nature of the discussion leading to (55) a more convincing argumentation may be desirable. A rigorous demonstration requires considerably more space and is beyond the scope of this paper.

V. CONCLUSIONS ON THE DESIGN OF A STABLE MULTIPLIER

Consider first a doubler. From (35), (44), and (54), one has

the output bandwidth should be broadband with respect to the input bandwidth. That is,   (56)

$$B_2 \gg B_1 \ .$$

The input circuit can have any arbitrary frequency dependence at those frequencies p which correspond to the passband of the output filter [see (44)]. In fact, this will not cause any instability.

Instability may arise *only if* both the input circuit and the output circuit satisfy simultaneously certain conditions. For instance it must be

$$X_g X_L \gtreqless 0, \qquad R_g < R_1, \qquad |\rho_L| > \tfrac{1}{3} \tag{57}$$

for some frequencies in order that instability may occur. From (57) one can see that spurious resonances of the input circuit do not produce instability if they cause

$$R_g > R_1.$$

Furthermore, in the design of the input and output circuits, those circuit configurations should be preferred which produce AM $\rightleftarrows$ PM conversion. Therefore, low-pass circuit configurations at the input, and high-pass circuit configurations at the output, are preferable to bandpass circuit configurations.

Finally, under certain circuit conditions, instability may arise even in the absence of spurious resonances. However, instability can be avoided by making the output circuit broadband enough with respect to the input circuit. This last part of the statement follows from (44).

The need for condition (56) is further emphasized by considering the stability conditions for a chain of doublers (or for a multiplier of order $N = 2^n$, which has been shown to be equivalent to a chain of doublers).

A chain of doublers will be stable if the following conditions are satisfied: each doubler is individually stable and satisfies (56); furthermore, each doubler is broadband enough with respect to the preceding one [see (54)].   (58)

If (58) is not satisfied, (53) shows that instability may arise even in the simplest case in which the input and output circuits of each doubler consist of (single-tuned) series resonant circuits.

Note that (56), (57) etc. apply also to the design of a multiplier of order $N = 2^n$. In this case, one should design the input circuit by considering the doubler which consists of the cascade connection of: the input circuit, an "ideal" doubler, the first idler (connected in series to the output of the "ideal" doubler). The output circuit should satisfy the stability conditions of the doubler which consists of the last idler connected in series to the input of the "ideal doubler", and of the output

circuit connected to the output of the "ideal" doubler. Note that requirement (56) gives

$$B_1 \ll B_2 \ll \cdots \ll B_n \ll B_{n+1} , \tag{59}$$

where B_1 is the input circuit bandwidth, B_2 is the first idler bandwidth [see (29) and the Appendix], B_n is the last idler bandwidth, and B_{n+1} is the output circuit bandwidth.

It is important to emphasize that all these requirements have been derived in the case of no losses. The presence of losses reduces the limitations [such as (56) and (59)] placed by the condition of stability. If the losses are so large that the multiplier efficiency is less than $1/N$ [see (24) of Ref. 3] then the multiplier will be in general stable.

Note that it has always been implied that the impedance presented to the varactor terminals by the external circuit* is "large" at frequencies very far from the carriers. If this is not true, other types of instabilities may arise, such as subharmonic generation, etc.

APPENDIX

Equivalence of a Multiplier of Order $N = 2^n$ to a Chain of Doublers

Consider a multiplier of the type defined by (26). Its equivalent circuit is shown in Fig. 19. V_g is the input voltage generator of frequency
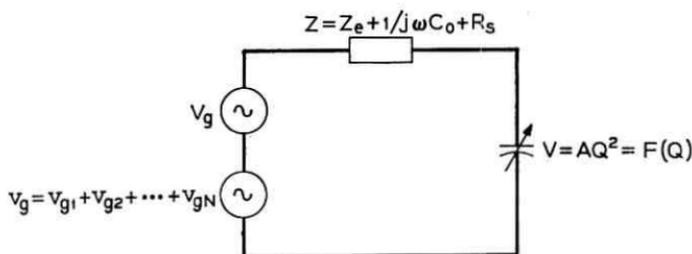

Fig. 19 — Multiplier equivalent circuit.

ω_0 . v_{g1} , v_{g2} , $\cdots$, v_{gN} are the noise terms present at the various sidebands $\omega_0 \pm p$, $2\omega_0 \pm p$, $\cdots$, $N\omega_0 \pm p$. The series resistance R_s and the average capacitance C_0 of the varactor are included in the impedance $Z(\omega)$. Therefore, the nonlinear capacitance shown in Fig. 19 represents the nonlinear part of the capacitance of the varactor and it has a Q–V

---

* The external circuit includes C_0, the average capacitance of the varactor, the inductance of the varactor, etc.

characteristic of the type: $V = F(Q) = AQ^2$ (see the Appendix of Ref. 3 for more details).

The hypothesis are

$$Z(r\omega_0) = R_r \quad (r = 2^s = 1, 2, \cdots, N) \tag{60a}$$

$$Z(\omega) = \infty \quad \text{for } \omega \text{ far from } r\omega_0 \quad (r = 2^s = 1, 2, \cdots, N). \tag{60b}$$

In (60a) it is assumed that the circuit is resonant at the idler frequencies and at the input and output frequency. R_1 is the impedance of the input generator; $R_2, \cdots, R_{N/2}$ represent the losses of the idlers; R_N is the impedance of the load at $N\omega_0$. In (60b) it is required that current flow be limited to the frequencies $\omega_0, 2\omega_0, \cdots, N\omega_0$ and their sidebands. In this appendix both positive and negative frequencies will be considered, as illustrated by the Fourier Series (61), (62), etc. In (60) consideration is confined to $\omega > 0$ because the case $\omega < 0$ is given by $Z(\omega) = Z(-\omega)^*$, where ( )* indicates the complex conjugate.

When the sidebands are absent (i.e., when $v_{g1} = v_{g2} = \cdots = v_{gN} = 0$) the charge of the nonlinear capacitance is of the type

$$Q(t) = \sum_{\substack{r = \pm 2^s \\ s = 0, \cdots, n}} Q_r \exp(jr\omega_0 t). \tag{61}$$

Notice that $Q_r = Q_{-r}^*$ because $Q(t)$ is real. The elastance $S(t)$ is given by

$$S(t) = 2AQ(t) = \sum_{\substack{r = \pm 2^s \\ s = 0, \cdots, n}} S_r \exp(jr\omega_0 t) \tag{62}$$

with

$$S_r = 2AQ_r. \tag{63}$$

Consider now the voltage $V(t)$ across the nonlinear capacitance. From $V = AQ^2$ and (61) one has

$$V(t) = A \sum_{i \neq r} Q_i Q_r \exp[j(r + i)\omega_0 t] = \sum V_l \exp[jl\omega_0 t]$$

$$(i = \pm 2^h; \quad r = \pm 2^s; \quad s, h = 0, \cdots, n)$$

from which one obtains the Fourier coefficient of $V(t)$ relative to $\omega = r\omega_0$

$$V_r = A[Q_{r/2}Q_{r/2} + 2Q_{2r}Q_{-r}] \quad (r = \pm 2^s; s = 0, \cdots, n). \tag{64}$$

The linear circuit connected to the nonlinear capacitance gives an additional expression for V_r,

$$V_r = -jr\omega_0 Q_r Z(r\omega_0) + V_{gr} \quad (r = \pm 2^s; s = 0, \cdots, n), \tag{65}$$

where V_{gr} is the Fourier coefficient of V_g relative to $\omega = r\omega_0$. Therefore, $V_{gr} = 0$ for $|r| \neq 1$. From (65), (64) one obtains

$$V_{gr} - jr\omega_0 Q_r Z(r\omega_0) = A(Q_{r/2}Q_{r/2} + 2Q_{2r}Q_{-r})$$

$$(r = \pm 2^s; s = 0, \cdots, n). \qquad (66)$$

Equation (66) gives the "equilibrium equations" of the circuit of Fig. 19 at the carriers $\pm\omega_0$, $\pm 2\omega_0$, $\cdots$, $\pm N\omega_0$.

Consider now the sidebands. Let

$$q(t) = \sum_{\substack{s=0,\cdots,n \\ r=\pm 2^s \\ i=\pm 1}} q_{r,i} \exp\left[j(r\omega_0 + ip)t\right] \qquad (67)$$

be the sidebands of the charge of the nonlinear capacitance. Then the voltage sidebands are given by

$$v(t) = S(t)q(t). \qquad (68)$$

By substituting (67), (62) in (68) one obtains $v_{r,i}$, the Fourier coefficient of $v(t)$ relative to $\omega = r\omega_0 + ip$

$$v_{r,i} = q_{r/2,i}S_{r/2} + q_{2r,i}S_{-r} + q_{-r,i}S_{2r} \qquad (i = \pm 1). \qquad (69)$$

For more details on the derivation of (69) see the Appendix of Ref. 3.

The linear circuit gives a second expression for $v_{r,i}$

$$v_{r,i} = (v_g)_{r,i} - j(r\omega_0 + ip)q_{r,i}Z(r\omega_0 + ip), \qquad (70)$$

where $(v_g)_{r,i}$ indicates the Fourier coefficient of $v_g = v_{g1} + \cdots + v_{gN}$ relative to $\omega = r\omega_0 + ip$. By combining (69) with (70) one obtains the "equilibrium equation" of the circuit of Fig. 19 at the sidefrequency $r\omega_0 + ip$

$$(v_g)_{r,i} = j(r\omega_0 + ip)q_{r,i}Z(r\omega_0 + ip)$$
$$+ q_{r/2,i}S_{r/2} + q_{2r,i}S_{-r} + q_{-r,i}S_{2r}. \qquad (71)$$

Consider now the chain illustrated in Fig. 20. Let the Q–V characteristic of the nonlinear capacitances of Fig. 20 be equal to that of Fig. 19. Furthermore, let $Z_r(\omega)$ be equal to $Z(\omega)$ for $\omega \cong r\omega_0$ and be infinite for ω far from $r\omega_0$. That is,

$$Z_r(\omega) = Z(\omega) \quad \text{for} \quad \omega = r\omega_0, r\omega_0 \pm p$$

$$Z_r(\omega) = \infty \quad \text{for} \quad \omega \text{ far from } r\omega_0 \quad (r = 2^s = 1, 2, \cdots, N), \qquad (72)$$

where only positive frequencies are considered because $Z_r(\omega) = Z_r^*(-\omega)$. Note that because of (72), the spectrum of the charge of the $(s + 1)$th
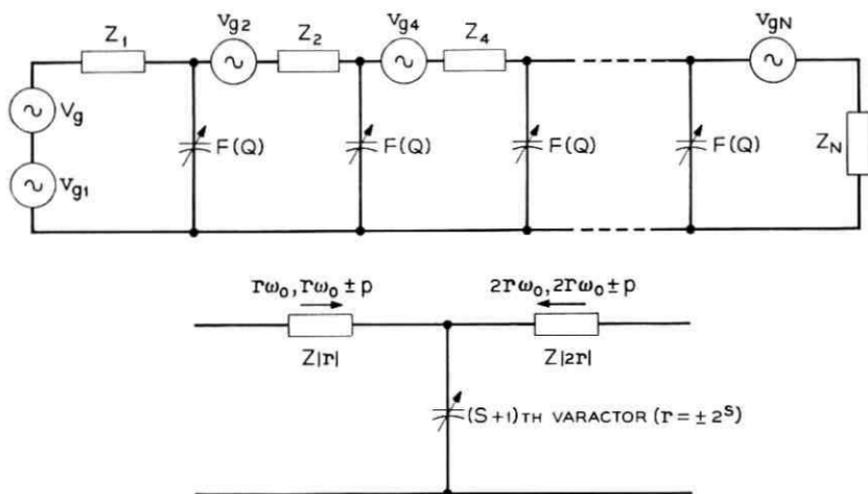
Fig. 20 — Chain of doublers equivalent to the multiplier of Fig. 19.

nonlinear capacitance of Fig. 20 is restricted to the carrier frequencies $r\omega_0$, $2r\omega_0$ ($r = \pm 2^s$) and to their sidebands $r\omega_0 \pm p$, $2r\omega_0 \pm p$. Let the symbol $(\ )'$ be used to distinguish the variables of the circuit of Fig. 19 from those of the circuit of Fig. 20. Then the charge components of the $(s + 1)$th capacitance of Fig. 20 are Q'_r, Q'_{2r}, $q'_{r,i}$, $q'_{2r,i}$, with $r = \pm 2^s$ and $i = \pm 1$.

It will be shown that, under these hypothesis, the chain of Fig. 20 is equivalent to the multiplier of Fig. 19. More precisely, it will be shown that the charges and voltages of the two circuits are equal.

Demonstration: Let first the fact that Q_1, $\cdots$, Q_N are equal to Q'_1, $\cdots$, Q'_N be demonstrated. Consider therefore, the "equilibrium equation" of the chain of Fig. 20 for $\omega = r\omega_0$ ($r = \pm 2^s$). With reference to Fig. 21, it is obtained by applying Kirchoff's law to the $(s + 1)$th loop.
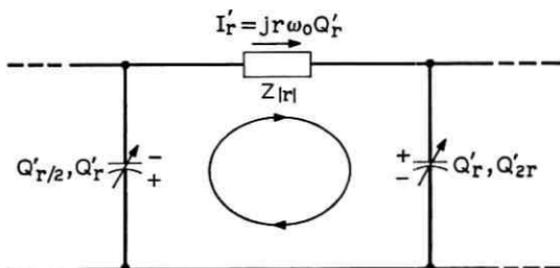
Across the first capacitance of Fig. 21, the voltage component of frequency $r\omega_0$ is

$$AQ'_{r/2}Q'_{r/2} \qquad (r = \pm 2^s, \mid r \mid\, \neq 1, N). \tag{73}$$

The voltage component of frequency $r\omega_0$ produced by the second capacitance is

$$2AQ'_{-r}Q'_{2r} \qquad (r = \pm 2^s, \mid r \mid\, \neq 1, N). \tag{74}$$

Since only the current $jr\omega_0 Q'_r$ is flowing through $Z_{\mid r \mid}$, the voltage

Fig. 21 — Sth loop ($r = 2^s$) of the chain of Fig. 20.

across $Z_{|r|}$ is

$$jr\omega_0 Q'_r Z_{|r|}(r\omega_0) = jr\omega_0 Q'_r Z(r\omega_0) \qquad (r = \pm 2^s, |r| \neq 1, N). \qquad (75)$$

Equilibrium requires that the sum of these three voltages be zero. Therefore, one obtains

$$0 = jr\omega_0 Q'_r Z(r\omega_0) + A[Q'_{r/2}Q'_{r/2} + 2Q'_{2r}Q'_{-r}] \qquad (76)$$

which is identical to the "equilibrium equation" of the circuit of Fig. 19 for $\omega = r\omega_0$ (with $r = \pm 2^s$, $|r| \neq 1, N$).

In a completely similar way, one finds that also for $|r| = 1$ and $|r| = N$ the "equilibrium equations" of the two circuits are identical. If $r = 1$, for instance, the loop consists of the voltage generator V_g in series with Z_1 and the first nonlinear capacitance of Fig. 20. One obtains

$$V_{g1} - j\omega_0 Q'_1 Z(\omega_0) = 2AQ'_2 Q'_{-1}$$

which is identical to (66) for $r = 1$. The case $r = N$ is obtained by considering the last loop and one finds

$$-jN\omega_0 Q'_N Z(N\omega_0) = AQ'_{N/2}Q'_{N/2}$$

which is identical to (66) for $r = N$. All this demonstrates that $Q_1, \cdots, Q_N$ are identical to $Q'_1, \cdots, Q'_N$. Therefore, the two circuits are equivalent at the carrier frequencies.

Consider now the sidebands. Notice that the elastance components of the $(s + 1)$th nonlinear capacitance of Fig. 20 are S_r, S_{2r} ($r = \pm 2^s$). Therefore, the equilibrium of the $(s + 1)$th loop for $\omega = r\omega_0 + ip$ requires

$$(v_g)_{r,i} = j(r\omega_0 + ip)q'_{r,i}Z(r\omega_0 + ip)$$
$$+ q'_{r/2,i}S_{r/2} + S_{2r}q'_{-r,i} + S_{-r}q'_{2r,i} \qquad (r = \pm 2^s, i = \pm 1). \qquad (77)$$

The first term of the second member represents the voltage component across $Z_{|r|}$, the second term corresponds to the voltage across the first capacitance, and the last two terms correspond to the voltage across the second capacitance. Equation (77) is equivalent to (71). Therefore, the two sets of charge components $(q_{r,i})$, $(q'_{r,i})$ ($|r| = 1, 2, \cdots, N$; $i = \pm 1$) are identical. This concludes the demonstration of the equivalence of the two circuits of Figs. 19 and 20. Note that the voltage across $Z_{|r|}$ is equal to the voltage components of Z due to the frequencies $\pm r\omega_0$ and their sidebands.

A.1 Discussion of the Circuit of Fig. 20

The circuit of Fig. 20 can be represented by the chain of doublers illustrated in Fig. 22. The equivalent circuit of the $(s + 1)$th doubler of Fig. 22 is shown in Fig. 23. It consists of an "ideal" input filter $F_{\text{in}}^{(s+1)}$ connected in series to the input, an "ideal" output filter $F_{\text{out}}^{(s+1)}$ connected in series to the output, and an "ideal" lossless varactor which has a Q-V characteristic of the type $V = AQ^2$. The filter $F_{\text{in}}^{(s+1)}$ has zero impedance at the input frequencies $r\omega_0$, $r\omega_0 \pm p$ and it has infinite impedance for ω far from $r\omega_0$ ($r = \pm 2^s$). Similarly, $F_{\text{out}}^{(s+1)}$ limits the output current to the frequencies $2r\omega_0$, $2r\omega_0 \pm p$. Consequently, the series connection of $F_{\text{out}}^{(s)}$, Z, $F_{\text{in}}^{(s+1)}$ is equivalent to $Z_{|r|}$, according to (72). Therefore, the two circuits of Figs. 22 and 20 are equivalent. By first approximation, the properties of the "ideal" doublers of Fig. 22 are independent of the modulation frequency p. More precisely, the frequency dependence of the "ideal" doublers can be neglected with respect to the frequency dependence of the impedances Z_1, Z_2, etc. This is more precisely explained by the following considerations.

Consider (77). Let $i_{r,i}$ indicate the Fourier coefficient of the current flowing through $Z_{|r|}$ for $\omega = r\omega_0 + ip$. Then
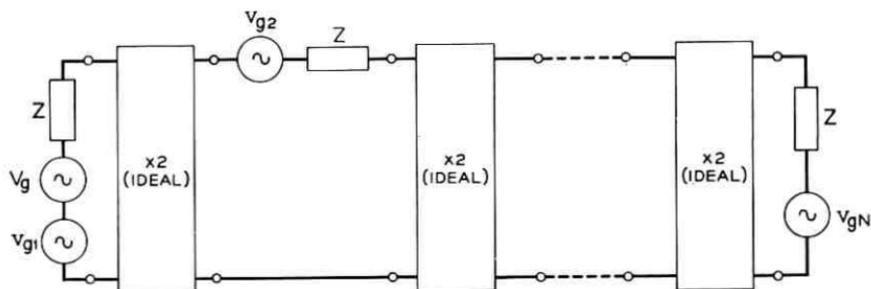

Fig. 22 — Chain of doublers equivalent to the circuit of Fig. 20.

$$i_{r,i} = j(r\omega_0 + ip)q_{r,i} \qquad (i = \pm 1). \tag{78}$$

If one approximates (78) by means of $i_{r,i} = jr\omega_0 q_{r,i}$, one obtains from (77)

$$Z(r\omega_0 + ip)i_{r,i} - \frac{2j}{r} S_{r/2}i_{r/2,i} + \frac{j}{r} S_{2r}i_{-r,i} - \frac{j}{2r} S_{-r}i_{2r,i}$$
$$= (v_s)_{r,i} . \tag{79}$$

Equation (79) gives the approximate version of (77) which is obtained by neglecting the frequency (p) dependence of the behavior of the nonlinear capacitances of Fig. 20. In fact, one can see that the only term in (79) which depends on p is the first term, and this term is caused by the impedance $Z_{|r|}$ of Fig. 20. In this analysis the approximate expression (79) is valid because it is assumed $Z(\omega) \cong \infty$ for ω far from $r\omega_0$. Therefore, at those frequencies p for which the approximation $i_{r,i} \cong r\omega_0 q_{r,i}$ becomes invalid, the only important term in (79) is the first, which does not contain any approximation. However, it is important to notice that (79) is exact only in the limiting case of a narrow-band multiplier. That is, when $Z_{|r|}(\omega)$ varies so rapidly with p that it is infinite outside a very narrow band around $\omega = r\omega_0$.

According to the preceding considerations, the properties of the "ideal" doublers of Fig. 22 can be assumed independent of the modulation frequency p, by first approximation.

In the following part of this Appendix, it will be shown that the impedance presented to the output of each "ideal" doubler by the following part of the chain is real at the carrier frequency. Furthermore, each doubler is lossless. Therefore, according to the results derived in Ref. 3, one concludes that each doubler does not produce AM $\rightleftarrows$ PM
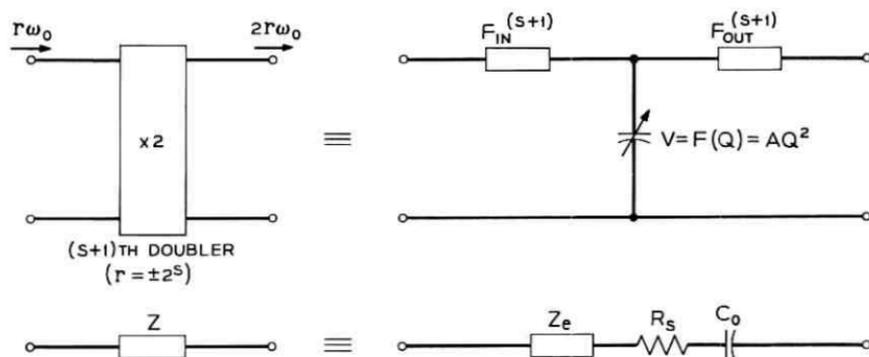

Fig. 23 — Constituents of the chain of Fig. 22.

and that it is characterized by the two scattering matrices given by (27). The characteristic impedances of the input and output ports of the various doublers are derived in the following part.

A.2 Calculation of the Output and Input Impedances of the Doublers of Fig. 22

At the carrier frequency $r\omega_0$, the input impedance of the $(s + 1)$th doubler $(r = 2^s)$ is given by

$$Z_{in}^{(s+1)} = \frac{V_{in}^{(s+1)}}{jr\omega_0 Q_r} = \frac{2AQ_{-r}Q_{2r}}{jr\omega_0 Q_r} \qquad (r = \pm 2^s), \qquad (80)$$

where $V_{in}^{(s+1)}$, the input voltage of the $(s + 1)$th doubler, has been obtained from (74).

The output impedance of the $(s + 1)$th doubler is given by

$$Z_{out}^{(s+1)} = Z(2r\omega_0) + Z_{in}^{(s+2)} \qquad (r = \pm 2^s). \qquad (81)$$

But one also has

$$Z_{out}^{(s+1)} = \frac{-V_{out}^{(s+1)}}{j2r\omega_0 Q_{2r}} = \frac{-AQ_r Q_r}{j2r\omega_0 Q_{2r}} \qquad (r = \pm 2^s), \qquad (82)$$

where $V_{out}^{(s+1)}$ is the output voltage of the $(s + 1)$th doubler and it is given by (73) with $r/2$ replaced with r. Therefore, from (80) and (82) one obtains

$$Z_{out}^{(s+1)} Z_{in}^{(s+1)} = \frac{A^2 |Q_r|^2}{r^2 \omega_0^2} \qquad (r = \pm 2^s). \qquad (83)$$

By combining (81) and (83), one has

$$Z_{in}^{(s+1)}[Z_{in}^{(s+2)} + Z(2r\omega_0)] = \frac{A^2 |Q_r|^2}{r^2 \omega_0^2} \qquad (r = \pm 2^s). \qquad (84)$$

The output impedance of the last doubler $(s = n)$ is

$$Z_{out}^{(n)} = Z(N\omega_0) = R_N. \qquad (85)$$

Equation (85) and the recurrent formulas (81) and (84) allow the output and input impedances of the various doublers to be readily calculated $(R_N, A, |Q_r|$ are known).

Note that Z is real at the frequencies ω_0, $2\omega_0$, $\cdots$, $N\omega_0$, because the circuit is resonant at these frequencies. $Z(2\omega_0)$, $\cdots$, $Z(N/2\omega_0)$ represent the idler losses. If, for instance, the varactor is the only lossy element of the circuit, then

$$Z(2\omega_0) = Z(4\omega_0) = \cdots = R_s, \qquad (86)$$

where R_s is the series resistance of the varactor. Therefore, since (85), (84), and (81) are real, one concludes that $Z_{in}^{(s)}$, $Z_{out}^{(s)}$ are real.

Finally, according to the convention of the first paper, the input and output characteristic impedances of the sth doubler are given by $Z_{in}^{(s)}$, $Z_{out}^{(s)}$, respectively.

REFERENCES

1. Hines, M. F., Bloisdell, A. A., Collins, F., and Priest, W., Special Problems in Microwave Harmonic Generator Chain, 1962, International Solid State Conference, Philadelphia, Pa.
2. Buck, D. C., Origin of Spurious Signals in a Varactor Tripler, Proc. IEEE (correspondence), October, 1965, p. 1677.
3. Dragone, C., Phase and Amplitude Modulation in High-Efficiency Varactor Frequency Multipliers—General Scattering Properties, B.S.T.J., this issue, pp. 775–796.
4. Prabhu, Vasant, Noise Performance of Abrupt-Junction Varactor Frequency Multipliers, Proc. IEEE (correspondence), February, 1966, pp. 285–287. Maunsell, H. I., private communication.
5. James, Nichols, Phillips, *Theory of Servomechanisms*, Radiation Laboratory Series, McGraw-Hill Book Co., New York, 1947.
6. Bode, H. W., *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand, New York, 1957.
7. Carlin, H. J., The Scattering Matric in Network Theory, Trans. IRE, *CT-3*, June, 1956, pp. 88–96. (See his extensive bibliography.)
8. Penfield, P. and Rafuse, R. P., *Varactor Applications*, MIT Press, 1962.

Contributors to This Issue

C. Dragone, Laurea in E.E., 1961, Padua University (Italy); Bell Telephone Laboratories, 1961—. Mr. Dragone has been engaged in experimental and theoretical work on microwave antennas and solid-state power sources. He is currently involved in solid-state radio systems experiments at the Crawford Hill Laboratory.

John M. Fraser, B.S. in E.E., 1945, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1934—. Mr. Fraser has been engaged in transmission systems engineering on a variety of carrier systems including submarine cable systems and TASI. He is currently responsible for a group working on the problems in transmitting *Picturephone*® signals between telephone offices. Senior member, IEEE; member, Sigma Xi, Tau Beta Pi, Eta Kappa Nu.

Detlef Gloge, Dipl. Ing., 1961, D.E.E., 1964, Braunschweig Technische Hochschule (Germany); research staff, Braunschweig Technische Hochschule 1961–1965; Bell Telephone Laboratories, 1965—. In Braunschweig, Mr. Gloge was engaged in research on lasers and optical components. At Bell Telephone Laboratories, he has concentrated in the study of optical transmission techniques. Member, VDE, IEEE.

H. H. Haas, B.S. in E.E., 1952, University of Washington; Bell Telephone Laboratories, 1952—. Mr. Haas completed the Communication Development Training Program in 1955 and has been concerned with system engineering aspects of radio systems. At present he is working on digital radio transmission. Member, IEEE, Sigma Xi, Tau Beta Pi.

Attilio J. Rainal, University of Alaska, University of Dayton, 1950–52; B.S.E.Sc., 1956, Pennsylvania State University; M.S.E.E., 1959, Drexel Institute of Technology; Dr. ENG., 1963, Johns Hopkins University; Bell Telephone Laboratories, 1964—. Mr. Rainal has been engaged in research on noise theory with application to radar

theory. Member, Tau Beta Pi, Eta Kappa Nu, Sigma Tau, Pi Mu Epsilon, Sigma Xi, IEEE.

G. H. ROBERTSON, B.Sc., 1943, University of Glasgow; after three years in the Royal Navy as an Air Radio Officer he returned to the University of Glasgow for two years and obtained a Post Graduate Certificate in Natural Philosophy; Bell Telephone Laboratories, 1948—. Until 1958 Mr. Robertson was engaged in electronics research and a variety of electron tube development projects. Since 1958 he has been working on signal propagation and processing studies in the Underwater Research and Systems Departments. Associate member, IEEE; member, AAAS.

MARSHALL G. SCHACHTMAN, B.S. and M.S. in E.E., 1958, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1955—. After early work on a new low-current telephone subscriber's set, Mr. Schachtman was engaged in designing circuits for use in test equipment for Radio Relay Systems. He later was engaged in systems engineering work on TASI and on system studies of the use of lasers in communications. He continued his systems engineering work on SD submarine cable systems and on studies of methods for improving the performance of HF radio telephone systems. He is currently working on systems studies in connection with the introduction of *Picturephone*® service. Member, IEEE, Eta Kappa Nu, Tau Beta Pi.